

CLASSIFICATION OF GEM MATERIALS USING MACHINE LEARNING

Matthew F. Hardman, Artitaya Homkrajae, Sally Eaton-Magaña, Christopher M. Breeding, Aaron C. Palke, and Ziyin Sun

Gemstones traded in markets globally can have extremely high values due to their physical appearance and, in some cases, their scarcity. Gemological laboratories provide identification services to determine gemstone species, growth origin, provenance, and the history of color- or clarity-enhancing treatments, all characteristics that can significantly affect gemstone values. These determinations are primarily conducted through microscopic analysis of the gem material and by identifying characteristic features in spectroscopic data acquired using non-destructive techniques. In addition, mildly destructive methods such as laser ablation–inductively coupled plasma–mass spectrometry may be employed to determine the trace element composition of a gem material. Traditionally, the diagnostic criteria used to determine geographic provenance or color/clarity treatment history are selected by trained specialists based on careful evaluation of large research datasets. Identifying the specific features that are useful for classification purposes can be time-consuming, and subtle diagnostic features may be overlooked. Machine learning enables rapid exploration of large, complex datasets in new ways. Hence, the application of machine learning algorithms may complement existing microscopic, spectroscopic, and geochemical approaches.

This study examines the application of several machine learning models to gemstone classification problems involving natural alexandrite, laboratory-grown diamonds, and natural saltwater pearls. Using machine learning in a number of test cases, the authors achieved classification error rates as low as 5% or less for determination of provenance and detection of color treatment and were able to reduce the number of samples classified as indeterminate using conventional techniques by more than 50%.

The market value of diamonds, pearls, and colored stones can be influenced strongly by the species of the material, its physical features (e.g., cut, color, clarity, and size), and its rarity. However, value can also be driven by other factors, such as the geographic provenance of the material, growth origin (natural or laboratory-grown), or its history of color- or clarity-enhancing treatments. Gemological laboratories offer identification services that provide consumers with this type of information, as it can aid in defining the value of a gem. This process also protects consumers by reducing the number of erroneously or fraudulently traded gems, such as laboratory-grown diamonds that are marketed as natural.

At GIA, geographic provenance and treatment history are determined by comparing unclassified gems against large curated databases of samples from the same species, for which the provenance or treatment history are independently known. Microscopic examination is generally the first stage of analysis, identifying diagnostic physical features (e.g., mineral inclusions indicative of certain localities; Palke et al., 2019a,b). The gem material can also be characterized using a variety of nondestructive analytical techniques including photoluminescence (PL), Raman scattering, ultraviolet/visible/near-infrared (UV-Vis-NIR) absorption, and Fourier-transform infrared (FTIR) absorption spectroscopy. GIA's traditional approaches to comparison of these data include visual inspection of spectra, identifying peaks and bands as well as their relative intensities, and in some cases using automated software for spectral matching (e.g., the WiRE package for Renishaw Raman devices). For diamonds, these features reflect atomic defects in the

See end of article for About the Authors and Acknowledgments.

GEMS & GEMOLOGY, Vol. 60, No. 3, pp. 306–329,

<http://dx.doi.org/10.5741/GEMS.60.3.306>

© 2024 Gemological Institute of America

lattice, which are created or destroyed during growth and color- or clarity-enhancing treatments (Martineau et al., 2004; D’Haenens-Johansson et al., 2022).

For some colored stones and pearls, minimally destructive methods such as laser ablation–inductively coupled plasma–mass spectrometry (LA-ICP-MS) can determine the elemental composition. This technique removes a small amount of material from the stone, which may leave a tiny crater approximately 55 μm in diameter. These chemical data are compared with gem databases using bivariate scatterplots or “selective plotting,” as research has shown that trace element compositions can indicate geographic origin (e.g., Groat et al., 2019; Palke et al., 2019a,b). Selective plotting is an advanced classification technique that can reduce the compositional overlap between gemstones in bivariate diagrams by filtering the reference data to only those samples with similar elemental concentrations as the unclassified stone, considering multiple trace elements simultaneously. For example, unclassified and reference stones might be plotted on a bivariate diagram using magnesium and iron concentrations; the reference stones shown on the plot would be filtered to those with similar compositions of other elements as well, such as titanium or gallium (see Palke et al., 2019a for a detailed overview). This approach reduces the compositional overlap between gemstones of different provenance when applied to alexandrite, ruby, and sapphire (Sun et al., 2019; Palke et al., 2019a,b).

