PAPER

Check for updates

Comparative analysis of LDA, PLS-DA, SVM, RF, and voting ensemble for discrimination origin in greenish-white to white nephrites using LIBS

Meiyu Shih, [ID] Ye Yuan and Guanghai Shi*

As there are distinct variations in economic value for greenish-white to white nephrites based on their geographical origin, it is crucial to develop a robust origin discrimination method for them. The reported correlation between the intensity of spectra and material properties gives us a clue that such a correlation may exist in nephrites worldwide. In this study, 364 pieces of greenish-white to white nephrite jades from different locations, including Qiemo, Qinghai, Xiuyan and Yecheng in China, South Korea, and Russia, were analyzed using laser-induced breakdown spectroscopy (LIBS). Four machine learning methods, including linear discriminant analysis (LDA), support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), random forest (RF), and an ensemble learning approach known as a voting classifier for origin discrimination were then employed. The results show a higher training accuracy of 99.81% (LDA), 94.01% (SVM), 100% (PLS-DA), 98.08% (RF), and 99.93% (voting classifier), with corresponding testing accuracies of 96.13%, 93.04%, 94.99%, 95.90%, and 99.93%, respectively. By appropriately selecting voting weights, the voting classifier effectively mitigates misclassification, achieving balanced accuracy for each origin. Therefore, the LIBS analyses could be utilized in the origin discrimination of greenish-white to white nephrite jades, offering valuable insights for accurately evaluating these gemstones, based on the successful application of various machine learning methods in the origin discrimination of nephrite jades. An integrated voting ensemble method was further introduced, providing new possibilities for rapid discrimination in diverse industries, including gemstone trading, manufacturing, archaeology, and more.

1 Introduction

Nephrite, a rock primarily composed of nearly monomineralic tremolite-actinolite $(Ca_2(Mg, Fe)_5Si_8O_{22}(OH)_2)$, is distributed worldwide.[1–8] It is classified into dolomite-related and serpentine-related types based on the parent and host rocks associated with ore formation. Both types of nephrites are formed through a process known as metasomatism.[6,9,10]

The geographical origin of nephrite jade carries significant cultural significance in various regions. In archaeology, distinguishing jades and gemstones from diverse origins aids in a deeper understanding of historical cultural exchanges.[11–13] Although examining the microstructure of dolomite-related nephrite offers some assistance, uncertainties remain.[14,15]

However, determining the origin of nephrite jade based solely on simple observations presents challenges. Researchers have achieved informative results by utilizing trace elements and hydrogen–oxygen isotopes to identify geographical origins.[16–18] Nevertheless, this often necessitates complex sample pretreatment, potentially damaging the samples.

School of Gemology, China University of Geosciences Beijing, Beijing 100083, China. E-mail: shigh@cugb.edu.cn

Nonetheless, distinguishing the origins of nephrite jade remains challenging, especially regarding rapid discrimination. Laser-induced breakdown spectroscopy (LIBS) is a technique that employs laser-induced photon emission from materials and detects it. LIBS offers advantages such as short testing times (less than 1 min) and no need for sample preparation. In the field of gemology, LIBS has been successfully applied to determine the origins of gemstones such as rubies, sapphires, diamonds, and others.[19–21]

Furthermore, the application of various multivariate models such as linear discriminant analysis (LDA), partial least squares (PLS), and their respective improved methods, as well as support vector machine (SVM), has become commonplace in the field of gemology.[22–26]

However, previous studies on nephrite jade discrimination have often been constrained by less than 200 samples; such a number is too small for modeling. As the dataset employed for modeling typically does not exceed 500 entries, integrated methods are seldom employed. Samples with different tones may reflect significant variations in mineralization environments, suggesting that their chemical compositions might vary as well.[27]

When dealing with a larger sample number (more than 100 samples), the selection process usually overlooks the variation

1560 | *J. Anal. At. Spectrom.*, 2024, 39, 1560–1570

This journal is © The Royal Society of Chemistry 2024

in color tones among the samples. Due to these factors, the necessity for comprehensive modeling for larger nephrite datasets is neglected. For instance, with chemical compositions, rare Earth elements, and isotopes, we can successfully distinguish between a few origins of nephrite; however, when faced with more diverse origins of samples, they may fall short.[28] In such cases, it becomes essential to explore a robust discrimination method to differentiate certain origins of nephrite that have similar tones accurately.

Previous studies have indicated that it is tough to differentiate the origin of chemically similar nephrite jades based on their major element contents. Fortunately, the content of minor elements, such as Be, Li, Na, Al, Mn, Sr, K, Zr, Fe, La, and Ce, has been reported to carry significant records in determining the origin.[23,29,30] All these elements could be analyzed using LIBS. In addition, using LIBS, variations in the intensity of major elements at the same emission lines could be determined among nephrites from different origins.[31] These advantages of using LIBS prove it to be a powerful tool for differentiating the origin although the exact reasons behind these differences remain unclear.

This study aims to analyze the discrimination of 364 pieces of greenish-white to white nephrite jades from different locations, including Qiemo, Qinghai, Xiuyan, and Yecheng in China, South Korea, and Russia, using a combination of multiple classifiers, including PLS-DA, LDA, SVM, and RF, based on LIBS, as well as a single voting ensemble method.

We explore the effectiveness of these methods in addressing the complexity of discriminating origins of nephrite, evaluate the performance of these classifiers, discuss potential factors influencing variations in the intensity of spectral lines, including differences in chemical composition and physical matrix effects such as crystallinity and transparency, and attempt to provide a novel ensemble approach for the origin discrimination of nephrite jades, which may contribute to the advancement of the field of geographical identification of gemstones and jades, and promoting similar research on other materials.

2 Methods

2.1 Samples

In this study, a total of 364 nephrite samples from various geographical origins were included for experimentation (Table 1): 60 samples from Chuncheon in Korea, 70 from the Lake Baikal region in Russia, 51 samples from Qiemo in China, 67 from Yecheng in China, 55 from Qinghai in China, and 61 from Xiuyan in China. All the samples used in this experiment were provided by a nephrite trading company with assistance from Prof. Yu Ming at the Chinese Jade Culture Research Center, Central Academy of Fine Arts. Representative specimens from different origins are shown in Fig. 1.

2.2 Experimental setup

The LIBS system (ChemReveal 3766, TSI Inc) at the Gemological Center of China University of Geosciences, Beijing, consisting of a pulsed Nd: YAG laser was operated at 1064 nm, emitting laser pulses with an energy of 50 mJ at a repetition rate of 1 Hz, in air at ambient pressure, in a spectral range from 188 to 980 nm, with a resolution of 0.1–0.2 nm. The laser spot size at the sample surface is approximately 100 μm.

2.3 Measurements and data processing

To ensure accurate and consistent testing results, we followed a systematic procedure to guarantee its plain face upwards: utilizing the real-time camera feature to capture a clear image within a field of view of approximately 500 μm × 500 μm, and then carefully assessed the image's clarity by adjusting the sample height to maintain a height deviation within 5 μm for optimal sharpness. All data collection spots were within the same focal plane to minimize possible errors due to surface irregularities.

