

From Machine Learning to Transfer Learning in Laser-Induced Breakdown Spectroscopy: the Case of Rock Analysis for Mars Exploration

Chen Sun

Shanghai Jiao Tong University, School of Physics and Astronomy

Weijie Xu

Shanghai Jiao Tong University, School of Physics and Astronomy

Yongqi Tan

Shanghai Jiao Tong University, School of Physics and Astronomy

Yuqing Zhang

Shanghai Jiao Tong University, School of Physics and Astronomy

Zengqi Yue

Shanghai Jiao Tong University, School of Physics and Astronomy

Long Zou

Shanghai Jiao Tong University, School of Physics and Astronomy

Sahar Shabbir

Shanghai Jiao Tong University, School of Physics and Astronomy

Mengting Wu

Shanghai Jiao Tong University, School of Physics and Astronomy

Fengye Chen

Shanghai Jiao Tong University, School of Physics and Astronomy

Jin Yu (✉ jin.yu@sjtu.edu.cn)

Shanghai Jiao Tong University, School of Physics and Astronomy

Research Article

Keywords: LIBS, Mars, powder pellet, glass or ceramic, TAS

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-400278/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

From Machine Learning to Transfer Learning in Laser-Induced Breakdown Spectroscopy: the Case of Rock Analysis for Mars Exploration

Chen Sun¹, Weijie Xu¹, Yongqi Tan¹, Yuqing Zhang¹, Zengqi Yue¹, Long Zou¹, Sahar Shabbir¹, Mengting Wu¹, Fengye Chen¹, and Jin Yu^{1,*}

¹Shanghai Jiao Tong University, School of Physics and Astronomy, Shanghai 200240, P. R. China

*jin.yu@sjtu.edu.cn

ABSTRACT

With the ChemCam instrument, laser-induced breakdown spectroscopy (LIBS) has successively contributed to Mars exploration by determining elemental compositions of soils, crusts and rocks. American Perseverance landed since Feb 18, 2021 on Mars and Chinese Tianwen 1 planned for landing soon, further increase the number of LIBS instruments on Mars. Such unprecedented situation requires a reinforced research effort on the methods of LIBS spectral data treatment. Although the matrix effects correspond to a general issue in LIBS, they become accentuated in the case of rock analysis for Mars exploration, because of the large variation of rock compositions leading to the chemical matrix effect, and the difference in surface physical properties between laboratory standards (in pressed powder pellet, glass or ceramic) used to establish calibration models and natural rocks encountered on Mars, leading to the physical matrix effect. The chemical matrix effect has been tackled in the ChemCam project with large sets of laboratory standards offering a good representation of various compositions of Mars rocks. The present work more specifically deals with the physical matrix effect which is still expecting a satisfactory solution. The approach consists in introducing transfer learning in LIBS data treatment. For the specific application of total alkali-silica (TAS) classification of rocks (either with a polished surface or in the raw state), the results show a significant improvement of the prediction ability of pellet-based models when trained together with suitable information from rocks in a procedure of transfer learning. The correct TAS classification rate increases from 25% for polished rocks and 33.3% for raw rocks with a machine learning model, to 83.3% with a transfer learning model for the both types of rock samples.

Introduction

It is generally considered that the matrix effects, both the chemical¹ and the physical² matrix effects, represent a critical issue in analysis with laser-induced breakdown spectroscopy (LIBS) for either qualitative classification or quantitative determination.³ Suitable solutions with respect to such consideration become primordially determinant for applications as important as LIBS analysis of rocks in Mars explorations,⁴ where the targeted scientific goals, searching the present and past water activities and the traces of the life as well as studying the habitability of Mars,⁵⁻⁷ rely, at least partially, on the reliability and the exactitude of the analytical data that one can extract from the LIBS spectra recorded by LIBS instruments embarked on Mars rovers.⁸ Certainly the diversity of chemical compositions of Mars rocks has been studied in the precedent missions, the absence of real samples from Mars, except meteors, requires a large number of laboratory rock standard samples to be prepared with Earth natural rocks or by mixing pure chemical compounds, in order to cover the chemical variety of Mars rocks. It was the purpose of the sets of laboratory standard rock samples prepared and used by the ChemCam team for training and validation of Mars LIBS spectral data processing models. The number of the involved samples was first 69,⁹ and was further increased to 408 in order to offer a more complete representation of the chemical and mineral compositions of Mars rocks.¹⁰ It is useful and important to point out that all the above mentioned laboratory rock standards were prepared in the forms of pressed powder pellet, glass, and ceramic to minimize the heterogeneity and the surface roughness of the samples in the scale of LIBS observations of typically several hundred μm . Such sample preparation leads to obvious differences in surface physical properties between laboratory standards and real rocks analyzed by LIBS instruments on Mars, differences from which the physical matrix effect can result. With this concern, the effects of sample surface roughness on the hydrogen emission line has been investigated.¹¹ Our recently published work¹² observed and analyzed the performance of a machine learning-based model¹³ trained with a set of pressed rock powder pellets for total alkali-silica (TAS) classification¹⁴ of rocks in their natural state. A significant degradation of the model prediction performance compared to the prediction for pellet samples has been observed. Such degradation prevents the models trained with laboratory standards from reliable predictions with LIBS spectra acquired on raw rock samples, a situation

that can become disappointing for *in situ* LIBS analysis of rocks on Mars, since we are not yet able to bring materials back from Mars.

In order to search a solution for the raised issue, in this work, transfer learning was introduced in LIBS spectral data treatment to more specifically overcome the physical matrix effect. Transfer learning is considered in machine learning when the knowledge gained while solving one problem is required to be applied to a different but related problem.¹⁵ Its necessity comes from the fact that a major assumption in machine learning data processing is that the training and the model-targeted samples to be analyzed should share the same feature space and have the same distribution.¹⁶ It is unfortunately not the case for the application scenario that we consider. Moreover, transfer learning has recently emerged as a new learning framework to address the problem of insufficient training data in an application (target domain) with the help of the knowledge learnt from a related application having the facility to get sufficient training data (source domain).¹⁷ Such strategy fits well the requirement of LIBS analysis of rocks on Mars, where sufficient laboratory standards can be prepared as the source domain, whereas real Mars rock samples are not yet available as the target domain. Simulations for their chemical as well as physical properties by terrestrial materials, whether natural or artificial, appear therefore a suitable solution. According to the specific contents of the “knowledge” to be transferred, we can distinguish feature-representation-transfer, where parts of relevant features respectively from the both source and target domains are merged and selected for their low sensitivity to the difference between the two domains, to form a common set of features contributing to the training of a transfer learning model.¹⁸ Instance-transfer is another specificity of transfer learning where data of the samples from the both source and target domains participate in the model training, with a conditional testing on the relevance of each sample from the source domain for its effectiveness in improving the performance of the model in a cross-validation process with the data from the target domain.¹⁸ A weight is then applied to each source domain sample participating the training, according to its contribution in improving the performance of the model for predicting with target domain data. We note that algorithms belonging to transfer learning, low rank alignment of manifolds or feature-based transfer learning for example, have been used respectively for calibration transfers between different LIBS instruments¹⁹ or metallic samples with different temperatures²⁰.

More specifically, in our experiment, on the basis of the LIBS spectra acquired from a set of laboratory standard samples in the form of pressed powder pellet, machine learning-based multivariate models were trained and used to predict the concentrations of major oxides necessary for TAS classification of rocks, SiO₂, Na₂O and K₂O, with LIBS spectra acquired from natural rocks. The purposes was first to observe the physical matrix effect due to the difference in surface states between a pressed powder pellet and the rock used for its preparation. Since for a LIBS measurement, such difference can be in particular due to surface hardness, heterogeneity or roughness of the rock, it was measured in its raw state and with a polished surface, in such way that the different contributions to the physical matrix effect can be investigated separately. Transfer learning-based models were then trained with the implementation of feature-representation-transfer and instance-transfer to effectively correct the physical matrix effect in the concentration prediction for rocks in their raw state or prepared with a polished surface, allowing their satisfactory TAS classifications.

Samples, Experimental Setup and Measurement Protocol

Samples. In this work, 20 natural terrestrial rocks were used as samples for LIBS analysis. The rocks were first washed using alcohol and distilled water before any further treatment. All the rocks were analyzed in 3 different forms. Raw rocks: LIBS measurements took place on the natural surface of each rock; Polished rocks: LIBS measurements took place on a polished flat surface of each rock (prepared using a 300-mesh sandpaper); Pellets: a part of each rock was crushed and ground into a powder by a laboratory mill and then sieved by a 300-mesh screen (grain size < 50 μm). A binder (microcrystalline cellulose powder) with a similar particle size was mixed with the rock powder at a weight ratio of 20%. One gram of the obtained powder was pressed under a pressure of 850 MPa for 30 minutes to form a pellet of 15 mm diameter and 2 mm thickness. The composition, with especially the concentrations of major oxides, SiO₂, Na₂O, K₂O, of each rock was determined with X-ray fluorescence spectroscopy (XRF) performed on the pellets. The detailed compositions and the geological names of the rocks are presented in the section “Methods” (Table 3), which allows presenting the rocks in a TAS diagram as shown in Fig. 1. The short notations of the 15 fields (surrounded by circles) are according to Reference 21. Eight rocks were selected as training samples: S3, S7, S8, S11, S13, S14, S18 and S19 (represented by red crosses in the figure). They joined the 20 pellet samples to train the transfer learning models. The rest of the 12 rocks were used as validation samples (represented by blue dots in the figure).

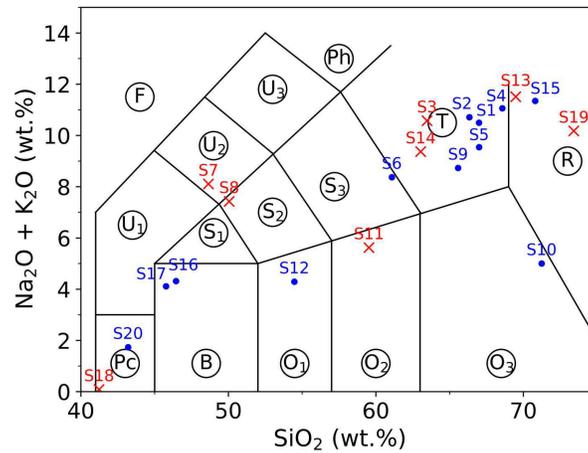


Figure 1. Presentation of the used rock samples in a TAS diagram according to their major oxide concentrations determined using XRF. The short notations of the 15 fields (surrounded by circle) are according to Reference 21. Eight rocks were selected as training samples together with the 20 pellets: S3, S7, S8, S11, S13, S14, S18 and S19 (represented by red crosses in the figure). The rest of the 12 samples were used as validation samples (represented by blue dots in the figure).

Experimental setup. A detailed description of the used experimental setup can be found elsewhere.¹² Briefly as shown in Fig. 2, a Q-switched Nd:YAG laser operated at a wavelength of 1064 nm with a pulse duration of 7 ns and a repetition of rate of 10 Hz, was used to ablate the samples with a pulse energy of 8 mJ. A lens of 50 mm focal length focused laser pulses about 0.86 mm below the surface of a sample. The diameter of the laser spot on the sample surface was estimated to 150 μm , leading to a laser fluence on the sample surface of about 45 J/cm², or an irradiance of about 6.5 GW/cm². The emission from a generated plasma was collected by a combination of two quartz lenses with a same focal length of 75 mm into an optical fiber of 50 μm core diameter. The output of the fiber was connected to the entrance of an echelle spectrometer equipped with a ICCD camera (Mechelle 5000 and iStar, Andor Technology) which provided a wide spectral range from 230 nm to 900 nm with spectral resolution power of 5000. The ICCD camera was triggered by laser pulses and set with a delay and a gate width of respectively 500 ns and 2000 ns. A lateral CCD camera (not shown in the figure) allowed capturing time-integrated plasma images as shown in the inset of Fig. 2. Samples were mounted on a 3D translation stage allowing recording replicate spectra on a sample surface with an ablation crater matrix, while keeping a constant distance between the focusing lens and the sample surface (approximately for a raw rock).

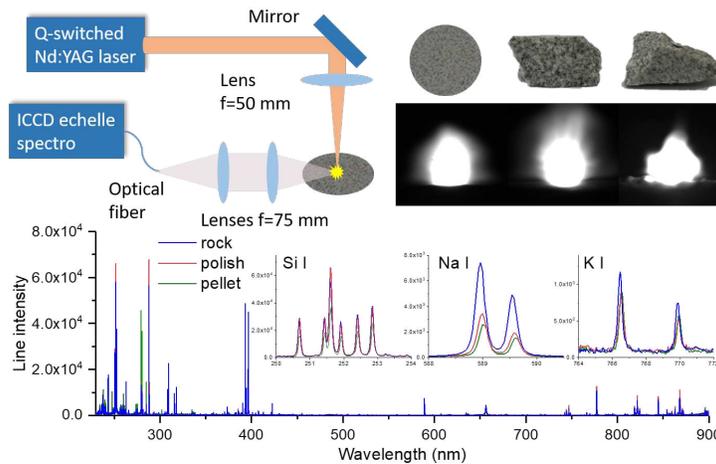


Figure 2. Schematic presentation of the used experimental setup, together with plasma images respectively induced on a pellet, a polished rock and a raw rock, and typical LIBS spectra showing differences in emission intensities of Si, Na and K between a pellet and the corresponding polished and raw rocks.

Spectrum recording. For each sample, 50 replicate spectra were taken on 50 different ablation craters, and each crater received 25 successive laser shots. The emission spectra induced by the first 5 laser shots were removed in order to avoid surface contamination, and those induced by the subsequent 20 laser shots were accumulated to produce a replicate spectrum. Such procedure also intended to reduce the influence of surface roughness for raw rock samples. In total, 3000 spectra were recorded for the 20 rocks with the 3 different forms for each of them. Typical spectra are shown in the inset of Fig. 2 for a pellet and the corresponding polished and raw rock samples, showing different line intensities of Si, Na and K. Such differences correspond well to those observed for plasma images induced on the different forms of a sample.