In Brief

- Machine learning methods including random forest, artificial neural networks, and support vector machines can enhance classification of gem materials.
- Machine learning methods are effective tools for determining the provenance of alexandrite and saltwater pearls.
- Subtle indicators of post-growth treatment in some laboratory-grown diamonds can be identified using machine learning.

To determine geographic provenance or treatment history using these approaches, each group of samples must have sufficiently distinct characteristics to allow separation. In practice, however, a single characteristic is generally not diagnostic enough for determination, and multiple forms of evidence are needed. This be-

comes much more difficult when samples lack a particular set of characteristic features (e.g., a lack of inclusions). When gemological specialists cannot observe clear diagnostic features, even after careful examination of all data, a classification of “undetermined” may be given.

To address this challenge, we investigate machine learning (ML) as a complementary tool to existing classification approaches. The criteria used for gemstone classification at GIA—physical features, spectroscopic characteristics, and elemental compositions—are selected based on years of detailed examination, resulting in recognizing patterns related to gemstone characteristics and their provenance or treatment history. Directing computer algorithms to identify patterns among large datasets, in a manner similar to that of a gem specialist, may provide new insights.

BACKGROUND

Gemological laboratories have acquired spectroscopic and trace element data for millions of diamonds, colored gemstones, and pearls collectively. By carefully inspecting these databases, trained specialists can identify characteristic features that can be used to determine geographic provenance or treatment history for new samples. However, the massive scale of available data, coupled with the complexity of different gem materials and the constantly evolving techniques for growing and treating them (e.g., Eaton-Magaña et al., 2024), make it time-consuming to identify the most useful classification criteria through manual inspection. As a result, subtle diagnostic features may be overlooked. Computers can rapidly parse large datasets, and some software already provides functions that enable users to match features in acquired data (such as Raman spectra) to extensive reference databases. Hence, computers are powerful tools that complement existing traditional approaches.

Artificial intelligence (AI) is the use of computer systems to emulate the way humans think, process information, and learn. ML is a subfield of AI associated with developing algorithms to learn from datasets in order to identify patterns or solve particular problems without requiring explicit programming (Xu et al., 2021). The application of ML in the physical sciences has increased in recent years, with successful uses in fields including the evaluation of the mantle and exploration for diamonds and kimberlites (Dawson and Stephens, 1975; Griffin et al., 2002; Hardman et al., 2018a,b), as well as mineral



Figure 1. Left: Faceted alexandrite with color-change effect, with a rough crystal for comparison. The green color of the faceted stone is visible when exposed to daylight-equivalent illumination, while the red color is visible when exposed to incandescent illumination. Photos by Robert Weldon; rough crystal courtesy of William F. Larson and faceted stone courtesy of GIA's Dr. Edward J. Gübelin collection, no. 34767. Right: World map indicating seven major alexandrite sampling localities. Other sampling localities are not discussed in this study (e.g., Sun et al., 2019).

chemistry and characterization (Schönig et al., 2021; Hazen and Morrison, 2022). The application of ML to gemological problems—including provenance determination—has also increased over time (e.g., Blodgett and Shen, 2011; Luo et al., 2015; Homkrajae et al., 2019; Krebs et al., 2020; Hardman et al., 2022; Bassoo et al., 2023; Bendinelli et al., 2024).

Applying ML to gemological research can yield significant benefits. For example, while the selective plotting approach surpasses bivariate scatterplots for classifying some colored stones, some samples still cannot be classified. ML allows the comparison of even more variables simultaneously and can reveal subtle classification criteria involving multiple variables with complex relationships that may not be apparent using bivariate scatterplots or selective plotting. Using these criteria, ML models can be automated to assist gemologists by suggesting a classification. This opinion can be quite valuable in making a confident judgment. However, some samples will still exhibit too much compositional overlap to be classifiable even with ML.