Before conducting LIBS measurements on our nephrite samples, we implemented a pre-analysis cleaning procedure involving two pre-ablation shots at the same location on each sample in order to eliminate any dust or potential surface pollution that may have accumulated during handling. These pre-ablation measurements were excluded from the subsequent data analysis. After this cleaning, three different areas were carefully selected on each sample. As a single shot may not capture the inhomogeneity of nephrite samples, ten laser analyses of each area were conducted, which were then averaged to create a single spectrum. As a result, three average spectra per sample were recorded.

Following the data collection, we applied a standard normal variance (SNV) transformation to all the LIBS spectra.[32–34] This transformation aimed to facilitate incorporating additional data into our database for future research endeavors. Next, we filtered out 370 characteristic peaks by utilizing an automatic peak-seeking program provided by the manufacturer, which operated within Chemlytics. This program helped us identify the characteristic peaks in nephrite spectra originating from six different origins. Subsequently, we excluded peaks with intensities exceeding 14 500 (the maximum detection range is 16 384), resulting in a total of 261 peaks. These characteristic peaks were input variables for LDA, PLS-DA, SVM, and RF models processed in Python 3.7. Our analysis randomly designated 819 spectra as the training set, while the remaining 273 spectra were allocated as the test set.

The training of these models involved using the LIBS spectra from the training set. Each spectrum was associated with a specific origin, determined by reference techniques. During

Table 1 Description of the 364 nephrite samples included in this case study

Numbers of samples	Provenance/origin	Code mark
60	Chuncheon in Korea	HL
70	Lake Baikal region in Russia	ELS
51	Qiemo of Xinjiang in China	QM
55	Qinghai in China	QH
61	Xiuyan of Liaoning in China	XY
67	Yecheng of Xinjiang in China	YC

This journal is © The Royal Society of Chemistry 2024

J. Anal. At. Spectrom., 2024, **39**, 1560–1570 | **1561**
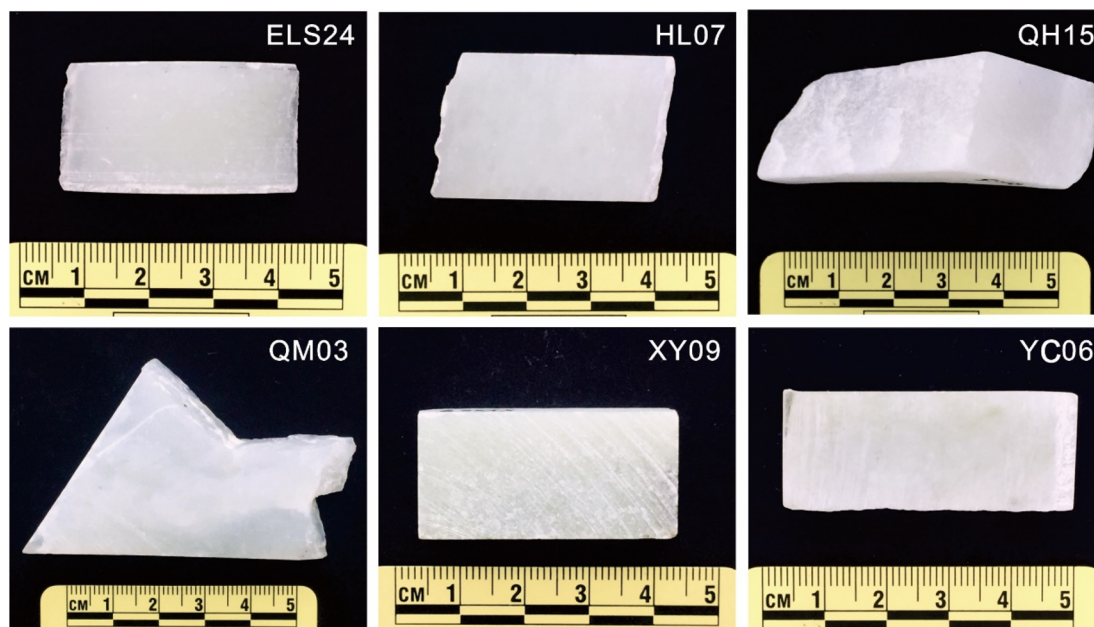
Fig. 1 The images of 6 representative greenish-white to white nephrite jades, one from each origin (denoted as ELS24, HL07, QH15, QM03, XY09 and YC03).

the training phase, each model learned the spectra associated with the origin of the sample. Once the training was completed using the 819 samples dedicated to this task, each resulting model was evaluated using the test set. Each sample from the test set was introduced into these models during this stage, and the model attributed an origin to each sample. By repeating this operation on the 273 spectra designated for testing, we could assess the model's performance in differentiating the spectra and, consequently, the origins of nephrites.

We fine-tuned model parameters such as the number of decision trees, maximum depth, and latent variables to enhance computational performance using ten-fold cross-validation and accuracy as the evaluation metrics. This selection process aimed to optimize the overall performance of the data analysis.

2.4 Machine learning methods

The ability to discriminate between nephrites from different origins based on their LIBS spectra was assessed using LDA, PLS-DA, SVM, and RF algorithms in Python 3.7. LDA[35] is a technique that projects high-dimensional sample data X into a vector space optimized for classification purposes, allowing for the extraction of classification information while reducing the dimensionality of the features. Its underlying principle is to maximize the distances between different classes while minimizing the distances within each class.

SVM is a statistical method to solve the separation hyperplane, which can divide the training data set correctly and has the maximum geometric interval.

Partial least squares discriminant analysis (PLS-DA)[34] is a discriminant method that employs partial least squares regression. This approach entails constructing a linear model to regress the sample data X against a categorical matrix Y, which

is subsequently utilized for classification purposes. Unlike traditional linear regression methods, PLS-DA projects both X and Y into a new space and extracts latent variables that maximize the covariance between X and Y.

Random forest (RF)[36] is an ensemble learning algorithm that enhances prediction accuracy and robustness by creating multiple decision trees and randomly selecting features for training. RF is capable of handling classification and regression problems that involve a large number of features and samples.

A voting ensemble classifier, often called a voting classifier,[37,38] is a machine-learning technique that combines predictions from multiple models or classifiers. Each model contributes its prediction, and the final prediction is determined by aggregating the votes of these models. This method enhances accuracy and robustness by leveraging diverse perspectives provided by individual models.

3 Results

3.1 The variation of intensities

The average spectra within the 200–320 nm range indicate that nephrite jade from six locations exhibits remarkable similarities (Fig. 2), whereas slight differences are observed in some spectral lines, as shown in the bar charts of the average spectral line intensity of the samples from each origin.

When examining several representative spectral lines (Fig. 3), such as Ca 364.4, Ca 409.8, Mg 202.6, Mg 292.8, Si 251.9, and Si 252.4, these bars in the histogram show somewhat similar distributions. This suggests a consistent pattern in the behavior of the emission of photons for that particular element in the same source, highlighting the variations in the excitation levels of the samples across different origins.

