Computers & Geosciences 80 (2015) 38–48

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Application of Decision Tree Algorithm for classification and identification of natural minerals using SEM–EDS

Efe Akkaş [a,*], Lutfiye Akin [a], H. Evren Çubukçu [a], Harun Artuner [b]

[a] Hacettepe University, Department of Geological Engineering, Beytepe, Ankara 06800, Turkey
[b] Hacettepe University, Department of Computer Engineering, Beytepe, Ankara 06800, Turkey

ARTICLE INFO

Article history:
Received 11 February 2015
Received in revised form
26 March 2015
Accepted 27 March 2015
Available online 28 March 2015

Keywords:
Mineral classification
Energy dispersive spectrometry
Decision Tree Algorithm
Scanning Electron Microscope
C5.0 Decision Tree

ABSTRACT

A mineral is a natural, homogeneous solid with a definite chemical composition and a highly ordered atomic arrangement. Recently, fast and accurate mineral identification/classification became a necessity. Energy Dispersive X-ray Spectrometers integrated with Scanning Electron Microscopes (SEM) are used to obtain rapid and reliable elemental analysis or chemical characterization of a solid. However, mineral identification is challenging since there is wide range of spectral dataset for natural minerals. The more mineralogical data acquired, time required for classification procedures increases. Moreover, applied instrumental conditions on a SEM–EDS differ for various applications, affecting the produced X-ray patterns even for the same mineral. This study aims to test whether C5.0 Decision Tree is a rapid and reliable method algorithm for classification and identification of various natural magmatic minerals.

Ten distinct mineral groups (olivine, orthopyroxene, clinopyroxene, apatite, amphibole, plagioclase, K-feldspar, zircon, magnetite, biotite) from different igneous rocks have been analyzed on SEM–EDS. 4601 elemental X-ray intensity data have been collected under various instrumental conditions. 2400 elemental data have been used to train and the remaining 2201 data have been tested to identify the minerals. The vast majority of the test data have been classified accurately. Additionally, high accuracy has been reached on the minerals with similar chemical composition, such as olivine ($(Mg,Fe)_2[SiO_4]$) and orthopyroxene ($(Mg,Fe)_2[SiO_6]$). Furthermore, two members from amphibole group (magnesiohastingsite, tschermakite) and two from clinopyroxene group (diopside, hedenbergite) have been accurately identified by the Decision Tree Algorithm. These results demonstrate that C5.0 Decision Tree Algorithm is an efficient method for mineral group classification and the identification of mineral members.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

A mineral is a natural, homogeneous solid having a definite chemical composition and a highly ordered atomic arrangement. Their physical and chemical properties provide a basis for characterizing the minerals and the mineral groups to which they belong. The characteristic physicochemical properties can be utilized for identification of minerals for various purposes focused on mineralogical data. There are numerous analytical methods for the identification and quantification of minerals which utilize their characteristic physicochemical properties. Among all these methods, a Scanning Electron Microscope (SEM) equipped with an Energy Dispersive X-ray Spectrometer (EDS) represents a convenient way to identify a mineral within a wide range of solid specimens.

Among spectrometric methods, SEM–EDS is a fast and a reliable tool for qualitative and quantitative microchemical analyses. SEM provides high-resolution raster data for different applications virtually on any solid. The electron source (or electron gun) in a SEM generates an incident focused electron beam which excites the atoms on the surface of the sample. The "bombardment" of the sample with an electron beam results in the emission of Auger electrons, secondary electrons, backscattered electrons, X-rays and photons (Goldstein et al., 2003; Seibert, 2004). Particles originating from the sample and their released energies are detected by various types of detectors in a typical SEM–EDS. Thus, digital information on the "structure" of the samples can be acquired. EDS "counts" the collected characteristic X-rays produced by electron transitions between energy shells following the removal of an inner shell electron upon excitation by the incident electron beam. The vacancy in the inner shell is filled by the transition of an outer shell electron. During such transition, released X-ray energy is represented by the difference of energy between inner and outer

* Corresponding author. Fax: +90 3122992034.
E-mail address: akkasefe@hacettepe.edu.tr (E. Akkaş).

http://dx.doi.org/10.1016/j.cageo.2015.03.015
0098-3004/© 2015 Elsevier Ltd. All rights reserved.

shell electrons, which are characteristic for each element (Goldstein et al., 2003). A typical EDS data is represented by a simple X-ray spectrum containing characteristic X-ray peaks and Bremsstrahlung (continuum X-rays) plot on a histogram where *y*-axis represents X-ray counts per second (cps)/energy (eV) and *x*-axis represents X-ray energies belong to each channel (keV) (Reed, 2005). For more detailed explanation on measuring peak intensities and calculations, readers should refer to Reed (2005) and Goldstein et al. (2003).

EDS is an analytical instrument used for the elemental analysis or chemical characterization of a solid. Obtaining elemental data by EDS technique is fast, which eases accurate and rapid characterization of minerals. But on the other hand, it is a challenging technique since there is wide range of spectral dataset for natural minerals. The diversity of chemical constituents of mineral species results in the spectral variability of the EDS datasets. Besides, accuracy of an SEM–EDS data can be affected by various factors. The electrical configuration of the instrument (e.g. accelerating voltage, beam current, spot size) during the data acquisition may alter spectral patterns. The physical parameters of the specimen (e.g. topography, electrical conductivity) may differ for each acquisition and analysis. These effects can cause drastic discrepancies in the elemental analysis and/or chemical characterization of even the same mineral sample. Moreover, the characterization of mineral analyses may become a challenging process, considering the effects of working conditions on spectral data. In addition, during mineralogical/petrological studies, recognition of a mineral member is more valuable than the group to which it belongs. Hence, correct identification of minerals according to their EDS data often requires a prior knowledge of the analyte. Once the spectral data obtained, then it can be treated for the calculation of elemental concentrations and may be expressed in terms of either % atomic mass or oxide weights. Upon the calculation of elemental concentrations, distribution of cations in the ideal crystal formulae is used to calculate proportions of end-members of a known mineral group, which are then used to identify the mineral member. This calculation, the basis for the proper identification of a mineral member, strictly requires the mineral group to be known. In other words, correct mineral identification is needed for classifying the mineralogical data correctly.

Although SEM–EDS provides speed and reliability in the analysis practice, considerable amount of time is needed for mineral identification due to the crystal chemical calculations. Regarding the spectral variations in elemental X-ray intensities that are affected by analytical conditions, probability of misidentification would increase on non-standard sample topography. In this study, we employed an automatic classification scheme using characteristic X-ray intensities based on rule induction principle which is independent of the physical working conditions and post-analysis data recalculation.

Numerous attempts to establish an automatic classifier/identifier for the minerals using various chemical, physical and the visual information have led the development of “smart systems” with Artificial Neural Network techniques (ANN) and Genetic Programming (GP). Ruisanchez et al. (1996) has used EDS spectral data with Kohonen ANN methods for automated classification of 12 standard mineral samples. Koujelev et al. (2010) has treated spectral data of Laser-Induced Breakdown Spectroscopy (LIBS) with an ANN method for classification of the several minerals. Thompson et al. (2001) and Baykan and Yılmaz (2010) have applied ANN; Ross et al. (2001) has used GP method for obtaining automated classification of visual information of minerals, acquired using rotating stage of a polarized light microscope. As a commercial success, a fully automated mineral identification platform, Mineral Liberation Analyzer (MLA), integrated with SEM and EDS has been developed by Gu and Napier-Munn (1997). MLA uses both visual grayscale information acquired via Backscatter Electron Detector (BSD) and EDS spectra as X-ray intensities. However, industrial applications of liberation data have been limited because the acquisition of mineral liberation data has been difficult and expensive (Gu, 2003). Among numerous studies on mineral identification processes, no decision tree methods have not been applied and evaluated yet.

