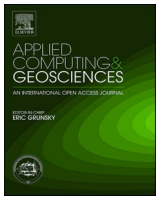
Applied Computing and Geosciences 23 (2024) 100186

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences



A supervised machine learning procedure for EPMA classification and plotting of mineral groups

R. Cossio [a], S. Ghignone [a,*], A. Borghi [a], A. Corno [a], G. Vaggelli [b]

[a] Dipartimento di Scienze Della Terra, Università Degli Studi di Torino, Via V. Caluso, 35, I-10123, Torino, Italy
[b] Istituto di Geoscienze e Georisorse, U.O. di Torino, CNR. Via V. Caluso, 35 I-10123 Torino, Italy

ARTICLE INFO

Keywords:
Machine learning
Minerals
EPMA
Petrology

ABSTRACT

An analytical method to automatically characterize rock samples for geological or petrological purposes is here proposed, by applying machine learning approach (ML) as a protocol for saving experimental times and costs.

Proper machine learning algorithms, applied to automatically acquired microanalytical data (i.e., Electron Probe Micro Analysis, EPMA), carried out with a SEM-EDS microprobe on randomly selected areas from a petrographic polished thin section, are trained, used, tested, and reported.

Learning and Validation phases are developed with literature mineral databases of electron microprobe analyses on 15 main rock-forming mineral groups. The Prediction phase is tested using an eclogite rock from the Western Alps, considered as an unknown sample: randomly selected areas are acquired as backscattered images whose intervals of gray levels, appropriately set in the gray level histogram, allow the automated particle mineral separation: automated separating Oxford Instruments Aztec Feature ® packages and a mineral plotting software are applied for mineral particle separation, crystal chemical formula calculation and plotting.

Finally, a microanalytical analysis is performed on each separated mineral particle. The crystal chemical formula is calculated, and the final classification plots are automatically produced for any determined mineral. The final results show good accuracy and analytical ease and assess the proper nature of the unknown eclogite rock sample. Therefore, the proposed analytical protocol is especially recommended in those scenarios where a large flow of microanalytical data is automatically acquired and needs to be processed.

1. Introduction

Modern petrological studies are classically based on chemical characterization of rocks and minerals, fundamental data to obtain reliable results and geological interpretations. Electron Probe Micro Analysis (EPMA) are classically obtained with SEM-EDS and WDS instruments, whose technology is exponentially increasing through time by improving their main features, such as resolution, speed of acquisition, automated and Machine Learning-based (ML) acquisition systems (e.g., Oxford Instruments Aztec Quantline, QuantMap and Feature ® packages). This allows to produce very large amounts of analyses in a shorter time. Consequently, in such large EPMA datasets each individual analyses are difficult to handle one-by-one. Moreover, verifying the reliability of single data point, recalculating and plotting all of them is highly time consuming.

Automated mineralogical recalculation has been developed over several decades (see e.g., Hagni, 2008; Sutherland et al., 1989; Pignolet-Brandom and Lapakko, 1990), producing interesting and promising results. Identification and quantification of mineral phases known as QEM-SCAN® was developed in Australia by the Commonwealth Scientific and Industrial Research Organization (Reid and Zuiderwyk, 1983).

To drastically reduce the time consumed to handle each analysis, a machine learning approach directly interconnected with a software that reports both the mineral formula calculation and classification plots (Walters, 2022), would prove to be extremely useful. Moreover, Petrelli (2024) hypothesizes that ML methods will soon support most current petrological techniques for reducing repetitive and time-consuming tasks. In this work, we applied some well-known Machine Learning (ML) algorithms (Random Forest, RF; Extremely Randomized Trees, ERT; Gradient Boosting, GB), for supervised learning classification of data, combining them with an automated mineral formula calculation and plotting software, i.e. MinPlot (Walters, 2022). This procedure consists of a first phase of i) Learning, followed by a ii) Validation phase,

* Corresponding author.
E-mail address: s.ghignone@unito.it (S. Ghignone).

https://doi.org/10.1016/j.acags.2024.100186
Received 16 May 2024; Received in revised form 15 July 2024; Accepted 9 August 2024
Available online 19 August 2024

and a iii) Prediction phase, based on experimental EPMA data, automatically acquired on a rock sample.

The most laborious and subtle phase to develop is Learning, by selecting a complete and reliable EPMA analysis database (WDS data) of rock-forming minerals. Nowadays, the most reliable database available is the RRUFF (see section 2.2 for details), which contains several EPMA analysis from different rock types with different chemical variations.

For testing the entire approach, we applied this procedure to a robust series of data automatically acquired (via Oxford Instruments Aztec Feature ®) on a rock sample, previously validated with standard stoichiometric recalculations. The sample selected for this study is an eclogitic rock from the Western Alps: the choice of this kind of metamorphic, poly-deformed rock was made because it represents a challenging task for testing the ML procedure. In fact, the selected sample encompasses numerous mineralogical phases, even compositionally similar to each other (e.g. amphibole and pyroxene mineral groups), and with different microstructural sites and shapes throughout the sample. Remarkably accurate results show that all the phases occurring in the rock sample have been identified, recalculated, and plotted into proper classification diagrams.