1562 | *J. Anal. At. Spectrom.*, 2024, **39**, 1560–1570

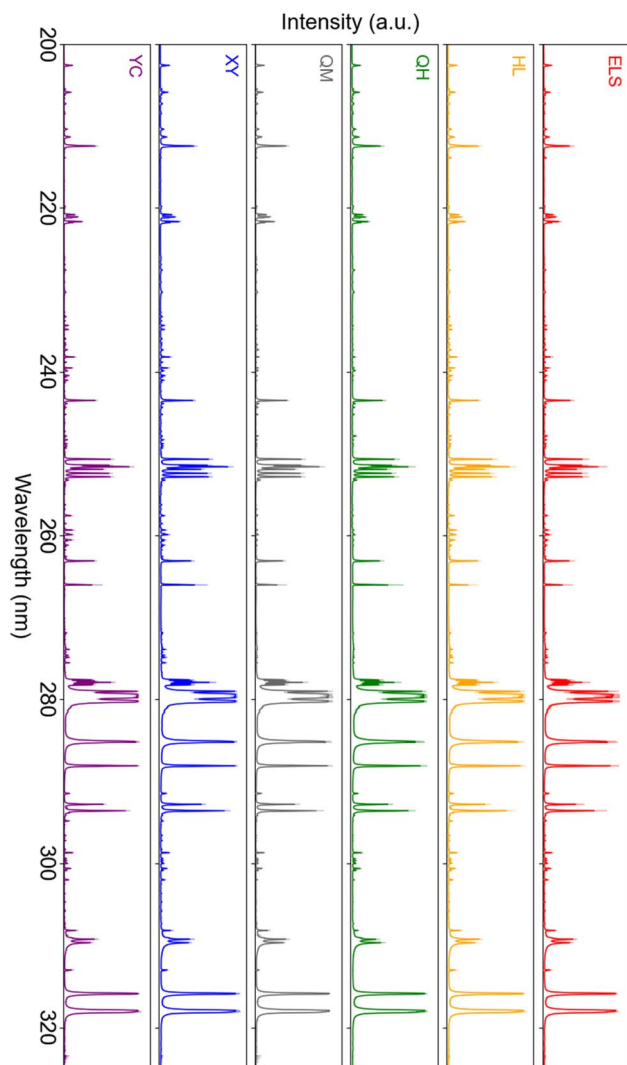This journal is © The Royal Society of Chemistry 2024

Fig. 2 The average LIBS spectra from nephrite jade samples of six different origins exhibit high similarities in the 200–320 nm wavelength range. The shaded area represents the standard deviation.

However, for spectral lines such as Al 308.2, Al 309.3, Be 234.8, and Be 313.0 (Fig. 3a), the heights exhibit relatively weaker average intensities in the Qinghai origins compared to the other origins. Mn 403.6, Mn 403.3, Sr 407.8, and Sr 421.6, show relative weakness in the Qiemo origins. For Li 670.7 and Na 819.5, the heights demonstrate relatively weaker average intensities in the Xiuyan origins. The heights of O 777.2 and H 656.3 exhibit no noticeable variations across different origins, suggesting that the test results are relatively consistent.

Plotting their intensities as scattered points across all samples by choosing the intensities observed in certain spectral lines of elements reveals a discernible clustering pattern for some origins (Fig. 4). The I_{Al}/I_{Be} ratio of samples from Qinghai exhibits a distinct characteristic of a left-lower clustering pattern relative to other origins (Fig. 4a). Such similar observations have been reported.[29] However, there is still a considerable overlap of scattered data points from various spectral lines *vs.* their origins in most cases (Fig. 4c), and *vice versa.*

There is a notable difference observed in an individual nephrite with higher and lower translucency from Qinghai (Fig. 5); in the higher-translucency region, a significant decrease in spectral line intensity was observed, particularly at $I_{Si\ 288.15\ nm}$ (Fig. 5b), with differences reaching up to 25%, whilst $I_{Mg\ 285.21\ nm}$ exhibited self-absorption in the lower-translucency portion (Fig. 5c).

3.2 Matrix effect of multivariate methods

The analysis of nephrite (Table 2) using LIBS spectra with four machine learning methods and a voting classifier achieved a total accuracy above 93% for all methods: a training accuracy of 99.81%, 94.01%, 100%, 98.08%, and 99.93%, and testing accuracy of 96.13%, 93.04%, 94.99%, 95.90%, and 99.93% for the LDA, PLS-DA, RF, and SVM algorithms, and voting classifier, respectively.

Some misclassifications appeared between specific methods and origins, resulting in a testing accuracy below 90% for specific categories. For example, higher error rates were occurred in RF for the Russian origin, and PLS-DA for the Korean origin.

In order to enhance the accuracy of specific origins, strengthen the overall robustness, and improve the reliability of the model, a voting ensemble method was conducted. In this approach, multiple models were employed to make predictions on the data, and majority voting was conducted based on their outcomes, by using which the final prediction was determined.

This methodology allowed for an effective balance of accuracy across various origins and yielded more dependable classification results. For instance, if the origin is Russia and classified as Qinghai in LDA, Russia in PLS-DA, Russia in RF, and Korea in SVM, the final prediction would be Russia. Then weights were adjusted as follows: SVM > RF > LDA > PLS-DA. Similarly, if the origin is Korea and classified as Korea in LDA, Russia in PLS-DA, Russia in RF, and Korea in SVM, the final prediction would be Korea. The voting classifier achieved a total testing accuracy of 99.93%, effectively increasing the testing accuracy to above 99% for each origin.

4 Discussion

The results from multivariate methods demonstrate highly effective discrimination among the six origins of Qiemo, Qinghai, Xiuyan, and Yecheng in China, South Korea, and Russia. In nearly all machine learning methods, results of Russian and Korean origins exhibited a relatively higher number of misclassifications compared to other origins. This observation is likely attributed to the similarities in their ore forming fluid origins (meteoric waters) and comparable microstructures.[39,40] And the misclassifications from Korea are higher in LDA and PLS-DA, but lower in SVM and RF. Conversely, misclassifications from Russia display the opposite trend, possibly reflecting differences in model adaptability.