Main objective of this study is to test whether Decision Tree Algorithms can be applied for the mineral classification/identification using raw, untreated X-ray intensity data without being affected by physical analytical conditions. In this study, Univariate Decision Tree (DT) algorithms were tested as the main classifier. The fastest and considerably accurate results have been obtained by using the C5.0 Decision Tree Algorithm (Quinlan, 2003).

2. Materials and methods

2.1. Data acquisition using SEM–EDS

Mineral samples in polished thin sections/epoxy mounts of natural magmatic rock specimens have been analyzed at the Electron Microscopy and Microanalysis Laboratory of Hacettepe University, Department of Geological Engineering. Carl-Zeiss EVO 50 EP Scanning Electron Microscope equipped with Bruker AXS X-Flash 3001 Silicon Drift Detector (SDD) Energy Dispersive X-ray Spectrometer (EDS). All samples have been carbon-coated prior to analyzing process. 4601 characteristic X-ray intensities (cps/eV: counts per second per electron volts) of Na, Mg, Al, Si, K, Ca, Ti, Mn, Fe, Zr and P for selected mineral groups were obtained using spot analysis method for 10–100 s. Table 1 shows the number of collected EDS data for each mineral used in this study.

Regarding the fact that the chemical compositions for analyzed minerals can vary according to the nature of the host rock, various types of magmatic rocks have been chosen. Ten mineral groups/species (olivine, orthopyroxene, clinopyroxene, plagioclase, K-feldspar, biotite, amphibole, magnetite, zircon and apatite) from different types of magmatic rocks (e.g. basalt, andesite, rhyolite, granite, gabbro, monzonite, diorite, and syenite) have been selected under either polarized light microscope or SEM. For each mineral, at least five points have been analyzed on different areas of each crystal. Thus, compositional variations of minerals in different rocks and elemental variations in different regions of the same crystals are incorporated into the dataset (see Section 4.2). Such approach is applied in order to facilitate the recognition of minerals having compositional variability by DT algorithm.

During data acquisition processes, the same analytical procedures under various physical conditions have been applied. All analyses have been performed under non-standard, random

Table 1
Number of performed analysis for each mineral groups and abbreviations used.

Mineral groups		
	Collected EDS data	Abreviations
Alkali Feldspar	496	Afs
Amphibole	499	Amp
Apatite	498	Ap
Biotite	400	Bt
Clinopyroxene	498	Cpx
Magnetite	493	Mag
Olivine	447	Ol
Orthopyroxene	413	Opx
Plagioclase	439	Pl
Zircon	418	Zr
Total	4601	

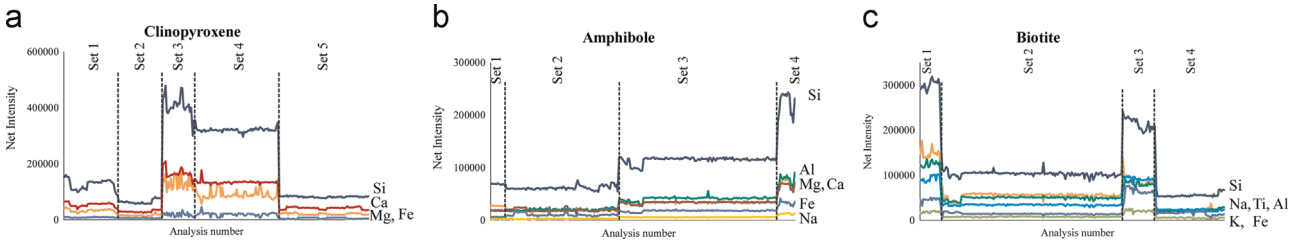

Fig. 1. Representative variation of elemental X-ray intensities (depicted with lines) for different instrumental conditions. (A) clinopyroxene, (B) amphibole and (C) biotite.

analytical conditions since each physical SEM parameter affects the spectral EDS pattern (Fig. 1), and consequently the calculated elemental concentrations for each mineral (see Section 4.1). The rationale behind analyzing under different SEM conditions is to help DT algorithm to characterize each mineral minimizing the physical effects on their chemical data. Besides, this approach reduces the total time for the standardization of SEM conditions. Table 2 shows the applied analytical condition sets during acquisitions for the training data for final DT classification.

2.2. Decision tree

A simple decision tree is a set of procedures in order to classify input training data into more homogenous subgroups using generated rules or decisions called nodes (Friedl and Brodley, 1997; Quinlan, 2003). During the training, DT aims to obtain maximum information and minimum entropy in the generated subsets of the tree (Quinlan, 2003). Fig. 2 shows an example of the workflow and basic parts of decision tree classification for three mineral families. All data have been gathered in a root node and divided into relatively more homogenous subsets called internal nodes (split) using feature values (thresholds). Partitioning/dividing process stops when DT reaches terminal nodes (leaf node). In this final step, labels are defined/assigned in the leaf nodes by allocation way (Pal and Mather, 2003).

The C5.0 Decision Tree Classifier uses the information gain ratio; the metric which tests each node and selects the subdivisions of the data in order to maximize the "Entropy Decrease" of the connected node (Friedl and Brodley, 1997). Thus, the best attributes (features) obtained are used to describe each case belonging to only one separate class (Friedl et al., 1999; Quinlan, 1999). These attributes are tested in the nodes of each split, and partitioning

Table 2
Instrumental condition sets applied during the acquisition of 240 training data for each mineral groups.

Training analyses of mineral groups			Apllied working conditions			
	Condition sets	Number of analyses	Probe current (nA)	Working distance (mm)	Filament current (A)	Accelerating voltage (kV)
Clinopyroxene	1	43	2	12	2300–2600	15
	2	36	1	15	2600–2750	15
	3	24	2	10	2300–2600	20
	4	71	3	10	2300–2600	17
	5	66	2	10	2300–2600	15
Orthopyroxene	1	14	3	15	2600–2750	15
	2	46	4	12	2300–2600	15
	3	180	4	10	2300–2600	15
Olivine	1	27	3	14	2600–2750	20
	2	12	2	10	2300–2600	15
	3	23	3	12	2600–2750	15
	4	178	5	10	2300–2600	15
Plagioclase	1	43	3	12	2600–2750	20
	2	29	1.5	15	2600–2750	15
	3	168	3	10	2300–2600	15
Alkali Feldspar	1	5	3	15	2600–2750	15
	2	70	3	12	2600–2750	20
	3	16	1.5	12	2600–2750	17
	4	28	1.8	15	2600–2750	15
	5	121	1.1	10	2300–2600	15
Amphibole	1	12	5	15	2600–2750	15
	2	26	5	17	2300–2600	15
	3	97	4	17	2300–2600	15
	4	105	2	12	2600–2750	20
Biotite	1	18	2	12	2600–2750	20
	2	142	3	11	2300–2600	15
	3	26	3	12	2600–2750	17
	4	54	1.5	12	2600–2750	15
Apatite	1	38	3	10	2600–2750	20
	2	41	1	12	2600–2750	15
	3	161	2	10	2300–2600	15
Magnetite	1	54	3	10	2600–2750	15
	2	29	3	12	2600–2750	17
	3	137	4	10	2300–2600	15
	4	20	2	10	2600–2750	20
Zircon	1	8	5	10	2600–2750	15
	2	172	5	10	2300–2600	15
	3	60	3	10	2300–2600	15