Data Treatment Method

The general data treatment flowchart used in this work is shown in Fig. 3. The procedure was respectively applied to the couples of sample types, pellets/polished rocks and pellets/raw rocks. Several steps can be distinguished: data pretreatment, feature selection, machine learning (ML) and transfer learning (TL) model trainings, as well as model validation. All the pellet samples were used as the training ones, while the rock samples were separated into a training set containing 8 samples and a validation data set containing 12 samples.

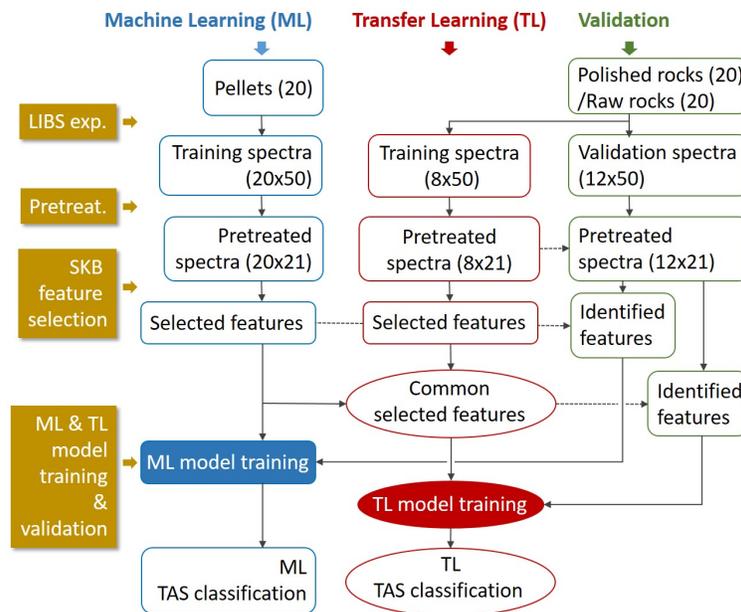


Figure 3. General flowchart used in this work allowing a comparative study between the performances of a machine learning (ML) model and those of a transfer learning (TL) model.

Data pretreatment. The pretreatment consisted in the following operations. i) Average in order to reduce experimental fluctuations and the influence of sample heterogeneity: For each sample, the 50 raw replicate spectra were averaged in a procedure where an averaged spectrum was calculated with a first group of randomly selected 30 spectra. The rest 20 spectra then replaced one by one, a spectrum in the first group, each time the new group of 30 spectra was averaged to generate 20 other average spectra. 21 average spectra were generated for each sample. ii) Baseline correction: an average spectrum was decomposed into a set of cubic spline of undecimated wavelet scales, the local minima were found, then the spline function was interpolated through the different minima to construct the spectral baseline which was removed.²² iii) Normalization: baseline-corrected average spectra were normalized with their respective total intensity (the area under the spectrum). iv) Standardization: standard normal variate (SNV) transformation was respectively applied to the normalized baseline-corrected average spectra of the pellets ($20 \times 21 = 420$ spectra) and the training set of the rock samples ($8 \times 21 = 168$ spectra). Within a given sample set, for each channel in a spectrum (22161 channels in total), the variation range of the intensity value over all the samples was transformed into a range with a mean value equal to 0 and a standard derivation (*SD*) equal to 1. The parameters determined in the standardization of the training set of the rock samples (the mean and the *SD*) were applied to the validation set of the rock samples ($12 \times 21 = 252$ spectra) by assuming a same statistical distribution of the data for all the rock samples. The

ensemble of the above operations generated pretreated spectra.

Spectral feature selection. SelectKBest algorithm²³ was respectively applied to the pretreated spectra of the pellets and the training set of the rocks, and successively for the 3 concerned oxides. Within a sample set, for each spectral channel, covariance was calculated between the channel intensity and the concentration of the concerned compound in the corresponding sample, over all the spectra of the sample set. A score was then calculated as a function of the covariance according to the definition given in Reference 13. A ranking index, $\rho_{i,j}$, was thus associated to each spectral channel according to its obtained score, with 2 index (i, j) and a value varying from 1 to 22161, which ranks the channels from the lowest score to the highest one. Such procedure was applied to the 2 sample sets ($i = 1$: pellets, $i = 2$: calibration set of the rocks) and the 3 concerned oxides ($j = 1$: SiO₂, $j = 2$: Na₂O, $j = 3$: K₂O). A feature selection procedure identified 100 highest ranked spectral channels respectively for each of the 3 oxides in each of the 2 sample sets. Pearson's correlation coefficient²⁴ related to the above mentioned covariance was calculated for the 6 groups of 100 selected features. The results showed that all the selected features had a Pearson's coefficient larger than 0.75.

As we can see in the Fig. 3, the 3 groups of 100 features selected for the 3 oxides for the pellet sample set were directly used to respectively train the calibration models for the 3 oxides base on a back-propagation neural network (BPNN). The training algorithm that involved stochastic gradient descent (SGD) and mini-batch stochastic gradient descent (MSGD) optimization iterations, as well as cross-validations with randomly generated statistically equivalent data configurations, has been presented in detail in Reference 13. We do not in this paper go into more details about such training algorithm.

For transfer learning model training, and according to the principle of feature-representation-transfer discussed above, an ensemble of common selected features was identified between the pellet and the training rock sample sets, by calculating a total ranking index $\rho_j = \rho_{1,j} + \rho_{2,j}$. One hundred highest ranked features according to the value of ρ_j from the highest one to the lowest one, were retained as the common selected features, respectively for the 3 oxides. These groups of features were then fed into the transfer learning model training algorithm. The results of feature selection for Na₂O with the couple of sample types pellet/raw rock, are shown in Fig. 4. Similar behaviors can be observed in the feature selections for the other 2 oxides and with the 2 couples of sample types pellet/polished rock and pellet/raw rock, the corresponding results are shown in the section "Methods" (Figs. 10 and 11).

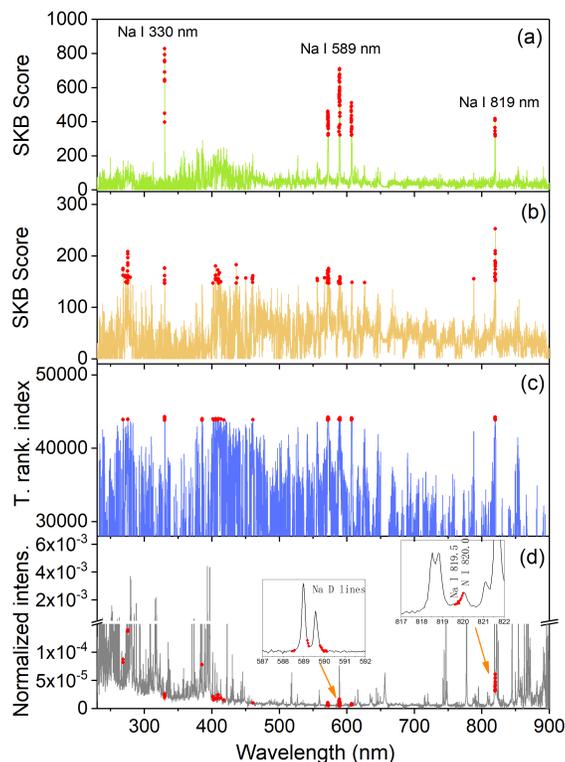


Figure 4. Results of feature selection for Na_2O : (a) for pellets and (b) for calibration raw rocks, SKB scores of all the spectral channels, and with in red dots the 100 selected features; (c) total ranking index of all the spectral channels, together with in red dots those of the 100 common selected features; (d) a typical normalized average spectrum from a pellet sample, together with in red dots the 100 common selected features, with 2 insets showing enlarged parts of the spectrum around 589 nm and 820 nm.

In Fig. 4 (a), we can see that for the pellets, the spectral channels with high SKB scores are clearly concentrated around the several Na emission lines: Na I 330.24 nm and 330.30 nm lines, Na I 588.99 nm and 589.59 nm lines (with 2 groups of ghost lines around 572.1 nm and 606.9 nm), Na I 818.33 nm and 819.48 nm lines. For the calibration raw rocks in Fig. 4 (b), the selected features are distributed also among other channels with a significant decrease of the scores for all the important features. This means that the physical matrix effect perturbs the inherent correlation between the emission line intensities of an element and its concentration in the material, and reduces therefore the importance of the intensities in the concentration determination. At the same time, other spectral channels, such as those around 275 nm and between 410 nm and 460 nm, get relatively higher scores. This means that they become important in the determination of elemental concentration when using a model based on the calibration set of the rock samples. These features, representative of the rock samples, are thus included in the common selected features for transfer learning model training. Fig. 4 (c) shows in red dots, the total ranking index of the 100 common selected features for Na_2O . These features are indicated in a typical spectrum in Fig. 4 (d) in red dots. We can see that, beside the features related to the Na emission lines, some features important for the rock samples are included. A more detailed peak identification using the NIST database,²⁵ shows the contributions from Fe II 268.475 nm and 275.57 nm lines, Si II 385.366 nm and 385.602 nm lines, and the probable contributions from K I 404.414 nm and 404.721 nm lines, Ca I 409.85 nm lines, and Si II 412.807 nm and 413.089 nm lines. A selected feature around 461 nm cannot have easy interpretation.

In the insets of Fig. 4 (d), 2 parts of the spectrum are enlarged. The inset around 589 nm shows the sodium D lines together with the selected features in red dots. We can see that the selected features are located in the side parts of the line profiles, whereas the central parts of the lines are not retained by the feature selection algorithm. This is due to self-absorption of the strong resonant Na D lines, which affects much more the central part of the spectral lines. This observation shows the ability of the feature selection process to reduce the influence of self-absorption. The second inset in Fig. 4(d) shows an enlarged part of the spectrum around 820 nm, where we can see the selected features related to the Na I 819.5 nm line in red dots. Due to the spectral interference with the N I 820.0 nm line, only the short wavelength part of the spectral profile around 820 nm is included

in the selected features, showing the effectiveness of the feature selection to avoid the influence of spectral interference.

Transfer learning-based calibration model training. A transfer learning model training algorithm was developed in this work on the basis of that used for machine learning model training presented in detail in our previous publication¹³ and used in various application scenarios.^{12,26–30} The flowchart of transfer learning model training is shown in Fig. 5. We can distinguish 3 main steps: data formatting, model training by optimization through iteration loops, and model validation. Training was performed for the 2 couples of sample types, pellets/polished rocks and pellets/raw rocks and the 3 concerned oxides.

Optimization and assessment of the models were performed in this work using a certain number of indicators specified below: determination coefficient of a regression r^2 indicating the correlation of the calibration data with the regression model, limit of detection LOD of a model, average relative error of calibration $REC(\%)$ assessing the trueness of a calibration model to be tested, average relative error of test $RET(\%)$ assessing the trueness of a tested model to be validated, average relative error of prediction $REP(\%)$ assessing the trueness of the model-predicted concentrations for the validation samples, average relative standard deviation $RSD(\%)$ assessing the precision of the model-predicted concentrations for the validation samples. The mathematical definitions of these parameters can be found elsewhere, in particular in References 13 and 31. Moreover, root-mean-square error ($RMSE$) is also used to assess the trueness of the model-predicted values for the calibration data ($RMSEC$), cross-validation test data ($RMSET$) and validations data ($RMSEP$).

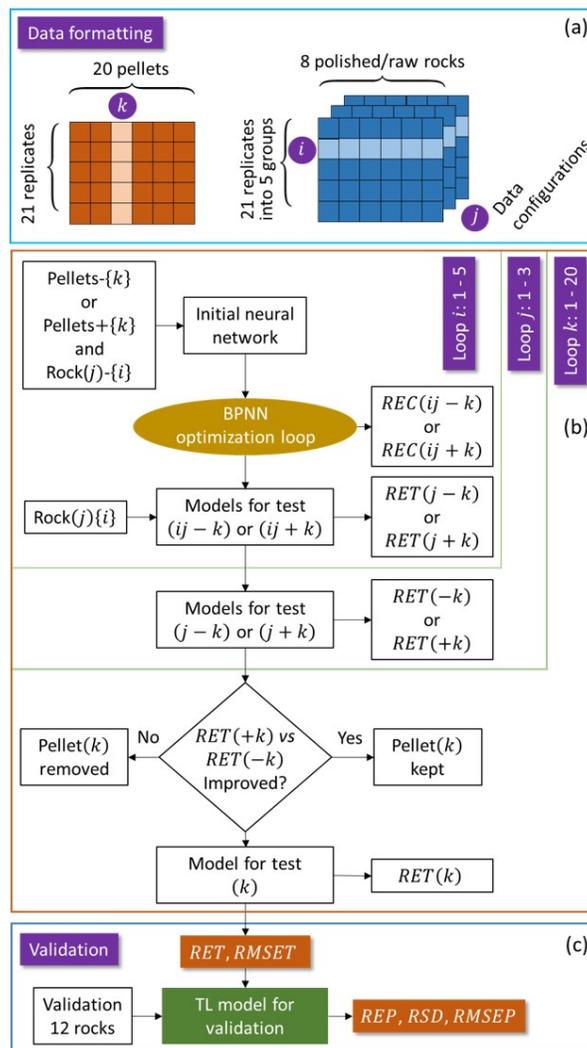


Figure 5. Flowchart of transfer learning model training with the implementation of feature-representation-transfer and instance-transfer.

