This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

Correction of MS 8083R

Olivine in picrites from Continental Flood Basalt provinces classified using machine learning

Lilu Cheng[1], Yu Wang[2], Zongfeng Yang[2]

1 Earth Observatory of Singapore, Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore

2 State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Beijing, 100083, China

Picrites, dominantly composed of highly forsteritic olivine, can serve as important constraints on primary magma composition and eruption dynamic processes in global Continental Flood Basalt (CFB) provinces. Picrites are commonly divided into high-Ti and low-Ti groups based on whole-rock TiO_2 content or Ti/Y ratio. Here, we use an Artificial Neural Network (ANN) to classify the individual olivine in picrites from global CFB provinces according to whether their parental magma is high-Ti or low-Ti to better understand the primary origin and magmatic processes. After training the ANN on one thousand olivine major element compositions data points, the network was able to differentiate chemical patterns for high-Ti and low-Ti olivine, and classify olivine into correct types with an accuracy of >95 %. Moreover, we find that two types of olivine mix in some single samples from Etendeka, Emeishan, and Karoo CFB provinces. Combining the results with chemical markers of source

lithology, we suggest that the two types of olivine originate from two different sources and their olivine populations mixed during the ascent. This mixing then makes the spatial and temporal variation of picrites types in some CFB provinces unclear.

Keywords: Olivine; Machine Learning; Picrites; Chemical Composition; Classification

PICRITE IN CONTINENTAL FLOOD BASALT PROVINCES

High-Ti (HT) and low-Ti (LT) types of picrites are commonly observed in Continental Flood Basalts (CFBs) or Large Igneous Provinces (LIPs). The classification is based on the values of TiO_2 or Ti/Y of whole-rock compositions (e.g. Ewart et al., 2004; Kamenetsky et al., 2017; Peate et al., 1999; Xiao et al., 2004; Xu et al., 2001). Both types are found either in separate locations or within the same succession of a CFB. For example, in Karoo province, LT picrites (TiO_2<1.5 wt%) are found mainly south of 26 ° S, whereas HT picrites predominate north of this latitude (Galerne et al., 2008). The different locations may indicate the two types were produced from different magmatic sources (e.g. Heinonen et al., 2013; Heinonen and Luttinen, 2008; Howarth and Harris, 2017; Kamenetsky et al., 2012). On the other hand, in Emeishan province, the two types are found within the Binchuan succession (e.g.(Cheng et al., 2014; Xu et al., 2001)). Why the two types of picrites occur in many LIPs is unclear.

Neither is it clear why the two types occur both separately and together. Since picrites are predominantly composed of highly forsteritic olivine, HT samples possibly consist of HT olivine and LT samples consist of LT olivine. In addition to HT and LT samples, some picrites have intermediate Ti/Y value (IT), and do not show clear characteristics of either HT or LT in Emeishan LIP (Kamenetsky et al. 2012). Existence of IT samples may indicate mixing of multiple olivine populations which should be confirmed, because the composition of picrite are usually used to constrain on the source composition (e.g., Zhang et al., 2006). Source compositions cannot be constrained correctly from multiple olivine populations.

To address these issues, we need to analyse the chemical patterns of olivine in picrites. For this study, we have compiled thousands of compositions of olivines from published picrites samples and built an Artificial Neural Networks (ANN) to investigate the chemical characteristics of olivine from HT and LT picrites. Further, we determined the links between olivine populations and their sources and answered whether olivine populations mixed during magmatic processes.

DATABASE OF OLIVINE COMPOSITIONS IN PICRITES

We have collected thousands of major elements data points of olivine as well as their whole-rock compositions of picrites from six CFBs (Emeishan LIP, Etendeka province, Ethiopian CFB, Karoo LIP, North Atlantic

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

province, and Siberia CFB) from the open-access and comprehensive global petrological database GEOROC (http://georoc.mpch-mainz.gwdg.de/georoc/). These CFB provinces are located in different parts of the earth, and have been well studied (Fig. 1). Within the data, Ti/Y values of picrites range from 250 to 1400 (Fig. 2). The picrites from the Ethiopian province have the highest Ti/Y values of 1400, while the maximal values of other CFBs are around 800 (Fig. 2). Although Xu et al. (2001) suggested that the Ti/Y value boundary between HT and LT is 500 for the Emeishan LIP, the boundaries for different provinces may vary between 350 and 600 based on the gaps in data in Figure 2. Thus, we assumed that samples with Ti/Y of more than 600 were HT end-members, samples with Ti/Y of less than 350 were LT end-members, and samples with Ti/Y from 350 to 600 were the intermediate Ti group (IT).

We determined the quality of olivine dataset, which comprised 2898 major element compositions of olivine from picritic samples of CFBs (Table 1s in appendix), by calculating the cardinality, minimum, mean, median, maximum and standard deviation of each element's concentration (Table 1). Note, that in this study we assume all the collected data could represent the true composition of whole olivine by measuring composition of the core of olivine without the effect of diffusion (e.g. Cheng et al., 2020; Costa, 2020; Costa et al., 2020) and the random cut effect in thin sections (e.g. Cheng et al., 2017; Cheng and Costa, 2019). The cardinality measures the number of distinct values. The cardinalities of TiO_2 were much lower than 1000, and the minimum value and mean value were lower than 0.01, meaning that many TiO_2 values were 0. The difference between the mean and median of each composition showed the outliers. Most of the elements differed only a little (0-1 wt%), which showed that there were not too many outliers. The standard deviations of the FeO and MgO values (3-4 wt%) were much larger than the standard deviations of the other elements (less than 1 wt%), showing the large FeO and MgO variation. However, the maximum values of several elements were much larger than the normal values of olivine. For example, the maximum of SiO_2 was 63 wt% and that of Al_2O_3 was 5 wt%. This indicated existence of several low-quality data points to be filtered out—those with analytical totals out of the range 98-101 and with stoichiometric ratios of (Mg+Fe)/Si out of the range of 1.95-2.05). Furthermore, low-quality training data to filter out were ones with SiO_2 values out of the range of 38-42 wt%, Al_2O_3 values out of the range of 0-0.15 wt%, Cr_2O_3 values out of the range 0-0.2 wt%, and NiO values out of the range of 0.1-0.6 wt%. The remaining total of 1002 training data points were high-quality (Table 1s in the appendix). The same approach was taken to obtain high-quality data for IT samples.

To avoid the high linear relationship between features, we also calculated the correlation coefficient between each two elements (Fig. 3). The correlation coefficients showed that MgO had a negative linear relationship

DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

with FeO^T and MnO, whose correlation coefficients were -0.98 and -0.84, respectively. There was a positive linear relationship between MgO and SiO_2, whose correlation coefficient was 0.86. It meant MgO could represent SiO_2, FeO^T and MnO. There were no clear positive or negative linear relationships between Al_2O_3, MgO, CaO, NiO, or Cr_2O_3.

ARTIFICIAL NEURAL NETWORKS

There are several models of machine learning such as Support Vector Machine (SVM), Random Forest (RF), Naïve-Bayes Classifier (NBC), Artificial Neural Networks (ANN), Convolution Neural Network (CNN) and Recurrent Neural Networks (RNN) (Ardabili et al., 2020; Bergen et al., 2019; Bishop, 2006). They all can convert experience into expertise or knowledge during learning (Ardabili et al., 2020; Bergen et al., 2019). Several ML models have been used in earth science studies (e.g. Hazen, 2014; Morrison et al., 2017; Petrelli and Perugini, 2016).

In this study, the ANN model was selected, since our data (major elements compositions of olivine) are tabular and ANN is powerful. But, we also compared the performance of our ANN with other traditional ML methods including SVM and NBC (see Discussion). We built a supervised ANN according to the types of picrites. The first goal of our ANN was to produce good prediction results by learning the chemical characteristics of the training data. The second goal was to use the ANN to classify the olivine data from IT samples which was a much more difficult task than predictions from training data. We used only Al_2O_3, MgO, CaO, NiO, or Cr_2O_3 as features as they had no clear linear relationship between each other, as we mentioned above (Fig. 4). Our ANN used one hidden layer with 10 nodes. We chose 'Levenberg-Marquardt backpropagation' ('trainlm' function in Matlab) as the training algorithm as it is often the fastest back propagation algorithm and gives us the most accurate results, although it does require more memory than other algorithms.