2. Material and methods

2.1. Machine learning algorithms

The Machine Learning approach allows to process large EPMA datasets combining several existing algorithms. Random Forests (RF - Breiman, 2001) refers to several learning algorithms consisting of multiple Decision Trees (DT – Quinlan, 1996). During the training step, RF constructs multiple decision trees (a forest) over different subsets of the training data: based on the results obtained from all the trees, the final prediction will be obtained through majority voting for classification or averaging for regression (Breiman, 2001). Random Forest and Extremely Randomized Trees (ERT - Geurts et al., 2006) belong to ensemble learning algorithms class, which use the power of many learning algorithms altogether. The concept behind all this is that using multiple learning algorithms can lead to a better final prediction. While RF works well on classification, the sampling of subsets may introduce some bias, ERT construct multiple trees sampling from the entire dataset and hence reduce both the bias and variance by randomized splitting of nodes. Gradient boosting (GB) produces highly interpretable procedures for classification, especially for extracting less than clean data (Friedman, 2001). Homogeneous ensembles use the same learning algorithm to train base estimators, and ensemble diversity can be achieved through different combination of input data (i.e. the measured chemical composition and the stoichiometric recalculations). Inhomogeneous ensembles use different learning algorithms to achieve ensemble diversity. For classification tasks, the “majority vote”, also known as the statistical mode, is used to aggregate predictions of individual base learners. This mode is simply the most frequently occurring element where base estimator has equal weight, and it is a statistical parameter like the mean or the median (Kunapuli, 2023). In order to use described ML algorithms, the SharpLearning package was used (an open-source ML library for C#.Net.; https://github.com/mdabros/SharpLearning). SharpLearning provides easy access to machine learning algorithms, focusing on supervised learning for classification. To obtain the best results from the 3 algorithms, we used the RandomSearchOptimizer function, present in the SharpLearning package, to tune Hyperparameters (Table S1, Supplementary Material) with their optimized values, for each algorithm. For this optimization we randomly split the RRUFF database, assigning 70% of the observations to the training dataset and 30% to the test dataset.

2.2. Stoichiometric calculations and plotting

For the purposes of this work, the stoichiometric recalculation and the related plots with numerous WDS and EDS quantitative spot analyses (belonging to different mineral groups), were tested with MinPlot (Walters, 2022), obtaining encouraging results.

MinPlot (Walters, 2022) is a MATLAB®-based software widely used for the recalculation of the mineral structural formula and the compositional plotting of the results. As input, the software uses the weight percent of 19 main oxides (Wt%) from EPMA data (Table S2, Supplementary Material) and provides the mineral formula for 15 different mineral groups (amphibole, apatite, chlorite, chloritoid, cordierite, epidote, feldspar, garnet, ilmenite, mica, olivine, pyroxene, serpentine, talc, titanite).

After testing the EPMA, the use of some well-known Machine Learning (ML) algorithms seemed to be appropriate for the building of the procedure object of this work. We used supervised learning, where the training data are accompanied by labels indicating the true class. To develop these methods, a complete and reliable database of microprobe spot analyses (expressed as wt.% oxides) was necessary for the training phase. The RRUFF chemistry database ((Chemistry–RRUFF); https://rruff.info/zipped_data_files/chemistry) proved to be most suiting for this aim, including 2103 RRUFF samples and about 20.000 WDS microprobe analyses. Some other EPMA data (WDS) in the learning phases are from Rigby et al. (2008); Lanari et al. (2014) and Ketcham (2015).

2.3. Analytical data acquisition

The thin section of an eclogite, used for the testing phase, was firstly analyzed via optical microscopy to identify mineral phases and microstructural sites suitable for the required microprobe analyses. The thin section was carbon coated to ensure a reliable beam current integration and was analyzed by EPMA, to determine major element distribution and to provide a suggestion on the heterogeneity of the crystals.

Microprobe analyses were acquired with a JEOL JSM-IT300LV scanning electron microscope supplied with an EDS Oxford Instruments X-act silicon drift detector hosted by the Dipartimento di Scienze della Terra (University of Turin). For this kind of detector, the Aztec Energy version 6.0 SP1 provided with Oxford Instruments Aztec Feature ® package was used. Sixteen disjoint areas (corresponding to about 20 mm^2) on the entire sample were selected and then acquired at 100 magnifications, arranged along a regular 4x4 grid. The analytical conditions were as follows: accelerating voltage = 15 kV, probe current = 2 nA, working distance = 10 mm, LiveTime (LT) = 3 s. Considering a count-rate of 10^5 counts per second (CPS), we obtain about $3x10^5$ counts in each recorded spectrum, implying an analytical precision around 0.1 wt% for each major element. Every spectrum was then quantified using a calibration with standards approach and expressed as Wt% ox.

A quantitative ZAF correction and a filtered least square treatment (Aztec Energy 6.0- Tru-Q) were used for the full quantitative analyses. Astimex Scientific Limited® natural mineral standards for microanalysis were used: [albite (Na), periclase (Mg), almandine garnet (Al and Fe), quartz (Si), apatite (P), sanidine (K), wollastonite (Ca), and rutile (Ti)]. Back-scattered electrons (BSE) gray levels images were calibrated prior to the run of analyses, in order to obtain consistent brightness and contrast for all the analyzed areas. Acquisition with Oxford Instruments Aztec Feature ® mode offers the possibility to analyze a frame after the identification of the particles (i.e., the mineral phases), via BSE imaging. This is fundamental for the system because it allows to make a single analysis in the central point of the particle, previously detected by the system. If a large number of mineral groups (>20) needs to be evaluated at the same time, a large number of ranges in the BSE (Mean Atomic Number Ranges) grayscale histogram needs to be taken into account. For some groups with a wide compositional zoning (i.e. amphibole, etc.) larger ranges or, in some cases, more ranges are required. To avoid an excessive number of mineral particles and to have greater accuracy at the point of analysis, features with Equivalent Circle Diameter (ECD) < 15 μm have been arbitrarily excluded.

Finally, we chose to automatically acquire just one microanalytical point for each homogeneous feature located at the center of the particle itself (i.e. the midpoint of the longest chord constituting the particle area).

According with Newbury and Ritchie (2015), in this case we used SEM-EDS instrument to perform our analysis, sure to obtain reliable, accurate and precise analytical results comparable to WDS microprobe, respecting appropriate analytical conditions, as reported above. In addition, we want to show that our procedure could be valid on any kind of EPMA data, obtained both with EDS and with WDS.

3. Results and discussion

3.1. Machine learning procedure

3.1.1. Experimental phase

The analyses reported in Table 1 were selected from the RRUFF database (Table S2; Supplementary Material): taking into account the main rock-forming minerals, we have investigated 15 mineral groups, 48 mineral phases and 150 RRUFF samples, for a total of 2103 spot analyses. To handle such amount of data more easily, the Microsoft DataSet and Datatable format were used within a Visual Studio program: in this way, the data were imported directly from RRUFF-chemistry Excel sheets.

Table 1
Selected microprobe analysis from Chemistry RRUFF database, in brackets (n.of Analyses)(n.of Samples).