Data formatting. According to the above discussed principles of feature-representation-transfer and instance transfer in transfer learning, spectra from the pellet samples (the source domain) and those from the calibration set of the rock samples (the target domain), with their 100 common selected features, participated in the training process. All the 20 pellet samples with their pretreated replicate spectra were initially involved in the training data set. These spectra were organized in a given data configuration where the replicate spectra for each sample were arranged in an arbitrarily given order. The effectiveness of each pellet was tested within an iteration loop where the *RET*'s with and without the spectra from the pellet were compared in order to decide the exclusion or the definitive inclusion of the pellet in the transfer learning model training sample set. It was why the ensemble of pretreated replicate spectra associated to one of the 20 pellets was indexed with k that went from 1 to 20 (Fig. 5 (a)). Eight chosen rock samples contributed to the transfer learning model training data set (S3, S7, S8, S11, S13, S14, S18 and S19 in Fig. 1). In particular, they were used in a cross-validation process during the optimization of the neural network. It was why the spectra of this data set were first organized in different data configurations where each configuration j corresponded to a certain arrangement of pretreated replicate spectra for each rock (Fig. 5 (a)). The data configurations were all statistically equivalent since the order of a replicate spectrum of a sample was a dummy index. The number of different data configurations were limited to 3 in this work because more configurations did not bring further improvement of the model as tested in the experiment. For a given configuration j , the pretreated replicate spectra of each sample were further organized into 5 groups containing respectively 4, 4, 4, 4 and 5 spectra, respectively. A new index i was introduced to designate ensemble of the groups of pretreated replicate spectra of all the rock calibration samples as shown in Fig. 5 (a). In the model training process, the index i went from 1 to 5 within an iteration loop of cross-validation, indicating each time the validation ensemble of the groups of pretreated replicate spectra.

Model training by optimization. A 3-layer back-propagation neural network (BPNN) similar to that used in Reference 13 was employed in this work for the transfer learning model. The network was composed of an input layer of 100 neurons corresponding to the 100 common selected features of each input spectrum; a hidden layer 5 neurons and an output layer with a single neuron corresponding to the targeted compound concentration. The function of the network was therefore to map an input spectrum (a vector of 100 dimensions) to a scalar which can be considered as the module of a vector in a hyperspace. The accuracy of the mapping was improved during the training process through different iteration loops under the supervision of the targeted concentration using the model performance indication parameters specified above.

As shown in Fig. 5 (b), 3 hierarchized iteration loops, i, j, k , among them i, j are doubled loops for a given k ($\pm k$) surrounding the BPNN optimization loop, performed the supervised optimization of the model.

- A doubled inner loop for $i = 1$ to 5: for the double cases of a given sample k among the pellet samples being excluded ($-k$) or included ($+k$) in the training data set, and a given data configuration j of the rock spectra, the network was optimized within a cross-validation process where the model was trained using 4 ensemble of groups of replicate spectra, of for example, $i = 2, 3, 4, 5$ with respectively 4, 4, 4 and 5 spectra for each sample. The resulted $REC(ij - k)$ and $REC(ij + k)$ were calculated for the respectively optimized models for test ($ij - k$) and ($ij + k$). These models were then tested using the rest ensemble of groups of replicate spectra, $i = 1$ for instance, generating $RET(j - k)$ and $RET(j + k)$, together with the optimized models for test ($j - k$) and ($j + k$).

- A doubled intermediate loop for $j = 1$ to 3: in this loop, the above discussed loop i was executed with 3 independent rock data configurations for the 2 cases of a given sample k among the pellets being excluded from or included in the training data set. The model was further optimized. Corresponding calculation of *RET* resulted in $RET(-k)$ and $RET(+k)$.

- An outer loop for $k = 1$ to 20: in this loop the above discussed loop i and loop j were executed for each of the 20 pellet samples successively assigned as the pellet k . For a given pellet k , $RET(-k)$ and $RET(+k)$ were compared. If an improvement was observed with the sample, it was kept in the training sample set, otherwise it was rejected. This loop generated a model for test (k) for each considered pellet sample with the corresponding $RET(k)$. The optimization process finally generated a model for validation with a minimized *RET* and *RMSET*.

Model validation. The resulted transfer learning model was validated by the pretreated spectra from the validation set of the rocks (12 samples) with the identified features according to the common selected features between the pellet sample set and the training set of the rock samples. The parameters assessing the performance of the model for prediction, *REP*, *RMSEP* and *RSD* were calculated. These parameters indicate the performance of the model when used for predictions with LIBS spectra from unknown rock samples.

Results and Discussions

Analytical performances with the machine learning model. We first present the results obtained with the models trained with the 20 pellet samples and validated with the 12 validation rock samples respectively for the 3 concerned oxides, SiO_2 , Na_2O and K_2O . The training method described in Reference 13 was implemented in this work to train a neural network. The training procedure was similar to the inner (loop i) and the intermediate (loop j) iteration loops used in the transfer learning model training (Fig. 5 (b)) with a similar neural network structure. As shown in Fig. 3, the input variables were the 100 selected features in a pretreated spectrum of a pellet sample for the training and the 100 identified features in a pretreated spectrum of a validation rock sample for the validation. For the cross-validation optimization in the training process, similarly as for the transfer learning model training, 3×5 -fold iterations were performed with 3 randomly organized pellet spectrum data configurations and 5 replicate groups of respectively 4, 4, 4, 4 and 5 spectra for each sample in a data configuration. A larger number of data configurations did not lead to improvement of the model as shown by tests in our experiment. The training process was executed for the 3 concerned compounds resulting in 3 prediction models respectively for SiO_2 , Na_2O and K_2O . The obtained results are shown in Fig. 6. The extracted parameters for assessment of model performances are presented in Table 1.

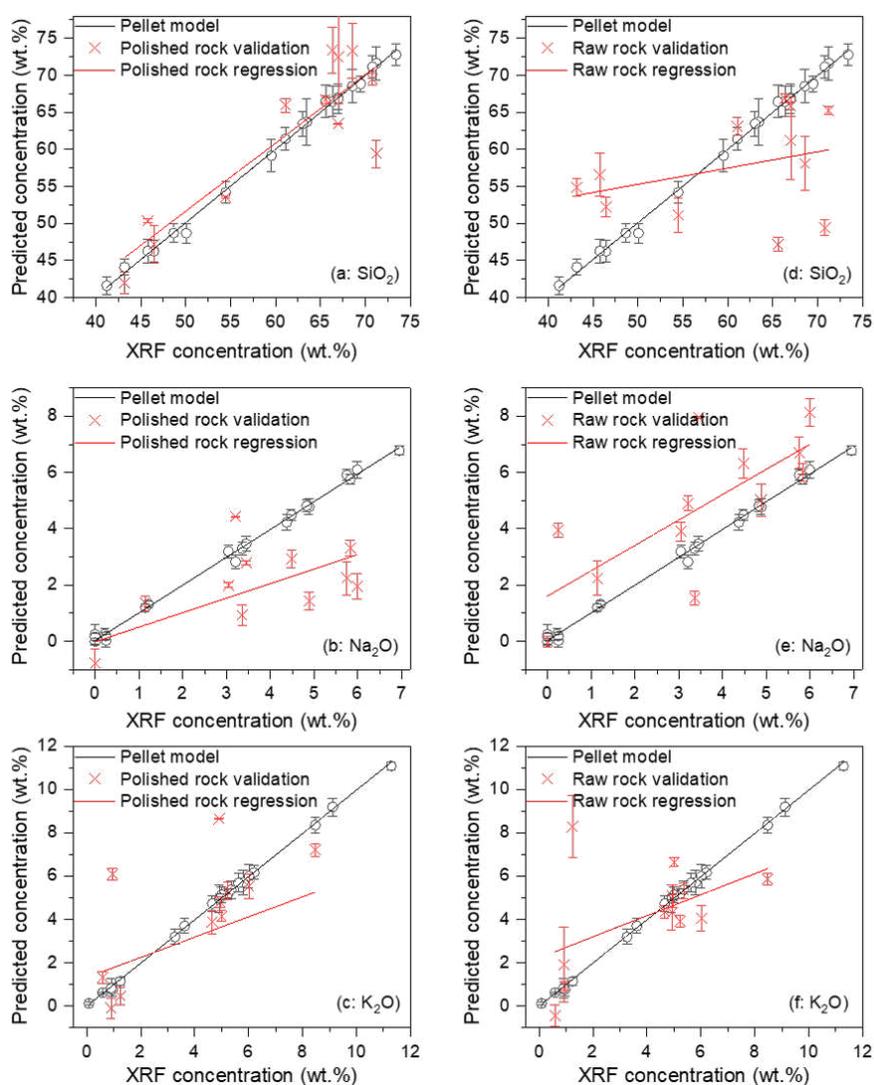


Figure 6. Machine learning-based calibration models trained with the spectra from the pellet samples (black lines) together with the calibration data (black open cycles) for the 3 compounds SiO_2 (a, d), Na_2O (b, e) and K_2O (c, f). Validation data from polished rock validation samples (a, b, c) and raw rock validation samples (d, e, f) are presented in red crosses, their linear regressions in red lines for the 3 compounds SiO_2 (a, d), Na_2O (b, e) and K_2O (c, f). The error bars of the presented data correspond to the standard deviations ($\pm SD$) of the predicted concentrations over the 21 pretreated spectra for a given sample.

In Fig. 6 and Table 1, we can see that the machine learning calibration models trained with the pellet samples present a good performance in terms of the usual assessment parameters including r^2 , LOD , REC , RET , and $RMSE$. This indicates an effective correction of the chemical matrix effect with machine learning regression. At the same time, we can remark a large degradation of the performance when the model was tested using the validation rock samples, in terms of REP , $RMSEP$ and RSD due to the influence of the physical matrix effect. In particular, Fig. 6 shows that use of the pellet models for prediction with the spectra from validation rock samples can lead to bias, with a shift of the linear regression of the validation data with respect to the model, as well as variance, with a rotation of the linear regression of the validation data with respect to the model. We can also remark that the model performance degradation observed with polished rock samples is in general, further aggravated with raw rock samples, also as indicated by Table 1 where we can see increased average REP and $RMSEP$ when one passes from polished rocks to raw rocks. This means that the physical matrix effect arises due to different surface hardness and heterogeneity of a polished rock with respect to its corresponding pressed powder pellet. Surface roughness of a raw rock introduces additional influence leading to in general, larger prediction uncertainties. A detailed look on the validation performances in Table 1 however shows that the influence due to surface roughness remains smaller than that due to surface hardness and heterogeneity, which contributes to the largest part of the physical matrix effect.

As a consequence of the physical matrix effect, the TAS classification of the validation rock samples with the pellet machine learning models led to an unsatisfactory result as shown in Fig. 7. In this figure, the reference position in the TAS diagram of each rock sample determined by the compound concentrations measured using XRF (as shown in Fig. 1 and Table 3) is indicated with a colored solid circular point. With the same color, the position predicted by the pellet machine learning models for the same sample is represented by a cross with error bars. More precisely, the cross represents the mean position calculated with the 21 pretreated replicate spectra. The error bars represent the standard deviations ($\pm SD$) of the concentrations over the replicate spectra, in particular the vertical error bars were obtained by summing the SD s for the 2 concerned compounds. A dash-dot line further links the reference and the predicted positions of a same rock sample in order to explicitly indicate their correspondence. Such presentation thus allows calculating the rate of correct classification. If the pellet model-predicted position of a rock sample stays in the same TAS field as its XRF reference position, it is correctly classified. For the polished rock samples in Fig. 7 (a), we can see 3 correct classifications (S12, S17 and S20), corresponding to a correct classification rate of 25%. For the raw rock samples in Fig. 7 (b), we can see 4 correct classifications (S1, S2, S5 and S10), corresponding to a correct classification rate of 33.3%. Here we can see an even lower rate of correct classification for polished rocks, confirming a dominant contribution to the physical matrix effect by a change of sample surface hardness and heterogeneity, comparing to the influence of surface roughness.

Compound		SiO ₂	Na ₂ O	K ₂ O	Average	
Calibration	Pellets	r^2	0.997	0.996	0.999	0.997
		<i>Slope</i>	0.994	0.988	0.997	0.993
		<i>LOD</i> (%)	5.14	0.62	1.01	2.26
		<i>REC</i> (%)	5.61	9.84	3.75	6.40
		<i>RET</i> (%)	7.42	10.5	5.20	7.71
		<i>RMSEC</i> (wt.%)	1.61	0.84	0.75	1.07
		<i>RMSET</i> (wt.%)	1.92	0.51	0.44	0.96
Validation	Polished rocks	<i>REP</i> (%)	6.27	98.9	103	69.4
		<i>RMSEP</i> (wt.%)	3.95	1.92	2.44	2.77
		<i>RSD</i> (%)	182	28.5	56.9	89.1
	Raw rocks	<i>REP</i> (%)	13.8	176	82.3	90.7
		<i>RMSEP</i> (wt.%)	8.11	1.57	1.48	3.72
		<i>RSD</i> (%)	21.9	9.44	272	101

Table 1. Parameters assessing the calibration and prediction performances of the machine learning calibration models for SiO₂, Na₂O and K₂O.

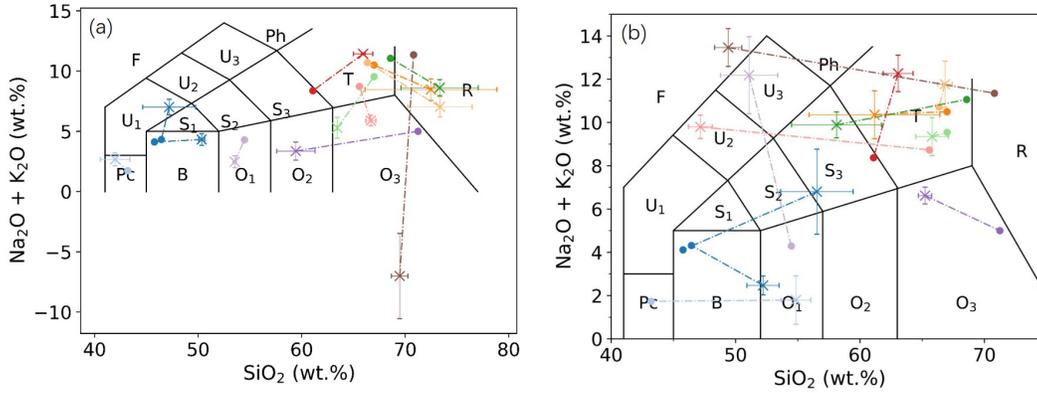


Figure 7. TAS classifications of the validation polished rock samples (a) and raw rock samples (b) using machine learning models. The positions determined by XRF are presented in colored solid circles, the corresponding model-predicted positions are presented in the same color in crosses. A dashed line links the XRF reference position and the model-predicted position of a same rock sample. The error bars on the predicted position are calculated over the different pretreated replicate spectra of a given sample.