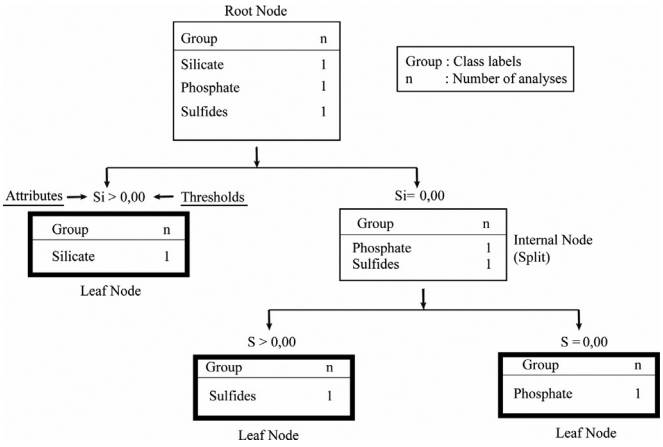

Fig. 2. Typical workflow of a simple Decision Tree Algorithm for the classification of three distinct mineral families (phosphates, silicates, sulfides).

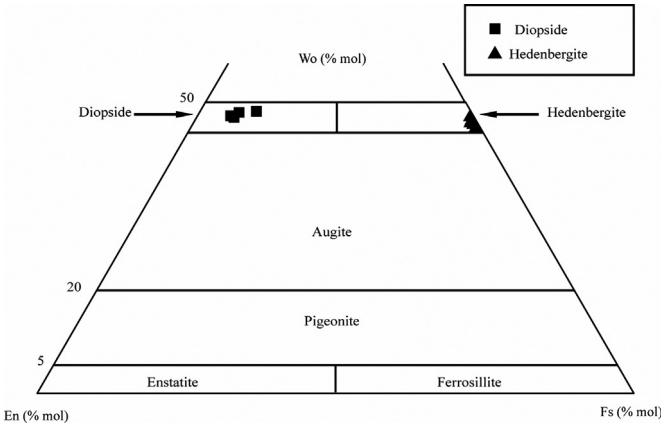

Fig. 3. Wo–En–Fs ternary diagram for nomenclature of Ca–Mg–Fe pyroxenes according to Morimoto (1988).

processes are continued until reaching to a terminal (leaf) node downwards the tree (Pal and Mather, 2003). For this study, C5.0 Decision Tree Algorithm, the successor of C4.5 algorithm produced by Quinlan (2003), has been used for the classification/identification of mineralogical EDS data. For a detailed description of the C5.0 Decision Tree, the reader should refer to Quinlan (2003).

The results obtained by the training procedure may perform poorly due to errors in the dataset, which generate noise producing false splits. In this case, these splits can lead to inaccurate results, especially for the new data (Brown de Colstoun, 2003; Friedl et al., 1999; Quinlan, 2003; Weiss and Kulikowski, 1991). In order to reduce the errors during processing the new data, DT has to be pruned back. This is the most important step in all of the decision tree methods (Friedl and Brodley, 1997). There are several methods for pruning, based on the separation of the noise from the leaf nodes and also from the tree branches. In C5.0, pruning back starts automatically on leaf nodes and spreads upward to the whole tree using the information gain ratio. Further details on the pruning methods are described in the following sources: Breiman et al. (1984), Mingers (1989) and Quinlan (1999)).

Freund and Schapire (1996)'s Boosting Method is another way of increasing the predictive capability of C5.0. The Boosting Method focuses on the "weak" classifier within the learning algorithm. In this method, equal weights are set for each event of the training data based on the first tree produced. After setting the first tree, all weights are adjusted according to the misclassified events of the training data. Therefore, the next trial weights are changed proportionally according to the error for each event, which provides a model focused on the complicated classification. Each new classification tree generated by those trials, contains correction data against misclassification (Freund and Schapire, 1996; Quinlan, 1996).

2.3. Mineral classification

Minerals have well-defined chemical composition and crystal structures; thus these properties provide subdivisions into families, groups and subgroups. Upon the acquisition of the elemental data, a reliable distinction between the members of a mineral group can be established by calculating the chemical formula. Further detail about calculation of mineral formulae have been explained by Deer et al. (2013). Regarding the major element compositions and recalculation of a mineral formula, silicon drift detector (SDD)–EDS produces comparable results with that of a WDS (Çubukçu et al., 2008). Calculating the chemical formula of a mineral is based on the ideal proportions of its structural anions and cations. The number of cations in the mineral composition should be balanced by an appropriate number of oxygen anions. The number of balancing oxygen anions is different for virtually most of the mineral groups, depending on their crystal structure. Therefore, prior to the calculation of a mineral formula, identification of the mineral/mineral groups is compulsory. However, during SEM–EDS studies on natural rocks, coming across an "unknown" mineral with an indistinguishable EDS spectrum is not infrequent.

In order to provide a basis for efficient mineral identification using X-ray intensity data, we employed C5.0 Decision Tree Algorithm on 4601 spectral data acquired from known mineral sets. For the purpose of generating reliable training data for the testing stage, single-algorithm filtering, Brodley and Friedl (1999) has been applied on the identified mineral EDS datasets prior to the final classification. Training and testing blocks were generated for final classification, after this filtering process.

In order to identify mineral species within chemically complex mineral groups (i.e. amphibole and pyroxene in the dataset) further classification procedure has been applied after the final classification, which identified the mineral groups only. Thus, two members (magnesiohastingsite and tschermakite) from amphibole group and two members (diopside and hedenbergite) from clinopyroxene group have been selected, where 50% their EDS data has been used for training. Mentioned members of each group have been named according to their chemical formulae using appropriate nomenclature. Cationic distribution of pyroxene members have been calculated for 6 oxygen per formula units (Table 3) and molar concentrations of three end-member compositions (wollastonite (Wo; $Ca_2Si_2O_6$), enstatite (En; $Mg_2Si_2O_6$) and ferrosillite (Fs; $Fe_2Si_2O_6$)) in each analysis have been plotted on Ca–Mg–Fe ternary diagram according to Morimoto (1988). Hence, these pyroxene minerals are named as hedenbergite and diopside (Fig. 3). Cationic distribution in amphibole members have been calculated for 23 oxygen units and the mineral members have been named as magnesiohastingsite and tschermakite according to Leake et al. (2004) (Table 4).

3. Results

3.1. Filtering process

Efficiency of single-algorithm filtering used by decision trees (C4.5, C5.0) has been demonstrated by several studies (Brodley and Friedl, 1999; Brown de Colstoun, 2003). In order to filter X-ray intensity spectral dataset, 10-fold cross-validation based on C5.0

Table 3
Major oxide concentrations and calculated cationic proportions 10 clinopyroxene analyses.