To train the ANN, we selected the olivine composition of an HT end-member to represent the HT group, and the composition of an LT end-member to represent the LT group. We had about 1000 high-quality olivine compositions from all the global CFBs. To avoid overfitting, we randomly split our data into three groups: the training set, the validation set and the test set, using the ratios of 70:15:15. Thus, we had about 700 training samples, about 150 validation samples, and about 150 test samples.

RESULTS

The results show that our ANN has achieved high accuracy (>95%) for all three sets (Fig. 5). For the training data, the ANN determined 233 data points as belonging to LT, out of which 229 predictions were correct and 473 data points as HT, out of which 468 predictions were correct. The accuracy was thus 99 % (Fig. 5a). For the validation set, our ANN have

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

determined that 38 data points belonged to LT, out of which 36 predictions were correct and 112 points as HT, out of which 109 predictions were correct. The accuracy achieved was 96.7% (Fig. 5b). For the test set, the ANN labelled 57 data points as LT with 54 predictions correct, and 93 data points were labelled as HT with 89 predictions correct. Thus, the accuracy was 95.3 % (Fig. 5c). The high and close accuracy values of the three sets means there are no overfitting problems. In total, the ANN's accuracy is 98.1 % for all of the data (Fig. 5d).

After training the ANN, we input all of the olivine data from olivine crystals contained in the IT samples. The olivine from the IT samples were determined by the ANN to belong to LT was labelled as IT-L, and the olivines determined to belong to HT were labelled as IT-H (Fig. 6). We find that most of LT and HT olivine are located in different part of NiO-MgO plot, but there are still many overlapping data points when MgO ranges from 40 to 45 wt% (Fig. 6a). The overlaps suggest that a simple relation such as NiO-MgO cannot accurately classify olivine crystals. However, the ANN model can classify them highly accurately (>95%), although several IT-L points whose MgO is about 40 wt% are located in the area of HT and IT-H, which suggests they are probably classified incorrectly (Fig. 6a). The LT and IT-L olivine show mostly higher CaO (Fig. 6b), higher Al_2O_3 (Fig. 6c) compared to HT and IT-H. Most LT and IT-L data points range from 0.3 to 0.6 wt% CaO, while HT and IT-H vary

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

from 0.1 to 0.4 wt%. However, LT and HT olivine overlap more in their relationships between Al_2O_3, CaO vs. MgO compared to NiO vs. MgO. There are also many overlapping points around 0.3 wt% CaO and 0.04 wt% Al_2O_3 (Fig. 6c & 6d). The HT and LT samples overlap mostly in the relationship between Cr_2O_3 and MgO compared to the rest (Fig.6d). Both HT and LT samples range from 0 to 0.16 wt%. Thus, we emphasize again that it is almost impossible to classify these overlapping data using simple relations. Overall, the ANN has correctly classified the olivine from the IT sample into the right group: most of IT-L points overlap the IT samples and most of IT-H points overlap the HT samples (Fig. 6). Moreover, ANN has performed well on the overlapping data points which could not be classified by the simple relations.

DISCUSSION

Comparison with other machine learning methods

In the previous section, we demonstrated the good performance of our ANN on IT olivine and on the overlapping data points where olivine crystals cannot be accurately classified by simple relations such as NiO-MgO. Although more complex relations such as NiO-MgO-Al_2O_3 could be used for overlapping data points, much longer time would be required, and distinguishing differences in three-dimensional space would be more difficult. In comparison, our ANN was able to directly predict olivine populations with high accuracy for such cases within several minutes. Moreover, we compared the performance of the ANN to other traditional machine learning methods such as the Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC). The SVM is a non-parametric classifier that finds a linear vector (if a linear kernel is used) to separate classes. It has been used in Earth science research. For example, Petrelli and Perugini (2016) outlined an SVM with a Gaussian kernel function for tectonic discrimination based on geochemical and isotopic data. We used the 'linear' function in Matlab as the kernel function. The NBC is a simple 'probabilistic classifier' based on Bayes' theorem, with strong independence assumptions between features (e.g., Ren et al., 2019). The NBC can be trained efficiently in a supervised learning setting, such as the one required in this study. We used the 'kernel' function in Matlab with a mean kernel smoothing density estimate. The SVM and NBC are highly accurate (96.5% and 94.5%, respectively), and capable of distinguishing between the two types of samples (Fig.7). However, since our ANN's accuracy remains the highest, we suggest using ANN model (Figs. 5 and 7).

The sources of olivine populations

We have shown that different machine learning models such as ANN and SVM are able to classify the olivine populations accurately. It indicates

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

that these results show that the olivine from two end-members are different. To investigate the link between olivine population and source lithology, we compare NiO concentrations of two olivine populations. We find that NiO of HT is much higher than that of LT at the given MgO value (Fig. 8a). High NiO in olivine has, for example, been suggested to indicate a pyroxenitic source and thus determine the mafic sources for Hawaiian magma (Sobolev et al. 2005). It has been widely used to argue for an olivine-free pyroxenite source for both continental and oceanic basalts (Herzberg, 2011, 2006; Sobolev et al., 2007; Xu et al., 2012). However, high-Ni olivine could crystallize under low-temperature conditions from high-temperature peridotite melts without contributions from a pyroxenite source as shown in Fig. 8a (Matzen et al., 2017). Minor and trace element characteristics in olivine strongly depend on pressure, temperature and melt composition, as suggested by experimental petrology (Matzen et al., 2013). Examples such as high-Ni and low-Mn olivine in Karoo was suggested as the result of temperature and pressure variations (Heinonen and Fusswinkel, 2017). However, we found a relatively narrow range of LT olivine and there is a linear trend between Fo and NiO for the global data, while that of HT is quite large and the Fo-NiO relationship not easily determined. These HT olivine points are further divided into two groups. For example, the olivine from Ethiopian CFB has higher NiO content than that of LT, but a bit lower NiO than that of olivine from other CFBs. Note, that olivine from other CFBs is located in the Ethiopian area (Fig. 8a). If the two populations are both from peridotite, we would expect similar trends even under different pressures or temperatures, and thus we suggest the two types of olivine are from different sources. However, we need to emphasize that the link between an olivine population and a source depends on the choice of markers of source lithology, and many markers have their own limitations (e.g. Yang et al., 2019, 2016). Thus, we also applied the last chemical marker suggested by Yang et al. (2019). This marker, combines FCAKANTMS) =ln (FeO/CaO) − 0.08 * ln(K_2O/Al_2O_3) − 0.052 * ln (TiO_2/Na_2O) − 0.036 * ln (Na_2O/K_2O) * ln (Na_2O/TiO_2) − 0.062 * (ln (MgO/SiO_2))^3 – 0.641 * (ln (MgO/SiO_2))^2 – 1.871 * ln (MgO/SiO_2) − 1.473; all major elements in wt%) with ln (SiO_2/(CaO+Na_2O+TiO_2) and was able to distinguish approximately 80% and 50% low to moderate degree (Fo<60%) partial melts of mafic sources from those of peridotite and transitional lithologies. We calculated the FCKANTMS values for the picrites in this study and the global melts compiled from Yang et al. (2019). We found that most of LT samples are much closer to area of peridotite source compared to HT samples (Fig. 8b). Data points of HT samples are located in two quite different areas: many data points have much lower ln (SiO_2/(CaO+Na_2O+TiO_2) which are all from Ethiopian CFB, while many have much higher FCKANTMS and middle ln(SiO_2/(CaO+Na_2O+TiO_2) that are from North, Karoo LIPs. The former is close to the carbonated mafic source and the latter is close to the mafic source. Two main areas of HT samples are consistent with the two main trends of HT olivines in Fig. 8a. Samples from Emeishan LIP are located in both areas, which shows the complexity of their source lithology. We propose that these HT olivines and picrites represent mafic sources and LT indicate peridotite sources. Further, the source of HT olivine could be divided into mafic (HT1) and carbonated mafic source (HT2). Our ANN has further determined the three groups linked to three sources by training the ANN on data with three labels using the same five features (Fig. 9). We have found that the accuracy for all of the data is 95%, which is lower than the accuracy of the ANN model with two outputs (Fig. 5). Clearly, classifying samples from HT1 and HT2 is more challenging due to their similarity.