Analytical performances with the transfer learning model. The calibrations models resulted from transfer learning are shown in Fig. 8 with a similar format as those resulted from machine learning, in order to review the improvements by comparison. The extracted parameters for assessment of the model performances are presented in Table 2. In Table 2, we can see that although the transfer learning models present in general, a slightly lower calibration performance in terms of r^2 , LOD , REC , RET and $RMSE$, the prediction performance for polished and raw rock samples are significantly improved, especially for REP and $RMSEP$. This means that the participation of the 8 rock samples in the training data set together with the retained pellet samples with common selected features, effectively takes into account the physical matrix effect and reinforces the robustness of the models for prediction for rock samples. We remark in particular, the prediction performances for the both polished and raw rocks are simultaneously improved, showing the effectiveness of the transfer learning models in the correction of physical matrix effects of different origins. Correspondingly in Fig. 8, we can see significant reductions of bias and variance of the predicted concentrations for the validation rock samples with respect to the transfer calibration models trained with a part of the pellet samples and the training rock samples. In particular, for SiO_2 , 13 pellet samples were retained in the training data set among the 20 ones by the optimization loop during the model training process to combine with the training polished rock samples, while this number becomes 18 when combined with the training raw rock samples. The retained pellet samples were respectively 14 and 13, and 15 and 14 for Na_2O and K_2O to combine respectively with the training polished and raw rock samples.

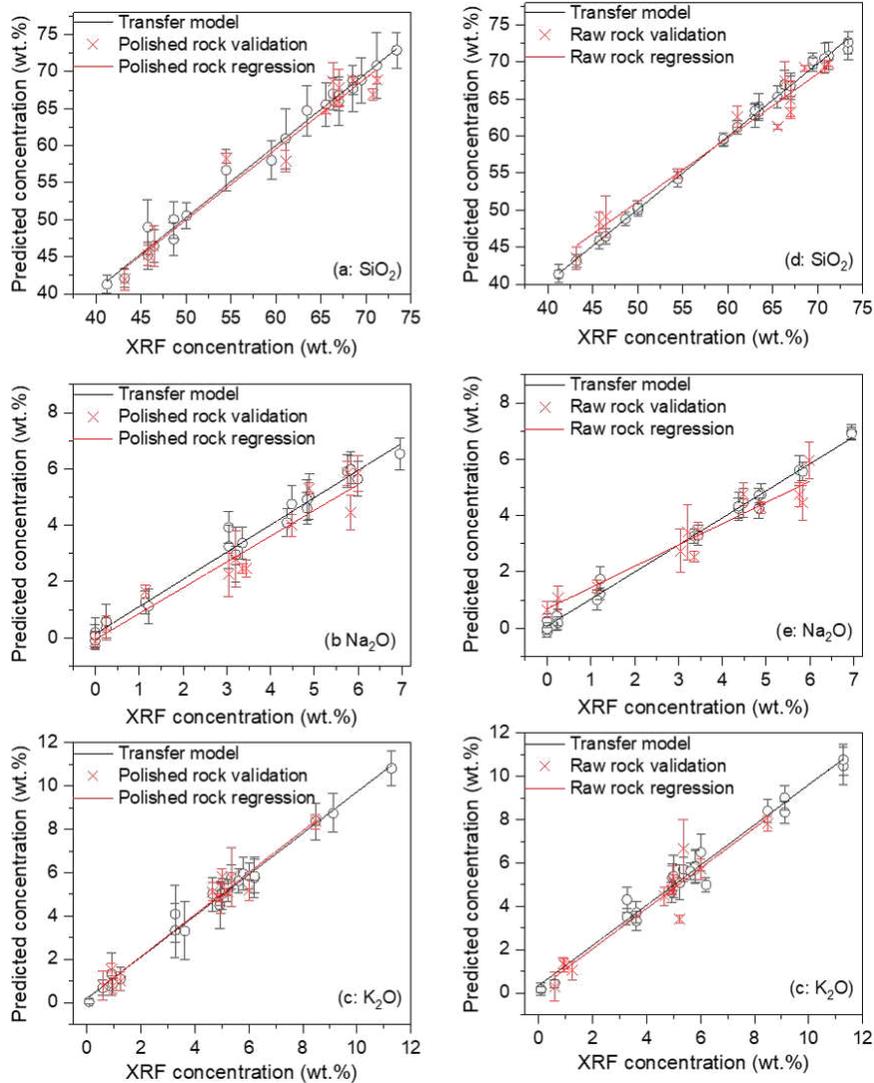


Figure 8. Transfer learning-based calibration models trained with the pellet samples and the training set of the rock samples (black lines) together with the calibration data (black open cycles) for the 3 compounds SiO_2 (a, d), Na_2O (b, e) and K_2O (c, f). Validation data from the polished rock samples (a, b, c) and the raw rock samples (d, e, f) are presented in red crosses, their linear regressions in red lines. The error bars of the presented data correspond to the standard deviations ($\pm SD$) of the predicted concentrations over the 21 pretreated spectra of a given sample.

The calibration models shown in Fig. 8 were used to present the validation rock samples in a TAS diagram. The obtained results are shown in Fig. 9 (a) for polished rock samples and Fig. 9 (b) for raw rock samples using the same symbols as in Fig. 7. We can see a significantly improved result conforming the improved performances of the transfer learning models shown in Fig. 8 and Table 2. A detailed counting shows 10 correctly classified validation samples for the both polished and raw rocks. Only two samples were classified into a wrong field (S12 and S15 for polished rocks, S4 and S6 for raw rocks). The rate of correct classification can thus be determined to be 83.3% in the both cases. These results show the effectiveness of the developed method to correct the physical matrix effect.

Compound		SiO ₂	Na ₂ O	K ₂ O	Average
Polished rocks	r^2	0.994	0.985	0.994	0.993
	<i>Slope</i>	0.989	1.017	0.993	0.999
	<i>LOD</i> (%)	4.53	2.03	2.45	3.00
	Calibration <i>REC</i> (%)	1.52	15.1	10.4	9.01
	<i>RET</i> (%)	1.90	15.71	9.72	9.11
	<i>RMSEC</i> (wt.%)	2.61	0.75	0.80	1.39
	<i>RMSET</i> (wt.%)	2.40	0.62	0.82	1.28
	Validation <i>REP</i> (%)	2.76	20.6	16.4	13.3
	<i>RMSEP</i> (wt.%)	1.68	0.53	0.35	0.85
	<i>RSD</i> (%)	128	45.8	47.5	73.8
Raw rocks	r^2	0.998	0.993	0.997	0.996
	<i>Slope</i>	0.985	0.961	0.993	0.980
	<i>LOD</i> (%)	3.70	1.03	2.26	2.33
	Calibration <i>REC</i> (%)	5.61	15.7	6.40	9.24
	<i>RET</i> (%)	4.90	16.2	7.72	9.60
	<i>RMSEC</i> (wt.%)	2.44	0.75	0.88	1.36
	<i>RMSET</i> (wt.%)	2.37	0.49	0.72	1.19
	Validation <i>REP</i> (%)	3.11	42.0	19.8	21.6
	<i>RMSEP</i> (wt.%)	1.85	0.54	0.50	0.96
	<i>RSD</i> (%)	24.1	21.3	101	48.8

Table 2. Parameters assessing the calibration and prediction performances of the transfer learning calibration models for SiO₂, Na₂O and K₂O.

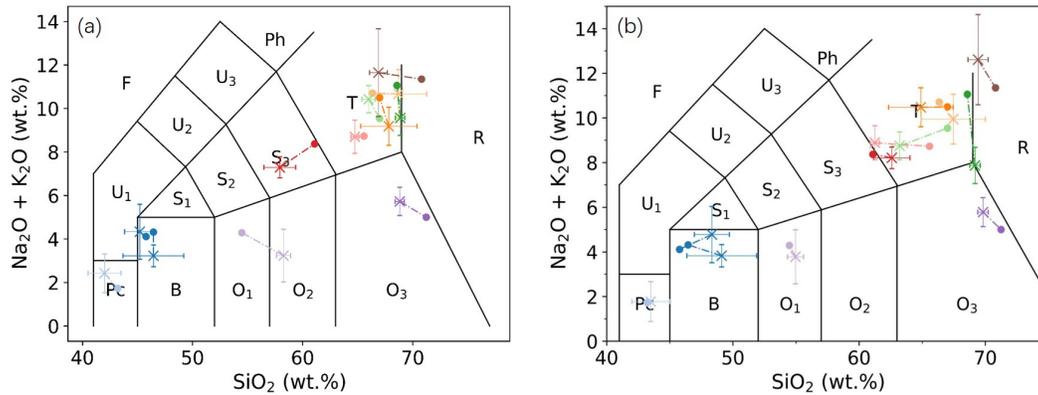


Figure 9. TAS classification using transfer learning models of the validation rock samples, (a) for polished rocks and (b) for raw rocks. The same symbols are used as in Fig. 7.

Conclusions

In this work, within a specific application of classification of rocks using the TAS diagram, we have introduced transfer learning in LIBS spectral data treatment to improve the performance of the models trained using laboratory standard samples in the form of pressed powder pellet, when used for prediction with LIBS spectra acquired from natural rocks. Such scenario corresponds to, for example, important applications of analysis of rocks with LIBS for Mars exploration. In particular, feature-representation-transfer and instance-transfer as the two important processes of transfer learning were implemented in the LIBS spectral data treatment. The performances of the transfer learning models were compared with those of the machine learning models. Significant improvements have been realized for prediction with LIBS spectra acquired on polished and raw rock samples for the 3 concerned compounds involved in the TAS classification, SiO₂, Na₂O and K₂O. The rate of correct TAS classification has been improved from 25% for polished rocks and 33.3% for raw rocks with machine learning models to 83.3% for the both types of samples with transfer learning models. Beyond analysis of rock with LIBS in Mars exploration, our findings in this work can also have more general interests in the development of LIBS technique for various applications

involving sets of samples with different surface physical properties.

Methods

Detailed chemical compositions of the samples.

The detailed chemical compositions of the rocks studied in this work are shown in Table 3.

Geological name	Sample name	Oxide compositions (wt. %)								Field in TAS (cf. Fig. 1)
		SiO ₂	Fe ₂ O ₃	Al ₂ O ₃	CaO	MgO	K ₂ O	Na ₂ O	Others	
Granite-Porphry	S1	66.99	2.39	16.04	1.98	0.74	6.01	4.48	1.37	T
Monzonitic-Granite	S2	66.34	1.94	17.11	2.19	0.50	4.96	5.75	1.21	T
Plagiogranite	S3	63.47	1.98	18.79	3.17	0.52	3.63	6.95	1.49	T
Moyite	S4	68.57	1.45	16.50	1.42	0.24	5.23	5.83	0.76	T
Biotite Admellite	S5	66.99	2.09	16.18	2.93	1.17	4.66	4.88	1.10	T
Shiyang Syenite	S6	61.08	6.85	14.23	5.05	1.66	4.92	3.45	2.76	S ₃
Granogabbro	S7	48.65	13.72	14.84	5.89	3.84	3.27	4.84	4.95	U ₂
Basaltglass	S8	50.05	15.00	14.82	6.24	3.73	6.21	1.22	2.73	S ₂
Mudstone	S9	65.57	3.81	18.33	0.31	2.18	8.48	0.25	1.07	T
Shale	S10	71.24	3.29	15.94	1.07	1.44	5.00	0.00	2.02	O ₃
Aleuritic texture shale	S11	59.53	13.62	17.32	0.49	1.57	5.62	0.00	1.85	O ₂
Packsand	S12	54.47	4.88	11.41	18.52	4.43	1.24	3.05	2.00	O ₁
Siltstone	S13	69.48	1.31	15.05	0.37	0.50	11.29	0.22	1.78	R
Marl	S14	63.04	5.57	17.73	0.91	2.05	9.12	0.25	1.33	T
Conglomerate	S15	70.80	1.41	14.87	0.79	0.16	5.36	5.99	0.62	R
Gabbro	S16	46.45	11.04	18.15	10.73	7.96	0.96	3.36	1.35	B
Angle flash green mudstone	S17	45.77	17.77	13.58	8.66	6.08	0.91	3.21	4.02	B
Module rock	S18	71.01	1.57	16.88	0.29	0.38	5.98	3.22	0.67	R
Red granite	S19	73.42	1.09	13.55	0.64	0.62	5.80	4.37	0.51	R
Pyroxenite	S20	43.20	15.92	8.48	14.09	14.09	0.59	1.14	2.49	Pc

Table 3. Composition (in wt.%) of the samples used in the experiment determined using XRF and their corresponding TAS field.

Feature Selection.

The results of feature selection for K₂O and SiO₂ for the couple of sample types pellet/raw rocks, are shown in Fig. 10. The results of feature selection for Na₂O, K₂O and SiO₂ for the couple of sample types pellet/polished rocks, are shown in Fig. 11.

