JAAS

Cite this: *J. Anal. At. Spectrom.*, 2011, **26**, 2419

www.rsc.org/jaas

PAPER

Accurate identification of geological samples using artificial neural network processing of laser-induced breakdown spectroscopy data

Siu-Lung Lui and Alexander Koujelev*

Received 9th March 2011, Accepted 27th July 2011
DOI: 10.1039/c1ja10093a

Mineral and rock identification is a fundamental analysis in geological study. It allows the retrieval of both physical and chemical information of an identified sample from an available database. In this study, laser-induced breakdown spectroscopy integrated with artificial neural network algorithm is proposed for geological sample identification. The training algorithm of the artificial neural network is modified from the conventional method which is used in our previous studies. The trained network is tested by a set of natural rock samples which include new rocks which are not in the certified training set. Despite the difference in surface texture and minor variation in chemical composition of the tested rocks as compared to the samples of the training set, the validation reports a higher correct identification rate. This demonstrates the robustness of the modified algorithm to assess the variation of samples and the readiness to recognize new samples for a detailed study.

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is a spectroscopic technique which identifies the elemental composition of a sample from its plasma emission.[1–3] The plasma is formed by focusing a high power laser beam, typically of GW cm^{-2} power density, onto the sample. When the excited atoms in the plasma relax, the photons featuring the atoms' unique identity are released. Compared with other established techniques, such as inductively coupled plasma-mass spectroscopy (ICP-MS), X-ray fluorescence (XRF), or instrumental neutron activation analysis (INAA), LIBS is easier to be implemented. It can probe remote targets at quasi-real time without any sample preparation. Owing to these advantages, LIBS finds its niche in many applications like space exploration, industrial quality assurance, environmental monitoring and biological identification.[1–5]

A LIBS spectrum is composed of emission lines and background continuum. Since each element has its own set of emission lines, LIBS is especially useful in discriminating the presence of particular elements as contaminants in a sample.[6,7] For classification and identification purpose, LIBS capability to distinguish materials makes it valuable in various applications.[8,9] Minerals and rocks, however, are relatively complex samples. Very often, geological samples have similar components and differ only by the ratio of their elemental content. Therefore, identification cannot be concluded by observing the presence of one or two elements. Complicating the issue, the matrix effect, self-absorption and overlapping of lines impose non-linear factors on line ratios. Therefore a more systematic and thorough algorithm for identification is required.

In recent years, many chemometric methods like principal components analysis (PCA), soft independent modeling of class analogy (SIMCA), partial least-squares discriminant analysis (PLS-DA), and artificial neural network (ANN) have been applied to LIBS spectra for classification and identification.[10–12] Except for ANN, the other methods are based on the concept of principal component (PC). Samples are classified according to the distribution of clusters in the PC-space and the identity of a test sample is associated to the nearest group. If a sample does not belong to any group, it is regarded as *unclassified*. To date, the best identification performance for geological samples is achieved by Gottfried *et al.*[13] They applied PCA and PLS-DA on LIBS spectra of four basic geological groups (carbonate, fluorite, silicate and soil) and demonstrated a correct identification rate of higher than 95%. In another study conducted by Harmon *et al.*, the success rate dropped from 98% to 90% when the mineral set of 19 samples was expanded to 40.[14] Clearly, the performance of these algorithms is correlated with the training set size and sample variety.

The analysis of LIBS spectrum with an ANN has been demonstrated in several applications. For quantitative assay, Inakollu *et al.* used an ANN to predict the element concentrations in aluminum alloys.[15] Ferreira *et al.* determined the concentration of copper in soil samples from LIBS spectra.[16] For qualitative analysis, Sattmann *et al.* discriminated PVC from other polymers by the chlorine 725.66 nm line.[17] Using an ANN Ramil *et al.* classified the LIBS spectra of a large set of archaeological ceramics into three groups.[11] Motto-Ros and Koujelev *et al.* evaluated the elemental composition of natural rocks with

Canadian Space Agency, 6767, route de l'Aéroport, Saint-Hubert, Québec, J3Y 8Y9, Canada. E-mail: Alexander.Koujelev@asc-csa.gc.ca

an ANN.[18,19] All these studies have proven the potential of ANN in LIBS applications.

In our earlier work, we reported the capability of geological sample identification with LIBS combined with an ANN.[20] With 19 certified samples, the success rate was at 96% using single-shot LIBS spectrum. When tested by new samples, the ANN was able to report an "un-identification" at a success rate of 78%. In this paper, we focus on three issues when implementing the technique in practical applications:

1. Capability to identify new samples:

According to the International Mineralogical Association, there are over 4000 known minerals.[26] Training a network with all known samples is not possible. In reality, we can only train a network with a limited number of samples. Therefore the samples should be carefully selected and balanced for each class or the performance of a network may decline.[13,14] Unless an application is defined (*e.g.*, identify olivine on the Moon or jarosite on Mars), the trained ANN will likely encounter a new sample. We will evaluate the capability of our ANN which is subjected to limited training samples to identify both learnt and new samples.

2. Sample similarity and variety:

Two samples from the same class can have variations in their composition. For instance, andesite AGV2 has twice the amount of K_2O than another andesite JA3. Therefore the algorithm should be precise to distinguish between these two andesites, and flexible to associate a new andesite to any andesites that have been learnt.

3. Sample texture and homogeneity:

Unlike those homogeneous certified powders used for training, rocks in the field are usually weathered and inhomogeneous. The laser-cleaning shots only remove the surface contamination but do not improve the homogeneity. Therefore, the spectrum obtained from a rock can be different from the training spectra. It is necessary to verify the applicability of an algorithm with hand samples from the field.

In the following sections, we will describe the experimental setup for the acquisition of LIBS spectra. The details of the optimization of an ANN which includes the selection of input information and the adoption of various training techniques will be discussed next. Finally the performance of the improved algorithm is verified using certified mineral powders and natural rock samples. To our knowledge, this is the first report of a trained ANN algorithm to be tested on non-treated rock samples from the field.