Analysis no.										
	I13-100-A-5	I13-100-A-14	I13-100-A-15	I13-100-A-17 m	I13-100-A-63 m	N256-12	N256-4	N256-6	N256-7	N256-25
Na_2O	0.14	0.16	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MgO	14.83	16.27	16.27	16.84	16.71	0.00	0.00	0.00	0.00	0.00
K_2O	0.01	0.00	0.00	0.00	0.00	0.17	0.21	0.25	0.19	0.20
CaO	24.22	24.67	24.67	24.26	24.14	20.54	20.98	20.83	20.93	20.96
TiO_2	0.38	0.53	0.53	0.23	0.07	0.31	0.05	0.25	0.43	0.36
Cr_2O_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FeO	5.79	4.12	4.12	3.42	3.86	31.11	30.88	29.69	30.01	30.27
MnO	0.00	0.00	0.00	0.00	0.00	0.85	0.64	0.68	0.48	0.58
NiO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SiO_2	50.59	51.38	51.38	52.92	53.62	46.86	47.00	48.22	47.17	46.89
Al_2O_3	3.16	2.54	2.54	2.07	1.93	0.00	0.00	0.00	0.00	0.00
Total	99.12	99.66	99.66	99.74	100.33	99.84	99.76	99.94	99.23	99.25
Cations based on 6 oxygens										
Si	1.88	1.88	1.88	1.93	1.95	1.95	1.96	2.00	1.97	1.96
Al	0.14	0.11	0.11	0.09	0.08	0.00	0.00	0.00	0.00	0.00
Fe_t	0.18	0.13	0.13	0.11	0.12	1.08	1.07	1.03	1.05	1.06
Mg	0.82	0.89	0.89	0.92	0.91	0.00	0.00	0.00	0.00	0.00
Ca	0.96	0.97	0.97	0.95	0.94	0.92	0.94	0.93	0.94	0.94
Na	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
K	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
Ti	0.01	0.02	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.01
Mn	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.02	0.02	0.02
Tot_Cat	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
Fe^3	0.10	0.11	0.11	0.03	0.01	0.08	0.10	0.00	0.04	0.07
Fe^2	0.09	0.02	0.02	0.07	0.11	1.00	0.98	1.03	1.01	0.99
Fe_2O_3	3.40	4.00	4.00	1.19	0.44	2.65	3.04	0.00	1.24	2.19
FeO	2.73	0.52	0.52	2.35	3.47	28.73	28.15	29.69	28.90	28.30
Al $^{(IV)}$	0.12	0.11	0.11	0.07	0.05	0.00	0.00	0.00	0.00	0.00
Al $^{(VI)}$	0.02	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00
Jd	1.62	0.00	0.00	2.21	3.41	0.00	0.00	0.00	0.00	0.00
Ae	9.61	10.36	10.36	3.32	1.20	3.65	5.23	0.05	1.14	2.99
Aug	88.78	89.64	89.64	94.47	95.39	96.35	94.77	99.95	98.86	97.01
Wo (%)	49.06	48.84	48.84	48.17	47.89	45.15	46.01	46.77	46.81	46.55
En (%)	41.77	44.80	44.80	46.50	46.11	0.00	0.00	0.00	0.00	0.00
Fs (%)	9.17	6.36	6.36	5.32	6.01	54.85	53.99	53.23	53.19	53.45
Mineral names	*Diopside*					*Hedenbergite*				

has been employed. Filtering has been applied to EDS data of predefined minerals. The rationale behind the filtering process is to determine misidentified data, so that a better training dataset can be constructed. Ten-fold cross validation technique requires a well-defined/labelled data to test the validity of each subset. For this reason, this process does not applicable for undefined/unlabelled datasets. For filtering, all datasets have been selected randomly and have been divided into 10 equal-sized subsets. For each run, one of the subsets has been used for testing and the remaining nine datasets have been used for training. After 10 runs, each tree has been evaluated using independent training and test subsets. As a result of cross-validation, only 8 of 4601 spectra have been misclassified. Table 5 shows the results of user-identified mineral groups and produced classification with misclassified data after 10-fold cross-validation process by C5.0. The determined 8 misclassification data were re-evaluated after the filtering process, but no subjective misclassification error has been identified. Therefore, in order to obtain the best results or to prevent probable noise in the training data, mentioned 8 EDS data have been discarded from the dataset before the selection of reference dataset for final classification.

3.2. Classification into mineral groups

Upon filtering process, remaining 4593 X-ray intensity spectral have been divided into “training” and “test” datasets. The effects of analytical working conditions and chemical variation of the minerals have been considered for the selection of the training data (Fig. 1). A total of 2400 characteristic spectral data in 10 sets of groups with 240 analyses, have been selected as the training dataset (Table 1). Remaining 2193 X-ray intensity data have been used for testing and classified using a boosted decision tree. Table 6 shows the confusion matrix of boosted decision tree. After the final classification, 4 analyses in the test dataset were misclassified and remaining 2189 X-ray intensity data were classified correctly. Specified misclassifications and the reference analyses are shown in Table 7. The measured elemental intensities might be lower or higher than the reference intensities due to the aforementioned effects. Orthopyroxene is misclassified as amphibole in the test data, due to considerably high calcium content acquired during the analysis of orthopyroxene. Generally the Ca intensity of an orthopyroxene lies between the ranges of 4000–7000 cps/eV. Considerably high Ca X-ray intensity (16,580 cps/eV) might be the result of either the impurity of the surface of the crystal or coexistence of a Ca-bearing mineral adjacent to the beam spot. Besides these, the determined two misclassifications between the olivine–orthopyroxene testing blocks seem to be caused by the subjective errors. During the evaluation of the two misclassified results within the orthopyroxene and olivine test blocks, it was recognized that the results and the reference elemental intensities were matching. In the final boosted decision tree, considerable high accuracy of the 2400 training data against 2193 test data were achieved after 10 trials. This demonstrates the effectiveness of the boosting technique.

Table 4
Major oxides concentrations and calculated cationic proportions 10 amphibole analyses.