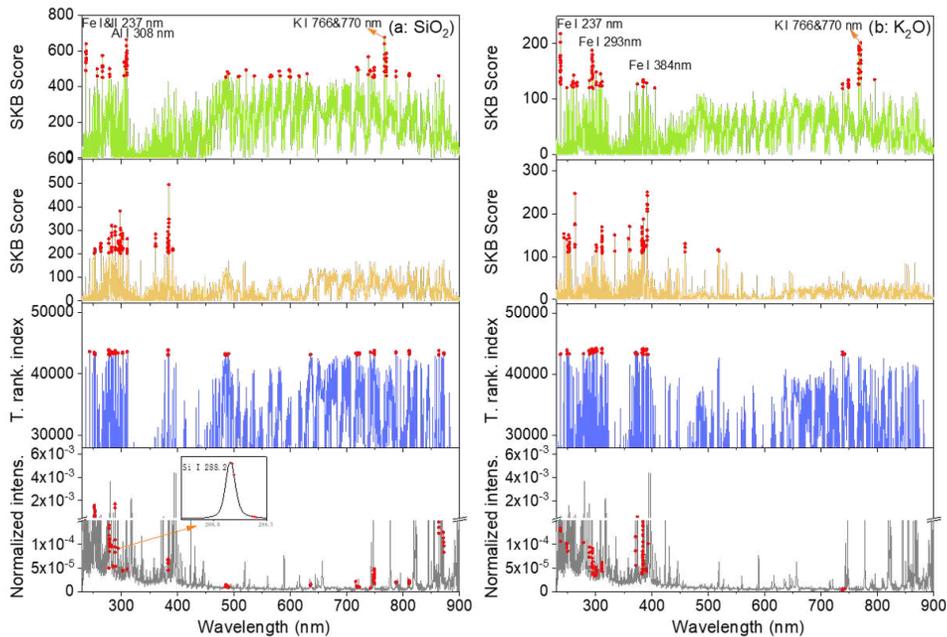


Figure 10. Results of feature selection for the couple of sample types pellet/raw rocks: (a) for SiO_2 and (b) for K_2O . A detailed caption can be found with Figure 4.

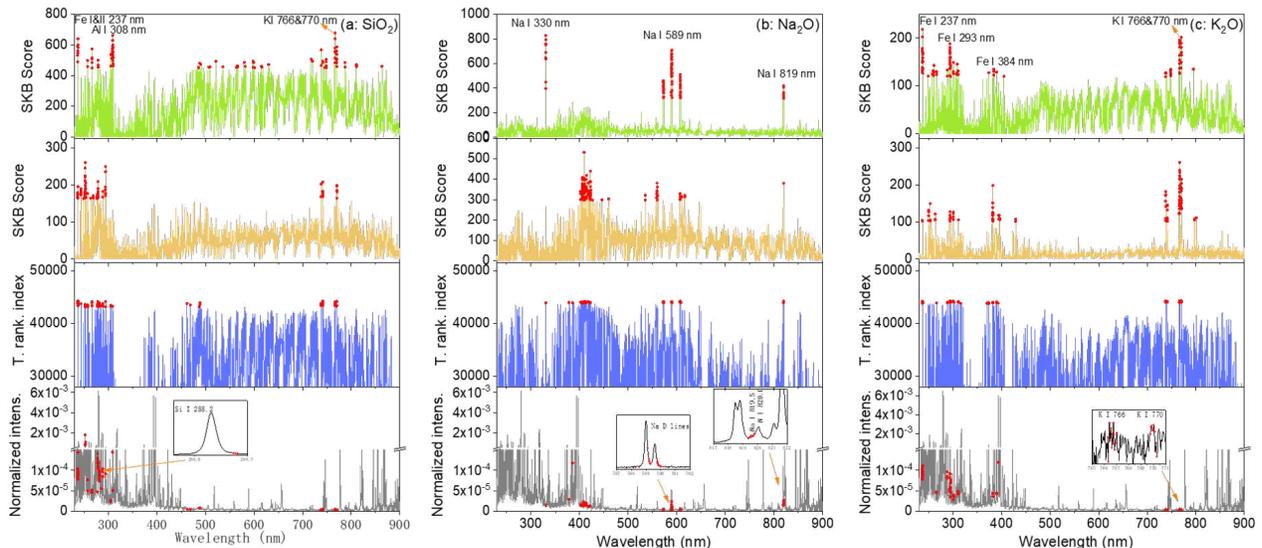


Figure 11. Results of feature selection for the couple of sample types pellet/polished rocks: (a) for SiO_2 , (b) for Na_2O and (c) for K_2O . A detailed caption can be found with Figure 4.

Software.

The program was written using Python version 3.7 and the BPNN was implemented using Keras framework based on TensorFlow 1.14.0. Scikit-learn and NumPy were used. In addition, Origin Pro 8.0 (Origin Lab Corporation, Northampton, MA, USA) was used to draw the figures. All the programs were run on a PC (CPU: Intel(R) Xeon(R) Silver 4214 @2.20 GHz, RAM: 128.00

References

1. Zaytsev, S. M., Krylov, I. N., Popov, A. M., Zorov, N. B. & Labutin, T. A. Accuracy enhancement of a multivariate calibration for lead determination in soils by laser induced breakdown spectroscopy. *Spectrochim. Acta B* **140**, 65–72, DOI: [10.1016/j.sab.2017.12.005](https://doi.org/10.1016/j.sab.2017.12.005) (2018).
2. Segnini, A., Xavier, A., Otaviani-Junior, P., Ferreira, E. & et. al. Physical and chemical matrix effects in soil carbon quantification using laser-induced breakdown spectroscopy. *Am. J. Anal. Chem.* **5**, 722–729 (2014).
3. Hahn, D. W. & Omenetto, N. Laser-induced breakdown spectroscopy (libs), part ii: Review of instrumental and methodological approaches to material analysis and applications to different fields. *Appl. Spectrosc.* **66**, 347–419, DOI: [10.1366/11-06574](https://doi.org/10.1366/11-06574) (2012).
4. Castelvechchi, D. The science events to watch in 2020. *Nature* **577**, 15–16 (2020).
5. Grotzinger, J. P., Crisp, J., Vasavada, A. R., Anderson, R. C. & et. al. Mars science laboratory mission and science investigation. *Space Sci. Rev.* **170**, 5–56, DOI: [10.1007/s11214-012-9892-2](https://doi.org/10.1007/s11214-012-9892-2) (2012).
6. Meslin, P. Y., Gasnault, O., Forni, O., Schroeder, S. & et. al. Soil diversity and hydration as observed by chemcam at gale crater, mars. *Science* **341**, 1238670, DOI: [10.1126/science.1238670](https://doi.org/10.1126/science.1238670) (2013).
7. Grotzinger, J. P., Sumner, D., Kah, L. & et. al. A Habitable Fluvio-Lacustrine Environment at Yellowknife Bay, Gale. *Science* **343**, 1242777, DOI: [10.1126/science.aac7575](https://doi.org/10.1126/science.aac7575) (2014).
8. Maurice, S., Wiens, R. C., Saccoccio, M., Barraclough, B. & et. al. The chemcam instrument suite on the mars science laboratory (msl) rover: Science objectives and mast unit description. *Space Sci. Rev.* **170**, 95–166, DOI: [10.1007/s11214-012-9912-2](https://doi.org/10.1007/s11214-012-9912-2) (2012).
9. Wiens, R. C., Maurice, S., Lasue, J., Forni, O. & et. al. Pre-flight calibration and initial data processing for the Chem Cam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover. *Spectrochim. Acta B* **82**, 1–27, DOI: [10.1016/j.sab.2013.02.003](https://doi.org/10.1016/j.sab.2013.02.003) (2013).
10. Clegg, S. M., Wiens, R. C., Anderson, R., Forni, O. & et. al. Recalibration of the Mars Science Laboratory ChemCam instrument with an expanded geochemical database. *Spectrochim. Acta B* **129**, 64–85, DOI: [10.1016/j.sab.2016.12.003](https://doi.org/10.1016/j.sab.2016.12.003) (2017).
11. Rapin, W., Bousquet, B., Lasue, J., Meslin, P.-Y. & et. al. Roughness effects on the hydrogen signal in laser-induced breakdown spectroscopy. *Spectrochim. Acta B* **137**, 13–22, DOI: <https://doi.org/10.1016/j.sab.2017.09.003> (2017).
12. Xu, W., Sun, C., Tan, Y., Gao, L. & et. al. Total alkali silica classification of rocks with libs: influences of the chemical and physical matrix effects. *J. Anal. At. Spectrom.* **35**, 1641–1653, DOI: [10.1039/d0ja00157k](https://doi.org/10.1039/d0ja00157k) (2020).
13. Sun, C., Tian, Y., Gao, L., Niu, Y. & et. al. Machine learning allows calibration models to predict trace element concentration in soils with generalized libs spectra. *Sci. Reports* **9**, 11363, DOI: [10.1038/s41598-019-47751-y](https://doi.org/10.1038/s41598-019-47751-y) (2019).
14. Le Bas, M., Le Maitre, R., Streckeisen, A. & Zanettin, B. A chemical classification of volcanic rocks based on the total alkali-silica diagram. *Petrol, J.* **3**, 745–750, DOI: [10.1180/minmag.1986.050.356.01](https://doi.org/10.1180/minmag.1986.050.356.01) (1986).
15. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J. & et. al. Top 10 algorithms in data mining. *Knowledge and Information Systems* **14**, 1–37, DOI: [10.1007/s10115-007-0114-2](https://doi.org/10.1007/s10115-007-0114-2) (2008).
16. Dai, W., Yang, Q., Xue, G. & Yu, Y. Boosting for transfer learning. *ICML 07: Proc. 24th international conference on Mach. learning* 193–200 (2007).
17. Dai, W., Xue, G.-R., Yang, Q. & Yu, Y. Co-clustering based classification for out-of-domain documents. *Proc. 13th ACM SIGKDD Int. Conf. on Knowl. Discov. Data mining* DOI: [10.1145/1281192.1281218](https://doi.org/10.1145/1281192.1281218) (2007).
18. Pan, S. J. & Yang, Q. A Survey on Transfer Learning. *IEEE TRANSACTIONS on Knowledge and Data engineering* **22**, 1345–1359, DOI: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191) (2010).
19. Boucher, T., Carey, C. J., Mahadevan, S. & Dyar, M. D. Aligning Mixed Manifolds. *Proceedings of the twenty-ninth AAAI Conference On Artificial Intelligence* 2511–2517 (2015).
20. Yang, J., Li, X., Lu, H., Xu, J. & Li, H. An libs quantitative analysis method for alloy steel at high temperature based on transfer learning. *J. Anal. At. Spectrom.* **33**, 1184–1195, DOI: [10.1039/C8JA00069G](https://doi.org/10.1039/C8JA00069G) (2018).
21. Le Maitre, R. W. *A Classification of Igneous Rocks and Glossary of Terms*, ed. (Blackwell Scientific Publications, Oxford, 1989).

22. Zhang, Z.-M., Chen, S., Liang, Y.-Z., Liu, Z.-X. & et. al. An intelligent background-correction algorithm for highly fluorescent samples in raman spectroscopy. *J. Raman Spectrosc.* **41**, 659–669, DOI: [10.1002/jrs.2500](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.2500) (2010). <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrs.2500>.
23. Cormen, T., Leiserson, C. & Rivest, R. *Introduction to Algorithms* (The MIT Press, Cambridge, MA, USA, 2009).
24. P. Bruce, P. & Bruce, A. *Practical Statistics for Data Scientists* (O'Reilly Media, Inc., 2017).
25. <https://www.nist.gov/pml/atomic-spectra-database>.
26. Zhang, Y., Sun, C., Gao, L., Yue, Z. & et. al. Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms. *Spectrochim. Acta B* **166**, 105802, DOI: [10.1016/j.sab.2020.105802](https://doi.org/10.1016/j.sab.2020.105802) (2020).
27. Yue, Z., Sun, C., Gao, L., Zhang, Y. & et. al. Machine learning efficiently corrects libs spectrum variation due to change of laser fluence. *Opt. Express* **28**, 14345–14356, DOI: [10.1364/OE.392176](https://doi.org/10.1364/OE.392176) (2020).
28. Zhang, Y., Sun, C., Yue, Z., Shabbir, S. & et. al. Correlation-based carbon determination in steel without explicitly involving carbon-related emission lines in a libs spectrum. *Opt. Express* **28**, 32019–32032, DOI: [10.1364/OE.404722](https://doi.org/10.1364/OE.404722) (2020).
29. Zou, L., Sun, C., Wu, M., Zhang, Y. & et. al. Online simultaneous determination of h₂o and kcl in potash with libs coupled to convolutional and back-propagation neural networks. *J. Anal. At. Spectrom.* **36**, 303–313, DOI: [10.1039/D0JA00431F](https://doi.org/10.1039/D0JA00431F) (2021).
30. Yue, Z., Sun, C., Chen, F., Zhang, Y. & et. al. Machine learning-based libs spectrum analysis of human blood plasma allows ovarian cancer diagnosis. *Biomed. Opt. Express* **12**, 2559–2574, DOI: [10.1364/BOE.421961](https://doi.org/10.1364/BOE.421961) (2021).
31. Sirven, J., Bousquet, B., Canioni, L., Sarger, L. & et. al. Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis. *Anal. Bioanal. Chem.* **385**, 256–262, DOI: [10.1007/s00216-006-0322-8](https://doi.org/10.1007/s00216-006-0322-8) (2006).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grants 11574209, 11805126, 61975190], the Startup Fund for Youngman Research at SJTU.

Author contributions statement

CS studied and developed the data treatment method, wrote the corresponding computer programs, and wrote the draft of the paper. WX prepared the samples and acquired the LIBS spectra. YT participated in the development of the feature selection algorithm. YZ, ZY, SS, MW, LZ, FC participated in the experimental setup development and LIBS spectrum acquisition. JY supervised the research program and wrote the paper. All authors have approved the final version of the manuscript.