This study provides an overview of ML approaches to gem classification, including some that are in active use at GIA. We examine the applications of ML to three distinct gem materials for which GIA offers lab-

oratory report services (alexandrite, laboratory-grown diamonds, and natural saltwater pearls) and discuss the advantages of ML over traditional approaches.

MATERIALS

Alexandrite. Alexandrite, a variety of the mineral chrysoberyl (BeAl_2O_4), has a unique color-change property when exposed to different light sources (for example, from green or green-blue in daylight to red in incandescent lighting; figure 1, left). This property is caused by trace concentrations of Cr^{3+} replacing Al^{3+} in the crystal lattice (Gübelin and Schmetzer, 1982). GIA provides services for determining the geographic origin of alexandrite (figure 1, right).

Alexandrite provenance determination is conducted at GIA using microscopic analysis to identify physical features such as mineral inclusions that may be characteristic of different sampling localities (Sun et al., 2019). Trace element compositions offer additional information. While alexandrite from different localities can have very similar compositions, possibly due to similar geological processes, the compositions can also vary significantly between localities (Sun et al., 2019). Data acquired for unknown stones are compared to curated databases of alexan-

TABLE 1. Summary of gem materials classified in this study.^a

Gem	Locality	Samples	Analyses
Alexandrite ^b	Brazil	174	528
	India	27	81
	Madagascar	38	132
	Russia	42	153
	Sri Lanka	33	170
	Tanzania	27	80
	Zimbabwe	5	13
CVD-grown diamond ^c	Treatment	Samples/spectra	
	As-grown/untreated	300	
	Post-growth treated	1,795	
Natural saltwater pearl ^d	Locality	Samples	Analyses
	Bahrain	268	843
	Kuwait	188	588
	Oman	148	444

^aBetween 3 and 5 spots were measured on each pearl, and between 2 and 10 spots on each alexandrite. A single spectrum was acquired for each CVD-grown diamond.

^bSun et al. (2019)

^cHardman et al. (2022)

^dThis study

drite with known provenance, using bivariate scatterplots and selective plotting. As it is uncommon for one or even two chemical variables to separate a group of samples from all others, selective plotting offers significant advantages over bivariate techniques by enabling comparison of more variables simultaneously. Despite these techniques, the provenance of some stones remains unclear.

To test the capability of ML to improve alexandrite provenance determination, the authors compiled the concentrations of the trace elements boron, magnesium, vanadium, chromium, iron, gallium, germanium, and tin for samples from mines located in seven countries. These alexandrites, detailed by Sun et al. (2019), represent a portion of the samples in GIA's colored stone reference collection, consisting of stones with known sampling locality that were loaned to GIA by individual donors. The number of stones from each sampling locality is given in table

1. Trace element compositions were acquired by LA-ICP-MS, originally reported by Sun et al. (2019) and summarized in table 2. We have compiled data for 10 additional samples with chemical compositions yielding inconclusive provenance determinations using traditional bivariate scatterplot and selective plotting approaches. We calibrated new ML models using samples with known provenance and applied these models to classify the population of "undetermined" alexandrites.

CVD-Grown Diamonds. Diamonds can be grown using chemical vapor deposition (CVD) by placing a diamond substrate within a vacuum chamber at moderate temperatures (approximately 700° to 1300°C) and subatmospheric pressures (generally 20–500 mbar; figure 2); see Arnault et al. (2022) and D'Haenens-Johansson et al. (2022) for reviews on CVD growth. CVD-grown diamonds can be treated

TABLE 2. Trace element concentrations (in ppmw) of alexandrite from global localities, determined by LA-ICP-MS.^a