2. Method

2.1. LIBS spectrum acquisition

The LIBS system in this study is similar to the one described in our previous work. The same instruments are used but the configuration is modified (Fig. 1). Briefly, a 20 mJ laser pulse from a Q-switch Nd:YAG laser (Spectra Physics, LPY150, 1064nm, 7 ns) passes through a pierced fibre mount and a pierced lens, and is focused onto the sample by a 7.5 cm lens. The emission from the plasma is collected by the lenses to the optical fibre and analyzed by the Ocean Optics LIBS 2000 spectrometer (200–970 nm, 0.1 nm resolution). This coaxial arrangement

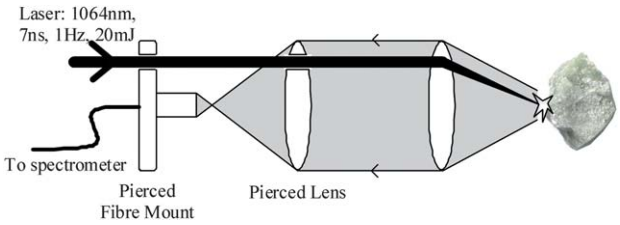

Fig. 1 Experimental configuration of the LIBS system.

emulated the situation in field applications in which the laser system and the spectrometer are put together in one unit in coaxial configuration. The integration delay of the spectrometer is 2 μs after the ablation and the integration time is 2 ms. The study is carried out under atmospheric condition. The recorded signal is sensitive to various experimental conditions such as laser intensity, ablation angle, sampling distance and collection optics. These factors affect the line intensity which is the input of the ANN.

Both certified geological powders and natural rocks are tested. Brammer Standard Company Inc. supplies 41 certified rock and mineral powders (Table 1). These certified materials are pressed into tablets for easy handling and sampling. Although some of them belong to same classes (*e.g.*, andesites, basalts, or anorthosites), they are considered as different samples in order to examine the 2nd issue listed in the previous section. Fourteen natural rocks, also known as hand samples, are purchased from Miners Inc. Their surfaces are not cleaned or polished. These rocks are to be classified by the trained network.

Five cleaning shots are fired at the samples in order to remove dusts and contaminants from the surface before data acquisition. For all acquisition, no more than 20 shots are fired at the same spot to avoid the crater effect. The acquired spectra are divided into three sets:

-Training set: spectra obtained from the certified powder tablets to train the ANN. For each tablet, 100 spectra (20 shots × 5 locations) are recorded;

-Validation set: another set of spectra obtained from the same tablets but at new locations. They are used to verify the training. For each tablet, 100 spectra (10 shots × 10 locations) are recorded. The shot number is reduced to 10 per location to preserve the samples because some of them are cracked during the acquisition of the training set data;

-Test set: spectra obtained from the hand samples. They are used to test the trained ANN. For each rock, 50 spectra (10 shots × 5 locations) are recorded.

2.2. The artificial neural network

The ANN reported in our earlier work is a conventional three-layer structure.[19,20] The weights and biases are optimized through the feed-forward back-propagation process. Each neuron is governed by the log-sigmoid function: $1/(1 + \exp(-u))$. This ANN is used as a basic structure and is trained with the modified training method.

Analyzing LIBS spectrum with an ANN is a pattern recognition task. The pattern to be recognized is the spectral lines and the continuum background which characterize a spectrum. For any rock sample, a representative spectrum can span a spectral

Table 1 List of certified geological samples and hand samples. The certified data of the samples can be found at www.brammerstandard.com. The certified powders are numbered for the presentation of results in later sections

Certified geological samples					
1	Andesite, AGV2	*2*	Andesite, JA3	*3*	Andesite, JA2
4	Andesite, JA1	*5*	Anorthosite, 2120	*6*	Anorthosite, 1042
7	Basalt, BCR2	*8*	Basalt, BHVO2	*9*	Basalt, JB2
10	Black soil, CRM2507	*11*	Borax frit, CERAM AN30	*12*	Coulsonite, DC9003a
13	Copper-molybdenum, USZ3-85	*14*	Flint clay, SRM97b	*15*	Granite, JG2
16	Graphite powder, KD2	*17*	Grey soil, CRM2504	*18*	Ilmenite, SARM59
19	Iron ore, DC11013	*20*	Kaolin, GBW03121	*21*	Potassium feldspar, GBW03116
22	Manganese ore, CMSI1691	*23*	Obsidian rock, SRM278	*24*	Olivine, DH4912
25	Orthoclase gabbro, RM1046-97	*26*	Pyroxenite, SARM5	*27*	Red clay, VS5372-90
28	Red soil, CRM 2501	*29*	Rhyolite, GBW07113	*30*	Dolomite, GBW07114
31	Andesite, GBW07104	*32*	Iron rock, CAN FeR2	*33*	Alumosilicate loose sediments, CRM3486-86
34	Shale, US Sco-1	*35*	Sillimanite, BCS309	*36*	Sulphide ore, RTS-4
37	Syenite, JSy1	*38*	Syenite, SARM2	*39*	Talc, BCS203a
40	Ultrabasic rock, GBW07101	*41*	Wollastonite, GBW03123		
Hand samples					
Talc, Hematite, Graphite, Sulfide mixture, Olivine, Dolomite, Kaolinite, Andesite, Gabbro, Basalt, Obsidian, Shale, Molybdenite, Fluorite					

range from UV to near IR. The resolution of the spectrum should also be fine enough to resolve the lines. As a result, a typical broadband spectrometer used in LIBS study can contain thousands of channels. Inputting the *entire* spectrum to an ANN will increase the network complexity and computation time. To simplify the task, intensities of the selected elemental lines are used. The intensity is defined by the summation of readings from channels above the full width half maximum (FWHM) points of a line. Using 41 training spectra, high quality lines which are free from instrumental limitations are selected. The selection is based on four criteria: (1) Good signal to noise ratio; (2) minimal overlapping with other lines; (3) minimal self-absorption; and (4) non-saturated.

Some spectral lines which are commonly used by other spectroscopic techniques are not selected due to these criteria. For instance, the sodium 589 nm doublet saturates the spectrometer easily and the carbon 247.9 nm and iron 248.3 nm lines are too close to be resolved clearly by our spectrometer. On the other hand, although the Mg 881 nm line is weaker than the Mg 285 nm line, it is selected since it is located in a region with less interference of other lines. During the line selection, we do not ignore any line which possesses non-linear property. Fig. 2 shows three calibration plots which explain the importance of the criteria and indicate how the non-linear features can help distinguish between the samples. After a careful inspection, a total of 139 lines are selected across the available spectral range. The FWHM intensity of each line is associated to one neuron in the input layer. To minimize signal variation due to instrumental fluctuation, each spectrum is normalized by the intensity of its oxygen 777 nm line.

The number of neurons in the hidden layer is adjusted for faster processing and more accurate prediction. After careful adjustment, it is set to be the same as the number of input neurons.