RRUFF&ox (2103)	
Groups(15)	**Minerals(48)**
amphibole(283)	actinolite(15)(1)
	eckermannite(15)(1)
	ferroactinolite(15)(1)
	glaucophane(12)(1)
	grunerite(28)(3)
	hastingsite(18)(2)
	magnesioriebeckite(10)(1)
	pargasite(76)(4)
	riebeckite(20)(1)
	tremolite(57)(4)
	winchite(17)(2)
apatite(46)	fluoroapatite(46)(3)
chlorite(42)	chamosite(15)(1)
	clinochlore(27)(2)
chloritoid(30)	chloritoid(30)(2)
cordierite(60)	cordierite(60)(3)
epidote(39)	epidote(39)(2)
feldspar(201)	albite(47)(4)
	anorthite(77)(6)
	orthoclase(63)(5)
	sanidine(14)(1)
garnet(616)	almandine(99)(7)
	andradite(133)(10)
	grossular(153)(10)
	pyrope(111)(8)
	spessartine(90)(6)
	uvarovite(30)(2)
ilmenite(8)	ilmenite(8)(1)
mica(209)	annite(32)(2)
	biotite(13)(1)
	muscovite(117)(7)
	phlogopite(35)(5)
	paragonite(12)(3)
olivine(143)	fayalite(37)(3)
	forsterite(96)(6)
	tephroite(10)(1)
pyroxene(382)	aegirine(44)(3)
	augite(42)(3)
	diopside(128)(9)
	enstatite(79)(6)
	ferrosilite(22)(2)
	hedenbergite(28)(2)
	jadeite(12)(1)
	omphacite(8)(1)
	wollastonite(19)(1)
talc(9)	talc(9)(3)
titanite(20)	titanite(20)(1)
serpentine(15)	lizardite(15)(1)

For all considered data, it is possible to perform the stoichiometric recalculation with MinPlot using the values of any chemical elements reported as weight percentage Oxides (Wt. % Ox). MinPlot Matlab ® functions are called directly from our program and then output results are redirected into our custom Dataset. For an easy comparison, useful graphs were faithfully replicated inside our program following MinPlot program instructions (Walters, 2022): for our graphic output, the package "(ScottPlot): an Interactive Plotting Library for.NET" (https://scottplot.net/) was used.

Once precision and accuracy of RRUFF data had been verified through MinPlot recalculation and plotting, it was possible to move on to the Learning phase.

3.1.2. Learning phase

The first step involved the training of the program, so that it had as reference a wide range of analyses of the most common and representative minerals of magmatic and metamorphic rocks (e.g. RRUFF database, (Chemistry–RRUFF); https://rruff.info/zipped_data_files/chemistry).

The Flow Chart of Fig. 1 is obtained by using all the available data: both the oxides (circle 1) and the 15 stoichiometric recalculations (Table 1, circles 2 to 16: column Group), for each analysis point in the database, are then used as input data (Parallel Homogeneous Ensemble). Each of the 16 datasets is then ready to be applied to the 3 ML (Parallel Inhomogeneous Ensemble) algorithms: RF= Random Forest, ERT = Extremely Randomized Trees and GB = Gradient Boosting, together with their output data (mineral groups). A total of 48 models is thus obtained which, after "training" the input/output data, is saved on disk with all the references to the conditions of use (input variables, algorithm used, output classes), to be used consistently in the Prediction phase.

3.1.3. Validation phase

To validate our system, we employed another set of known point analysis, different from those used in the Learning stage, which would cover as many as possible of the previously instructed classes. The Analysis database for testing, reported in Table 2 and Table S3 (Supplementary Material), is taken from literature (Borghi, 1989; Mellini and Viti, 1994; Rigby et al., 2008; Vaggelli et al., 2009; Lanari et al., 2014;

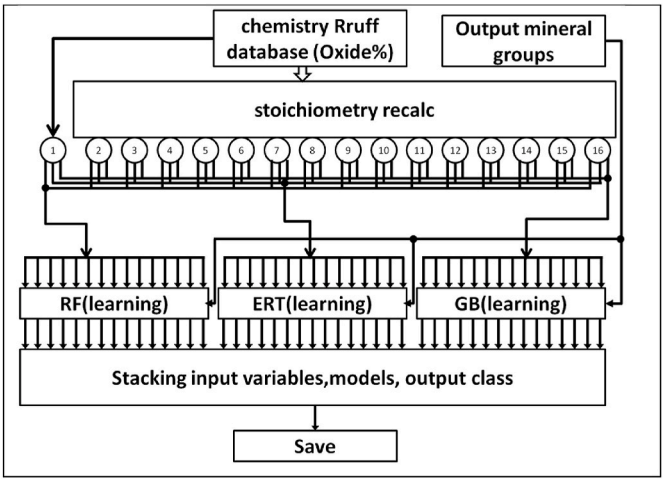


Fig. 1. Learning Flow Chart: 1: Ox. (Wt%), 2 ÷ 16: 15 different stoichiometry calculations (Table 1, Groups column); RF=Random Forest; ERT = Extremely Randomized Trees; GB = Gradient Boosting.

Table 2
Analysis testing database, in brackets (n.of Analyses)(n.of Samples). Analysis source: see text for references.

Analysis (1610)	
Groups (15)	**Minerals (49)**
Amphibole (168)	Act(3)(1)
	Brs(3)(1)
	Ed_Mhb(1)(1)
	Gln(47)(8)
	Gln_Fgl(8)(3)
	Gln_Mrbk(5)(2)
	Ktp(6)(3)
	Mhb(7)(3)
	Mhb_Act(8)(3)
	Prg_Fhb(3)(1)
	Prg_Fts (6)(1)
	Prg_Mhb(12)(2)
	Prg_Ts(3)(1)
	Tr(17)(3)
	Tr_Act(23)(9)
	Tr_Mhb(2)(2)
	Trm(5)(2)
	Wnc(9)(3)
Apatite (6)	Ap(6)(1)
Chlorite (181)	Chm(52)(9)
	Clc(129)(16)
Chloritoid (56)	Cld(56)(5)
Cordierite (10)	Crd(10)(1)
Epidote (52)	Czo(16)(4)
	Ep(36)(7)
Feldspar (146)	Ab(68)(10)
	Or(6)(1)
	Pl(72)(4)
Garnet (167)	Alm(149)(8)
	Sps(18)(2)
Ilmenite (8)	Ilm(8)(2)
Mica (412)	Alcel(201)(16)
	Ann(17)(4)
	FeAlcel(6)(3)
	Mrg(2)(1)
	Ms(155)(20)
	Pg(20)(8)
	Phl(11)(5)
Olivine (64)	Fa(37)(2)
	Fo(27)(1)
Pyroxene (275)	Aeg_Aug(53)(7)
	Aug(72)(8)
	Di(14)(2)
	En(23)(4)
	Jd(6)(1)
	Omp(107)(7)
Serpentine (46)	PHS(46)(2)
Talc (11)	Tlc(11)(2)
Titanite (8)	Ttn(8)(1)