Element	Brazil (174 samples, 528 analyses)		India (27 samples, 81 analyses)		Madagascar (38 samples, 132 analyses)	
	Range	Average $\pm 2\sigma$	Range	Average $\pm 2\sigma$	Range	Average $\pm 2\sigma$
B	bdl–42.7	2.5 \pm 12.5	bdl–68.9	10.9 \pm 29.6	0.78–225	54.9 \pm 84.4
Mg	2.5–46.7	11.9 \pm 10.8	38.4–203	104.3 \pm 55.9	4.92–337	64.5 \pm 125.1
V	56.8–201	136.1 \pm 52.7	314–1210	722.0 \pm 365.6	13.6–468	132.6 \pm 187.2
Cr	424–6050	2955 \pm 1993	192–2120	1090 \pm 907	43.4–3430	875 \pm 1206
Fe	4180–11500	6352 \pm 2048	829–4610	2898 \pm 2108	1060–12100	5771 \pm 4221
Ga	53.3–145	72.8 \pm 21.1	168–284	199.4 \pm 41.8	50.9–1070	383.8 \pm 385.4
Ge	bdl–2.55	0.71 \pm 0.77	bdl–1.68	0.32 \pm 0.45	bdl–2.55	0.38 \pm 0.94
Sn	87.4–4470	870.6 \pm 824.2	0.28–6.01	1.7 \pm 2.4	0.82–327	43.8 \pm 133.4

^aBetween 2 and 10 spots were measured on each alexandrite.

after growth under high-pressure, high-temperature (HPHT) or low-pressure, high-temperature (LPHT) conditions to improve their color grade. Several of GIA’s grading reports for laboratory-grown dia-

monds document any evidence of post-growth treatment.

Lattice defects form during CVD growth and may manifest as peaks or bands in PL spectra (e.g., figure 3;

Figure 2. Left: A 0.21 ct HPHT-treated CVD-grown diamond. Photo by Robison McMurtry. Right: Schematic diagram of a CVD reactor. From D’Haenens-Johansson et al. (2022).

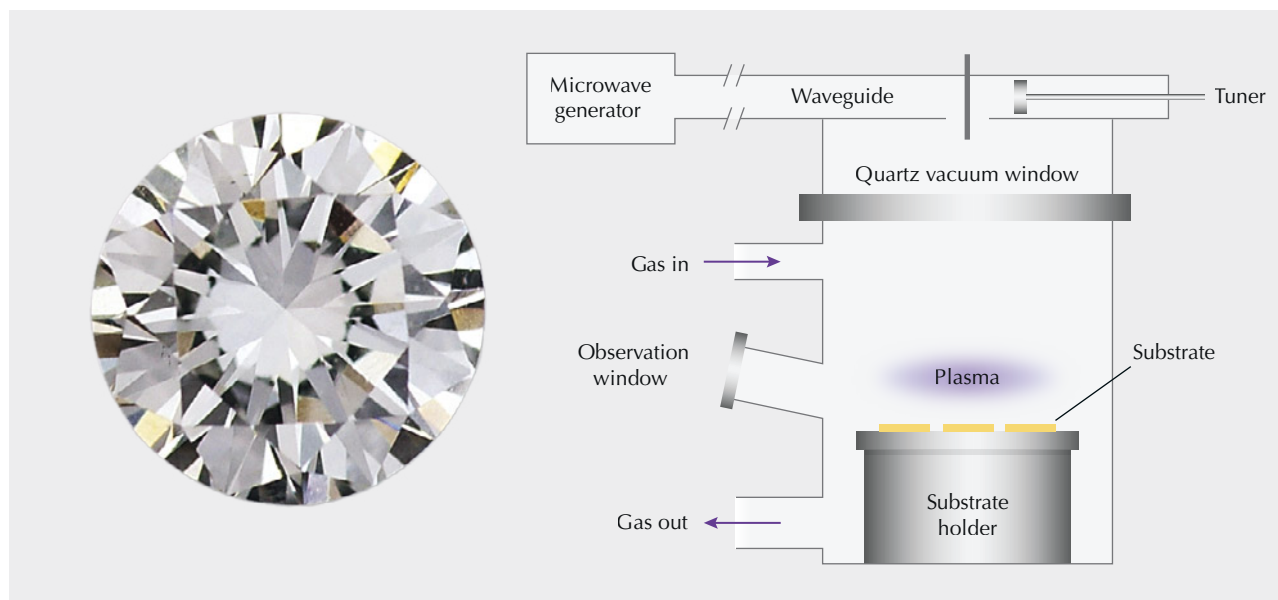


TABLE 2 (continued). Trace element concentrations (in ppmw) of alexandrite from global localities, determined by LA-ICP-MS.^a