	Analysis no.									
	I13-013-1	I13-013-8	I13-013-16	I13-025-5	I13-025-13	Amp-exp-3-6	Amp-exp-3-9	Amp-exp-3-12	Amp-exp-3-14	Amp-exp-3-36
SiO_2	39.99	41.56	38.87	38.20	38.17	45.64	43.38	44.92	44.62	45.04
TiO_2	2.81	1.98	2.37	2.14	2.79	1.77	1.48	1.98	1.69	1.96
Al_2O_3	11.39	13.03	11.41	11.65	12.79	9.19	9.16	9.62	9.58	10.82
MnO	0.11	0.44	0.39	0.40	0.18	0.19	0.52	0.00	0.39	0.05
FeO	16.32	16.93	16.85	16.66	18.77	13.82	14.25	12.02	12.58	10.01
MgO	10.25	11.25	10.76	10.91	9.18	14.80	13.29	14.88	14.67	15.52
CaO	11.56	10.23	10.67	11.24	10.93	11.22	11.05	10.95	10.59	10.42
Na_2O	1.62	2.29	2.13	2.01	1.86	1.46	1.91	1.63	1.76	1.81
K_2O	1.78	1.51	1.72	1.87	1.91	0.62	0.58	0.47	0.47	0.43
Total	95.83	99.20	95.16	95.07	96.60	98.71	95.61	96.46	96.36	96.05
Cations based on 23 oxygens										
Si	6.20	6.18	6.10	6.02	5.95	6.65	6.59	6.64	6.63	6.60
Ti	0.33	0.22	0.28	0.25	0.33	0.19	0.17	0.22	0.19	0.22
Al	2.08	2.28	2.11	2.16	2.35	1.58	1.64	1.68	1.68	1.87
Mn^2	0.01	0.05	0.05	0.05	0.02	0.02	0.07	0.00	0.05	0.01
Fe^2	2.11	2.11	2.21	2.20	2.45	1.68	1.81	1.48	1.56	1.23
Mg	2.37	2.50	2.52	2.56	2.13	3.22	3.01	3.28	3.25	3.39
Ca	1.92	1.63	1.79	1.90	1.83	1.75	1.80	1.73	1.69	1.64
Na	0.49	0.66	0.65	0.61	0.56	0.41	0.56	0.47	0.51	0.51
K	0.35	0.29	0.34	0.38	0.38	0.12	0.11	0.09	0.09	0.08
Sum	15.86	15.93	16.06	16.14	16.01	15.63	15.76	15.58	15.64	15.54
Si	6.15	6.02	5.98	5.91	5.85	6.48	6.45	6.49	6.45	6.45
Al^{iv}	1.85	1.98	2.02	2.09	2.15	1.52	1.55	1.51	1.55	1.55
SumT	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00
Al^{vi}	0.21	0.25	0.04	0.03	0.15	0.02	0.05	0.13	0.08	0.27
Ti	0.32	0.22	0.27	0.25	0.32	0.19	0.17	0.21	0.18	0.21
Fe^3	0.34	1.17	0.93	0.85	0.83	1.17	0.97	1.00	1.20	1.06
Mg	2.35	2.43	2.47	2.51	2.10	3.13	2.95	3.20	3.16	3.31
Fe^2	1.75	0.88	1.24	1.30	1.58	0.47	0.80	0.46	0.32	0.14
Mn^2	0.01	0.05	0.05	0.05	0.02	0.02	0.07	0.00	0.05	0.01
SumC	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Ca	1.90	1.59	1.76	1.86	1.79	1.71	1.76	1.69	1.64	1.60
Na	0.10	0.41	0.24	0.14	0.21	0.29	0.24	0.31	0.36	0.40
SumB	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Na	0.39	0.23	0.39	0.46	0.35	0.11	0.31	0.15	0.14	0.10
K	0.35	0.28	0.34	0.37	0.37	0.11	0.11	0.09	0.09	0.08
SumA	0.74	0.51	0.73	0.83	0.72	0.22	0.42	0.24	0.22	0.18
Mineral names	*Magnesiohastingsite*					*Tschermakite*				

3.3. Identification of mineral members

For the classification of mineral groups, 100 X-ray intensities (25 intensities for each of magnesiohastingsite, tschermakite, diopside and hedenbergite) have been used for filtering process. After running 10-fold cross-validation process according to the generated confusion matrix one magnesiohastingsite misclassification was found within the tschermakite data. The analysis was re-evaluated in the cationic proportion process and identified as the magnesiohastingsite within calcic amphibole classification diagram. The cause of the misclassification was the occurring of the slight chemical variations between magnesiohastingsite and tschermakite compositions within the same crystal structure. On the contrary, compared to the group characterization, the small chemical differences may lead to considerable differences for the characterization of mineral members. This result indicated that all analyses used for training the member characterization, should also be subjected to the cationic calculations.

After the filtering, the site intensity for the magnesiohastingsite has been replaced within its member training site. For the final

Table 5
Confusion matrix after 10-fold cross-validation filtering procedures.

	Afs	**Amp**	**Ap**	**Bt**	**Cpx**	**Mag**	**Ol**	**Opx**	**Pl**	**Zr**	**Total misclassifications**
Afs	494	0	0	0	0	0	0	0	2	0	2
Amp	0	496	0	1	2	0	0	0	0	0	3
Ap	0	0	498	0	0	0	0	0	0	0	0
Bt	1	0	0	399	0	0	0	0	0	0	1
Cpx	0	0	0	0	498	0	0	0	0	0	0
Mag	0	0	0	0	0	493	0	0	0	0	0
Ol	0	0	0	0	0	0	445	2	0	0	2
Opx	0	0	0	0	0	0	0	413	0	0	0
Pl	0	0	0	0	0	0	0	0	439	0	0
Zr	0	0	0	0	0	0	0	0	0	418	0

Note: The classified names have been shown in the columns and user identified names have been shown in the rows.

Table 6
Confusion matrix of boosted decision tree for final mineral group classification used with 2193 test data.

	Afs	Amp	Ap	Bt	Cpx	Mag	Ol	Opx	Pl	Zr	Total misclassifications
Afs	252	0	0	0	0	0	0	0	1	0	1
Amp	0	256	0	0	0	0	0	0	0	0	0
Ap	0	0	258	0	0	0	0	0	0	0	0
Bt	0	0	0	159	0	0	0	0	0	0	0
Cpx	0	0	0	0	258	0	0	0	0	0	0
Mag	0	0	0	0	0	253	0	0	0	0	0
Ol	0	0	0	0	0	0	204	1	0	0	1
Opx	0	1	0	0	0	0	1	171	0	0	2
Pl	0	0	0	0	0	0	0	0	200	0	0
Zr	0	0	0	0	0	0	0	0	0	178	0

Note: The classified names have been shown in the columns and user identified names have been shown in the rows.

Table 7
Reference and misclassified analyses of final group classification.

Testing block		X-ray intensity										Final labelling
		Na	Mg	Al	Si	K	Ca	Ti	Mn	Fe	O	
Afs	Reference	19,998	1	59,371	156,891	13,945	2986	271	1	295	63,297	Afs
	Misclassified	9969	1	28,925	78,141	6768	1583	85	1	176	32,689	Pl
Opx	Reference	1	84,349	4448	149,513	696	5637	690	298	18,917	79,068	Opx
	Misclassified	1	69,700	5452	148,494	818	16,746	1436	547	19,355	64,799	Amp
	Misclassified	1	59,948	2135	157,332	323	7670	1291	1117	31,858	58,989	Ol
Ol	Reference	1	106,477	1	102,670	252	744	1	865	32,486	93,980	Ol
	Misclassified	605	70,438	73	142,948	625	1142	299	334	23,414	80,109	Opx

Table 8
Confusion matrix of final boosted decision tree for mineral member identification for 40 test data.

	Diopside	Hedenbergite	Magnesiohastingsite	Tschermakite
Diopside	10	0	0	0
Hedenbergite	0	10	0	0
Magnesiohastingsite	0	0	11	0
Tschermakite	0	0	0	9

Note: The classified names have been shown in the columns and user identified names have been shown in the rows.

boosted decision tree; 40 test data were classified using 60 training data (15 intensities per member). Generated tree and confusion matrix have demonstrated that the efficiency of C5.0 algorithm was achieved within 100% accuracy level for 40 testing intensities (see Table 8).

4. Discussion

4.1. Effects of working conditions on EDS analyses

Even though the "standardless" analysis is one of the most important advantages of EDS instruments; Statham (2002) measured that, the X-ray energies and intensities of the same mineral, even at the same point, might show discrepancies due to effects of different working conditions of SEM. Maintaining standard working conditions and analytical standards on SEM–EDS is a challenging process considering the diversity of solid analytes. Additionally, special analyzing procedures may be required for some mineral groups such as amphibole, biotite and feldspar. For example; due to the migration of alkalis (Na, K), feldspars are sensitive for long measurement durations, high probe currents and accelerating voltages. Similarly, the alkaline and OH contents of amphibole and mica minerals are also affected by high probe currents and prolonged analyses. Therefore, different working conditions should be applied as per the type of the studies and minerals.