The number of neurons at the output layer is determined by the application and how the users classify the training samples. If the Dana classification system is used, the minerals are divided into nine groups.[27] There will be nine output neurons so more than one sample will be assigned to each group. This improves the accuracy of classification since more samples are available. Our certified samples consist of both minerals and rocks so the Dana classification system is not applicable. Moreover, as we intend to evaluate the ANN's capability to identify similar samples, new samples and new classes, we decide to treat our 41 samples as individual samples during the training. As a result there are 41 output neurons each corresponds to one sample and returns an ANN output P of value between 0 and 1. The identity of a test sample corresponds to the neuron with the highest output. Note that as we regard all samples as different samples, one should be aware of the interpretation of an *identified* and an *un-identified* result. An identified result means that the test sample can be a same sample or a similar sample of a known class. For an un-identified result, the test sample can be a different sample of a known class or from a completely new class.

2.3. ANN training and validation

As a rule of thumb, a trained ANN should be applicable to both the training set data as well as any test data. To avoid over-fitting of the training data, the number of training cases must be sufficiently large and each class should be represented by a sufficient number of representative cases.[21] Since all training samples are considered as different samples, each output neuron has only one representative case. If this ANN is trained by the average spectrum of each sample (*i.e.*, 41 spectra), then the ANN can be over-fitted. Owing to the complexity of the network and the nature of our application, we cannot simply use a single algorithm to avoid overfitting while maintaining the accuracy of the network. Instead, we apply a combination of the following protocols and methods:

1. High quality training spectra:

Using sequential training (see below), each training spectrum in one subset is an average of 20 spectra and is normalized with the oxygen line so individual fluctuation is minimized.

2. Precaution with line selection:

With the line selection criteria we imposed, high quality input data can be assured.

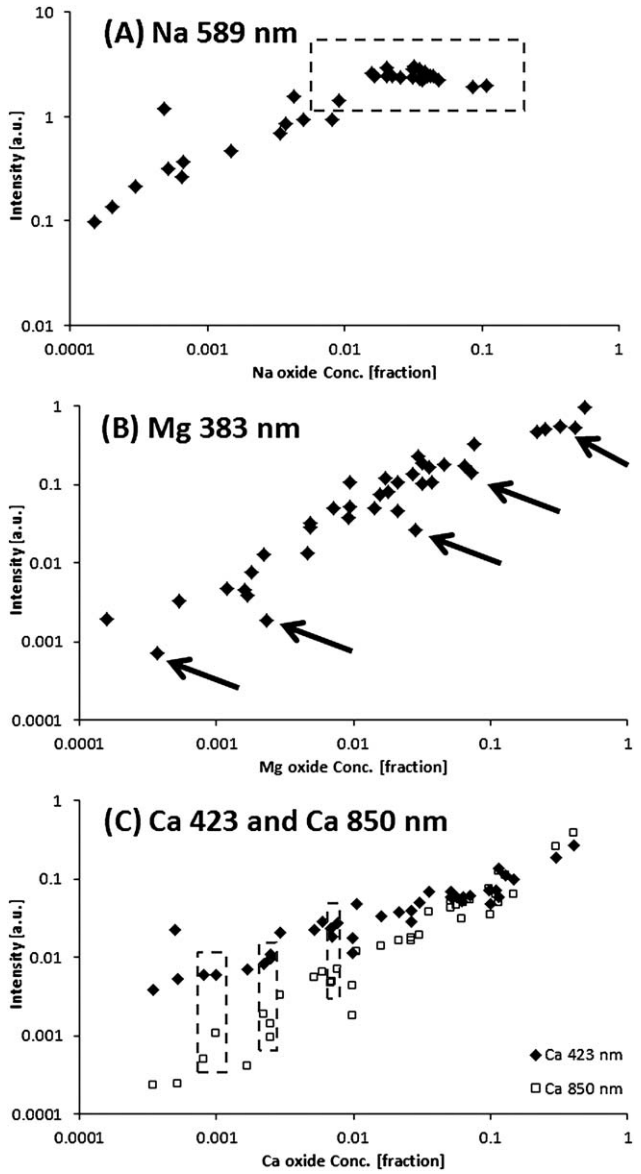

Fig. 2 Calibration plots of different elemental lines. (A) The box highlights the region where Na 589 nm line is saturated and no conclusive information can be provided. (B) The Mg 383 nm does not show any instrumental limitation. The fluctuating data points can provide extra features for identification. (C) Two lines from the same element can exhibit different non-linear properties. The boxes show different behaviours between the Ca 423 and 850 nm lines.

3. Early stopping:

In early stopping, the training is monitored simultaneously by another set and the training is stopped when the error of this set starts to go up.[22] Early stopping is a common solution to overfitting when the number of training cases is limited.[28] In one example, Nelson and Illingworth demonstrate it using just 50 cases to train a network with 16 219 weights.[32] In our work, besides the validation set, there is a 3rd set: the rock, to test the ANN. In the results section, we will show that the high correct identification rate in both sets suggests that overfitting is not observed in the ANN.

4. Sequential training:

The conceptual idea of sequential training is illustrated in Fig. 3.[23–25] Five spectra are obtained by averaging the spectra from each sampling spot of a tablet. These average spectra from all samples are divided into five subsets which contain one spectrum per sample. Then an ANN is trained by a subset with the early stopping criterion. The optimized weights and biases are transferred as the initial values for the second training with a new subset. By separating the training data into five subsets, we provide not only more training cases, but also more "variations" for the ANN to learn from. Although there are reports showing that the performance of sequential learning is only comparable to that of batch learning,[29–31] in our identification of the rock samples reported in the results section, the result will justify that the network trained by sequential method is more generalized.

After training, the network is evaluated by the validation set. For validation, ten subsets are established by averaging the spectra at each ablation spot on the sample. At the output layer, each neuron reports a predicted value P between 0 (*complete mismatch*) and 1 (*perfect match*). The identity of a sample is associated with the neuron with the highest output above a threshold of 70%.[20] If all outputs are below the threshold, the sample is regarded as "*unidentified*".

Table 2 compares the results of the improved ANN with those by the ANN trained with the conventional method described in ref. 20. The prediction of the LIBS-ANN improves with the number of sequential trainings. After the 5th training, the rate of correct identification increases from 82.4% to 90.7%, while the incorrect identification rate drops from 2% to 0.5%. This is equivalent to only two false identifications out of 410 test spectra. Compared to the ANN trained by the conventional method, its performance is slightly better. Taking into account the variety of the certified samples and the experimental shot-to-shot fluctuation, LIBS-ANN has the potential to become an excellent tool for mineral and rock identification.