Element	Russia (42 samples, 153 analyses)		Sri Lanka (33 samples, 170 analyses)		Tanzania (27 samples, 80 analyses)		Zimbabwe (5 samples, 13 analyses)		Detection limit (ppmw)
	Range	Average ± 2σ	Range	Average ± 2σ	Range	Average ± 2σ	Range	Average ± 2σ	
B	15.5–208	45.0 ± 83.6	23.1–400	81.6 ± 114.2	4.21–87.8	28.3 ± 33.9	bdl–12.4	7.95 ± 8.14	0.12
Mg	4.47–3140	93.0 ± 718.7	2.95–987	169.7 ± 390.3	24–270	53.3 ± 67.6	10.9–38.3	20.5 ± 16.5	0.024
V	21.5–347	101.7 ± 132.1	29.8–1190	210.9 ± 373.0	33.2–241	76.1 ± 75.1	19.1–50.6	31.2 ± 21.4	0.015
Cr	156–14200	3092 ± 6352	24.1–3200	702 ± 1202	232–8270	1991 ± 2763	2550–15600	8704 ± 8807	0.21
Fe	1390–8610	2784 ± 2359	1720–12200	7676 ± 4801	2470–9210	5623 ± 2732	7480–25000	14064 ± 10164	1.3
Ga	171–406	262.9 ± 107.4	246–1580	731.4 ± 720.5	54.9–212	94.1 ± 59.4	269–343	312.5 ± 48.2	0.019
Ge	1.08–76.4	10.3 ± 32.1	0.1–2.7	0.59 ± 1.08	bdl–4.48	0.87 ± 1.49	bdl–3.19	1.81 ± 2.14	0.074
Sn	37.9–1550	324.1 ± 523.5	7.35–785	57.9 ± 217.0	3.56–192	22.9 ± 66.4	338–2110	1204 ± 1144	0.046

^aBetween 2 and 10 spots were measured on each alexandrite.

Martineau et al., 2004). The post-growth treatment process may create, destroy, or alter the concentration or distribution of lattice defects (figure 3; Martineau et al., 2004; D’Haenens-Johansson et al., 2022). Therefore, treatment can be detected by inter-

preting PL spectroscopic data. However, the full suite of potential treatments applied to CVD-grown diamonds is not always disclosed by manufacturers, and these processes can change over time. GIA seeks to identify all forms of post-growth treatment. This

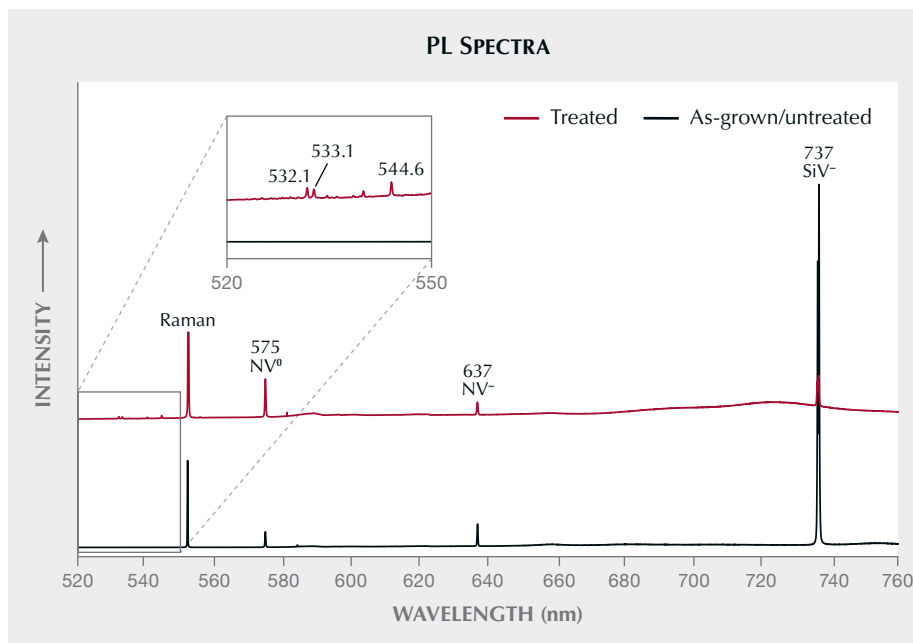


Figure 3. Example PL spectra for different as-grown (untreated) and post-growth treated CVD diamonds acquired at liquid nitrogen temperature (approximately 77 K) using 514 nm laser excitation. The spectra are offset vertically for clarity and scaled so that the Raman peaks (at 552.4 nm) have equal intensity. Inset: An enlarged view highlights the occurrence of multiple weak features in some diamond spectra in the 520–550 nm range. NV = nitrogen-vacancy and SiV = silicon-vacancy defects, with neutral (⁰) or negative (⁻) charge states.