For electron microanalysis; analytical working conditions depend on electrical/physical parameters (accelerating voltage, filament/beam/probe currents, chamber vacuum), sample features (structure, working distance/geometry, surface topography) and the duration of analysis. These are the most important parameters on the acquired X-ray intensity patterns. The effects of measurement duration, chamber vacuum, electrical conductivity or surface topography of the sample on the quality of analyses are not included in this study. However, it must be noted that; qualitative and quantitative EDS measurements are affected by all of these analytical conditions.

During all the analyses, we kept the "emission current", the leaving part of the electron beam from the gun, constant as 100 μA. The smaller portion of beam current passes through the anode, the smaller it becomes passing through the column apertures leading to insufficient energy (Goldstein et al., 2003). For the electron microprobe analysis, an accelerating voltage between 15 kV and 25 kV is generally accepted. Thus, sufficient voltage and current are obtained to excite the atoms having atomic numbers as much as $Z=83$ (Small, 2002). The accelerating voltages during the analyses for this study were between 15 kV and 20 kV.

Table 9
Elemental and X-ray intensity variations for 3 analyses from one orthopyroxene crystal under different filament currents.

Analysis no.	Working conditions				X-ray intensity								Oxide values							
	Filament current (A)	Probe current (nA)	EHT target (kV)	Working distance (mm)	Mg	Al	Si	K	Ca	Ti	Mn	Fe	O	MgO	Al_2O_3	SiO_2	K_2O	CaO	TiO_2	MnO	Fe_2O_3
opx – 1	2483	2.5	15	10	44,516	3015	77,788	589	2888	474	210	9043	41,789	24.8	1.9	52.7	0.3	1.9	0.5	0.3	17.6
opx – 3	2713	2.5	15	10	43,658	2886	76,587	231	2568	270	188	9164	40,847	24.5	1.8	51.6	0.1	1.7	0.3	0.3	19.6
opx – 5	2713	2	15	10	44,287	2523	77,024	99	2548	362	208	8808	41,575	24.6	1.6	51.2	0.1	1.7	0.4	0.3	20.1

For X-ray detection, the working distance between the sample and the detector should also be optimum; neither too close nor too far away (Reed, 2005). The working distances during the analyses have varied between 9 and 17 mm. Probe current of the focused electron beam provides emission of backscattered electrons, secondary electrons and X-rays from a specimen (Sun-Jong and Chan-Hong, 2008). For microanalysis in a thermionic system, Goldstein et al. (2003) have defined the efficient probe current as 10 nA. However, Si-drift X-ray detectors have high X-ray collection efficiency which are convenient even at 1 nA probe current (Ware and Reed, 1973). The ranges of probe current values in this study have changed between 1.1 nA and 5 nA. Thus, sufficient amount of X-ray energy and backscattered electrons have been emitted for the analyses and for visualization. The duration of analyses have been kept constant at 40 s.

Respective working conditions applied (condition sets) during analyses have been given in Table 2. Fig. 1 depicts the observed elemental concentrations of training sets of three mineral group, which were affected by condition sets as depicted in Table 2.

All of the working conditions have been changed randomly by the analyst throughout data acquisition. However, the filament heating current and the beam current have been kept constant throughout the study. Saturation of the filament heating current means that electrons are concentrated in a tight bundle (Goldstein et al., 2003). Saturation value mostly affects the lifetime of the filament and can be changed by the user during the analyses. The effects of varying filament currents for the three analysis of one orthopyroxene are depicted in Table 9. Especially, the oxide values calculated using X-ray energies exhibit significant difference than the X-ray intensity values.

Arranging standard physical/electrical working conditions on SEM requires considerably difficult preparation procedures. Both uncontrolled (filament life, chamber vacuum etc.) and user controlled conditions (beam current, accelerating voltage etc.) lead to change either X-ray intensities for each element or major oxides in mineral formula. In order to either determine spectral effects or testing the effectiveness of C5.0 DT method, different conditional sets are used/arranged during different analyzing period for each mineral group (Table 2). Although, all differences in collected mineral chemical dataset, DT method obtains high accuracy on classification of mentioned mineral groups and identification of end-members. Effectiveness of DT method for classifying on nonlinear spectral or intensity dataset has been demonstrated to be qualified for rapid characterization of minerals. Moreover, no standardization of analytical conditions is required for the accurate output of DT method.

4.2. *Effects of variation in mineral chemistry*

Most of the natural crystals exhibit chemical variations within their structure. Elemental X-ray intensity spectral patterns exhibit spatial variation in the crystal as a result of ionic substitution or solid solution. These differences occur during crystallization or after the crystallization process (Klein et al., 1993). As a result of ionic substitution in a mineral, different compositional zones can occur in the crystal. Especially magmatic solid solution minerals such as plagioclase, pyroxenes and olivine show compositional zoning since the crystals react continuously with the surrounding melt or liquid during crystallization or cooling of magma (Vernon, 2004). In some cases this zoning is generated by the changing physical conditions (e.g. pressure or temperature fluctuations in the magma) during crystal growth (Humphreys et al., 2006; Streck, 2008). Spatial chemical variation in a crystal often makes it difficult for the analyst to characterize mineral groups or members at the first glance. If these chemical variations had not been recognized, erroneous misclassification would occur. Subjective

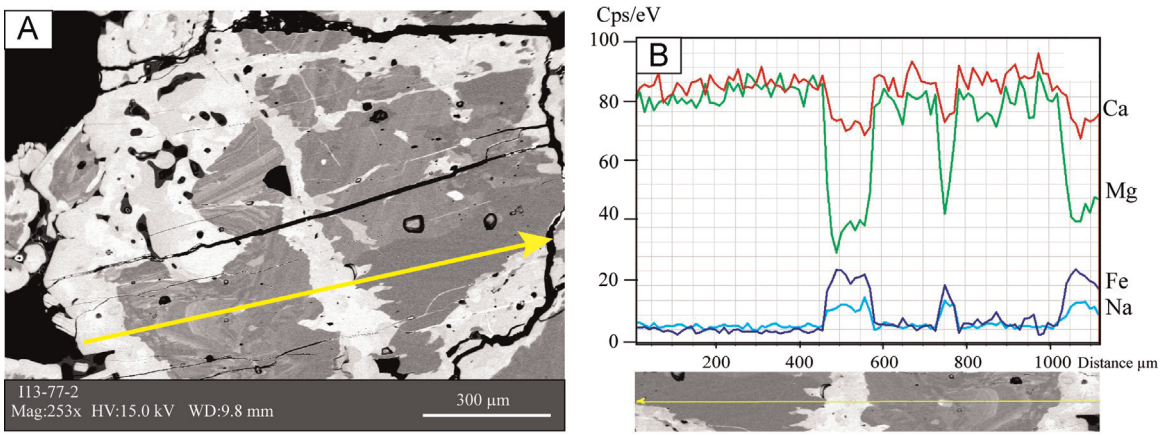

Fig. 4. Elemental variation along a profile within a compositionally zoned clinopyroxene crystal. (A) Electron backscatter detector image on SEM of clinopyroxene crystal (analysis profile is given as an arrow). (B) Cps/eV (X-ray counts per electron volts) variations of Ca, Mg, Fe and Na along the profile.

misclassification of two olivines, one orthopyroxene (in Section 3.2) and one magnesiohastingsite in tschermakite test blocks (in Section 3.3) are due to the chemical compositions between mineral groups and the members. Significant chemical variation in a clinopyroxene mineral is exemplified in Fig. 4, which indicates that varying elemental concentrations of Na, Mg, Ca and Fe in the clinopyroxene crystal cause respective X-ray energies (Cps/eV) to reflect the effects of compositional zoning.