Ketcham, 2015; Ghignone, 2019; Ghignone et al., 2021; Ghignone et al., 2023; Corno, 2023; Corno et al., 2021) it includes 1610 analysis points representing the 15 mineral groups (amphibole, apatite, chlorite, chloritoid, cordierite, epidote, feldspar, garnet, ilmenite, mica, olivine, pyroxene, serpentine, talc, titanite). When the flow chart of Fig. 2 is applied, similarly to the learning phase the 48 ML models are loaded from the disk, the points of the Analysis database (oxides) are stoichiometrically recalculated with the same functions of the Learning phase and applied to the same models already present in memory, obtaining a prediction (mineral group) for each model. These 48 predictions, for each input analysis point, are then combined with the "majority vote" process to obtain the response (mineral group) for each phase. The performance of the three different algorithms used (ERT, RF, GB) is highlighted in Fig. 3a (Mean Vote) and Fig. 3b (Mean Probability), reported in the Supplementary Material. The mean Majority Vote remains essentially the same for the three algorithms when the number of observations is high. Contrarily, GB performance decreases with a decrease of observations (Fig. 3a and Table S4).

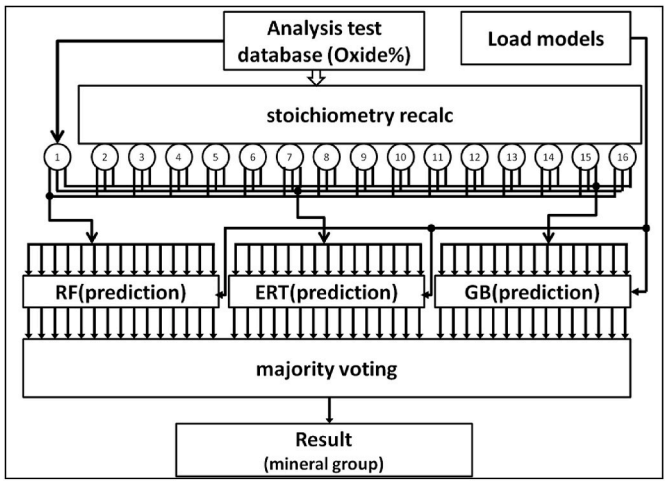


Fig. 2. Prediction flow chart: 1: Ox. (Wt%), 2 ÷ 16: 15 different stoichiometry calculations (Table 1, Groups column); RF=Random Forest; ERT = Extremely Randomized Trees; GB = Gradient Boosting.

Opposite trend was observed for the Mean Probability (Fig. 3b and Table S4): GB mean probability is always higher with major differences with respect to the two other algorithms when the number of observations is low. It would appear that the three algorithms offset each other thus providing a reliable overall determination even in cases of limited statistical observations. Table 3 shows the results of the Validation phase divided by mineral group: the percentage of total error for each mineral group is very low (<1%) and the "majority vote" is quite high, where all mineral groups have a Majority Vote >40 except for amphibole and cordierite. For some groups, the error percentage can increase and/or the majority vote can decrease based on the quantity and/or distribution of analysis points used in the Learning phase, or the quality of the analysis points used in these Validation phase, verified through classification plots.

3.2. *Prediction phase*

In this phase, our model was applied to a completely unknown single analysis point (i.e. a suitable list of values expressed as wt.% oxides), obtaining as an answer the correct mineral group to which it belongs.

3.2.1. *Sample description*

To test the reliability of the ML procedure, a metamorphic rock sample of a meta-ophiolitic eclogite rock (ZAC2; Fig. 4) was collected and selected from the Western Alps. The sample is a petrographic thin section (~30 μm in thickness), polished and coated with graphite, with a surface area of about 600 mm^2. The rock consists of a medium-to coarse-grained, partly re-equilibrated eclogite rock, preserving a classic eclogitic assemblage, overprinted by a quite pervasive re-equilibration in greenschist facies conditions. This sample is suitable to check our procedure because its textural, mineralogical, and chemical features represent a good test of complexity; indeed, this lithology offers a wide mineralogical variety, important differences in grain size, and chemical variability of the phases.

Optical microscope observations (Fig. 4) show that the rock consists of an eclogitic assemblage made of garnet, omphacite (clinopyroxene), glaucophane (Na-amphibole), while the greenschist mineral assemblage is characterized by chlorite, epidote, albite (plagioclase), tremolite (Ca-amphibole).

Garnet is mm-sized, partly fractured and slightly re-equilibrated in chlorite along rims and fractures. Clinopyroxene crystals are cm-sized, partly re-equilibrated at the rim and along the cleavage planes by symplectite textures, made of green amphibole + albite. Amphiboles widely occur in the rock, with different grain sizes and compositions.