Olivine populations in a single picritic sample

We selected olivine from a single IT sample of each CFB and input them into our ANN model. Here, we used the ANN model with three outputs to determine whether LT, HT1 and HT2 olivine populations mix. Note, that we did not have IT olivine data from Karoo and Ethiopian CFBs. The results show that no single sample has all three olivine populations: some samples have both LT and HT1, and some samples only have LT or HT1. For example, for the Siberian flood basalt province, the ANN determined that all the olivine from sample SU33 with Ti/Y=350 belongs to the LT group, and all the olivine from SU50 whose Ti/Y is 474 belongs to the HT1 group (Fig.10). However, the results show that: (1) 3+/-10% and 4+/-8% of the olivine are HT1 from two IT samples from Etendeka province (97SB41 and 97SB62); (2) 11+/-5%, 10+/-3% and 20+/-5% of the olivine from three IT samples of Emeishan LIP are HT1 (13-EJH08,7-EJH08, and 1a-EJH06); and (3) 10 +/-11%,11 +/-9% , 7+/-4% of olivine are HT1 from three samples from North Atlantic province (400457, 400230, and 340740). We found that mixing of two olivine types was found in IT samples from Etendeka, Emeishan LIP, and North Atlantic provinces. The possible reasons are: (1) In some provinces, both types of picrite are found in the same location (e.g., in Binchuan succession in Emeishan LIP, Xu et al., 2001). The two types of olivine are relatively easier to mix when they both occur in the same location in these CFBs compare to different locations. (2) The proportion of HT1 olivine in these IT samples is low, which means a significant amount of olivine data is required to find HT1 olivine. But enough olivine data of intermediate samples whose Ti/Y is between 300 and 500 in other CFBs are not available now. If more olivine data of samples with Ti/Y of about 300-400 are available, olivine population mixing could probably be found in other CFBs.

With our crystal-scale classification, we are able to explain why two types of picrites are common in all the CFBs, yet, in some CFBs, the spatial and temporal variations of different picrite types become unclear. Although the two or three types of olivine are produced by different sources at the beginning of flood basalt eruptions, during magma ascent, these olivine types may mix. Thus, if we consider only the classification of rocks, the two types of olivine (LT and HT) in the samples will be classified into one group. However, with the powerful machine learning tool, olivine from IT samples of CFBs such as Etendeka, Emeishan, and North Atlantic provinces can still be classified into different groups. Overall, the existence of different types of picrites may be better explained by variations in sources rather than crystallization processes. Their existance in different CFBs may indicate that the origin in the global CFBs is linked to the different lithologies: peridotitic or mafic sources (e.g. Heinonen et al., 2013; Heinonen and Luttinen, 2008; Kamenetsky et al., 2012; Li et al., 2014; Sobolev et al., 2005). The plume in these CFBs may start from the one source such as peridotite, but with the huge volume of magma, as the plume rises another source, such as mafic, starts melting and is mixed into the magma. Thus, our results show both peridotitic and mafic sources are involved in CFB provinces and olivine populations from two sources may mix during ascent.

IMPLICATIONS

Our study shows a clear link between olivine populations and two types of picrites, with the olivine from different types of picrites showing varying chemical characteristics. Using the source markers including NiO-Fo and FCKANTMS, we propose HT and LT olivines represent mafic and peridotite sources, respectively. The source of HT olivine could be further divided into mafic and carbonated mafic sources. We have also built an ANN model with three outputs which are linked to three different sources. Using the ANN model, we are able to find olivine populations within a single sample, which indicates that olivine populations mixed in CFB.

The ANN model we have built enables grouping the olivine from the picrites found in these CFBs without knowing the whole-rock composition of the picrites. Thus, our ANN for global CFB provinces could help classify the multitude of data on olivine published without the rocks' compositions, which could provide the information about source lithology. Another application is the recognition of multiple olivine populations in a single sample. As we mentioned above, the picritic samples are commonly used to constrain the source compositions which requires that the samples provide direct source information without any effect from crystal populations mixing. Thus, it is necessary to check the populations using our ANN model. Moreover, the application of ANN is much more simple and efficient compare to human judgment as we mentioned above. It only requires users to prepare the olivine composition with the five features and input them into the ANN models we provided (Matlab scripts and training data in appendix). The ANN model will determine the olivine type correctly within several minutes.

Last but not least, there are many whole-rocks or mineral major elements, trace elements, or radiogenic isotope compositions, it is difficult for people to distinguish the hidden patterns of massive data. ML methods can exam these large and varied data sets to uncover information including hidden patterns, unknown correlations. Our study shows that ML models offer the potential to make more data-driven decisions such as classification at more high accuracy rates for these data sets.

Acknowledgements

The reviews of S. Morrison, two anonymous reviewers, and the editors M. Ballmer and D. Baker are appreciated and helped to largely clarify the structure and content of the paper. L.L. thanks R. Made, F. Costa for their comments about machine learning, and J. Herrin for the suggestions on data cleaning. We also thank P. Adamek for his language suggestions on the manuscript. This research was supported by the National Basic Research Program of China (973 Program NO. 2011CB808901) and the National Research Foundation of Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative, and a

National Research Foundation Singapore Investigatorship Award (NRF-NRFI2017-06).

Figures

Figure 1 A map showing the location of the main Continental Flood Basalts (CFBs) around the world. The CFBs highlighted in the rectangle provide many olivine data points from picrites for this study. This map is adapted from (Bryan and Ernst, 2008).

Figure 2 Ti/Y of samples ranging from 200 to 1400 from the six CFBs. The gray area represents intermediate Ti/Y (IT) samples.

Figure 3 The correlations between major elements compositions of olivine. The diagonal shows the elements, the lower triangle is the diagram of two related elements, and the upper triangle shows the correlation coefficient of the two related elements. For example, 0.25 in the second column of the first row is the correlation coefficient of SiO_2 and Al_2O_3, and their plots are shown in the first column of the second row.

DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

Figure 4 The general structure of the ANN. It has five features including Al_2O_3, MgO, CaO, NiO and Cr_2O_3. There are 10 nodes in a hidden layer. The output of the ANN is LT and HT.

Figure 5 A confusion matrix showing the classification accuracy of the ANN. (a) In the confusion matrix of the training data, the blue color represents a correct prediction and the red a wrong prediction. The number inside is the total number of samples. (b) The confusion matrix of the validation data. (c) The confusion matrix of test data. (d) The confusion matrix of all the data.

Figure 6 Variation of NiO, CaO, Al_2O_3, and Cr_2O_3 versus MgO for olivine from global CFBs. It shows that a portion of IT samples whose major element content is similar to that of LT samples is determined as IT-L, and others similar to HT samples are determined as IT-H by our ANN. The grey triangle and circles represent olivines from HT and LT samples, respectively. The dark-yellow triangle and light-blue circles represent olivine from IT samples, which are classified as HT and LT by our ANN, respectively. The dashed line represents the area with most HT and IT-H olivines, and the solid line represents the area with most LT and IT-L olivine. Note, that many contents of Al_2O_3 and Cr_2O_3 in 6c and 6d are zero which may mean they were below the detection threshold.

DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

Figure 7 A confusion matrix of all the data using an SVM and an NBC. It shows that the accuracy of the ANN shown in Fig.5 is better than the accuracy achieved by either an SVM or an NBC. The blue color represents a correct prediction and the red a wrong prediction. The number inside is the total number of samples which are same as in Fig.5.

Figure 8 (a) NiO-Fo relationship of olivine populations in picrites from the six CFBs. The pink field is for phenocrystals from mid-ocean-ridge basalts (after Sobolev et al., 2005). The light-green field is the NiO_{89} content range produced by peridotite primary melts (Matzen et al. 2017). The white dashed circle is the data from Ethiopian CFB. (b) FCKANTMS vs. $\ln(SiO_2/(CaO+Na_2O+TiO_2))$ of whole-rock compositions of picrites which we collected in this study. Mafic, peridotite, and carbonatite mafic melts data were compiled by Yang et al. (2019). EM43 is one LT sample from Emeishan LIP. The black dashed circle are samples from Ethiopian LIP.

Figure 9 A confusion matrix showing the classification accuracy of the ANN with three outputs for all the data. HT1 and HT2 represent for mafic and carbonatite mafic source, respectively. The blue color represents a correct prediction and the red a wrong prediction. The numbers inside are the total numbers of samples, and the percentage numbers inside are the ratios.

Figure 10 The classification results of a single IT sample for each CFB (except Karoo and Ethiopian CFBs, which do not have any IT samples). Both HT1 and LT olivine are found in the same samples from Etendeka, North Atlantic, and Emeishan CFBs.

Reference:

Ardabili, S., Mosavi, A., Dehghani, M., Várkonyi-Kóczy, A.R., 2020. Deep Learning and Machine Learning in Hydrological Processes Climate Change and Earth Systems a Systematic Review, in: Várkonyi-Kóczy, A.R. (Ed.), Engineering for Sustainable Future. Springer International Publishing, Cham, pp. 52–62.