Additional information

Competing Interests: The authors declare no competing interests.

Figures

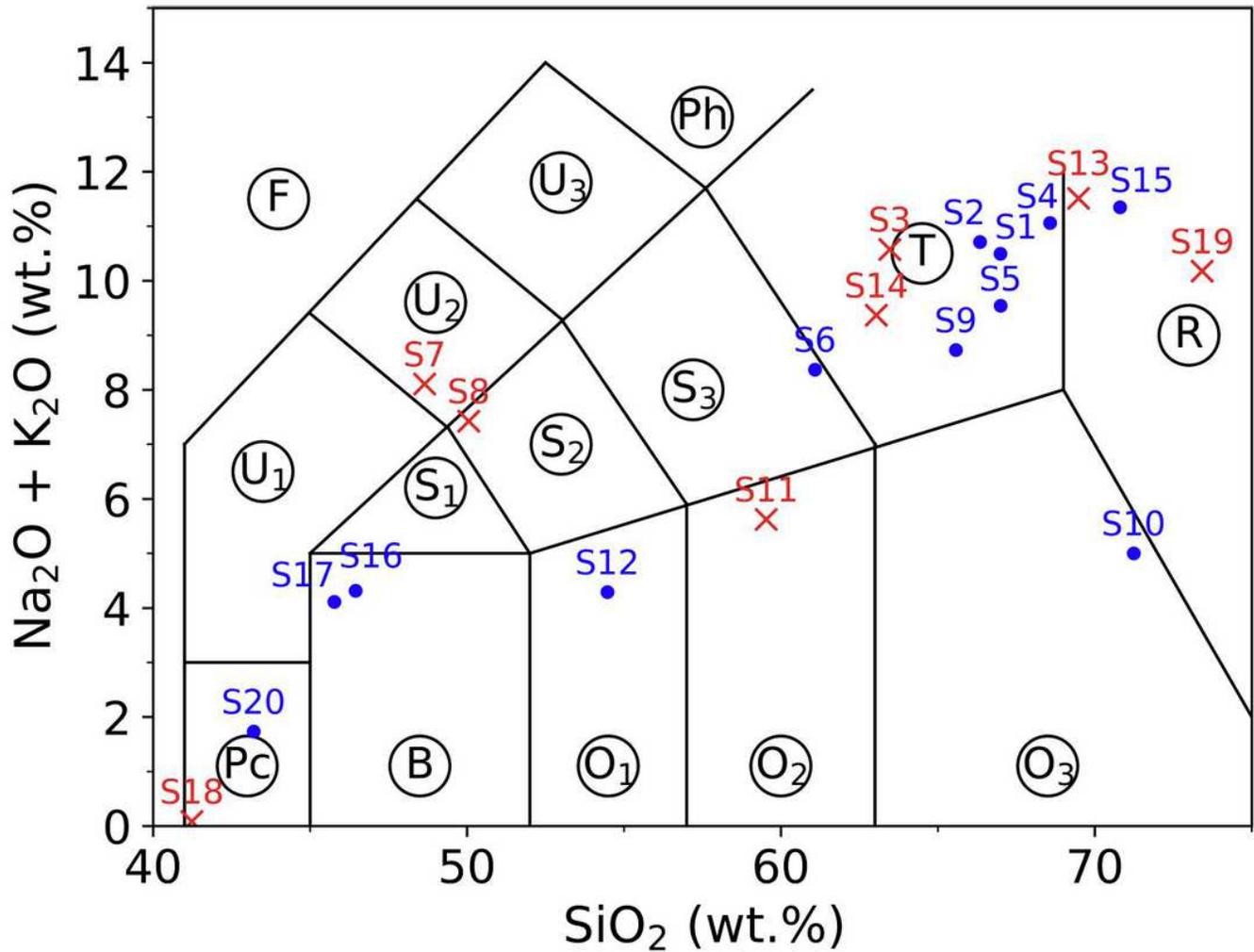


Figure 1

Presentation of the used rock samples in a TAS diagram according to their major oxide concentrations determined using XRF. The short notations of the 15 fields (surrounded by circle) are according to Reference 21. Eight rocks were selected as training samples together with the 20 pellets: S3, S7, S8, S11, S13, S14, S18 and S19 (represented by red crosses in the figure). The rest of the 12 samples were used as validation samples (represented by blue dots in the figure).

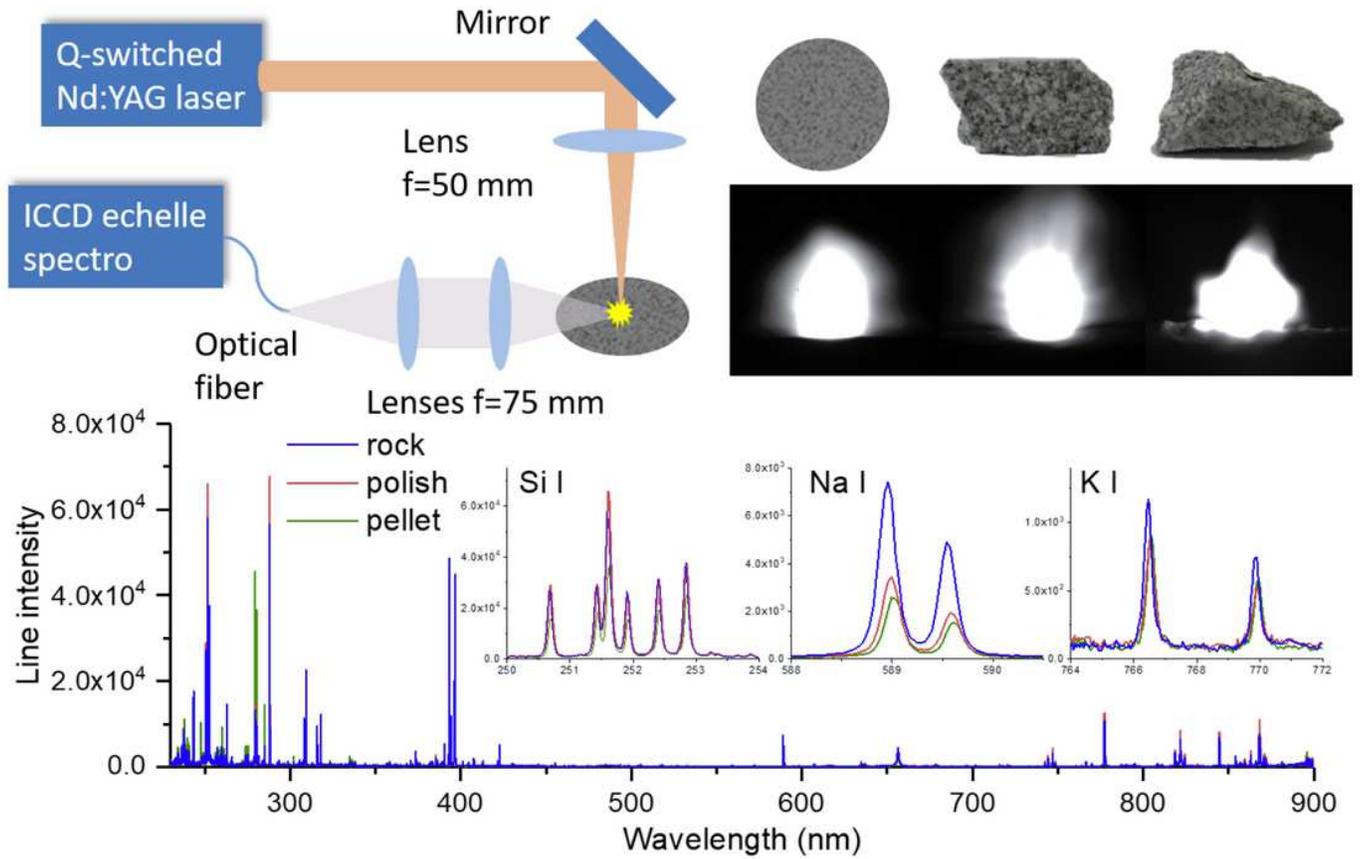


Figure 2

Schematic presentation of the used experimental setup, together with plasma images respectively induced on a pellet, a polished rock and a raw rock, and typical LIBS spectra showing differences in emission intensities of Si, Na and K between a pellet and the corresponding polished and raw rocks.

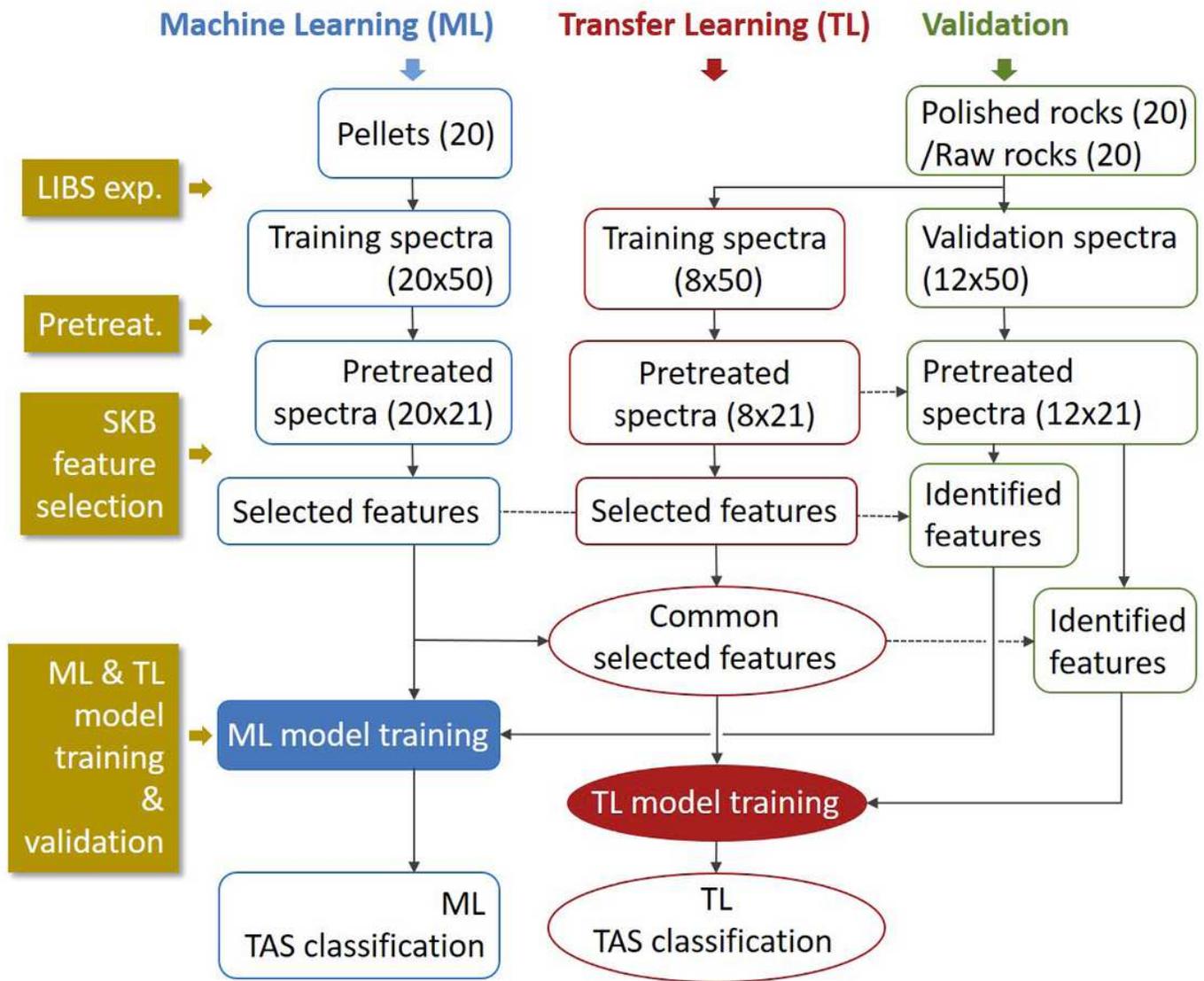


Figure 3

General flowchart used in this work allowing a comparative study between the performances of a machine learning (ML) model and those of a transfer learning (TL) model.