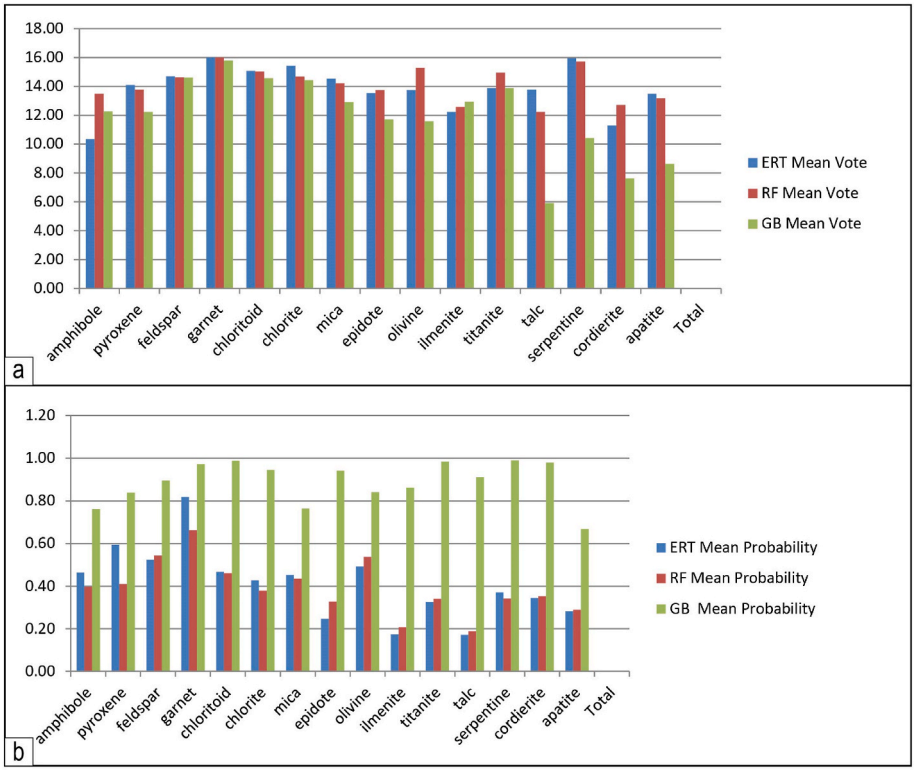


Fig. 3. a) Mean Majority vote diagram for each mineral group. b) Mean Majority probability diagram for each mineral group. RF=Random Forest; ERT = Extremely Randomized Trees; GB = Gradient Boosting.

Table 3
Majority Vote ensemble prediction results ordered by mineral groups.

Group	Number	False Negative	False Positive	MajorityVote (Mean: max = 48)
amphibole	168	0	0	38.08
apatite	6	0	0	43.83
chlorite	181	1	1	47.09
chloritoid	56	0	0	47.91
cordierite	10	0	0	36.40
epidote	52	0	0	41.44
feldspar	146	1	1	44.87
garnet	167	0	2	47.72
ilmenite	8	0	0	43.25
mica	412	2	0	44.12
olivine	64	0	0	46.34
pyroxene	275	0	1	40.87
serpentine	46	0	0	41.26
talc	11	0	0	38.73
titanite	8	0	0	43.38
Total	1610	4	5	

Na-amphibole (glaucophane) occurs in coarse-grained crystals with the typical elongated shape, while Ca-amphibole (and Na-Ca-amphibole) occurs as fine-grained matrix and aggregates.

3.2.2. *Eclogite rock microanalyses*

The rock sample (Fig. 4) was analyzed through an automated acquisition of point analyses via EDS microanalysis, using the Oxford Instrument Aztec Feature ® package.

An example area is shown in Fig. 5, where all recognized features are highlighted on the back-scattered electron image by a white contour lines. Prior to the automatic acquisition, some test frames were acquired in back scattered electrons mode: from these frames, appropriate intervals in the histograms of gray levels (corresponding to average atomic number ranges) were identified and selected with the Oxford Instrument

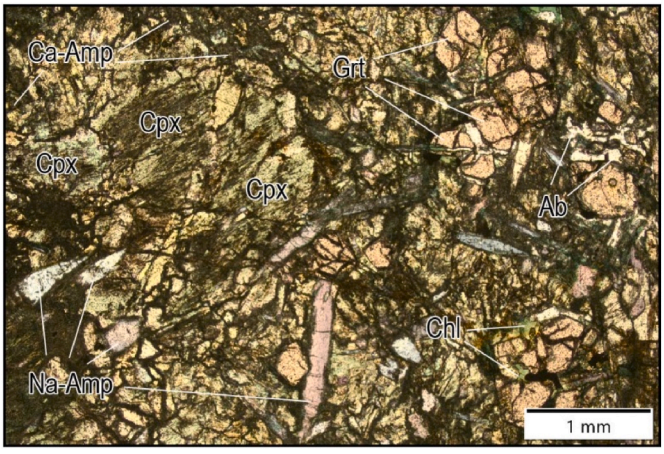


Fig. 4. Optical microscope microphotograph (Plane Polarized Light) of the studied eclogite rock (ZAC2): main texture of the sample and the principal constituents (Mineral abbreviations in the text, figures and tables are from Warr (2021): Ca-amp = calcic amphibole; Na-amp = sodic amphibole; Cpx = pyroxene; Grt = garnet; Chl = chlorite; Ab = Albite (feldspar).

Aztec Feature ® package. That is to effectively contour any "feature" representing individual mineral particles with homogeneous composition and therefore single mineral phase of interest (Fig. 5b and c).

The calculated oxides of each spectrum were then introduced into the ML algorithm to predict the correct group of mineral phases (Fig. 5b and c). The first results of the ML algorithm in the investigated total area of 20 mm^2 is shown in Table 4, where 4193 features were identified, analyzed, and partitioned in the appropriate mineral group previously identified using the ML protocol. The appropriate stoichiometric recalculation was carried out later by applying the MinPlot software and