Figure 4. Left: Representative natural saltwater nacreous (light-colored) and non-nacreous (dark-colored) pearls produced by *Pinctada radiata* mollusks from Kuwait in the Persian (Arabian) Gulf. Photo by Robert Weldon; courtesy of Ahmad Abdullah Alomaish Alajmi. Right: Map showing the location of Bahrain, Kuwait, and Oman.

study assesses the capability of ML in establishing new criteria for the identification of post-growth treatment in CVD-grown diamonds. PL spectra were compiled for 300 as-grown (untreated) and 1,795 HPHT-treated CVD-grown diamonds (D–Z color range), measured during routine analysis of diamonds submitted by clients to GIA. Each determination of treated or untreated was made by trained specialists based on inspection and comparison of PL spectra and surface fluorescence patterns. These spectra were acquired with 514 nm laser excitation at liquid nitrogen temperature (approximately ~77 K) using a Renishaw inVia Raman microscope at 5× magnification. They were previously compiled by Hardman et al. (2022).

Natural Saltwater Pearls. Pearls differ from many other gem materials in that they are formed organically from a mollusk. Shells and pearls are composed of calcium carbonate (CaCO_3) together with organic substances called *conchiolin*, plus a small volume of water. Pearls have a wide variety of trace element compositions that can be inherited from different growth-related processes and the environment in which they formed (Fuge et al., 1993). They can also have nacreous surfaces (featuring a layered structure of aragonite platelets in “brick-and-mortar” formation and commonly with pearly luster) or non-nacreous surfaces (lacking pearly luster; figure 4). Calcite and aragonite are polymorphs, minerals having the same chemical composition (i.e., CaCO_3) but different structure. Pearls are produced by a variety of different mollusk species and can form in a variety of freshwater and saltwater environments globally. They can form naturally or be cultured, grown through human intervention. This complex set of features increases

the difficulty of evaluating the relationships between pearl compositions and growth conditions.

GIA’s pearl identification services include a determination of natural or cultured identity, mollusk species, saltwater or freshwater formation environment, and the presence of any treatment. ML has been previously applied to determine the provenance of freshwater pearls using trace element compositions (Homkrajae et al., 2019), while pearls formed in saltwater environments have not been tested as thoroughly. To evaluate whether the provenance of saltwater pearls can be determined, the authors compiled trace element compositions measured by LA-ICP-MS for a set of 604 new natural saltwater pearls sampled from Oman, Bahrain, and Kuwait (table 3). Pearls from all three localities were sourced from local pearl divers and classified as C1 samples according to the GIA pearl classification codes (see table 1 in Homkrajae et al., 2021). The Bahrain and Kuwait pearls reportedly originated from *Pinctada radiata* mollusks from the Persian (Arabian) Gulf, while the Oman pearls were derived from *Pinctada margaritifera* mollusks in the Gulf of Oman region. Trace element compositions were determined using a Thermo Fisher Scientific iCAP Qc ICP-MS coupled with a New Wave Research UP-213 laser ablation unit using the same analytical parameters as Homkrajae et al. (2019). We analyzed between three and five spots on each sample and report the compositions of six elements (sodium, magnesium, potassium, manganese, strontium, and barium) in table 3.