The ANN outputs P from the same powders are averaged and the results are displayed in a matrix format in Fig. 4. The outstanding bright boxes arranged diagonally across the matrices indicate a perfect identification by both training methods. The matrix for the sequential training, however, has less background

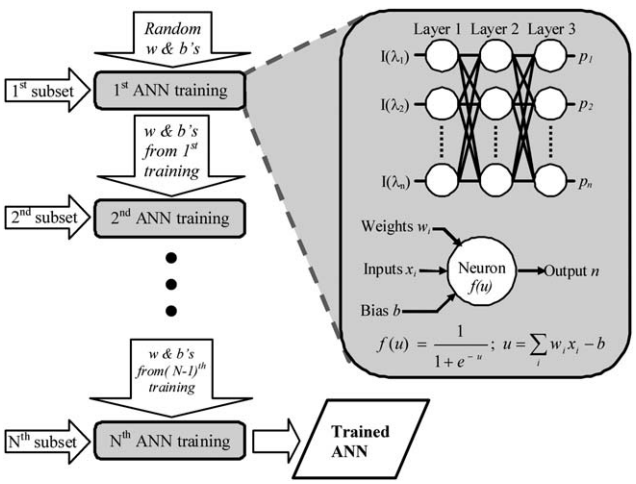

Fig. 3 Illustration of the concept of sequential training.

Table 2 Validation result of the ANN trained by sequential and conventional methods. The total number of spectra for validation is 410

		Average rate (%)			
		Classified			
Training method		Correct	Misidentified	Success within classified samples	Unidentified
Conventional		*87.1*	*2.0*	*97.9*	*11.0*
After 1st	Sequential training	82.4	2.0	96.7	15.6
After 3rd		88.5	1.7	97.5	9.8
After 5th		*90.7*	*0.5*	*99.5*	*8.8*

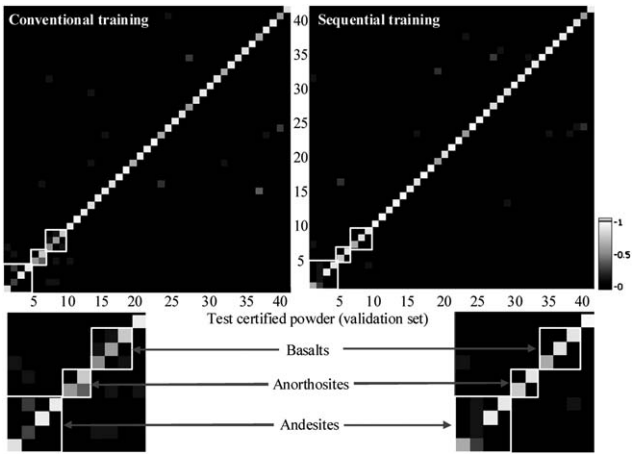

Fig. 4 Matrices showing the prediction of validation data sets analyzed by the LIBS-ANN trained with conventional (left) and sequential method (right). The vertical axis is the sample number of the training set defined in Table 1. The horizontal axis is the category of the validation sample which uses the same numbering scheme. Each cell in the matrices represents an ANN output *P*. The boxes group similar sample of the same classes. Two zoomed views are added below the matrices.

noise. This observation is especially clear for andesite, anorthosite and basalt samples (marked by square boxes) which have different samples of the same class. Therefore, non-zero outputs from their variances are expected. The off-diagonal cells show lower *P* values, suggesting that the ANN with sequential training is more subtle to handle similar samples of the same classes. Note that both training methods misidentify sample 2 (*i.e.*, andesite JA3) as another andesite. According to the certified data, the concentrations of major oxides for JA3 are very similar to those of other andesites so there are no distinct spectral features to differentiate JA3.

3. Results

Once the network is trained and verified by the certified samples, it can be tested by the hand samples. Fig. 5 shows an olivine tablet and a hand sample of olivine. The microscopic views of their surfaces are completely different. This can impose variations to the LIBS spectra and affect the accuracy of the ANN prediction. Fig. 6 compares the average spectra for the certified tablets (bottom traces) with the corresponding hand samples (top traces). Their LIBS spectra are different even if the certified and hand samples are from the same classes. For instance, the olivine hand sample contains a higher concentration of iron. The graphite tablet has a higher amount of sodium.

During identification, the hand samples being tested can be *identical* or *similar to* one of those known samples. In the case where a hand sample doesn't have a similar sample in the training set, it should be identified as a new sample of a known class or a new class. As mentioned in the Introduction, unless an ANN is trained for *all* minerals, the simple *above-the-threshold* scheme is not sufficient to determine the identity of a sample and their variances. Therefore the rules are redefined in order to classify learned, similar or new samples:

-Known sample: If the highest *P* output is above 0.70, the test sample is classified to the known sample;[20]

-Similar sample: If the highest *P* output lies between 0.45 and 0.70, the test sample is classified as a similar sample. The 0.45 threshold is determined by the validation result of the samples which have similar samples of the same class (*i.e.*, andesites, anorthosites, basalts and syenites). Detailed procedure for obtaining this threshold is described in Table 3.

-New sample: If the highest *P* falls below 0.45, the sample is regarded as an unknown. It can be a new class (*e.g.*, fluorite) or a new sample (*e.g.*, andesite JA4) which is different from the known one in the training set.

The identification result of the hand samples based on the redefined rules is summarized in Table 4. The spectra at each spot are averaged so each hand sample has five data. Among the hand samples, talc, graphite, olivine, dolomite, andesite, basalt,

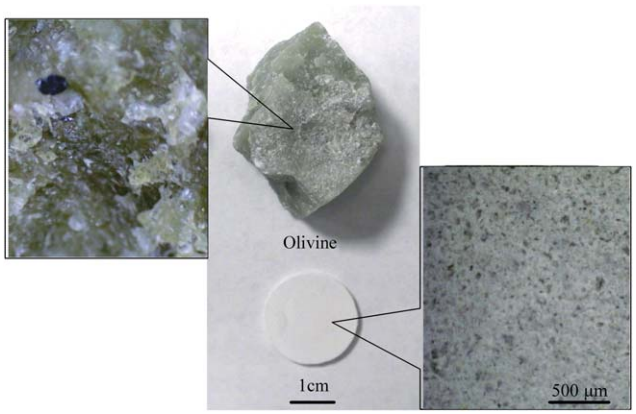

Fig. 5 Comparison of an olivine hand sample (top) and an olivine tablet (bottom). The microscopic views show the hand sample is heterogeneous and the surface is rough as opposed to the homogeneity and smoothness of the tablet.