Bergen, K.J., Johnson, P.A., Maarten, V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience. Science 363.

Bishop, C.M., 2006. Pattern recognition and machine learning. springer.

Bryan, S.E., Ernst, R.E., 2008. Revised definition of Large Igneous Provinces (LIPs). Earth-Sci. Rev. 86, 175–202. https://doi.org/10.1016/j.earscirev.2007.08.008

Cheng, L., Costa, F., 2019. Statistical analysis of crystal populations and links to volcano deformation for more robust estimates of magma replenishment volumes. Geology. https://doi.org/10.1130/G46826.1

Cheng, L., Costa, F., Bergantz, G., 2020. Linking fluid dynamics and olivine crystal scale zoning during simulated magma intrusion. Contrib. Mineral. Petrol. 175, 53. https://doi.org/10.1007/s00410-020-01691-3

Cheng, L., Costa, F., Carniel, R., 2017. Unraveling the presence of multiple plagioclase populations and identification of representative two-dimensional sections using a statistical and numerical approach. Am. Mineral. 102, 1894–1905.

Cheng, L.-L., Yang, Z.-F., Zeng, L., Wang, Y., Luo, Z.-H., 2014. Giant plagioclase growth during storage of basaltic magma in Emeishan Large Igneous Province, SW China. Contrib. Mineral. Petrol. 167, 1–20. https://doi.org/10.1007/s00410-014-0971-0

Costa, F., 2020. Clocks in Magmatic Rocks. Annu. Rev. Earth Planet. Sci. https://doi.org/10.1146/annurev-earth-080320-060708

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2022-8083. http://www.minsocam.org/

Costa, F., Shea, T., Ubide, T., 2020. Diffusion chronometry and the timescales of magmatic processes. Nat. Rev. Earth Environ. https://doi.org/10.1038/s43017-020-0038-x

Ewart, A., Marsh, J.S., Milner, S.C., Duncan, A.R., Kamber, B.S., Armstrong, R.A., 2004. Petrology and Geochemistry of Early Cretaceous Bimodal Continental Flood Volcanism of the NW Etendeka, Namibia. Part 1: Introduction, Mafic Lavas and Re-evaluation of Mantle Source Components. J. Petrol. 45, 59–105. https://doi.org/10.1093/petrology/egg083

Galerne, C.Y., Neumann, E.-R., Planke, S., 2008. Emplacement mechanisms of sill complexes: Information from the geochemical architecture of the Golden Valley Sill Complex, South Africa. J. Volcanol. Geotherm. Res. 177, 425–440.

Hazen, R.M., 2014. Data-driven abductive discovery in mineralogy. Am. Mineral. 99, 2165–2170.

Heinonen, J.S., Fusswinkel, T., 2017. High Ni and low Mn/Fe in olivine phenocrysts of the Karoo meimechites do not reflect pyroxenitic mantle sources. Chem. Geol. 467, 134–142.

Heinonen, J.S., Luttinen, A.V., 2008. Jurassic dikes of Vestfjella, western Dronning Maud Land, Antarctica: Geochemical tracing of ferropicrite sources. Lithos 105, 347–364. https://doi.org/10.1016/j.lithos.2008.05.010

Heinonen, J.S., Luttinen, A.V., Riley, T.R., Michallik, R.M., 2013. Mixed pyroxenite–peridotite sources for mafic and ultramafic dikes from the Antarctic segment of the Karoo continental flood basalt province. Lithos 177, 366–380. https://doi.org/10.1016/j.lithos.2013.05.015

Herzberg, C., 2011. Identification of Source Lithology in the Hawaiian and Canary Islands: Implications for Origins. J. Petrol. 52, 113–146. https://doi.org/10.1093/petrology/egq075

Herzberg, C., 2006. Petrology and thermal structure of the Hawaiian plume from Mauna Kea volcano. Nature 444, 605. https://doi.org/10.1038/nature05254

Howarth, G.H., Harris, C., 2017. Discriminating between pyroxenite and peridotite sources for continental flood basalts (CFB) in southern Africa using olivine chemistry. Earth Planet. Sci. Lett. 475, 143–151. https://doi.org/10.1016/j.epsl.2017.07.043

Kamenetsky, V.S., Chung, S.-L., Kamenetsky, M.B., Kuzmin, D.V., 2012. Picrites from the Emeishan Large Igneous Province, SW China: a Compositional Continuum in Primitive Magmas and their Respective Mantle Sources. J. Petrol. 53, 2095–2113. https://doi.org/10.1093/petrology/egs045

Kamenetsky, V.S., Maas, R., Kamenetsky, M.B., Yaxley, G.M., Ehrig, K., Zellmer, G.F., Bindeman, I.N., Sobolev, A.V., Kuzmin, D.V., Ivanov, A.V., Woodhead, J., Schilling, J.-G., 2017. Multiple mantle sources of continental magmatism: Insights from “high-Ti” picrites of Karoo and other large igneous provinces. Chem. Geol. 455, 22–31. https://doi.org/10.1016/j.chemgeo.2016.08.034

Li, J., Wang, X.-C., Ren, Z.-Y., Xu, J.-F., He, B., Xu, Y.-G., 2014. Chemical heterogeneity of the Emeishan mantle plume: Evidence from highly siderophile element abundances in picrites. J. Asian Earth Sci. 79, 191–205.

Matzen, A.K., Baker, M.B., Beckett, J.R., Stolper, E.M., 2013. The Temperature and Pressure Dependence of Nickel Partitioning between Olivine and Silicate Melt. J. Petrol. 54, 2521–2545. https://doi.org/10.1093/petrology/egt055

Matzen, A.K., Wood, B.J., Baker, M.B., Stolper, E.M., 2017. The roles of pyroxenite and peridotite in the mantle sources of oceanic basalts. Nat. Geosci. 10, 530–535.

Morrison, S.M., Liu, C., Eleish, A., Prabhu, A., Li, C., Ralph, J., Downs, R.T., Golden, J.J., Fox, P., Hummer, D.R., Meyer, M.B., Hazen, R.M., 2017. Network analysis of mineralogical systems. Am. Mineral. 102, 1588–1596. https://doi.org/10.2138/am-2017-6104CCBYNCND

Peate, D.W., Hawkesworth, C.J., Mantovani, M.M., Rogers, N.W., Turner, S.P., 1999. Petrogenesis and stratigraphy of the high-Ti/Y Urubici magma type in the Paraná flood basalt province and implications for the nature of 'Dupal'-type mantle in the South Atlantic region. J. Petrol. 40, 451–473.

Petrelli, M., Perugini, D., 2016. Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. Contrib. Mineral. Petrol. 171, 81. https://doi.org/10.1007/s00410-016-1292-2

Ren, Q., Li, M., Han, S., 2019. Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives. Big Earth Data 3, 8–25.

Sobolev, A.V., Hofmann, A.W., Kuzmin, D.V., Yaxley, G.M., Arndt, N.T., Chung, S.-L., Danyushevsky, L.V., Elliott, T., Frey, F.A., Garcia, M.O., Gurenko, A.A., Kamenetsky, V.S., Kerr, A.C., Krivolutskaya, N.A., Matvienkov, V.V., Nikogosian, I.K., Rocholl, A., Sigurdsson, I.A., Sushchevskaya, N.M., Teklay, M., 2007. The Amount of Recycled Crust in Sources of Mantle-Derived Melts. Science 316, 412–417. https://doi.org/10.1126/science. 1138113

Sobolev, A.V., Hofmann, A.W., Sobolev, S.V., Nikogosian, I.K., 2005. An olivine-free mantle source of Hawaiian shield basalts. Nature 434, 590.