The heights of the spectral lines (Fig. 2) from Russia demonstrate a significantly higher error rate in most bars, suggesting that samples from this origin exhibit substantial variability and share certain characteristics with other origins,

This journal is © The Royal Society of Chemistry 2024

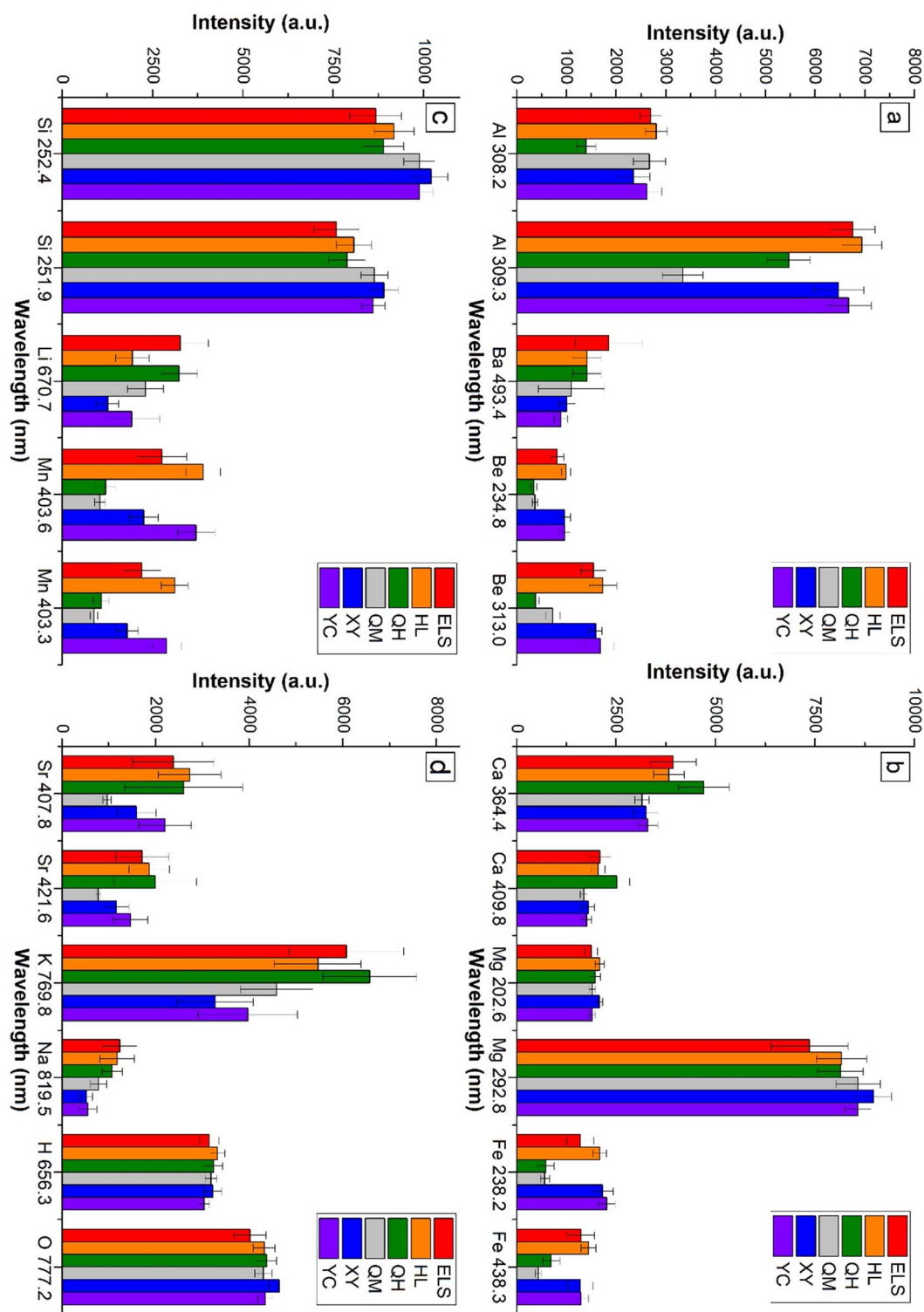*J. Anal. At. Spectrom.,* 2024, **39**, 1560–1570 | **1563**

Fig. 3   The characteristic peaks include: Al, Ba, and Be (a); Ca, Mg, and Fe (b); Si, Li, and Mn (c); Sr, K, Na, H, and O (d).

and therefore the likelihood of misclassifying Russian samples into different origins increases. Additionally, the uneven distribution of sample numbers within each origin may also contribute to an increased probability of misclassification.