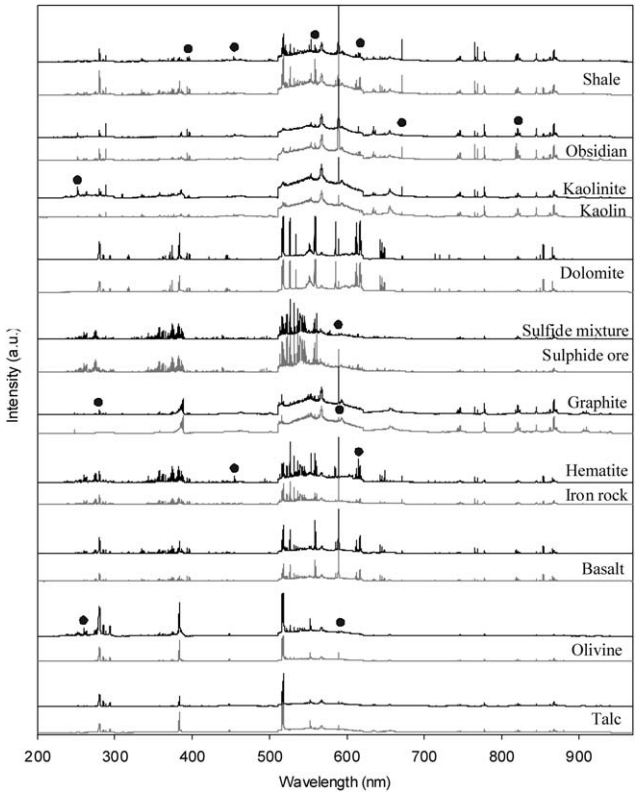

Fig. 6 Comparison of the LIBS spectra between the certified powder (grey spectra, lower) and the hand samples (black spectra, upper). The dots mark the major differences. The spectra are normalized with the oxygen 777 nm line to minimize the effect of instrumental fluctuation.

obsidian, and shale have their references (samples of the same class) in the training set. Other samples like hematite, sulphide mixture, and gabbro have similar samples of the same classes as well. Only molybdenite and fluorite are hand samples which do not have a corresponding class in the trained ANN. Thus the identification of these two samples should report an "unidentify". The prediction of most samples returns accurate results with high *P*s. For andesite and gabbro hand samples, the interpretation of the prediction is not so straight-forward. Andesite and basalt are both volcanic rocks with andesite slightly richer in SiO_2 and Fe_2O_3, and poorer in MgO. Gabbro, on the other hand, is the intrusive equivalent of the extrusive basalt thus these two types have the same range of chemical compositions. In other words, "basalt" for these two hand samples can be regarded as *correct*. To support the arguments, these two hand samples are certified with ICP by Activation Laboratories Ltd (Actlabs). The certification indicates that, based on chemical composition, the andesite and gabbro hand samples can possibly be identified as basalt since their major components: SiO_2, Al_2O_3 and MgO/CaO, have similar concentrations (Table 5).

In Table 4, the correct and incorrect identifications are indicated by a symbol "√" and "×", respectively. The symbol "○" indicates an *undetermined result* which means all *P*s fall below the thresholds and the test sample can be a new class or a sample which is different from the members of the given one in the training set. In practice, a second measurement can always be performed to clarify any uncertainty. Out of the 70 hand sample averaged spectra, 61 are identified correctly. This corresponds to a success rate of 87.1%. Two spectra are identified incorrectly and seven spectra are undetermined. The incorrect identifications are from hematite and sulphide mixture which are classified as basalt and iron ore, respectively. By inspecting the other four identification results of hematite, one can observe that the 1st, 2nd and the 5th spots are clearly identified as iron rock with high value of *P*s. The 4th spot is identified as unknown due to its low value of *P*. Therefore we suspect that near the 3rd spot area there is a basalt grain which affects the identification result of the 3rd and 4th spots. In the case of sulfide mixture, if one considers the highest and 2nd highest *P*s for all five spots, one can conclude that the "mixture" is possibly an iron sulphide compound. Unfortunately, the certified data for this sample is unavailable. These examples suggest that the identification should be done on different locations of a sample in order to reveal its true identity, especially for an inhomogeneous hand sample.

Table 3 Illustration showing the determination of the 0.45 threshold for the definition of a similar sample. Only the *unidentified* results (*i.e.*, $P < 0.7$) from the validation test of andesites, anorthosites, basalt and syenites are considered. If the highest *P* of a sample corresponds to a neuron of the same class, then this *P* is selected. We average all the selected *P* values and obtain an average of 0.45

		P					
Test sample	Validation subset	Basalt BCR2	Basalt BHVO2	Basalt JB2	Talc ⋯	Identified as	Result with $P < 0.7$
Basalt BCR2	1	**0.8**	0.1	0.2	0	Basalt BCR2	—
	2	0.1	**0.4**	0.1	0	Unidentified	⇨ 0.4
	3	0.2	0.3	**0.6**	0	Unidentified	⇨ 0.6
	4	**0.9**	0.3	0.3	0	Basalt BCR2	—
	5	**0.6**	0.1	0.1	0	Unidentified	⇨ 0.6
	6	0.4	0.2	**0.6**	0	Unidentified	⇨ 0.6
	7	0.4	0.3	**0.6**	0	Unidentified	⇨ 0.6
	8	**0.9**	0.2	0.5	0	Basalt BCR2	—
	9	**0.8**	0.5	0.6	0	Basalt BCR2	—
	10	**0.8**	0.2	0.3	0	Basalt BCR2	—
Basalt BHVO2	1	**0.4**	0.1	0.1	0	Unidentified	⇨ 0.4
	2	0.2	0.5	**0.6**	0	Unidentified	⇨ 0.6
	3	0.1	**0.9**	0.5	0	Basalt BHVO2	—
⋮			⋮				⋮
						Average:	**0.45**

Table 4 Identification results of hand samples. Each row represents the prediction of a sampling spot. The *P* values in the brackets are percentages. The last column is the prediction result obtained from an ANN trained by a conventional procedure