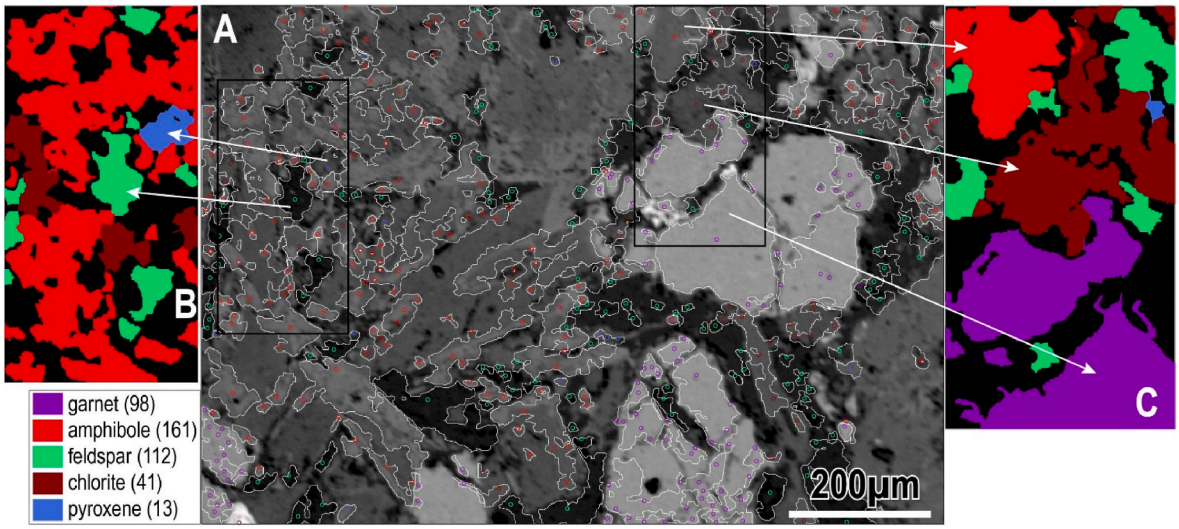


Fig. 5. Example of an investigated area. A) BSE (back-scattered electron) image of the selected area with the identified features (white line). The position of the spot analysis is indicated by the small circles: the features of the same mineral "group" are grouped with the same colour; B) and C) enlarged inlets displaying features in false colours after mineral group attribution. Arrows indicate the corresponding mineral feature. In brackets, the number of the identified and analyzed minerals from the investigated area.

Table 4
Number of analyses automatically performed by Aztec Feature Package and classified by ML algorithm to the correct group of mineral phases.

Group	N.Analysis
garnet	1280
pyroxene	784
titanite	136
amphibole	1341
rutile	103
feldspar	287
chlorite	160
epidote	82
mica	12
ilmenite	4
quartz	4
Total	**4193**

finally plotted on the classification diagrams selected for each individual mineral group.

The identified features were allocated in the main mineral groups as garnet, amphibole, pyroxene (Fig. 5a and b), which represent the main phases of an eclogite rock followed by feldspar, chlorite, titanite, rutile, epidote, and by some accessory minerals (i.e. occurring in less than 15 particles in total). When considering the whole investigated area, the identified features (i.e. the mineral particles with homogeneous compositions) are summarized as reported in Table 4.

All the chemical compositions of the identified groups were then moved into MinPlot, which recalculated the Crystallochemical Formula and plotted the results in the Diagrams of Fig. 6. Stoichiometry output and plotting obtained by MinPlot show the chemical variability of the different phases occurring in the sample (Fig. 6).

Garnet chemical composition exhibits weak zoning, consisting of a feeble chemical variation of the almandine and grossular + spessartine content, amongst the different analyzed features (Fig. 6a).

Clinopyroxene chemical composition shows a weak zonation between Na and Ca (jadeite-QUAD) content, in the omphacite field (Fig. 6b).

The chemical composition of amphiboles shows important variations on the Na, Ca, Al (and slightly on Fe^{3+}) contents. Automated MinPlot-based classification shows the occurrence in the rock sample of tremolite, hornblende, pargasite, sadanagaite and glaucophane (Fig. 5c and d), undetected with optical microscope and SEM-EDS analyses.

The chlorite composition is almost homogeneous, in the clinochlore field, no significant zoning among the different end-members has been observed through different microstructural sites of the sample (Fig. 6e).

Plagioclase crystals show a weak zonation, with cores of almost pure albite and tiny rims recording an enrichment in Ca, in the oligoclase compositional field (Fig. 6f). SEM-EDS analyses of the main mineral phases and their mineral recalculations are reported in Table S5 (Supplementary material).

3.3. *Accuracy and reproducibility of the method*

The proposed analytical protocol allows to characterize 15 main mineral phases of any rock sample, either for geological or for petrological purposes by an automated microanalytical method, which interrogates a wide chemical database with main statistical occurrences.

This method provides a large amount of data on a single rock sample with a relatively low cost and in short times, and therefore may be applied with low facilities to many samples with good accuracy. The cause of any group attribution (i.e. the accuracy of the group attribution) may be calculated for any mineral attribution.

More in details, the reliability of the group allocation is evaluable by the "Majority Vote" procedure, whose maximum value is 48 (corresponding to the number of trained groups multiplied for the three used ML models: 16 x 3). The obtained value is different from mineral-to-mineral group, depending both on the number of the group-analyses available and used in the training phase of the ML, and by the complexity of the Stoichiometric formula calculations. However, all attributions with a Majority Vote $>$ of 24 (i.e. $>$ of the half maximum value) can be exhaustively considered as proper attribution, as shown in Fig. 7 where the Majority Vote for most minerals of the tested eclogite rock is $\geq$ 30 (See also Table S5 in Supplementary Material).

4. Conclusions

The proposed analytical method is recommended for providing all mineral-chemical information of a sample, in reasonable experimental times and with good accuracy. By combing well-known ML algorithms and the MinPlot classification software (Walters, 2022), the proposed procedure allows to characterize any rock sample, either for geological or for petrological purposes. This automated microanalytical method