According to the algorithms, certain spectral lines originating from trace elements such as Fe, Mn, Al, Sr, Be, Li, Na, and K carry more significant weights in discriminating between different origins. Fig. 3a shows that the intensity of the Be and Al area is noticeably lower in Qinghai compared to other locations. This result is closely related to the ore-forming mechanisms, as the lower Be intensity in Qinghai was reported to be attributed to the contact rocks being mafic rocks (gabbro),[41] being unlike the
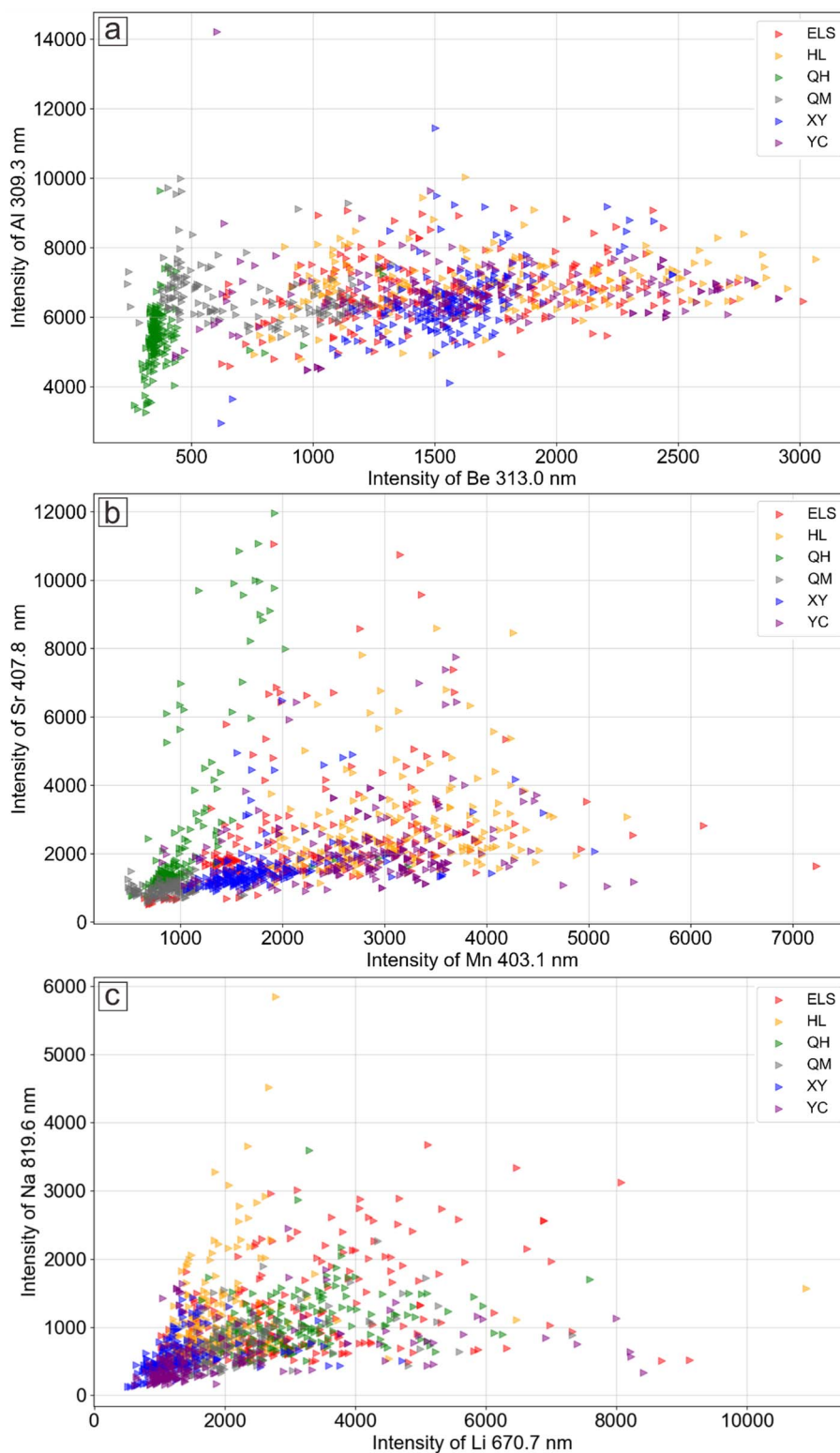
Fig. 4 Combined three scatter plots showing the relationship between two spectral lines and intensity: Be 313.0 and Al 309.3 (A), Sr 407.8 and Mn 403.1 (B), and Na 819.6 and Li 670.7 (C). The scatter plots illustrate the intensity variations and highlight the distinctive left-clustered distribution observed in the Qinghai region (A) and Qiemo region (B), which is distinct from that of other origins. However, in most cases, there are still some overlapping clustering distributions that are not apparent (C).The characteristic peaks include: Al, Ba, and Be (a); Ca, Mg, and Fe (b); Si, Li, and Mn (c); Sr, K, Na, H, and O (d).
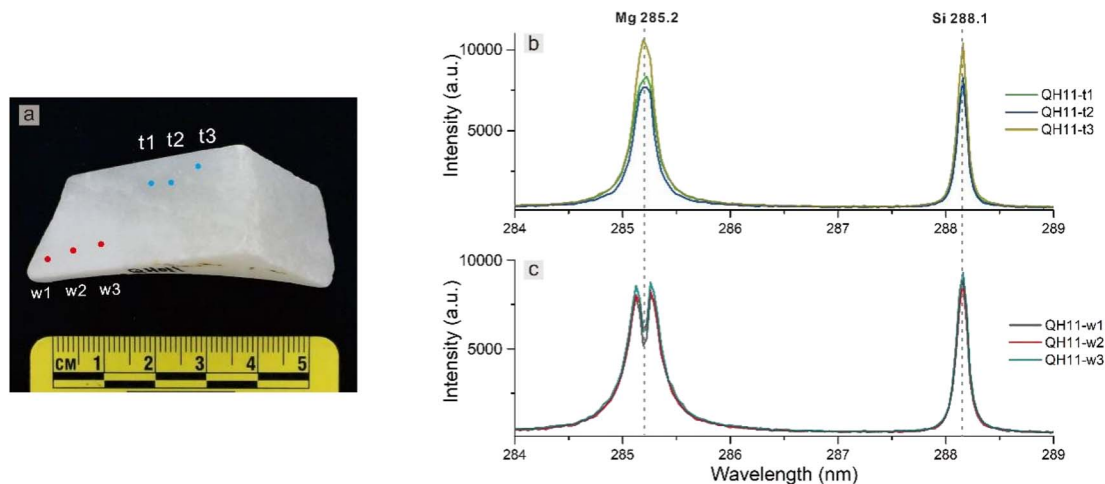
Fig. 5 Spectra of Qinghai nephrite jade and its test point. (a) The sample and its test points, spectra from the higher translucency portion (b), and spectra from the lower translucency portion (c). The dashed lines point to the characteristic peak.

intermediate-acidic rocks found in other areas that contain higher concentrations of Be. The relatively lower intensity of Al may be attributed to multiple stages of hydrothermal processes.[42]

As the intensity of emission peaks does not necessarily have a linear relationship with element concentrations, various physical matrix effects such as grain size, cohesive forces, sample roughness, crystallinity and the thermal and optical properties of the target material could contribute to this variation,[43] leading to corresponding variations in spectral lines.

Nephrite jade exhibits various microstructures and grain sizes, such as felted and fibrous textures. In their microstructural study, white jade in the Qinghai region often displays a distinctive fibroblastic texture,[44,45] and has higher crystallinity compared to that in Russia, Korea, and Xinjiang.[46] This texture can consist of weakly oriented fibers of similar sizes or the interweaving of fine fibers (>200 μm) with even finer fibers (<50 μm). As the laser spot size used in the test is approximately 100 μm, the potential to cause slight variations in spectral patterns may exist when ablating samples with different grain sizes. Nephrite jades from Korea and Russia often exhibit similar crystalline microstructures inherited from carbonatites, which are less commonly found in other origins.[39,40] Such a similarity might be the reason for the higher rates of misclassification observed.

As the intensity of Na and K spectral lines can be influenced by physical matrix effects associated with smaller grain sizes ranging from 250–500 μm,[43] the grain size of nephrite in our samples is less than 500 μm, mostly with size below 200 μm; this may explain why algorithms assigned higher weights to the spectral lines associated with Na and K.

The reduction in intensity at $I_{Si\ 288.15\ nm}$ (Fig. 5b) can be attributed to the loss of irradiance as the laser beam passes through the front surface and traverses the body of the higher-translucency sample.[47] This can be explained by the optical thickness of the plasma generated during high-energy pulsed laser ablation of the sample. When photons generated within the plasma propagate outward, they may be absorbed by similar

atoms or ions along their path. The self-absorption phenomena (Fig. 5c) are more likely to occur for transitions involving lower energy levels. This suggests that less homogeneous samples are more appropriate for collecting data from multiple locations on the sample surface.

In order to avoid the misclassification assigned through local optimality that the machine learning method encountered in some origins, the voting classifier was introduced. Although the total accuracy of the voting classifier was slightly lower than that of SVM in the test set, it achieved the highest accuracy in the training set. The voting classifier proves to be a powerful tool for enhancing the robustness of the model for origin discrimination.

The misclassifications can be attributed not only to the chemical composition but also to the material characteristics. Several factors would influence the effectiveness of origin discrimination, including similarities in the nephritization processes between the origins, heterogeneity in the samples and the groups, the number of samples introduced, instrument sensitivity and/or instability, and the complexity and expressive capacity of different machine learning models. Minimizing these misclassifications is crucial to ensure the robustness and reliability of the results.