Xiao, L., Xu, Y.G., Mei, H.J., Zheng, Y.F., He, B., Pirajno, F., 2004. Distinct mantle sources of low-Ti and high-Ti basalts from the western Emeishan large igneous province, SW China: implications for plume–lithosphere interaction. Earth Planet. Sci. Lett. 228, 525–546. http://dx.doi.org/10.1016/j.epsl.2004.10.002

Xu, Y., Chung, S.-L., Jahn, B., Wu, G., 2001. Petrologic and geochemical constraints on the petrogenesis of Permian–Triassic Emeishan flood basalts in southwestern China. Lithos 58, 145–168. https://doi.org/10.1016/s0024-4937(01)00055-x

Xu, Z., Zhao, Z.-F., Zheng, Y.-F., 2012. Slab–mantle interaction for thinning of cratonic lithospheric mantle in North China: Geochemical evidence from Cenozoic continental basalts in central Shandong. Lithos 146–147, 202–217. https://doi.org/10.1016/j.lithos.2012.05.019

Yang, Z.-F., Li, J., Jiang, Q.-B., Xu, F., Guo, S.-Y., Li, Y., Zhang, J., 2019. Using Major Element Logratios to Recognize Compositional Patterns of Basalt: Implications for Source Lithological and Compositional Heterogeneities. J. Geophys. Res. Solid Earth 124, 3458–3490. https://doi.org/10.1029/2018JB016145

Yang, Z.-F., Li, J., Liang, W.-F., Luo, Z.-H., 2016. On the chemical markers of pyroxenite contributions in continental basalts in Eastern China:

Implications for source lithology and the origin of basalts. Earth-Sci. Rev. 157, 18–31. https://doi.org/10.1016/j.earscirev.2016.04.001

Zhang, Z., Mahoney, J.J., Mao, J., Wang, F., 2006. Geochemistry of Picritic and Associated Basalt Flows of the Western Emeishan Flood Basalt Province, China. J. Petrol. 47, 1997–2019. https://doi.org/10.1093/petrology/egl034

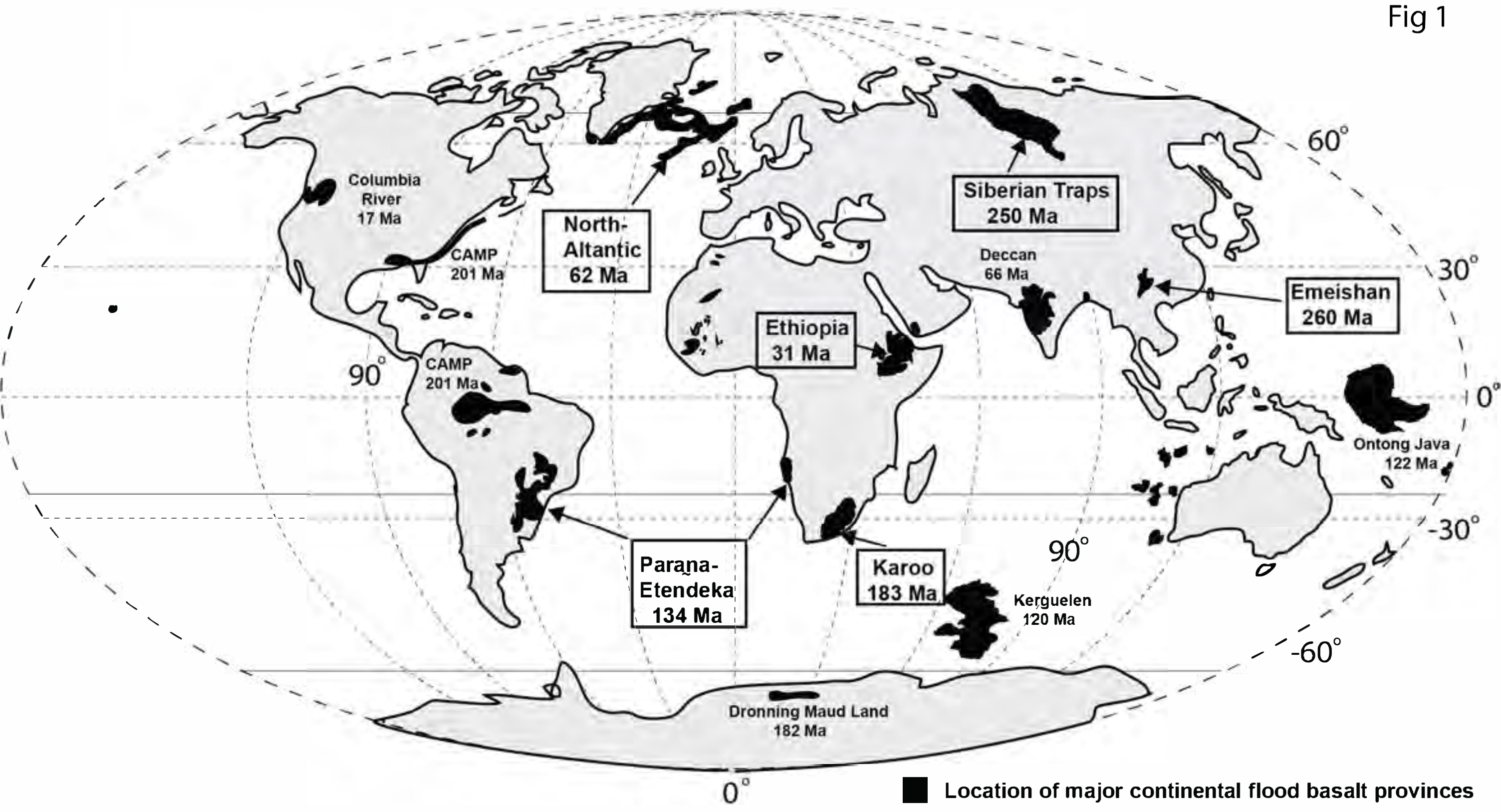
Table 1 Data describition of olivine used in this study

	Count	Card.	Min	1st Qrt	Mean	Median
SIO_2(wt%)	2898	2466	35.62	39.44	39.86	39.90
TiO_2(wt%)	2898	749	0.00	0.00	0.01	0.00
AL_2O_3(wt%)	2898	1042	0.00	0.02	0.04	0.04
CR_2O_3(wt%)	2898	1121	0.00	0.04	0.06	0.06
FEOT(wt%)	2898	2690	0.00	11.62	14.14	13.56
CAO(wt%)	2898	1644	0.00	0.26	0.31	0.32
MGO(wt%)	2898	2685	26.60	43.78	45.29	45.67
MNO(wt%)	2898	2048	0.00	0.17	0.20	0.20
NIO(wt%)	2898	1982	0.00	0.31	0.35	0.34

Card. Cardinality of each element, which measure the number of distinct values present

3rd Qrt.	Max	Std Dev.
40.29	62.52	1.02
0.01	0.36	0.01
0.06	5.00	0.11
0.08	0.49	0.04
16.16	31.37	4.08
0.36	4.51	0.12
47.51	52.88	3.32
0.23	0.48	0.06
0.40	0.63	0.09

Fig 1



Columbia River 17 Ma
CAMP 201 Ma
North-Altantic 62 Ma
Siberian Traps 250 Ma
Deccan 66 Ma
Emeishan 260 Ma
Ethiopia 31 Ma
CAMP 201 Ma
Ontong Java 122 Ma
Paraña-Etendeka 134 Ma
Karoo 183 Ma
Kerguelen 120 Ma
Dronning Maud Land 182 Ma
60°
30°
0°
-30°
-60°
90°
90°
0°
Location of major continental flood basalt provinces

Fig 2

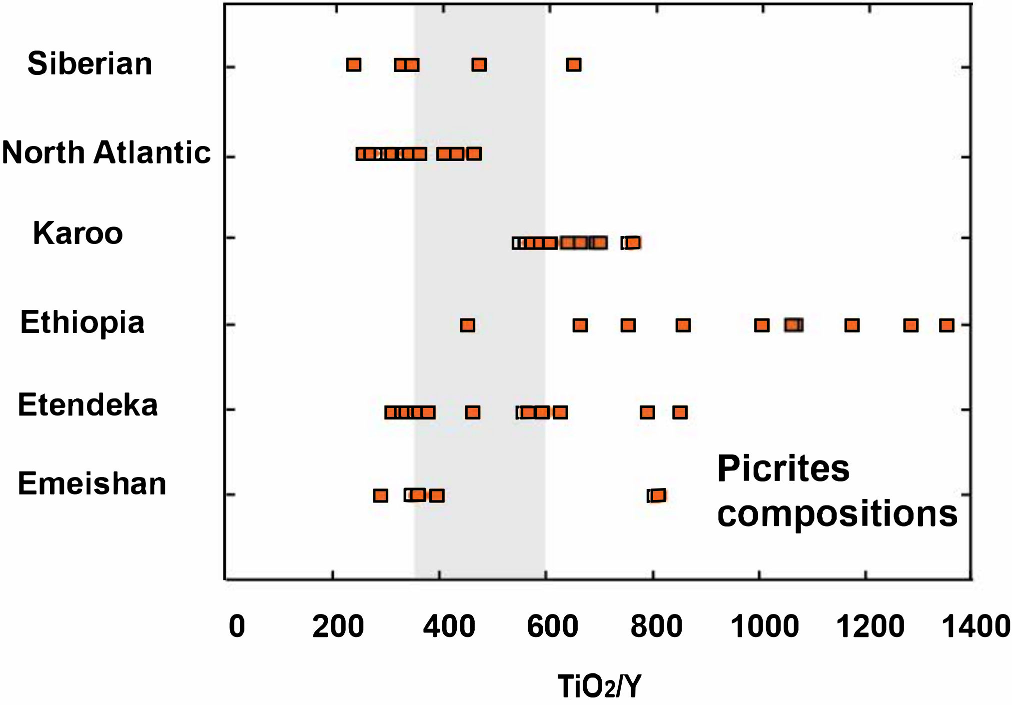

Siberian
North Atlantic
Karoo
Ethiopia
Etendeka
Emeishan
Picrites
compositions
0
200
400
600
800
1000
1200
1400
TiO2/Y