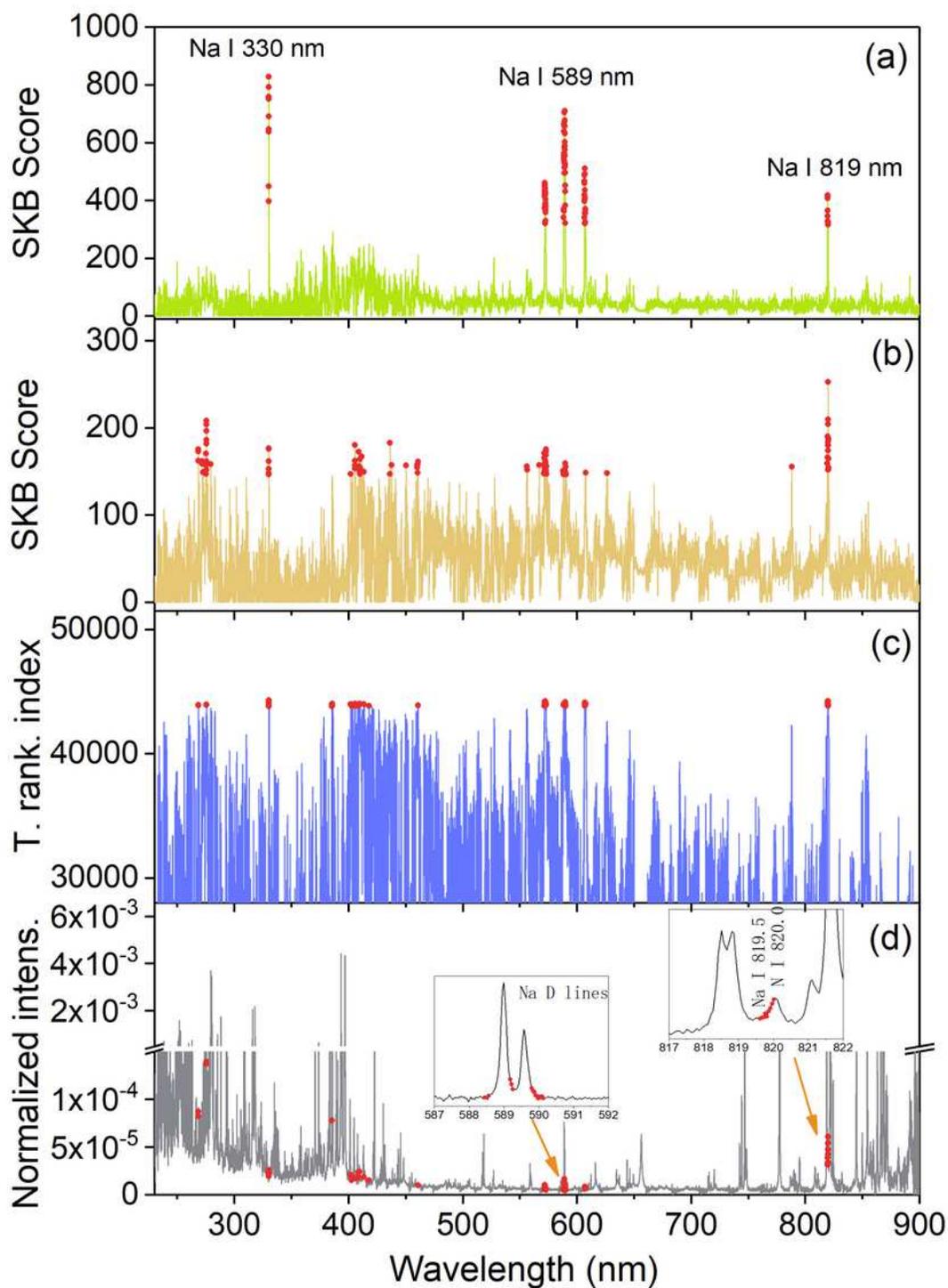


Figure 4

Results of feature selection for Na₂O: (a) for pellets and (b) for calibration raw rocks, SKB scores of all the spectral channels, and with in red dots the 100 selected features; (c) total ranking index of all the spectral channels, together with in red dots those of the 100 common selected features; (d) a typical normalized average spectrum from a pellet sample, together with in red dots the 100 common selected features, with 2 insets showing enlarged parts of the spectrum around 589 nm and 820 nm.

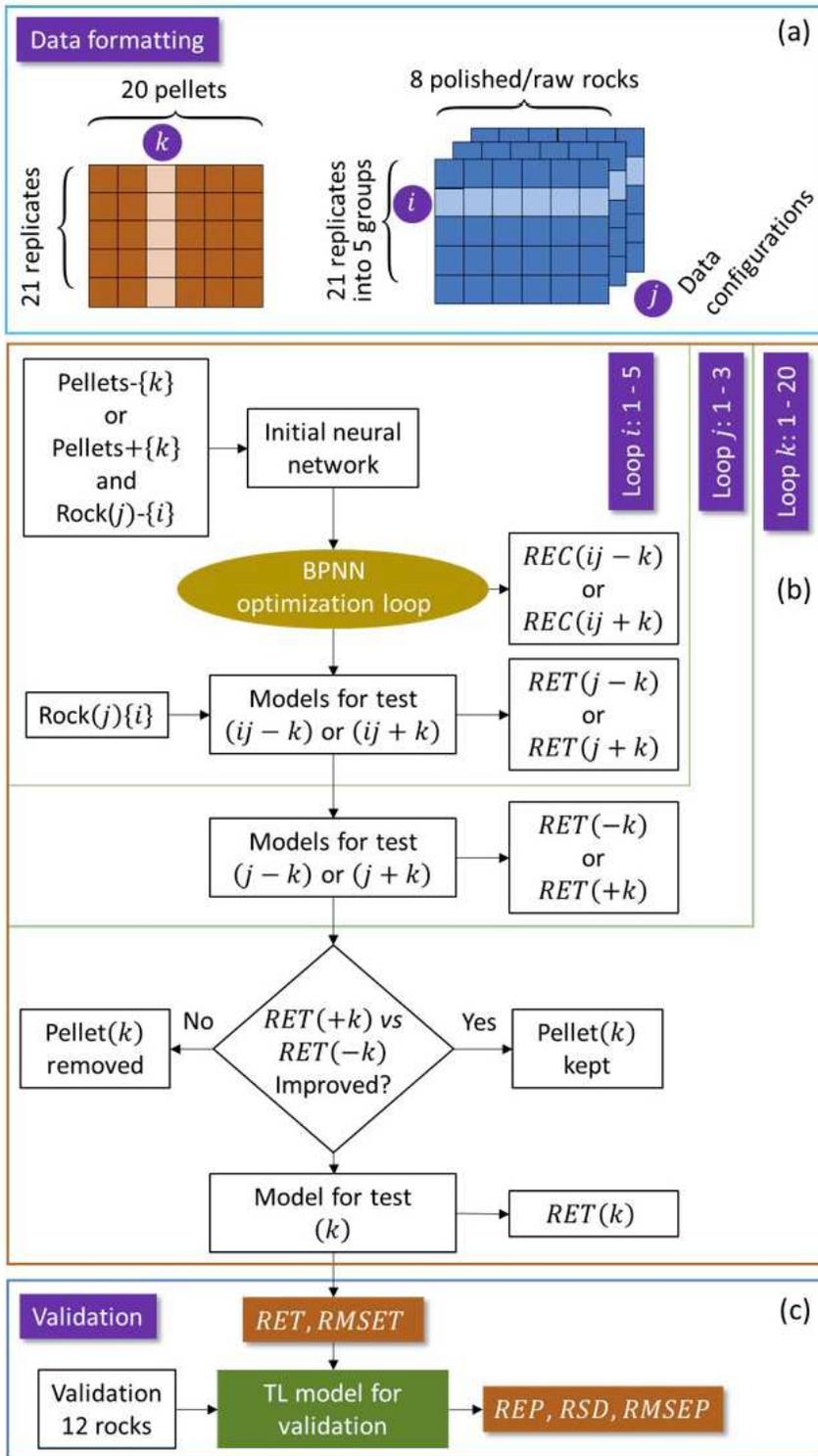


Figure 5

Flowchart of transfer learning model training with the implementation of feature-representation-transfer and instance-transfer.

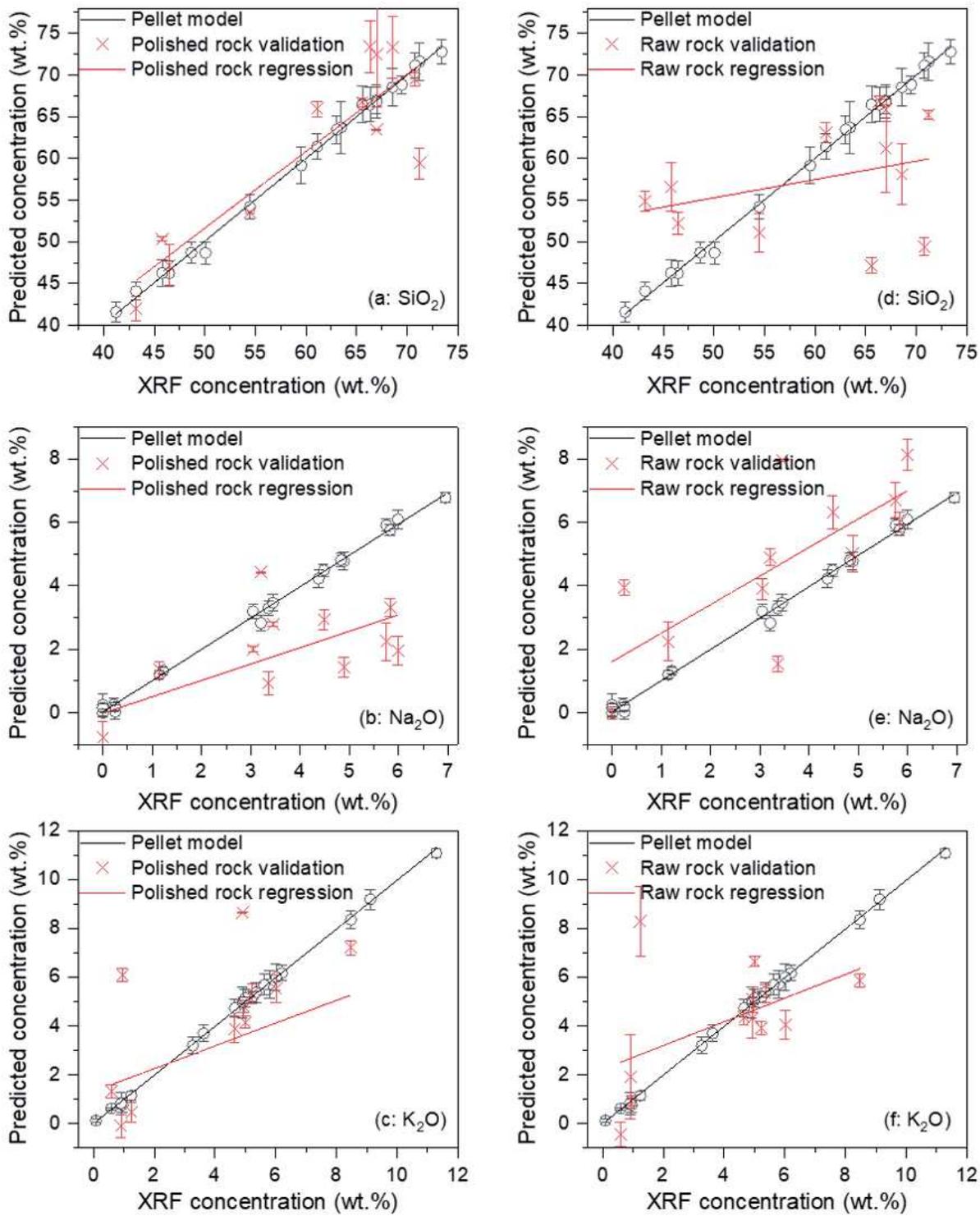


Figure 6

Machine learning-based calibration models trained with the spectra from the pellet samples (black lines) together with the calibration data (black open cycles) for the 3 compounds SiO₂ (a, d), Na₂O (b, e) and K₂O (c, f). Validation data from polished rock validation samples (a, b, c) and raw rock validation samples (d, e, f) are presented in red crosses, their linear regressions in red lines for the 3 compounds

SiO₂ (a, d), Na₂O (b, e) and K₂O (c,f). The error bars of the presented data correspond to the standard deviations (σ SD) of the predicted concentrations over the 21 pretreated spectra for a given sample.

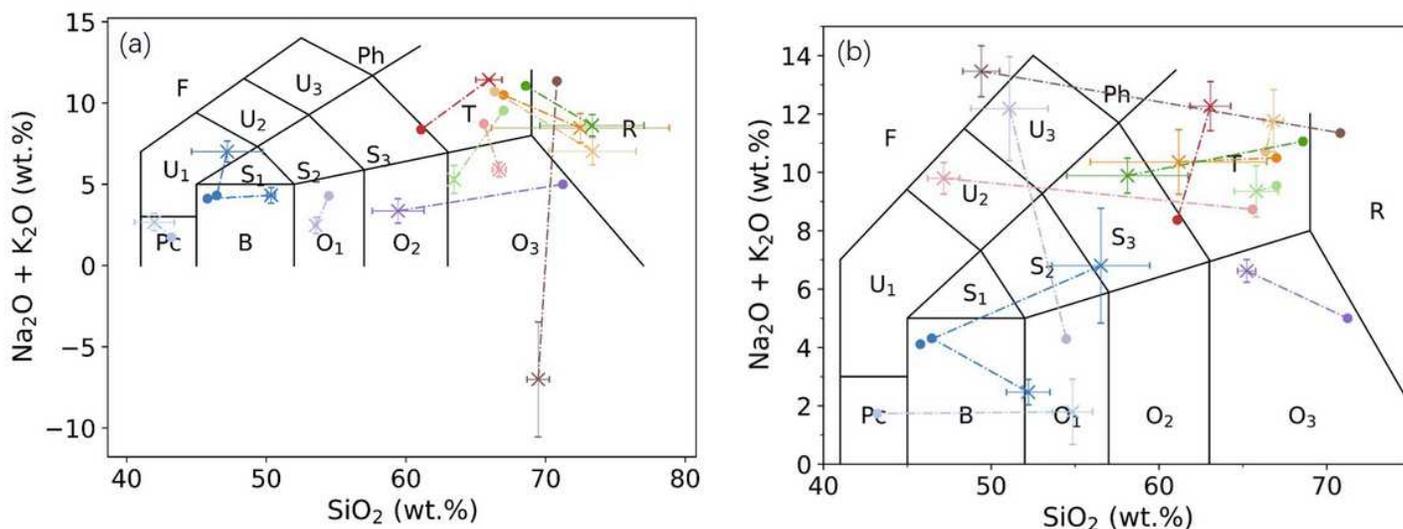


Figure 7

TAS classifications of the validation polished rock samples (a) and raw rock samples (b) using machine learning models. The positions determined by XRF are presented in colored solid circles, the corresponding model-predicted positions are presented in the same color in crosses. A dashed line links the XRF reference position and the model-predicted position of a same rock sample. The error bars on the predicted position are calculated over the different pretreated replicate spectra of a given sample.