Due to the limited detection of sensitivity for the LIBS instrument in this study, potentially valuable records of medium and heavy rare earth elements were prevented. To address this issue, one possible approach is to increase the number of tests conducted on each sample to obtain more comprehensive information. However, this may cause visible surface damage on the samples in such cases, which is unacceptable in many fields. Therefore, establishing a reliable method and a sufficiently large LIBS database would effectively reduce the testing numbers for samples in future research studies. Besides, we seek other methods to minimize the damage caused by LIBS analysis of the samples. Note that elements with low concentration such as Na, Sr, Be, and Li exhibit clear detectability in LIBS, but conversely, it is

1566 | J. Anal. At. Spectrom., 2024, 39, 1560–1570

This journal is © The Royal Society of Chemistry 2024

Paper

Table 2 Average ten-fold cross-validation confusion matrices for six origins, including four multivariate statistical methods and a voting classifier

Prediction results of the training set							Prediction results of the test set						
ELS	HL	QH	QM	XY	YC	Accuracy	ELS	HL	QH	QM	XY	YC	Accuracy
LDA							**LDA**						
ELS 156.9	0.1	0	0	0	0	99.94%	ELS 50.2	2.8	0	0	0	0	94.72%
HL 0	135	0	0	0	0	100%	HL 3	42	0	0	0	0	93.33%
QH 0	0.7	123.3	0	0	0	99.44%	QH 0	1.3	39.3	0	0.4	0	95.85%
QM 0.5	0	0	114.5	0	0	99.57%	QM 0.6	0	0	37.3	0.1	0	98.16%
XY 0	0	0	0	136.9	0.1	99.93%	XY 0	0	0	0	44.5	1.5	96.74%
YC 0	0	0	0	0	151	100%	YC 0	0	0	0	1	49	98.00%
Total correct rate for the training set						99.81%	Total correct rate for the testing set						96.13%

Prediction results of the training set							Prediction results of the test set						
ELS	HL	QH	QM	XY	YC	Accuracy	ELS	HL	QH	QM	XY	YC	Accuracy
PLD-DA							**PLD-DA**						
ELS 149.9	6.3	0	0.8	0	0	95.48%	ELS 49.6	3.1	0.1	0.2	0	0	93.58%
HL 11.1	121.8	0	0	1.9	0.2	90.22%	HL 5.4	38.6	0	0	0.9	0.1	85.78%
QH 0	4.2	119.8	0	0	0	96.61%	QH 0	1.3	39.7	0	0	0	96.83%
QM 4.5	0	0	109.7	0.3	0.5	95.39%	QM 0.7	0	0	37	0.1	0.2	97.37%
XY 0	2.1	0.5	0.1	121.2	13.1	88.47%	XY 0.1	0.7	0.4	0	40.7	4.1	88.48%
YC 0	1.7	0	0.2	1.3	147.8	97.88%	YC 0.1	0.9	0	0.2	0.7	48.1	96.20%
Total correct rate for the training set						94.01%	Total correct rate for the testing set						93.04%

Prediction results of the training set							Prediction results of the test set						
ELS	HL	QH	QM	XY	YC	Accuracy	ELS	HL	QH	QM	XY	YC	Accuracy
RF							**RF**						
ELS 157	0	0	0	0	0	100%	ELS 47.3	4.6	0.5	0.1	0.2	0.3	89.25%
HL 0	135	0	0	0	0	100%	HL 3.5	40.8	0	0	0.4	0.3	90.67%
QH 0	0	124	0	0	0	100%	QH 0	0.8	40.2	0	0	0	98.05%
QM 0	0	0	115	0	0	100%	QM 0	0	0.1	37.8	0.1	0	99.47%
XY 0	0	0	0	137	0	100%	XY 1.4	0.1	0	0	43.3	1.2	94.13%
YC 0	0	0	0	0	151	100%	YC 0.2	0.2	0	0	0.4	49.2	98.40%
Total correct rate for the training set						100%	Total correct rate for the testing set						94.99%

Prediction results of the training set							Prediction results of the test set						
ELS	HL	QH	QM	XY	YC	Accuracy	ELS	HL	QH	QM	XY	YC	Accuracy
SVM							**SVM**						
ELS 150.8	6	0	0	0	0.2	96.05%	ELS 48.9	3.3	0	0	0	0.8	92.26%
HL 3.9	131.1	0	0	0	0	97.11%	HL 2.8	42.1	0	0	0.1	0	93.56%
QH 0	0.3	123.7	0	0	0	99.76%	QH 0	0.5	40.5	0	0	0	98.78%
QM 0.2	0	0	114.8	0	0	99.83%	QM 0.7	0	0	37.1	0	0.2	97.63%
XY 0	1.7	0	0	131.3	4	95.84%	XY 0.1	0.4	0.1	0	43.4	2	94.35%
YC 0.2	0	0	0	0	150.8	99.87%	YC 0.6	0	0	0	0	49.4	98.80%
Total correct rate for the training set						98.08%	Total correct rate for the testing set						95.90%

Prediction results of the training set							Prediction results of the test set						
ELS	HL	QH	QM	XY	YC	Accuracy	ELS	HL	QH	QM	XY	YC	Accuracy
Voting							**Voting**						
ELS 157	0	0	0	0	0	100%	ELS 157	0	0	0	0	0	100%
HL 0	135	0	0	0	0	100%	HL 0	135	0	0	0	0	100%
QH 0	0.3	123.7	0	0	0	99.76%	QH 0	0.3	123.7	0	0	0	99.76%
QM 0.1	0	0	114.9	0	0	99.91%	QM 0.1	0	0	114.9	0	0	99.91%
XY 0	0	0	0	136.9	0.1	99.93%	XY 0	0	0	0	136.9	0.1	99.93%
YC 0	0	0	0	0	151	100%	YC 0	0	0	0	0	151	100%
Total correct rate for the training set						99.93%	Total correct rate for the testing set						99.93%

significantly lower in XRF. This may contribute to research on gemstones rich in sodium components, such as jadeite and albite. Meanwhile, analyzing data from other complementary techniques would also contribute to research progress on LIBS.

5 Conclusions

This study aimed to compare the classification performance of four methods for six origins of nephrites based on LIBS spectra. Furthermore, a voting classifier was utilized to integrate the results, enabling a comprehensive analysis. The analysis revealed that slight variations in the intensity of some spectral lines are related to the physical state, such as the microstructure, transparency, grain size of the samples, and the number of spectra included in the model. It is necessary to collect more spectra from multiple spots for each sample to establish an optimal model, especially when dealing with polycrystalline samples.

Using machine learning methods, the LIBS technique successfully differentiated the origins of greenish-white to white nephrite jades from six locations. Combining four algorithms using a voting ensemble approach effectively balanced the misclassification rates in certain origins within these algorithms. This method should be considered a key approach for rapid and reliable decision-making in sourcing greenish-white to white nephrite.

These preliminary results have motivated us to develop LIBS technology further and explore the advanced data processing methods described in this study for broader applications across diverse origins and materials. Therefore, LIBS is a feasible and promising option for analyzing the origins of nephrite jade.

It is more likely that LIBS could potentially become a routine reference technique in archaeological and jade analyses within current geographical and chrono-cultural frameworks. Finally, with the possibility of extending this method, it has the potential to encompass a wider array of gemstones, geological materials, and other sample types, providing enhanced analysis of different material backgrounds and sourcing applications.

Author contributions

Meiyu Shih collected and processed the experimental data, conducted data analysis and interpretation, and drafted the manuscript. Ye Yuan provided expert guidance and supervision throughout the research process, offering critical revisions and guidance. Guanghai Shi made significant contributions of valuable revisions, project administration, and guidance during the entire research process. All authors critically reviewed and approved the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 I. Adamo and R. Bocchio, Nephrite Jade from Val Malenco, Italy: Review and Update, *Gems Gemol.*, 2013, **49**, 98–106, DOI: **10.5741/GEMS.49.2.98**.