Fig 3


SiO2
0.26
0.48
-0.86
0.27
0.86
-0.74
0.46
Al2O3
0.42
-0.28
0.33
0.29
-0.19
0.02
Cr2O3
-0.58
0.16
0.57
-0.49
0.36
FeO
-0.2
-0.98
0.86
-0.6
CaO
0.24
-0.01
-0.24
MgO
-0.84
0.56
MnO
-0.76
NiO

Fig 4

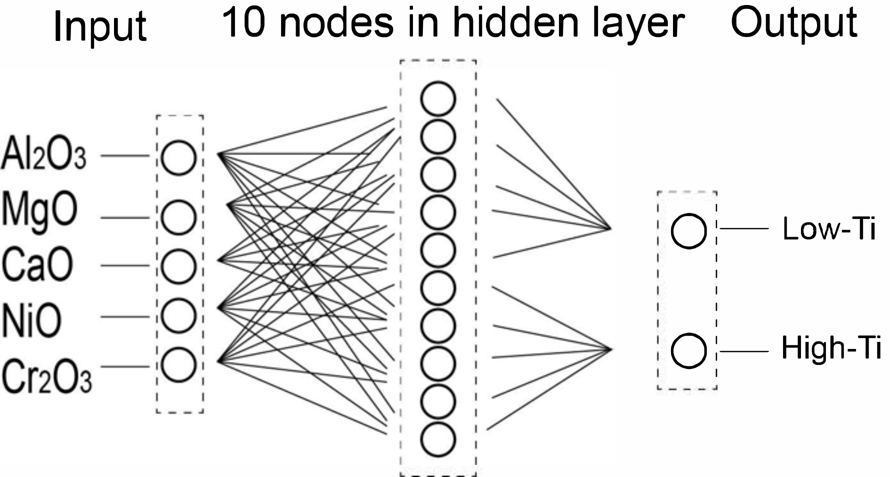

Input
10 nodes in hidden layer
Output
Al2O3
MgO
CaO
NiO
Cr2O3
Low-Ti
High-Ti

Fig 5

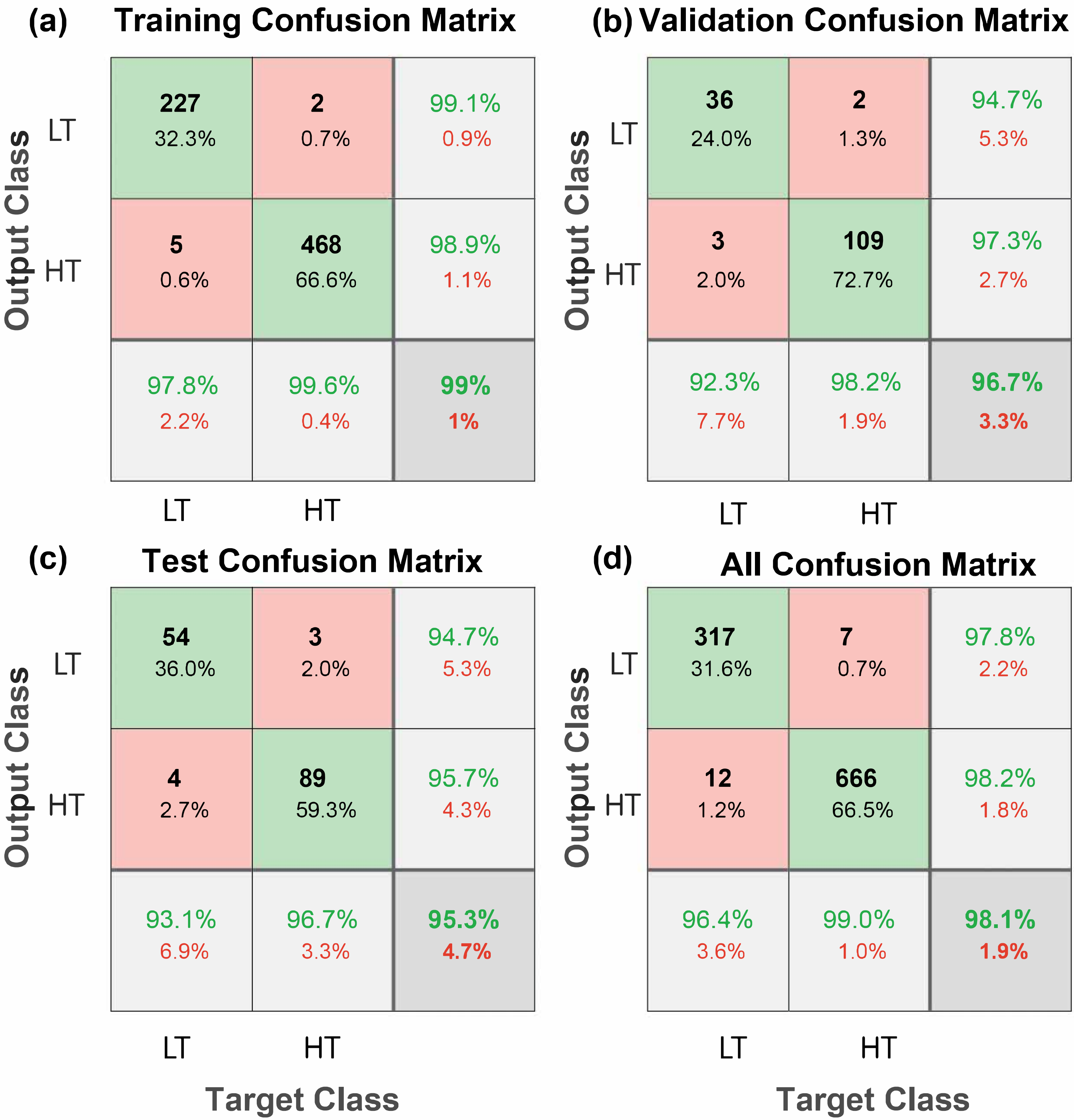
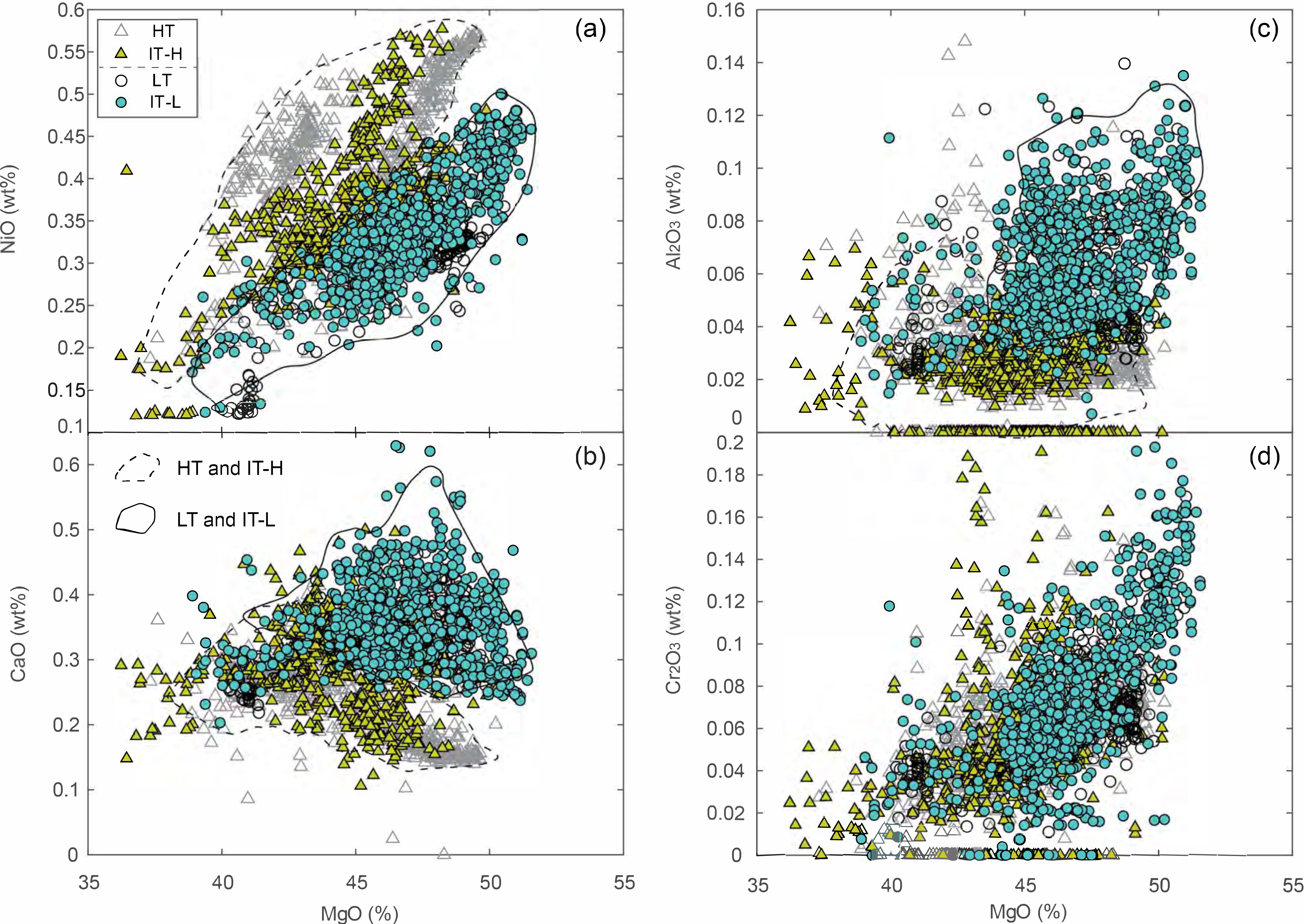