	P (*Sequential Training*)			Identified as	
Rock	Highest	2nd highest	3rd highest	Sequential	Conventional
Talc	Talc (52)	Graphite (15)	Shale (1)	√ *Talc-similar*	√ *Talc-similar* (*51*)
	Talc (87)	Kaolin (17)	Shale (2)	√ *Talc*	√ *Talc* (*84*)
	Talc (87)	Kaolin (14)	Shale (2)	√ *Talc*	√ *Talc* (*80*)
	Talc (88)	Kaolin (16)	Shale (2)	√ *Talc*	√ *Talc* (*81*)
	Talc (95)	Olivine (4)	Dolomite (1)	√ *Talc*	√ *Talc* (*92*)
Hematite	Fe-rock (93)	Mn-ore (14)	S-ore (11)	√ *Fe-rock*	√ *Fe-rock* (*70*)
	Fe-rock (97)	Mn-ore (6)	S-ore (5)	√ *Fe-rock*	√ *Fe-rock-similar* (*46*)
	Basalt (54)	Fe-rock (40)	Mn-ore (12)	× *Basalt-similar*	○ *Unknown*
	Basalt (8)	S-ore (3)	Mn-ore (3)	○ *Unknown*	○ *Unknown*
	Fe-rock (85)	Mn-ore (27)	S-ore (26)	√ *Fe-rock*	√ *Fe-rock -similar*(*60*)
Graphite	Graphite (46)	Sillimanite (17)	Obsidian (17)	√ *Graphite-similar*	√ *Graphite-similar* (*45*)
	Graphite (80)	Obsidian (46)	Anorthosite(11)	√ *Graphite*	√ *Graphite* (*78*)
	Graphite (96)	Kaolin (2)	—	√ *Graphite*	√ *Graphite* (*93*)
	Graphite (75)	Obsidian (43)	Anorthosite(13)	√ *Graphite*	√ *Graphite* (*70*)
	Graphite (9)	Kaolin (6)	Redsoil (4)	○ *Unknown*	○ *Unknown*
Sulfide mixture	S-ore (74)	Fe-ore (41)	Cumo (1)	√ *S-ore*	√ *S-ore* (*80*)
	Fe-ore (92)	S-ore (46)	Obsidian (2)	× *Fe-ore*	○ *Unknown* (*40*)
	S-ore (94)	Fe-ore (3)	Cu-Mo (1)	√ *S-ore*	√ *S-ore* (*95*)
	S-ore (98)	Ultrabasic (1)	Fe-ore (1)	√ *S-ore*	√ *S-ore* (*97*)
	S-ore (87)	Fe-ore (30)	Cu–Mo (1)	√ *S-ore*	√ *S-ore* (*81*)
Olivine	Olivine (97)	Fe-ore (11)	Ultrabasic (6)	√ *Olivine*	√ *Olivine* (*91*)
	Olivine (95)	Fe-ore (27)	Ultrabasic (10)	√ *Olivine*	√ *Olivine* (*90*)
	Olivine (97)	Ultrabasic (2)	Talc (1)	√ *Olivine*	√ *Olivine* (*95*)
	Olivine (94)	Fe-ore (33)	Ultrabasic (11)	√ *Olivine*	√ *Olivine* (*93*)
	Olivine (97)	Fe-ore (16)	Ultrabasic (8)	√ *Olivine*	√ *Olivine* (*92*)
Dolomite	Dolomite (98)	—	—	√ *Dolomite*	√ *Dolomite* (*90*)
	Dolomite (98)	—	—	√ *Dolomite*	√ *Dolomite* (*95*)
	Dolomite (98)	Basalt (1)	—	√ *Dolomite*	√ *Dolomite* (*92*)
	Dolomite (98)	Wollastonite (1)	—	√ *Dolomite*	√ *Dolomite* (*93*)
	Dolomite (98)	Wollastonite (1)	—	√ *Dolomite*	√ *Dolomite* (*92*)
Kaolinite	Kaolin (93)	Obsidian (13)	—	√ *Kaolin*	× *Obsidian* (*89*)
	Kaolin (96)	Obsidian (16)	Wollastonite (1)	√ *Kaolin*	× *Graphite* (*74*)
	Kaolin (97)	Redsoil (12)	Flintclay (2)	√ *Kaolin*	○ *Unknown*
	Kaolin (94)	Obsidian (11)	Wollastonite (1)	√ *Kaolin*	× *Graphite* (*76*)
	Kaolin (97)	Obsidian (10)	Blacksoil (1)	√ *Kaolin*	× *Graphite* (*75*)
Andesite	Basalt (70)	Basalt (48)	Coulsonite (2)	√ *Basalt-similar*	√ *Basalt-similar* (*62*)
	Basalt (10)	Orthogabbro (9)	Basalt (8)	○ *Unknown*	○ *Unknown*
	Andesite (43)	Basalt (2)	Orthogabbro (1)	○ *Unknown*	○ *Unknown*
	Basalt (52)	Orthogabbro (3)	Basalt (2)	√ *Basalt-similar*	√ *Basalt-similar* (*47*)
	Basalt (19)	Orthogabbro (11)	Basalt (3)	○ *Unknown*	○ *Unknown*
Gabbro	Basalt (87)	Coulsonite (11)	Dolomite (5)	√ *Basalt*	√ *Basalt* (*80*)
	Basalt (75)	Coulsonite (4)	Dolomite (3)	√ *Basalt*	× *Greysoil* (*90*)
	Basalt (86)	Coulsonite (17)	Dolomite (1)	√ *Basalt*	√ *Basalt* (*86*)
	Basalt (22)	Andesite (8)	Wollastonite (7)	○ *Unknown*	√ *Basalt* (*79*)
	Wollastonite(44)	Anorthosite (10)	Andesite (3)	○ *Unknown*	○ *Unknown*
Basalt	Basalt (94)	Pyroxenite (4)	Dolomite (2)	√ *Basalt*	× *Greysoil* (*87*)
	Basalt (62)	Basalt (8)	Coulsonite (4)	√ *Basalt-similar*	× *Greysoil* (*73*)
	Basalt (86)	Rhyolite (2)	Basalt (2)	√ *Basalt*	√ *Basalt* (*80*)
	Basalt (85)	Basalt (4)	Andesite (2)	√ *Basalt*	√ *Basalt-similar* (*52*)
	Basalt (60)	Rhyolite (9)	Basalt (2)	√ *Basalt-similar*	√ *Basalt-similar* (*62*)
Obsidian	Obsidian (98)	Sillimanite (3)	Kaolin (3)	√ *Obsidian*	○ *Unknown*
	Obsidian (99)	Kaolin (11)	Fe-ore (2)	√ *Obsidian*	× *Syenite-similar* (*46*)
	Obsidian (99)	Kaolin (14)	Fe-ore (2)	√ *Obsidian*	× *Syenite-similar* (*49*)
	Obsidian (99)	Kaolin (11)	Fe-ore (3)	√ *Obsidian*	○ *Unknown*
	Obsidian (99)	Kaolin (13)	Fe-ore (2)	√ *Obsidian*	○ *Unknown*
Shale	Shale (70)	Sillimanite (46)	Sediment (17)	√ *Shale*	○ *Unknown*
	Shale (70)	Sediment (23)	Redsoil (9)	√ *Shale*	○ *Unknown*
	Shale (97)	Sillimanite (24)	Sediment (5)	√ *Shale*	○ *Unknown*
	Shale (68)	Sediment (25)	Sillimanite (16)	√ *Shale-similar*	○ *Unknown*
	Shale (97)	Sediment (6)	Sillimanite (3)	√ *Shale*	○ *Unknown*
Molybdenite	Blacksoil (8)	Rhyolite (6)	Mn-ore (6)	√ *Unknown*	√ *Unknown*
	Kaolin (24)	Blacksoil (11)	Redsoil (2)	√ *Unknown*	√ *Unknown*
	Fe-rock (1)	Rhyolite (1)	Mn-ore (1)	√ *Unknown*	√ *Unknown*
	Blacksoil (3)	Orthogabbro (2)	Mn-ore (1)	√ *Unknown*	√ *Unknown*
	Blacksoil (1)	Basalt (1)	—	√ *Unknown*	√ *Unknown*
Fluorite	Wollastonite (4)	S-ore (1)	Dolomite (1)	√ *Unknown*	√ *Unknown*
	Wollastonite (4)	Dolomite (3)	S-ore (1)	√ *Unknown*	√ *Unknown*
	Wollastonite(10)	Blacksoil (2)	Dolomite (1)	√ *Unknown*	√ *Unknown*