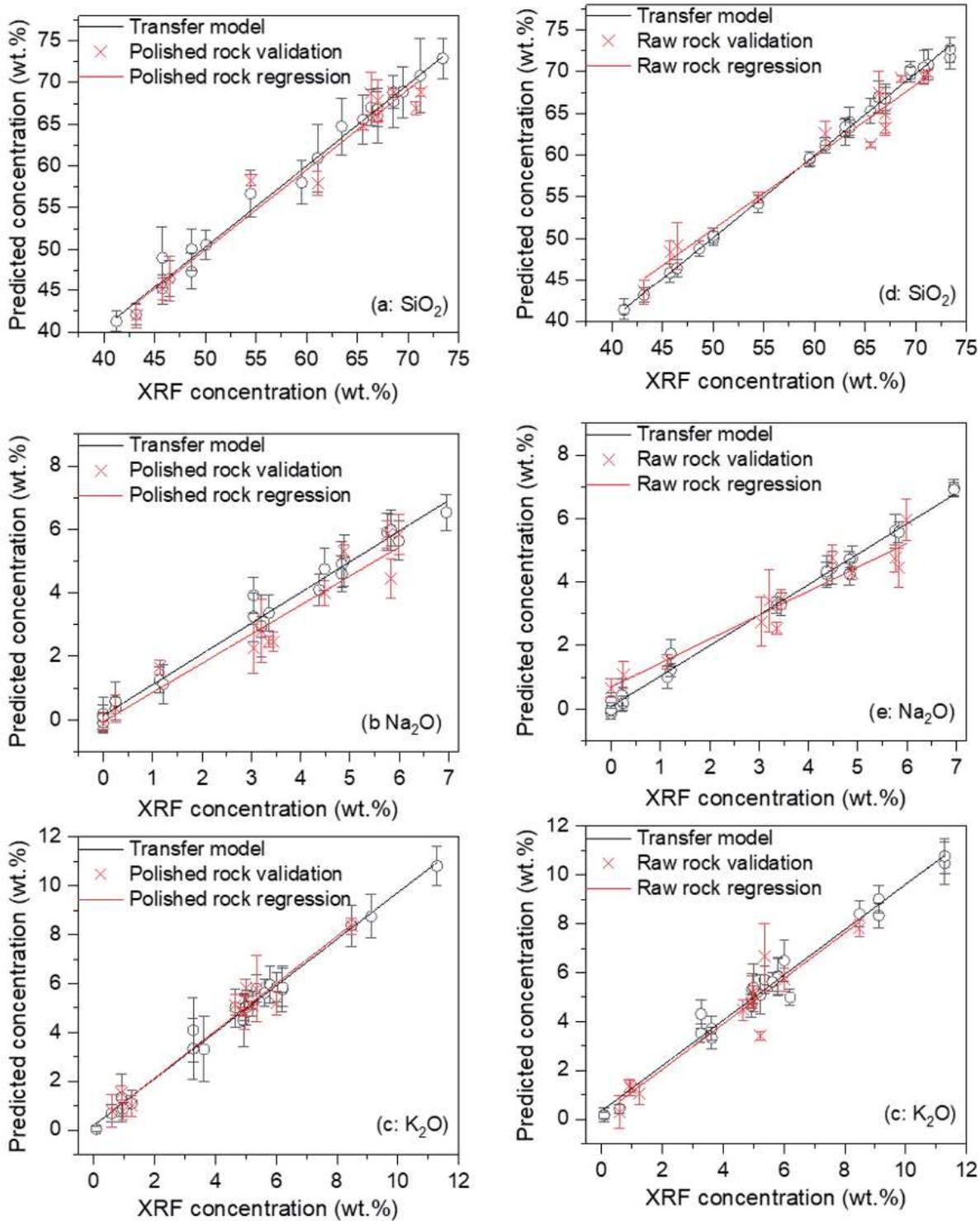


Figure 8

Transfer learning-based calibration models trained with the pellet samples and the training set of the rock samples (black lines) together with the calibration data (black open cycles) for the 3 compounds SiO₂ (a, d), Na₂O (b, e) and K₂O (c, f). Validation data from the polished rock samples (a, b, c) and the raw rock samples (d, e, f) are presented in red crosses, their linear regressions in red lines. The error bars of the

presented data correspond to the standard deviations (Σ SD) of the predicted concentrations over the 21 pretreated spectra of a given sample.

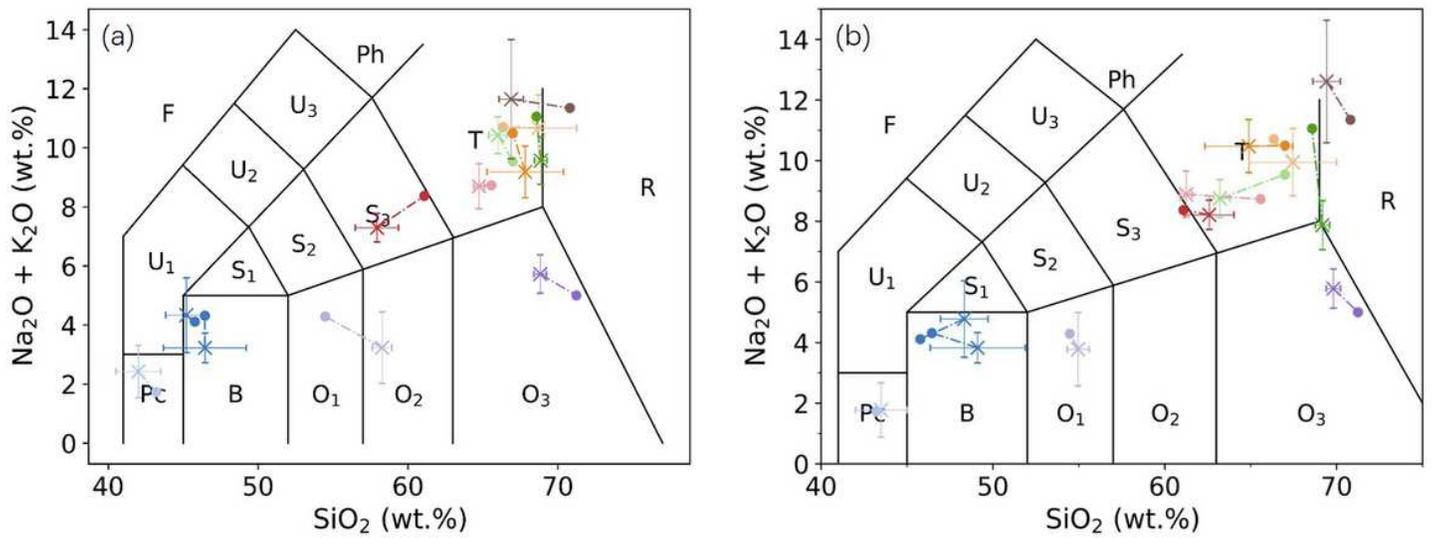


Figure 9

TAS classification using transfer learning models of the validation rock samples, (a) for polished rocks and (b) for raw rocks. The same symbols are used as in Fig. 7.

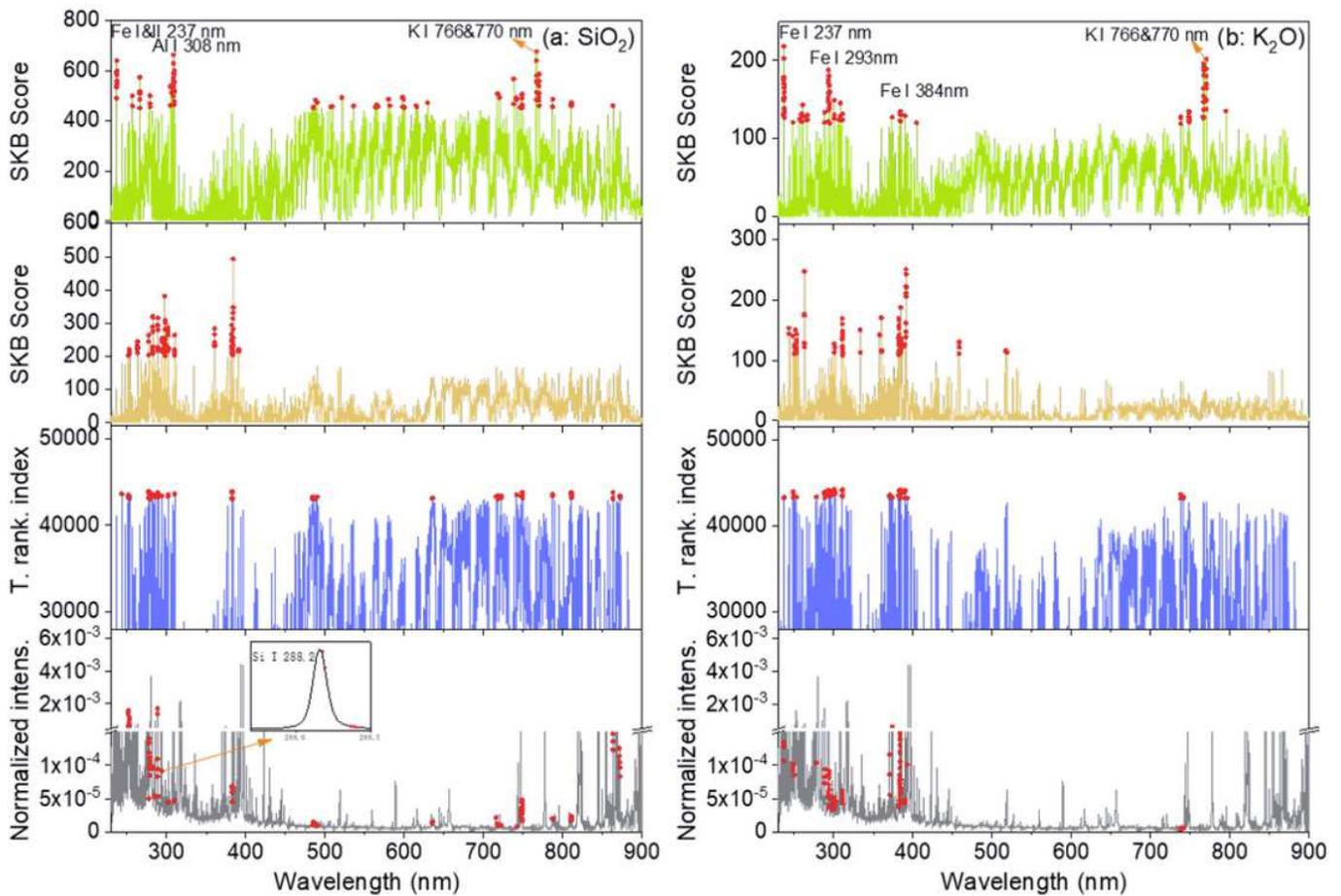


Figure 10

Results of feature selection for the couple of sample types pellet/raw rocks: (a) for SiO₂ and (b) for K₂O. A detailed caption can be found with Figure 4.

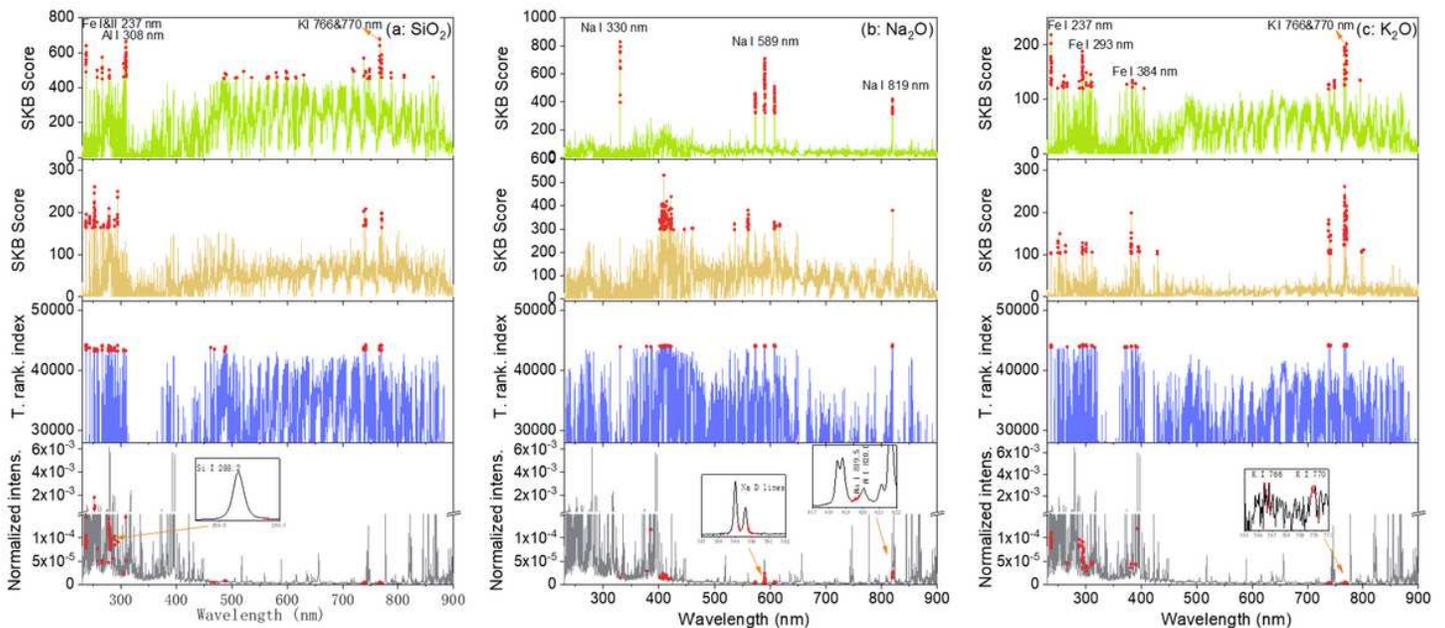


Figure 11

Results of feature selection for the couple of sample types pellet/polished rocks: (a) for SiO₂, (b) for Na₂O and (c) for K₂O. A detailed caption can be found with Figure 4.