(a)
Training Confusion Matrix
Output Class
LT
HT
227
32.3%
2
0.7%
99.1%
0.9%
5
0.6%
468
66.6%
98.9%
1.1%
97.8%
2.2%
99.6%
0.4%
99%
1%
LT
HT
(b)
Validation Confusion Matrix
Output Class
LT
HT
36
24.0%
2
1.3%
94.7%
5.3%
3
2.0%
109
72.7%
97.3%
2.7%
92.3%
7.7%
98.2%
1.9%
96.7%
3.3%
LT
HT
(c)
Test Confusion Matrix
Output Class
LT
HT
54
36.0%
3
2.0%
94.7%
5.3%
4
2.7%
89
59.3%
95.7%
4.3%
93.1%
6.9%
96.7%
3.3%
95.3%
4.7%
LT
HT
Target Class
(d)
All Confusion Matrix
Output Class
LT
HT
317
31.6%
7
0.7%
97.8%
2.2%
12
1.2%
666
66.5%
98.2%
1.8%
96.4%
3.6%
99.0%
1.0%
98.1%
1.9%
LT
HT
Target Class

Fig 6



(a)
HT
IT-H
LT
IT-L
NiO (wt%)
(b)
HT and IT-H
LT and IT-L
CaO (wt%)
MgO (%)
(c)
Al2O3 (wt%)
(d)
Cr2O3 (wt%)
MgO (%)

Fig 7

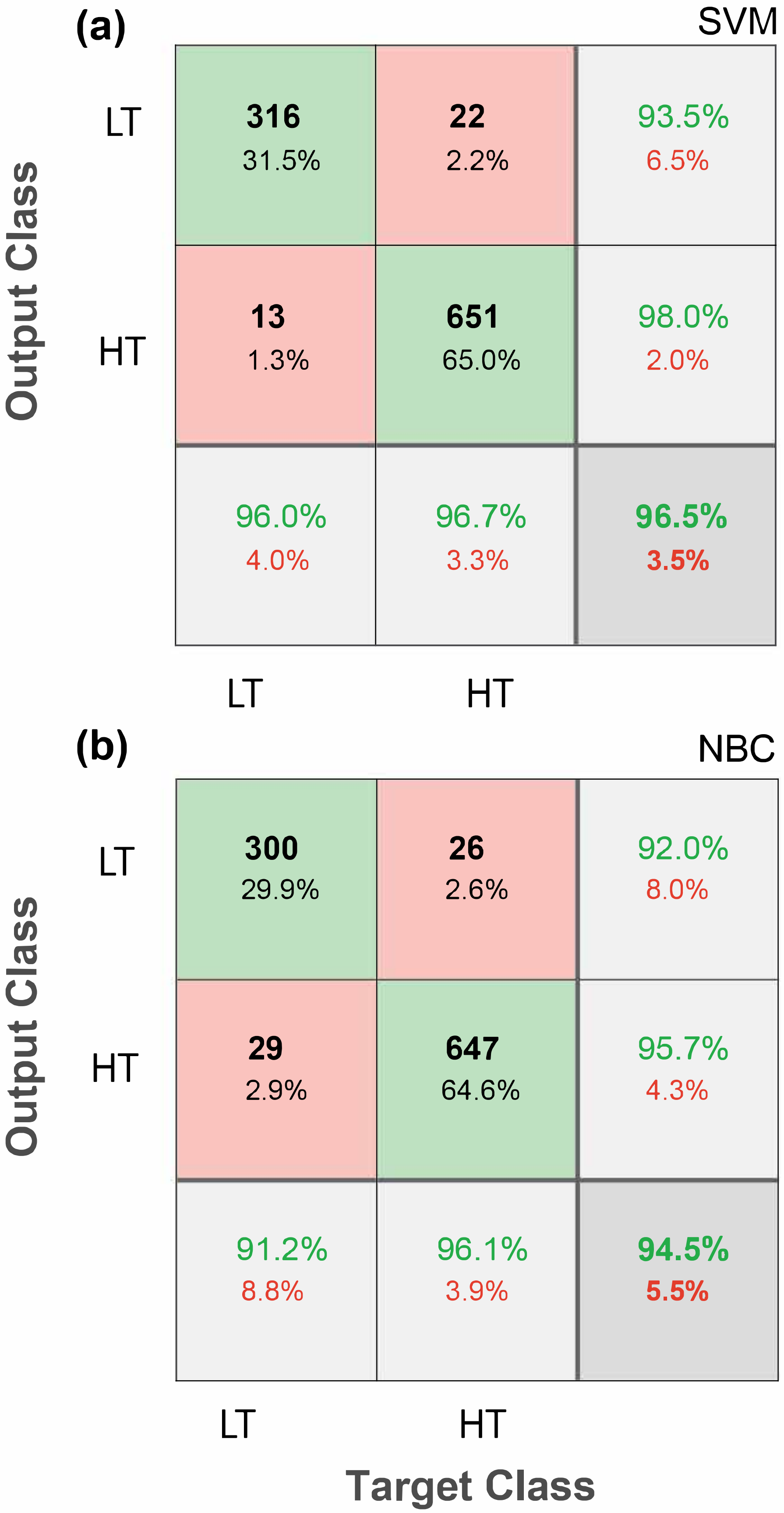
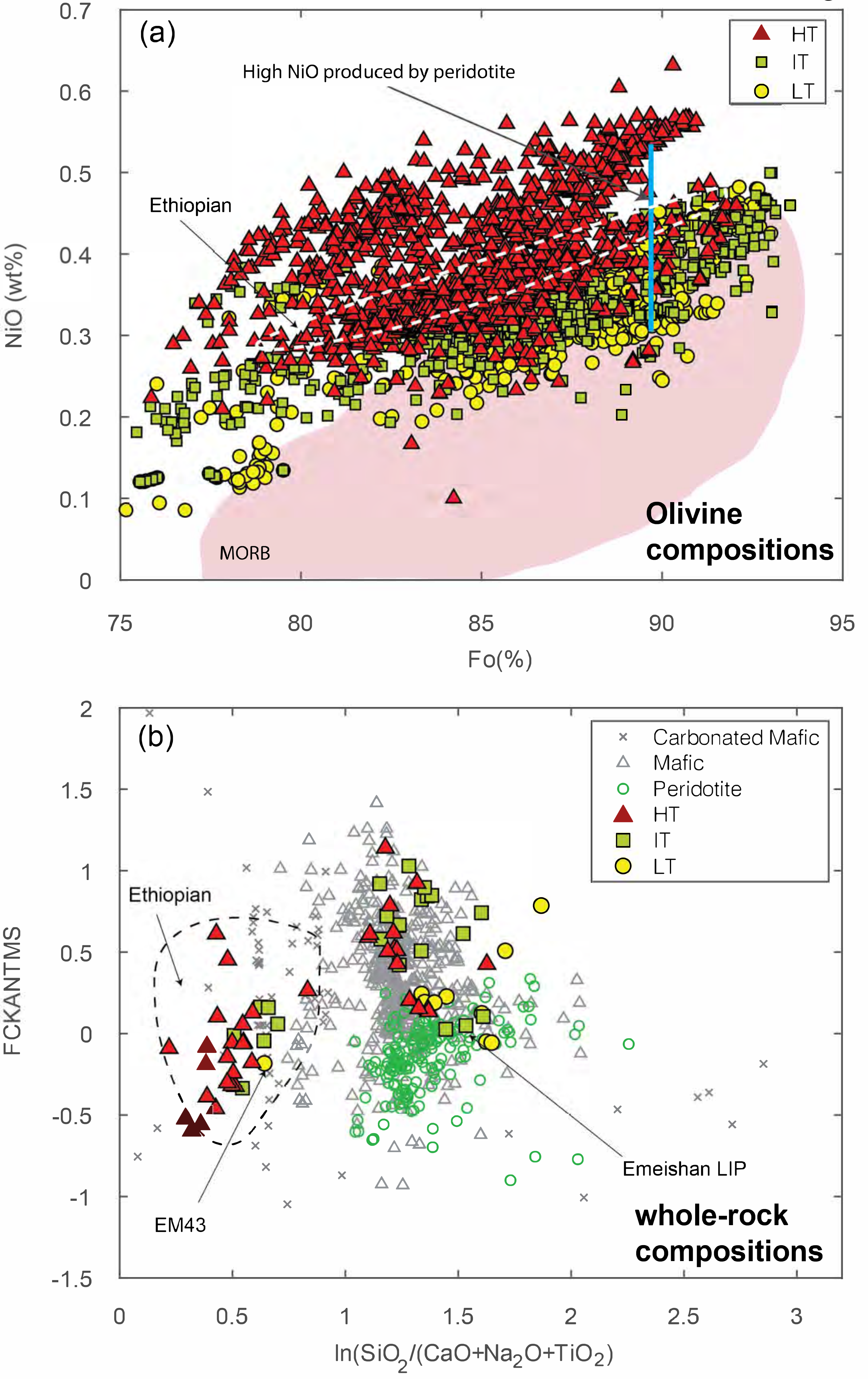
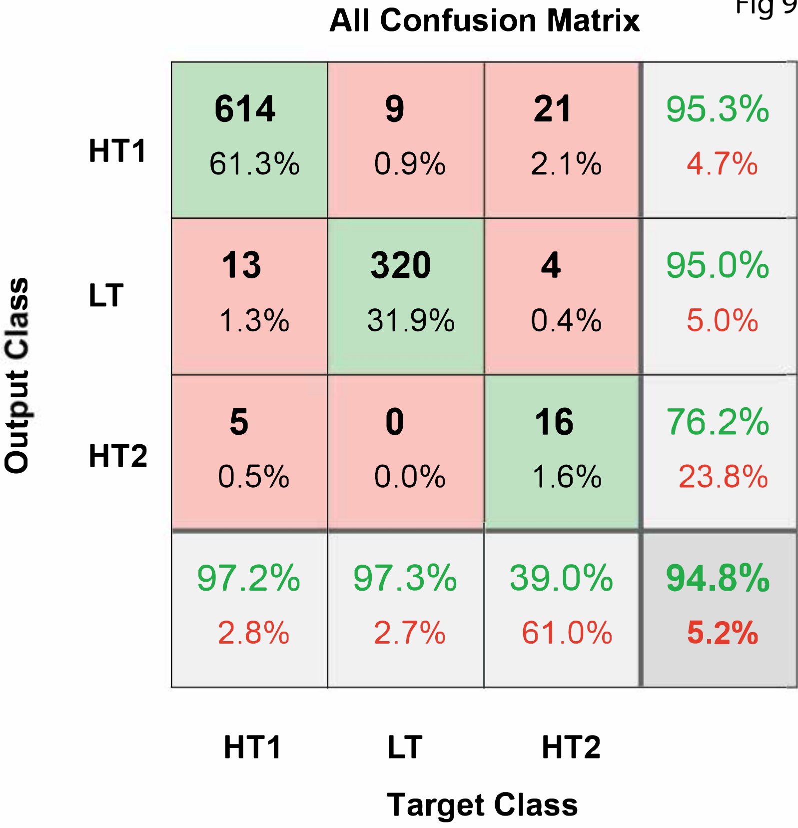