Table 4 (*Contd.*)

Rock	*P* (*Sequential Training*)			Identified as	
	Highest	2nd highest	3rd highest	Sequential	Conventional
	Wollastonite (7)	Dolomite (1)	Mn-ore (1)	√ *Unknown*	√ *Unknown*
	Wollastonite (2)	S-ore (1)	Dolomite (1)	√ *Unknown*	√ *Unknown*

Table 5 Comparison of the concentration (in wt.%) of the andesite and gabbro hand samples with the certified samples

		SiO_2	Al_2O_3	Fe_2O_3	MnO	MgO	CaO	Na_2O	K_2O	TiO_2
Andesite	Andesite hand sample[a]	53.57	16.43	8.89	0.14	5.96	7.80	3.84	1.20	0.34
	Certified basalt	53.25	14.64	14.25	0.22	4.62	9.82	2.04	0.42	1.19
	Certified andesite	63.97	15.22	7.07	0.16	1.57	5.70	3.84	0.77	0.85
Gabbro	Gabbro hand sample[a]	53.24	14.58	8.69	0.14	9.08	8.70	2.68	0.47	0.82
	Certified basalt	53.25	14.64	14.25	0.22	4.62	9.82	2.04	0.42	1.19
	Certified gabbro	40.79	17.60	12.35	0.15	6.46	14.62	2.05	0.75	3.39

[a] Measured by Actlabs.

The last column in Table 4 compares the performance of the improved ANN with an ANN trained by the conventional method. The success rate is dropped to 62.9% while the mis-identification is significantly increased. The ANN trained by the sequential method clearly performs better. From the example of hematite and sulphide mixture we discussed earlier, due to the spatial heterogeneity, the identity of a whole rock should better be represented by all its spectra. We average all 50 spectra of a hand sample and input the line intensities to the ANN. The correct identification rate for the sequential training method is 100% (Fig. 7). In conventional training, it is only 57% with the rest results regarded as "*undetermined*". In Fig. 7, the matrix obtained from the improved method shows a higher *P* for the correct prediction with a cleaner background. This contrast confirms the reliability of the prediction. The fluorite and molybdenite hand samples do not have a similar class in the training set so their *P*s are low. The outstanding performance of the improved ANN shows a better generalization of the algorithm.

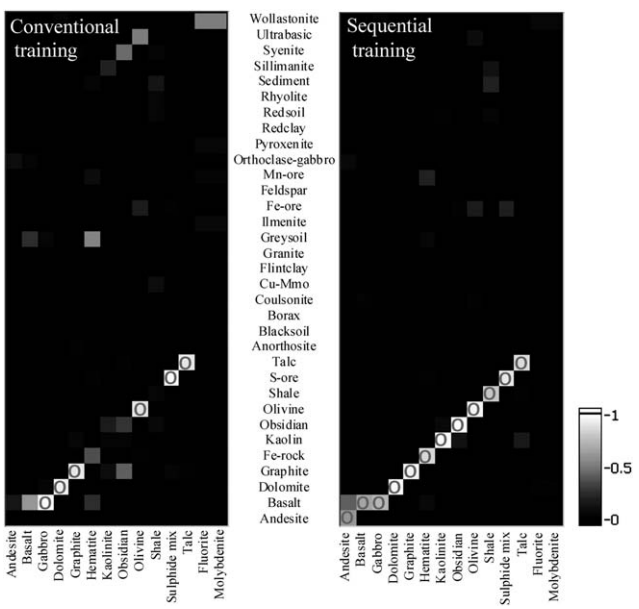

Fig. 7 Matrices displaying the identification performance of the average spectra of the hand samples by the LIBS-ANN trained by conventional method (left) and sequential method (right). The horizontal axis labels the hand samples and the vertical axis is the certified sample. Each cell represents a value *P* of an output neuron. A circle inside a cell indicates a correct identification of the ANN.

In summary, the identification result is exceptionally good if one considers the training of the network is accomplished by a set of certified powders while the test is performed with hand samples which are completely different in surface texture and homogeneity. Most importantly, even though our test consists of various classes which include silicates, carbonates, sulfide and oxides, the ANN can still perform well without suffering from the increased variety of samples. Nevertheless, the algorithm does not confuse new samples (*e.g.*, fluorite and molybdenite) with the known samples. These advantages promote the potential of this LIBS-ANN in planetary exploration.