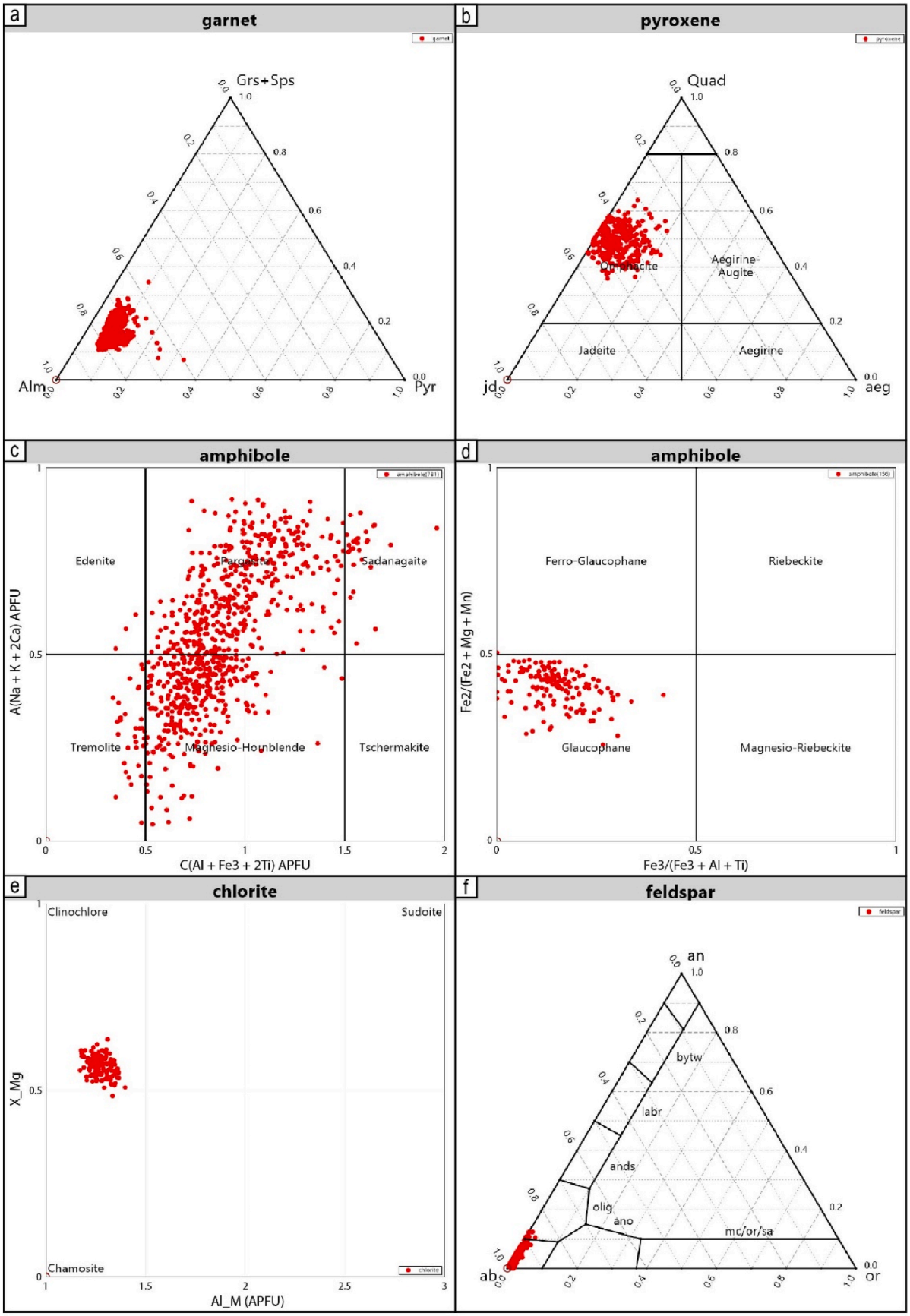


Fig. 6. Stoichiometry output and plotting obtained with MinPlot for the six selected main mineral phases (groups) of the tested eclogite sample, using the analyses classified by the proposed ML algorithm.

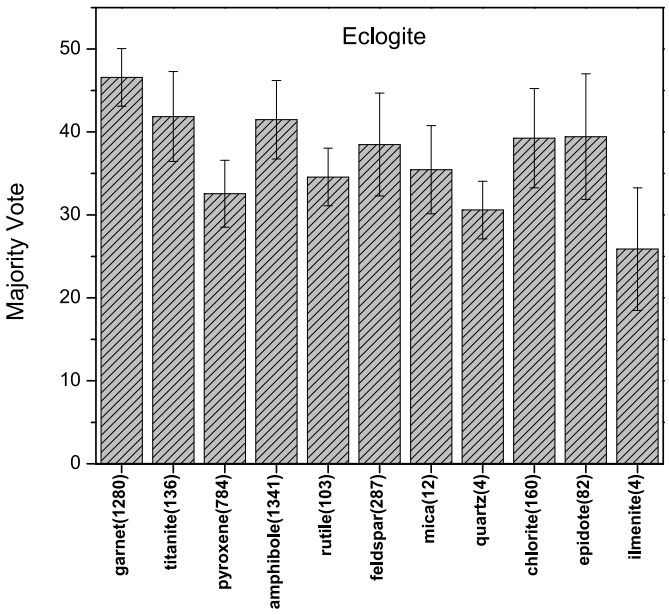


Fig. 7. The average of Majority Vote approach (i.e. the quality of the proper attribution to the group) plotted for any identified mineral group of the tested eclogite rock (standard deviation reported as vertical black line).

permits to obtain a wide chemical database, with main statistical occurrences. Moreover, the reliability of the group attribution is evaluable by the "Majority Vote" approach, being calculated for any mineral attribution. In particular, all attributions with a reliable "Majority Vote" can be exhaustively considered as correct attribution.

The real novelty is achieved combining EPMA measured chemical composition and the stoichiometric recalculations as input data to obtain homogeneous ensembles diversity. To apply this protocol, you need accurate analytical data and an excellent mineral formula recalculation software. We tested Minplot with RRUFF data and it proved to be a reliable recalculation software.

Our procedure was tested on an eclogite rock sample from the Western Alps, which proved to be a challenging task. This kind of rock is characterized by a complex mineralogical assemblage, made of minerals similar to each other in chemical composition and shape, often slightly changing both these features throughout the thin section. If the tested eclogite sample is considered, the identified features (4193 mineral particles with homogeneous compositions) can be summarized as garnet, pyroxene, and amphibole, i.e. the main rock-forming minerals for an eclogite rock with minor and accessory minerals as feldspar, chlorite, titanite, rutile, and epidote. Any chemical composition of the identified groups, expressed as Wt.% oxides, were moved into MinPlot, to recalculate the mineral formula (see Table S5 Supplementary Material). The MinPlot results were directly plotted in the many classification diagrams known in literature, thus better constraining the petrological information for the investigated rock sample.

In conclusion, our test, not only provided remarkable results but also validated the proposed analytical protocol which is recommended when a large flow of microanalytical data needs to be automatically acquired on a single rock and/or when numerous rock samples need to be investigated. Moreover, this approach does not only allow to manage a large database, but it is also suitable when in search of reasonable costs and short experimental times.