(a)
SVM
Output Class
LT
HT
316
31.5%
22
2.2%
93.5%
6.5%
13
1.3%
651
65.0%
98.0%
2.0%
96.0%
4.0%
96.7%
3.3%
96.5%
3.5%
LT
HT
(b)
NBC
Output Class
LT
HT
300
29.9%
26
2.6%
92.0%
8.0%
29
2.9%
647
64.6%
95.7%
4.3%
91.2%
8.8%
96.1%
3.9%
94.5%
5.5%
LT
HT
Target Class

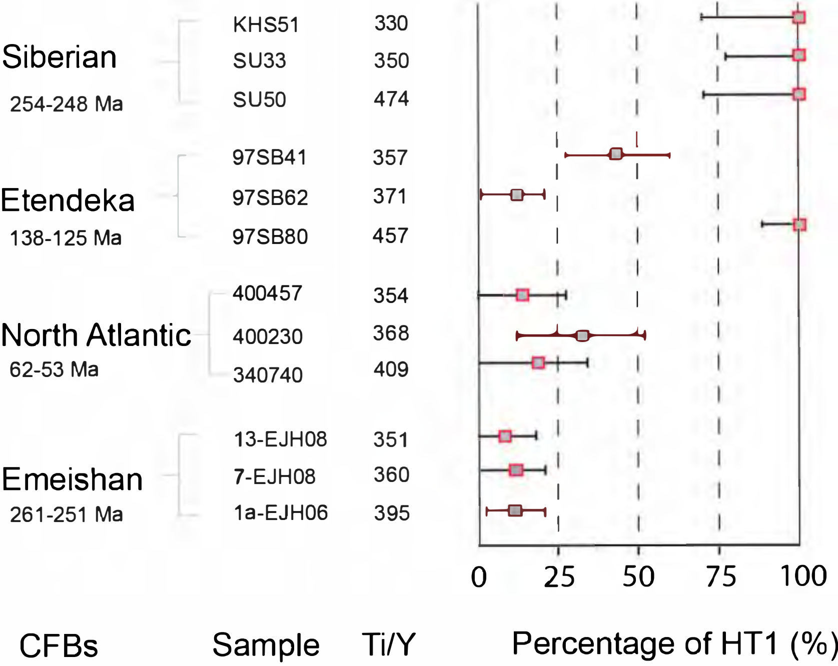
Fig 8



(a)
HT
IT
LT
High NiO produced by peridotite
Ethiopian
MORB
Olivine compositions
NiO (wt%)
Fo(%)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
75
80
85
90
95
(b)
Carbonated Mafic
Mafic
Peridotite
HT
IT
LT
Ethiopian
EM43
Emeishan LIP
whole-rock compositions
FCKANTMS
ln(SiO2/(CaO+Na2O+TiO2)
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
0.5
1
1.5
2
2.5
3

Fig 9


All Confusion Matrix
Output Class
HT1
LT
HT2
614
61.3%
9
0.9%
21
2.1%
95.3%
4.7%
13
1.3%
320
31.9%
4
0.4%
95.0%
5.0%
5
0.5%
0
0.0%
16
1.6%
76.2%
23.8%
97.2%
2.8%
97.3%
2.7%
39.0%
61.0%
94.8%
5.2%
HT1
LT
HT2
Target Class


Siberian
254-248 Ma
KHS51 330
SU33 350
SU50 474
Etendeka
138-125 Ma
97SB41 357
97SB62 371
97SB80 457
North Atlantic
62-53 Ma
400457 354
400230 368
340740 409
Emeishan
261-251 Ma
13-EJH08 351
7-EJH08 360
1a-EJH06 395
0
25
50
75
100
CFBs
Sample
Ti/Y
Percentage of HT1 (%)