2 G. L. Barnes, Understanding Chinese jade in a world context, *J. Br. Acad.*, 2018, **6**, 1–63, DOI: **10.5871/jba/006.001**.

3 G. Gil, J. D. Barnes, C. Boschi, P. Gunia, G. Sazkmány, Z. Bendő, P. Raczyński and B. Péterdi, Origin of serpentinite-related nephrite from Jordanów and adjacent areas (SW Poland) and its comparison with selected nephrite occurrences, *Geol. Q.*, 2015, **59**, 457–472, DOI: **10.7306/gq.1228**.

4 S. J. Kim, D. J. Lee and S. Chang, A Mineralogical and Gemological Characterization of the Korean Jade from Chuncheon: Korea, *J. Geol. Soc. Korea*, 1986, **22**(3), 278–288.

5 A. Kochnev, D. Krasnov and R. Ivanova, Experience Of Multifactor Local Forecasting On Example Of The Golubinsko-Ollaminskoe Nephrite Field (Republic Of Buryatia), *Earth Sciences and Subsoil Use*, 2018, **41**, 50–66, DOI: **10.21285/2541-9455-2018-41-4-50-66**.

6 T. F. Yui, H. W. Yeh and C. W. Lee, Stable isotope studies of nephrite deposits from Fengtien, Taiwan, *Geochim. Cosmochim. Acta*, 1988, **52**, 593–602, DOI: **10.1016/0016-7037(88)90321-3**.

7 X. Zhang, G. Shi, X. Zhang and K. Gao, Formation of the Nephrite Deposit with Five Mineral Assemblage Zones in the Central Western Kunlun Mountains, China, *J. Pet.*, 2022, **63**(11), 1–16, DOI: **10.1093/petrology/egac117**.

8 J. J. Hockley, Nephrite (jade) Occurrence in the Great Serpentine Belt of New South Wales, Australia, *Nature*, 1974, **247**, 364, DOI: **10.1038/247364a0**.

9 G. E. Harlow and S. S. Sorensen, Jade (Nephrite and Jadeitite) and Serpentinite: Metasomatic Connections, *Int. Geol. Rev.*, 2005, **47**, 113–146, DOI: **10.2747/0020-6814.47.2.113**.

10 T. Yanling, C. Baozhang and J. Renghua, *Chinese Hetian Jade*, Xinjiang People's Publishing House; Taiwan Earth Press, 1994.

11 D. Syvilay, B. Bousquet, R. Chapoulie, M. Orange and F. X. Le Bourdonnec, Advanced statistical analysis of LIBS spectra for the sourcing of obsidian samples, *J. Anal. At. Spectrom.*, 2019, **34**(5), 867–873, DOI: **10.1039/C8JA00340H**.

12 N. Tsydenova, M. V. Morozov, M. V. Rampilova, Y. A. Vasil'Ev, O. P. Matveeva and P. B. Konovalov, Chemical and spectroscopic study of nephrite artifacts from Transbaikalia, Russia: Geological sources and possible transportation routes, *Quatern. Int.*, 2015, **355**, 114–125, DOI: **10.1016/j.quaint.2014.07.065**.

13 Y. Qin, Y. Y. Xu, H. M. Li, S. X. Li and Q. F. Xi, Turquoise Mine and Artefact Correlation for Some Bronze Age Archaeological Sites in Hubei Province, China, *Archaeometry*, 2015, **57**(5), 788–802, DOI: **10.1111/arcm.12126**.

1568 | *J. Anal. At. Spectrom.*, 2024, **39**, 1560–1570

This journal is © The Royal Society of Chemistry 2024

14 D. Chen, M. Yu, W. Luo and C. Wang, Sub-microstructures of Nephrite from Five Sources Based on Multispectral Imaging and Effect Enhancement, *Asian J. Adv. Res. Rep.*, 2020, **12**(3), 13–24, DOI: **10.9734/AJARR/2020/v12i330288**.

15 S. Wang and L. Sun, Visual Identification of Tremolite Features of Five Origins in Today's Nephrite Jade Market, in *2013 China Gems & Jewelry Academic Conference*, Beijing, 2013, **https://kns.cnki.net/KCMS/detail/detail.aspx?filename=GTYS201310001036&dbname=CPFDTEM**.

16 Z. W. Zhang, Y. C. Xu, H. S. Cheng and F. X. Gan, Comparison of trace elements analysis of nephrite samples from different deposits by PIXE and ICP-AES, *X-Ray Spectrom.*, 2012, **41**(6), 367–370, DOI: **10.1002/xrs.2413**.

17 K. Gao, T. Fang, T. Lu, Y. Lan, Y. Zhang, Y. Wang and Y. Chang, Hydrogen and Oxygen Stable Isotope Ratios of Dolomite-Related Nephrite: Relevance for its Geographic Origin and Geological Significance, *Gems Gemol.*, 2020, **56**(2), 266–280, DOI: **10.5741/GEMS.56.2.266**.

18 J. Wang and G. Shi, Comparative Study on the Origin and Characteristics of Chinese (Manas) and Russian (East Sayan) Green Nephrites, *Minerals*, 2021, **11**(12), 1434, DOI: **10.3390/min11121434**.

19 C. E. Mcmanus, Determination Of Diamond Provenance Is Possible With Multivariate Analysis Of Libs Spectra, in *GSA Annual Meeting*, Baltimore Convention Center, 2015, **https://gsa.confex.com/gsa/2015AM/webprogram/Paper261650.html**.

20 K. A. Kochelek, N. J. McMillan, C. E. McManus and D. L. Daniel, Provenance determination of sapphires and rubies using laser-induced breakdown spectroscopy and multivariate analysis, *Am. Mineral.*, 2015, **100**(8), 1921–1931, DOI: **10.2138/am-2015-5185**.

21 P. Bao, Q. Chen, A. Zhao and Y. Ren, Identification of the Origin of Bluish White Nephrite Base on Laser-Induced Breakdown Spectroscopy and Artificial Neural Network Model, *Spectrosc. Spect. Anal.*, 2023, **43**(1), 25–30, DOI: **10.3964/j.issn.1000-0593(2023)01-0025-06**.

22 M. C. Ortiz, L. Sarabia, A. Jurado-López and L. D. Castro, Minimum value assured by a method to determine gold in alloys by using laser-induced breakdown spectroscopy and partial least-squares calibration model, *Anal. Chim. Acta*, 2004, **515**(1), 151–157, DOI: **10.1016/j.aca.2004.01.003**.

23 W. Han, L. Bi, J. Ke, H. Chen and T. Lu, Artificial Intelligence +Gem Identification: Origin Determination of White Nephrite Using Laser-induced Breakdown Spectroscopy and Support Vector Machines Algorithm, *2017 Proceeding China international Gems & Jewerly Academic Conferences*, Beijing, China, 2017, pp. 334–336, **https://d.wanfangdata.com.cn/conference/9924551**.