Code availability section

Name of the code/library: SharpLearning, MINplot.
Contact: Mark Dabros
Hardware requirements: Any Desktop PC.
Program language: C# (version 0.31.8)
Software required: Microsoft Visual Studio (2017)
Program size: 840 KB.
The source codes are available for downloading at the link: https://github.com/mdabros/SharpLearning.

Name of the code/library: MINPLOT.
Contact: Jesse B. Walters, thegeojesse@gmail.com.
Hardware requirements: Any Desktop PC.
Program language: MATLAB® (2021)
Software required: MATLAB® (2021)
Program size: 104 KB.
The source codes are available for downloading at the link: https://github.com/MinPlot.

Fundings

This research was funded by research Grants CASD_RILO_21_01 from University of Torino, Ricerca Locale "ex 60%" (A.B.)

CRediT authorship contribution statement

R. Cossio: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **S. Ghignone:** Writing – review & editing, Writing – original draft, Validation, Supervision, Formal analysis. **A. Borghi:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **A. Corno:** Writing – review & editing, Supervision. **G. Vaggelli:** Writing – review & editing, Writing – original draft, Validation, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to acknowledge Roberto Compagnoni ("professor emeritus" of University of Turin) for suggestions and encouragements, the Editor (Boyan Brodaric) for the editorial handling and the two anonymous reviewers for their insightful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.acags.2024.100186.

References

Borghi, A., 1989. L'evoluzione metamorfico-strutturale del settore Nord-orientale della Serie dei Laghi (Alpi Meridionali). Univ., Torino, p. 187. PhD Thesis.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Chemistry–RRUFF (https://rruff.info/zipped_data_files/chemistry).
Corno, A., 2023. Tectono-Stratigraphic Setting and Metamorphic Evolution of Ligurian-Piedmont Units in the Upper Susa and Chisone Valleys. Univ. Torino, p. 210. PhD thesis.
Corno, A., Groppo, C., Mosca, P., Borghi, Gattiglio, M., 2021. Eclogitic metamorphism in the Alpine far-west: petrological constraints on the Banchetta-Rognosa tectonic unit. Swiss J. Geosci. 114, 16.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annuals of Statistics 29 (5), 1189–1232.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42.
Ghignone, S., 2019. Structural and Petrological Constraints in the Exhumation Processes of the Western Alpine Meta-Ophiolite Units. Univ., Torino, p. 246. PhD thesis.

Ghignone, S., Borghi, A., Balestro, G., Castelli, D., Gattiglio, M., Groppo, C., 2021. HP tectono-metamorphic evolution of the Internal Piedmont Zone in Susa Valley (Western Alps): new petrologic insight from garnet+ chloritoid-bearing micaschists and Fe–Ti metagabbro. J. Metamorph. Geol. 39 (2), 391–416.
Ghignone, S., Scaramuzzo, E., Bruno, M., Livio, F.A., 2023. A new UHP unit in the western Alps: first occurrence of coesite from the monviso massif (Italy). Am. Mineral. 108 (7), 1368–1375.
Hagni, A.M., 2008. Phase identification, phase quantification, and phase association determinations utilizing automated mineralogy technology. J. Maps 60, 33–37.
Ketcham, R.A., 2015. Technical Note: calculation of stoichiometry from EMP data for apatite and other phases with mixing on monovalent anion sites. Am. Mineral. 100 (7), 1620–1623.
Kunapuli, G., 2023. Ensemble Methods for Machine Learning. Manning Pubns Co, p. 330.
Lanari, P., Vidal, O., De Andrade, V., Dubacq, B., Lewin, E., Schwartz, S., 2014. XMapTools a MATLAB-based graphic user interface for microprobe X-ray images processing. Comput. Geosci. 62, 227–240.
Mellini, M., Viti, C., 1994. Crystal structure of lizardite-l T from Elba, Italy. Am. Mineral. 79, 1194–1198.
Newbury, D.E., Ritchie, N.W.M., 2015. Performing elemental microanalysis with high accuracy and high precision by scanning electron microscopy/silicon drift detector energy-dispersive X-ray spectrometry (SEM/SDD-EDS). J. Mater. Sci. 50, 493–518.
Petrelli, M., 2024. Machine learning in petrology: state-of-the-art and future perspectives. J. Petrol., egae036 https://doi.org/10.1093/petrology/egae036.
Pignolet-Brandom, S., Lapakko, K.A., 1990. In: Petruk, W., et al. (Eds.), Process Mineralogy IX. TMS, Warrendale, PA, pp. 525–531, 1990.
Quinlan, J.R., 1996. Learning decision tree classifiers. ACM Comput. Surv. 28 (1), 71–72.
Reid, A.F., Zuiderwyk, M.A., 1983. QEM*SEM: automated image analysis and stereological applications to mineral processing and ore characterization. Acta Stereol. 2 (1), 205–208.
Rigby, M., Droop, G., Plant, D., Gräser, P., 2008. Electron probe micro-analysis of oxygen in cordierite: potential implications for the analysis of volatiles in minerals. S. Afr. J. Geol. 111, 239–250.
ScottPlot: an Interactive Plotting Library for .NET" (https://scottplot.net/).
SharpLearning, An opensource machine learning library for C# .Net. (https://github.com/mdabros/SharpLearning).
Sutherland, D.N., Wilkie, G., Johnson, C.R., 1989. Mineralogy-Petrology Symposium. The AusIMM Sydney Branch, Sydney, NSW, Australia, pp. 81–84, 1989.
Vaggelli, G., Pellegrini, M., Vougioukalakis, G., Innocenti, S., Francalanci, L., 2009. Highly Sr radiogenic tholeiitic magmas in the latest inter-Plinian activity of Santorini volcano, Greece. J. Geophys. Res. Solid Earth 114, B06201. https://doi.org/10.1029/2008JB005936.
Walters, J.B., 2022. MinPlot: a mineral formula recalculation and plotting program for electron probe microanalysis. Mineralogia 53 (1), 51–66.
Warr, L.N., 2021. Recommended abbreviations for the names of clay minerals and associated phases. Clay Miner. 55 (3), 261–264. https://doi.org/10.1180/clm.2020.30.