Additionally, a solid solution mineral exhibits spatial compositional difference. For instance, olivine group is a complete solid solution series between Mg-rich forsterite (Mg_2SiO_4) and Fe-rich fayalite (Fe_2SiO_4). Another solid solution is orthopyroxene group formed between enstatite ($MgSiO_3$) and ferrosillite ($FeSiO_3$). Elemental intensities (Table 10) and X-ray spectra (Fig. 5) which belong to the two olivine, and two orthopyroxene groups have been obtained under nearly identical working conditions. These two orthopyroxene and olivine analysis from different type of rocks have shown significant chemical discrepancy in the between Mg and Fe concentrations. These examples emphasize the chemical similarity between members and mineral groups.

In order to obtain various compositional dataset for each mineral group, different type of rock samples have been selected. Elemental similarities and varieties between same group minerals have been considered during data collection for either referencing/training or testing the chemical dataset. Obtained results have shown that DT method is considerably successful for the classification and identification of minerals having compositional varieties in similar group. The results have been demonstrated that DT method might be applicable for classifying minerals without being affected by compositional variations.

4.3. Accuracy of Decision Tree Algorithm on EDS spectra for mineral identification

Decision tree has several advantages over similar supervised classification procedures. These advantages have been proved especially in the remote sensing studies (e.g. Friedl et al., 1999; Brown de Colstoun, 2003; Pal and Mather, 2003).

Perhaps the most protruding advantages of decision trees are; being fast, providing flexibility over datasets and they are easy to interpret. Benefits of boosting, pruning and various filtering techniques on C5.0 Decision Tree are mentioned especially in image analysis for remote sensing studies. Although such advantages can be gained from available methods, the accuracy of supervised classifiers can be affected by the quality and the size of the training data sets. The quality of training data depends on the capability of representation of the test data. In brief, the quality of the characteristic data that were used to train has significant impacts on the accuracy of the classification (Pal and Mather, 2003). Similarly; the effects of quality, size and dimensions of the training data should contain adequate information to represent all classified data.

Achieving quality training data set was challenging considering the chemical variations for 10 mineral groups. As indicated previously in Section 3.1, in order to obtain best quality on the training data, a single filtering algorithm technique was applied. Consequently, 240 selected reference training intensities from each group for 10 mineral groups provided 99.77% accuracy in the final boosted decision tree classification. Achieving this accuracy implies that the training data can be accepted as representative. In order to determine the adequacy of training data, and to evaluate the effects of the dimension of the training set, five different subsets of training data have been generated by partitioning previous training data: Equal numbers of intensities from each mineral group for 10 groups (40, 80, 120, 160, 200), and total number of intensities for five subsets (400, 800, 1200, 1600, 2000) were used to train five distinct boosted and un-boosted decision trees. For each classification, the same 2193 intensities were used for the test groups.

Fig. 6 shows achieved accuracies for six distinct boosted and un-boosted decision trees when six different-sized training sets have been applied. The boosting techniques have shown slightly

Table 10
Chemical variations depicted as X-ray intensities of representative olivine and orthopyroxene analyses under nearly identical instrumental conditions.

Mineral and analysis no.											
	Mg	Al	Si	K	Ca	Ti	Mn	Fe	O	Working distance (mm)	Probe current–spot size (nA)
Olivine-381	73,062	1	100,818	316	1482	409	1370	45,492	95,721	11	4
Olivine-141	110,036	1	103,427	808	1526	560	670	30,454	88,503	10	5
Orthopyroxene-15	91,170	3193	156,428	1	5321	431	531	17,534	78,197	11	4
Orthopyroxene-274	68,373	2584	151,858	978	6990	1256	428	25,945	75,917	10	4

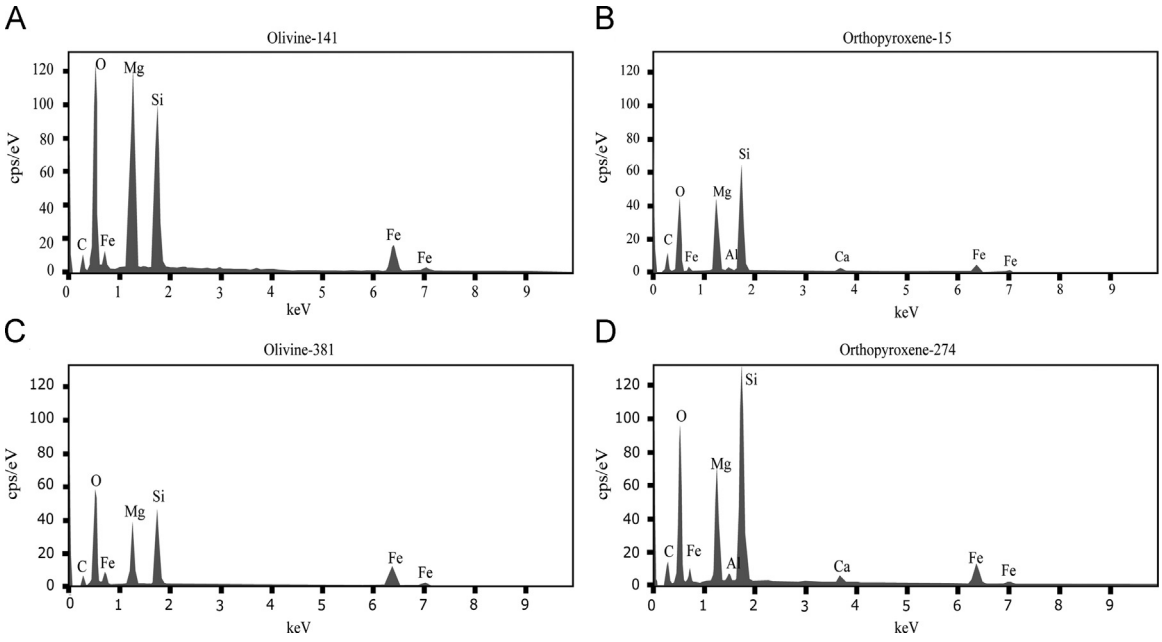

Fig. 5. (A–D) Elemental X-ray spectra of olivine and orthopyroxene analyses from different type of rock samples. (A) Olivine, analysis no 141, (B) orthopyroxene, analysis no: 15, (C) Olivine, analysis no 381, and (D) orthopyroxene, analysis no: 274.

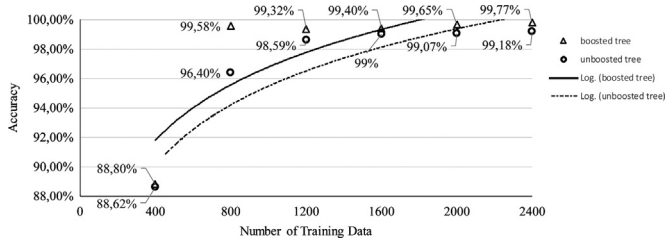

Fig. 6. Relationship between number of training data and the accuracy of boosted and unboosted decision trees.

higher accuracy compared to those of un-boosted decision trees. Second training set on boosted tree (80 intensities per group) was adequate to classify 2193 number of intensities. The accuracy increases steeply from 80 to 800 training intensities and then linearly when the number of training data is bigger than 800. As a result, 2400 intensity training data on boosted tree have shown the highest accuracy (99.77%) among all the decision trees for the final mineral group classification.