4. Conclusion

We introduce an improved training process for an ANN algorithm which is applied on LIBS spectra analysis of geological samples. The hand samples are used to test the ANN which is trained by a set of certified samples. The technique can achieve a success identification rate of 88% using 10-shot averaged spectra. A 100% accuracy has been attained with 50-shot-averaged spectra.

As far as planetary exploration is concerned, it should be noted that this study is performed under atmospheric environment. For the Lunar or Martian condition, where the ambience is replaced by vacuum or low pressure CO_2, respectively, the emission is affected. The line intensity can also change with the sampling distance. Further study is required to estimate the impact of these effects on the ANN performance.

As initiated by the "mis-identification" in the hand sample study, besides simple mineralogical identification the LIBS-ANN can easily extend to a more sophisticated mineralogical assay with minimal instrumental conversion. Since a rock can be composed of various minerals, probing its surfaces at different spots will allow one to identify the rocks chemical variation and to identify the present mineral phases. Typically, a LIBS probe-spot is tens to hundreds of microns in size. A two dimensional systematic LIBS scan of a rock will make it possible to obtain a surface mineralogical map with sub-millimetre resolution. This map can reveal the mineral distribution on the rock surface. Current work is to focus on expanding the mineral database for identification as well as advancing the algorithm for quantitative assay. The two predictions from mineralogical and oxide assay can crosscheck each other to increase the reliability of the results.

Acknowledgements

We thank Dr Caroline-Emmanuelle Morisset for her useful discussions. We also thank two anonymous reviewers for their valuable comments.

References

1 D. A. Cremers and L. J. Radziemski, *Handbook of Laser-induced Breakdown Spectroscopy*, John Wiley & Sons, 2006.
2 A. W. Miziolek, V. Palleschi and I. Schechter, *Laser Induced Breakdown Spectroscopy*. Cambridge University Press, 2006.
3 D. A. Cremers and R. C. Chinni, *Appl. Spectrosc. Rev.*, 2009, **44**, 457.
4 E. Gibb-Snyder, B. Gullett, S. Ryan, L. Oudejans and A. Touati, *Appl. Spectrosc.*, 2006, **60**(8), 860.
5 N. Lanza, R. C. Wiens, S. M. Clegg and the ChemCam Team (LA-UR 09-02368), presented in part at the 2nd North American Symposium on Laser-induced Breakdown Spectroscopy, New Orleans, July, 2009.
6 W. E. Ernst, D. F. Farson and D. J. Sames, *Appl. Spectrosc.*, 1996, **50** (3), 306.
7 R. Noll, H. Bette, A. Brysch, M. Kraushaar, I. Mönch, L. Peter and V. Sturm, *Spectrochim. Acta, Part B*, 2001, **56**, 637.
8 O. Samek, H. H. Telle and D. C. S. Beddows, *BMC Oral Health*, 2001, **1**, 1, DOI: 10.1186/1472-6831-1-1.
9 A. Jurado-López and M. D. Luque de Castro, *Spectrochim. Acta, Part B*, 2003, **58**, 1291.
10 J.-B. Sirven, B. Sallé, P. Mauchien, J.-L. Lacour, S. Maurice and G. Manhès, *J. Anal. At. Spectrom.*, 2007, **22**, 1471.
11 A. Ramil, A. J. López and A. Yáñez, *Appl. Phys. A: Mater. Sci. Process.*, 2008, **92**, 197.
12 S. M. Clegg, E. Sklute, M. D. Dyar, J. E. Barefield and R. G. Wien, *Spectrochim. Acta, Part B*, 2009, **64**, 79.
13 J. L. Gottfried, R. S. Harmon, F. C. De Lucia, Jr. and A. W. Miziolek, *Spectrochim. Acta, Part B*, 2009, **64**, 1009.
14 R. S. Harmon, J. Remus, N. J. McMillan, C. McManus, L. Collins, J. L. Gottfried, F. C. De Lucia, Jr. and A. W. Miziolek, *Appl. Geochem.*, 2009, **24**(6), 1125.
15 P. Inakollu, T. Philip, A. K. Rai, F.-Y. Yueh and J. P. Singh, *Spectrochim. Acta, Part B*, 2009, **64**, 99.
16 E. C. Ferreira, D. M. B. P. Milori, E. J. Ferreira, R. M. Da Silva and L. Martin-Neto, *Spectrochim. Acta, Part B*, 2008, **63**, 1216.
17 R. Sattmann, I. Mönch, H. Krause, R. Noll, S. Couris, A. Hatziapostolou, A. Mavromanolakis, C. Fotakis, E. Larrauri and R. Miguel, *Appl. Spectrosc.*, 1998, **52**(3), 456.
18 V. Motto-Ros, A. S. Koujelev, G. R. Osinski and A. E. Dudelzak, *J. Eur. Opt. Soc. Rapid Publications*, 2008, **3**, 08011.
19 A. Koujelev, V. Motto-Ros, D. Gratton and A. Dudelzak, *Can. Aeronaut. Space J.*, 2009, **55**, 97.
20 A. Koujelev, M. Sabsabi, V. Motto-Ros, S. Laville and S. L. Lui, *Planet. Space Sci.*, 2010, **58**, 682.
21 J. E. Moody, *Advances in neural information processing systems*, 1991, **4**, 847.
22 L. Prechelt, in *Neural Network: Tricks of the Trade, LNCS 1524*, ed. G. B. Orr and K.-R. Müller, Springer Verlag, Heidelberg, 1998, 55.
23 V. Kadirkamanathan and M. Niranjan, *Neural Comput.*, 1993, **5**(6), 954.
24 S. Rajasekaran, D. Suresh and G. A. Vijayalakshmi Pai, *Eng. Comput.*, 2002, **18**, 138.
25 S. Rajasekaran, K. Thiruvenkatasamy and T. L. Lee, *Appl. Math. Modell.*, 2006, **30**(1), 85.
26 http://www.ima-mineralogy.org/IMAlistmin.html.
27 http://webmineral.com/danaclass.shtml.
28 W. S. Sarle, *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, 1995, 352.
29 L. Carbonara and A. Borrowman, *Proceedings of PKDD*, 1998, 264.
30 C. M. Higgins and R. M. Goodman, *Proceedings of the Joint Conference of Neural Networks*, 1991, 875.
31 R. Polikar, L. Udpa, S. S. Udpa and V. Honavar, *IEEE Transactions on System, Man and Cybernetics*, 2001, 497.
32 M. C. Nelson and W. T. Illingworth, *A practical guide to neural nets*, MA, Addison-Wesley, 1991, 165.