24 Y. Wang, X. Yuan, B. Shi, Q. Zhang and T. Chen, Origins of Nephrite by Laser-Induced Breakdown Spectroscopy Using Partial Least Squares Discriminant Analysis, *Chin. J. Lasers*, 2016, **43**(12), 254–261, DOI: **10.3788/CJL201643.1211001**.

25 S. M. Clegg, E. Sklute, M. D. Dyar, J. E. Barefield and R. C. Wiens, Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques, *Spectrochim. Acta, Part B*, 2009, **64**(1), 79–88, DOI: **10.1016/j.sab.2008.10.045**.

26 A. Zhou, J. Jiang, C. Sun, X. Xu and X. Lü, Identification of Different Origins of Hetian Jade Based on Statistical Methods of Multi-Element Content, *Spectrosc. Spect. Anal.*, 2020, **40**(10), 3174–3178, DOI: **10.3964/j.issn.1000-0593(2020)10-3174-05**.

27 H. Yu, R. Wang, J. Guo, J. Li and X. Yang, Study of the minerogenetic mechanism and origin of Qinghai nephrite from Golmud, Qinghai, Northwest China, *Sci. China: Earth Sci.*, 2016, **59**(8), 1597–1609, DOI: **10.1007/s11430-015-0231-8**.

28 Y. Su and M. Yang, Combining Rare Earth Element Analysis and Chemometric Method to Determine the Geographical Origin of Nephrite, *Minerals*, 2022, **12**(11), 1399, DOI: **10.3390/min12111399**.

29 T. U. Cai, Y. Xin-qiang and Z. Zhao, Determination of Be and Al for Source Region Identification of White Nephrite Using Laser-Induced Breakdown Spectroscopy, *J. Rock Miner. Anal.*, 2012, **31**(2), 301–305, DOI: **10.3969/j.issn.0254-5357.2012.02.020**.

30 J. Li, R. Wu, X. Ling and Q. Li, Study on Chemical Compositions and Microstructures of Nephrite from Hetian, *J. Gems Gemmol.*, 2009, **11**(4), 9–14, DOI: **10.3969/j.issn.1008-214X.2009.04.003**.

31 J. Yu, Z. Hou, S. Sheta, J. Dong, W. Han, T. Lu and Z. Wang, Provenance classification of nephrite jades using multivariate LIBS: a comparative study, *Anal. Methods*, 2018, **1**(3), 281–289, DOI: **10.1039/c7ay02643a**.

32 M. S. Dhanoa, S. J. Lister, R. Sanderson and R. J. Barnes, The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra, *J. Near Infrared Spectrosc.*, 1994, **1**, 43–47, DOI: **10.1255/jnirs.30**.

33 R. J. Barnes, M. S. Dhanoa and S. J. Lister, Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra, *Appl. Spectrosc.*, 1989, **5**, 772–777, DOI: **10.1366/0003702894202201**.

34 M. Barker and W. Rayens, Partial least squares for discrimination, *J. Chemom.*, 2003, **17**(3), 166–173, DOI: **10.1002/cem.785**.

35 A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York, 2008, pp. 237–280, **https://link.springer.com/book/10.1007/978-0-387-78189-1**.

36 L. Breiman, Random forests, *Machine Learning*, 2001, vol. 45, 1, pp. 5–32, DOI: **10.1023/A:1010933404324**.

37 I. B. Gurevich, Selection of ensemble of features of recognition systems by the voting principle, *Sov. At. control*, 1974, **7**(5), 34–42.

38 R. Polikar, S. Krause and L. Burd, Ensemble of classifiers based incremental learning with dynamic voting weight update, *Proceedings of the International Joint Conference on Neural Networks*, 2003, vol. 4, pp. 2770–2775, DOI: **10.1109/IJCNN.2003.1224006**.

39 X. Zhang, R. Wu and L. Wang, Research on Petrologic Character of Nephrite Jade From Baikal Lake Region in

This journal is © The Royal Society of Chemistry 2024

J. Anal. At. Spectrom., 2024, **39**, 1560–1570 | **1569**

Russia, *J. Gems Gemmol.*, 2001, **3**(1), 12–17, DOI: **10.3969/j.issn.1008-214X.2001.01.003**.

40 X. Pei, Z. Qian and G. Shi, A mineralogical study of the Chuncheon nephrite, South Korea, *Acta Petrol. Mineral.*, 2011, **30**, 89–94, **https://www.yskw.ac.cn/yskwxzz/article/abstract/2011z116**.

41 T. Yang, M. Yang, H. Liu, Y. Wu and J. Li, New Understanding for Sanchahe Nephrite Deposit in East Kunlun, *J. Guilin Univ. Technol.*, 2013, **33**(22), 239–245, DOI: **10.3969/j.issn.1674-9057.2013.02.007**.

42 D. Ma, Characteristics and metallogenic law of the Sanchakou jade ore in Golmud,Qinghai province, Master thesis, China University of Geosciences, 2013, **https://kns.cnki.net/KCMS/detail/detail.aspx?filename=1016298823.nh&dbname=CMFDTEMP**.

43 G. David, P. Y. Meslin, E. Dehouck, O. Gasnault, A. Cousin, O. Forni, G. Berger, J. Lasue, P. Pinet, R. C. Wiens, S. Maurice, J. F. Fronton and W. Rapin, Laser-Induced Breakdown Spectroscopy (LIBS) characterization of granular soils: Implications for ChemCam analyses at Gale crater, Mars, *Icarus*, 2021, **365**, 114481, DOI: **10.1016/j.icarus.2021.114481**.

44 P. Zhang, X. Liu, J. Li and M. Chen, Comparative Analysis of Gemmological Characteristics of Brown-white Nephrites from Xinjiang, Qinghai and Russia, *J. Gems Gemmol.*, 2011, **13**(4), 31–38, **https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=cjfd2011&filename=bshb201104009&dbcode=cjfq**.

45 L. Han and H. Hong, Study on Mineral Components and Geological Background of Nephrites from Three Localities in China, *J. Gems Gemmol.*, 2009, **11**(3), 6–10, DOI: **10.15964/j.cnki.027jgg.2009.03.002**.

46 L. Lu, Z. Bian, F. Wang, J. Wei and X. Ran, Comparative Study on Mineral Components, Microstructures and Appearance Characteristics of Nephrite from Different Origins, *J. Gems Gemmol.*, 2014, **16**, 56–64, DOI: **10.15964/j.cnki.027jgg.2014.02.008**.

47 M. Abdel-Harith, A. Elhassan, Z. Abdel-Salam and M. F. Ali, Back-reflection-enhanced laser-induced breakdown spectroscopy (BRELIBS) on transparent materials: Application on archaeological glass, *Anal. Chim. Acta*, 2021, **1184**, 339024, DOI: **10.1016/j.aca.2021.339024**.

1570 | *J. Anal. At. Spectrom.*, 2024, **39**, 1560–1570

This journal is © The Royal Society of Chemistry 2024