For both boosted and un-boosted trees, the last five subsets produced nearly identical accuracy, and these results indicate that a considerable increase in the accuracy is also limited with the increasing size of the training dataset. This conclusion conforms to Pal and Mather (2003)'s Univariate Decision Tree results for the effects of training set size.

5. Conclusions

Generated trees and achieved results have demonstrated that C5.0 Decision Tree Algorithm stands as an accurate and rapid method for mineral classification/identification using characteristic X-ray intensities produced in a typical SEM–EDS without the need for standardized analytical conditions. Time spent on the procedures for the standard (manual) mineral classification and standard working conditions have been diminished by using decision tree models. Furthermore, the variety of working conditions are negligible for Decision Tree Algorithm during both mineral group and member classifications. These results indicate that, accurate and rapid mineral group/member classification might be accepted as possible without being dependent on instrumental working conditions.

The Decision Tree Algorithm showed that it is capable of identifying the mineral members as diopside, hedenbergite, tschermakite and magnesiohastingsite without calculating the cationic formulae which requires time and post-analysis data treatment. At the same time, the effects of quality and training set were evaluated for mineralogical EDS data by this study. As a result; the evaluation produced six boosted trees which have achieved high accuracy rates for training sets with 800 data, even though by using non-standard quality training sets.

The capability of C5.0 Decision Tree Algorithm for mineralogical datasets has been shown that it might be a promising tool for further development.

Acknowledgment

The authors would like to thank Prof. Jef Caers for the editorial handling and two anonymous reviewers for constructing suggestions. Asst. Prof. Inan Ulusoy is thanked for his comments.

References

Baykan, N.A., Yılmaz, N., 2010. Mineral identification using color spaces and artificial neural networks. Comput. Geosci. 36, 91–97.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification And Regression Trees. Wadsworth, Belmont, CA 358 pp..
Brodley, C.E., Friedl, M.A., 1999. Identifying mislabeled training data. J. Artif. Intell. Res. 11, 131–167.
Brown de Colstoun, E., 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. Remote Sens. Environ. 85, 316–327.
Çubukçu, H.E., Ersoy, O., Aydar, E., Çakir, U., 2008. WDS versus silicon drift detector EDS: a case report for the comparison of quantitative chemical analyses of natural silicate minerals. Micron 39, 88–94.
Deer, W.A., Howie, R.A., Zussman, J., 2013. An Introduction to the Rock-Forming Minerals, 3rd ed. Mimeralogical Society, London 498 pp..
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm, In: Proceedings of the Thirteenth International Conference on Machine Learning. San Francisco, pp. 148–156.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. 61, 399–409.
Friedl, M.A., Brodley, C.E., Strahler, A.H., 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Trans. Geosci. Remote Sens. 37, 969–977.
Goldstein, J., Newbury, D.E., Joy, D.C., Lyman, C.E., Echlin, P., Lifshin, E., Sawyer, L., Michael, J.R., 2003. Scanning electron microscopy and X-ray microanalysis, 3rd ed. Springer, New York, NY 689 pp..
Gu, Y., 2003. Automated scanning electron microscope based mineral liberation analysis an introduction to JKMRC/FEI mineral liberation analyser. J. Miner. Mater. Charact. Eng. 2, 33–41.
Gu, Y., Napier-Munn, T., 1997. JK/Philips mineral liberation analyzer – an introduction, In: Proceedings of the Minerals Processing '97 Conference. Cape Town, SA, p. 2.
Humphreys, M.C.S., Blundy, J.D., Sparks, R.S.J., 2006. Magma evolution and open-system processes at Shiveluch volcano: insights from phenocryst zoning. J. Petrol. 47, 2303–2334.
Klein, C., Hurlbut, C.S., Dana, J.D., 1993. Manual of Mineralogy, 21st ed. Wiley, New York, NY 681 pp..
Koujelev, A., Sabsabi, M., Motto-Ros, V., Laville, S., Lui, S.L., 2010. Laser-induced breakdown spectroscopy with artificial neural network processing for material identification. Planet. Space Sci. 58, 682–690.
Leake, B.E., Woolley, A.R., Birch, W.D., Burke, E.A., Ferraris, G., Grice, J.D., Hawthorne, F.C., Kisch, H.J., Krivovichev, V.G., Schumacher, J.C., 2004. Nomenclature of amphiboles: additions and revisions to the International Mineralogical Association's amphibole nomenclature. Am. Mineral. 89, 883–887.
Mingers, J., 1989. An empirical comparison of pruning methods for decision tree induction. Mach. Learn. 4, 227–243.
Morimoto, N., 1988. Nomenclature of pyroxenes. Mineral. Petrol. 39, 55–76.
Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. 86, 554–565.
Quinlan, J.R., 1996. Bagging, Boosting, and C4.5. In: Proceedings of the 13th National Conference on Artificial Intelligence. AAAI Press, Portland, pp. 725–730.
Quinlan, J.R., 1999. Simplifying decision trees. Int. J. Hum.-Comput. Stud. 51, 497–510.
Quinlan, J.R., 2003. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA, p. 302.
Reed, S.J.B., 2005. Electron Microprobe Analysis and Scanning Electron Microscopy in Geology, 2nd ed. Cambridge University Press, Cambridge 232 pp..
Ross, B.J., Fueten, F., Yashkir, D.Y., 2001. Automatic mineral identification using genetic programming. Mach. Vis. Appl. 13, 61–69.
Ruisanchez, I., Potokar, P., Zupan, J., Smolej, V., 1996. Classification of energy dispersion X-ray spectra of mineralogical samples by artificial neural networks. J. Chem. Inf. Comput. Sci. 36, 214–220.
Seibert, J.A., 2004. X-ray imaging physics for nuclear medicine technologists. Part 1: basic principles of X-ray production. J. Nucl. Med. Technol. 32, 139–147.
Small, J., 2002. The analysis of particles at low accelerating voltages (< 10 kV) with energy dispersive X-ray spectroscopy (EDS). J. Res. Natl. Inst. Stand. Technol. 107, 555–566.
Statham, P.J., 2002. Limitations to accuracy in extracting characteristic line intensities from X-ray spectra. J. Res. Natl. Inst. Stand. Technol. 107, 531–546.
Streck, M.J., 2008. Mineral textures and zoning as evidence for open system processes. Miner. Incl. Volcan. Process. 69, 595–622.
Sun-Jong, L., Chan-Hong, L., 2008. Analysis of probe current in scanning electron microscopy. In: Proceedings of the IEEE International Conference on Control, Automation and Systems. Seoul, pp. 1200–1203.
Thompson, S., Fueten, F., Bockus, D., 2001. Mineral identification using artificial neural networks and the rotating polarizer stage. Comput. Geosci. 27, 1081–1089.
Vernon, R.H., 2004. A Practical Guide to Rock Microstructure. Cambridge University Press, Cambridge, p. 594.
Ware, N.G., Reed, S.J.B., 1973. Background corrections for quantitative electron-microprobe analysis using a lithium drifted silicon X-ray detector. J. Phys. E – Sci. Instrum. 6, 286–288.
Weiss, S.M., Kulikowski, C.A., 1991. Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann Publishers Inc., San Mateo, CA 223 pp..