METHODS

Data Processing. For this study and for statistical analysis in general, datasets need to be reprocessed from their original form into a format that is better

TABLE 3. Trace element concentrations (in ppmw) of natural saltwater pearls from Bahrain, Kuwait, and Oman, determined by LA-ICP-MS.^a

Element	Bahrain (268 samples, 843 analyses)		Kuwait (188 samples, 588 analyses)		Oman (148 samples, 444 analyses)		Detection limit (ppmw)
	Range	Average $\pm 2\sigma$	Range	Average $\pm 2\sigma$	Range	Average $\pm 2\sigma$	
Na	2710–10000	5887 \pm 1700	2880–8050	6316 \pm 1779	3350–9910	5154 \pm 1162	3.9
Mg	27.8–12300	839.0 \pm 4458.3	26.2–13300	533.5 \pm 3564.5	43–7850	280 \pm 1111	0.25
K	bdl–289	48.19 \pm 64.62	bdl–501	51 \pm 83	bdl–153.0	24.4 \pm 21.5	1.3
Mn	bdl–42.8	1.35 \pm 1.72	0.05–12.30	1.13 \pm 3.40	bdl–17.7	3.0 \pm 4.3	0.05
Sr	604–2330	1012 \pm 423	610–3020	998 \pm 389	746–2040	1150 \pm 399	0.1
Ba	v0.12–4.83	0.66 \pm 1.18	0.11–8.81	0.78 \pm 1.16	bdl–1.19	0.37 \pm 0.34	0.005

^aBetween three and five spots were measured on each pearl.

suiting to a particular statistical method. Trace element concentrations reported in units of parts per million by weight (ppmw) can be used in comparisons, including in bivariate scatterplots. For some statistical models, however, trace elements with very low concentrations—near or below the instrument’s detection limit—must be identified. We replace these values with the respective detection limit for each element so they can be processed using statistical methods. This reduces uncertainties for values near the analytical detection limit, but it is worth noting that the data

distribution will be shifted slightly higher than if all such values were replaced with zero. During trace element analysis of pearls and colored gemstones, GIA typically tests three spots on each sample to ensure data quality and to assess the diversity of compositions that some samples may have: Some samples may have significant compositional heterogeneity that cannot be characterized with a single analytical spot.

Spectroscopic data is often presented graphically, with measured intensity (on the y-axis) for a variety of wavelengths (on the x-axis; e.g., figure 5). The

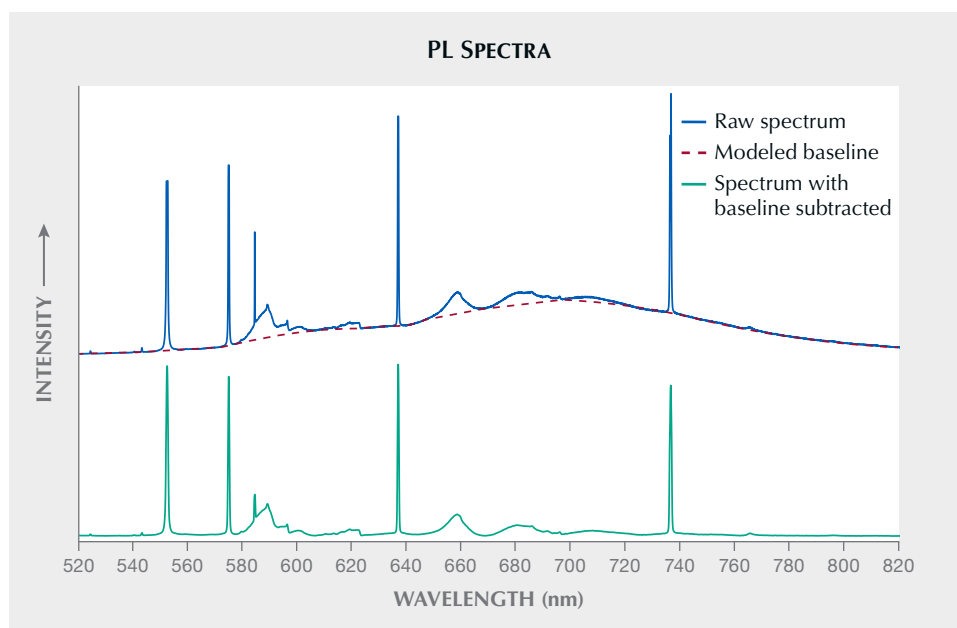


Figure 5. Example PL spectra measured using 514 nm laser excitation, before and after subtraction of a modeled baseline.

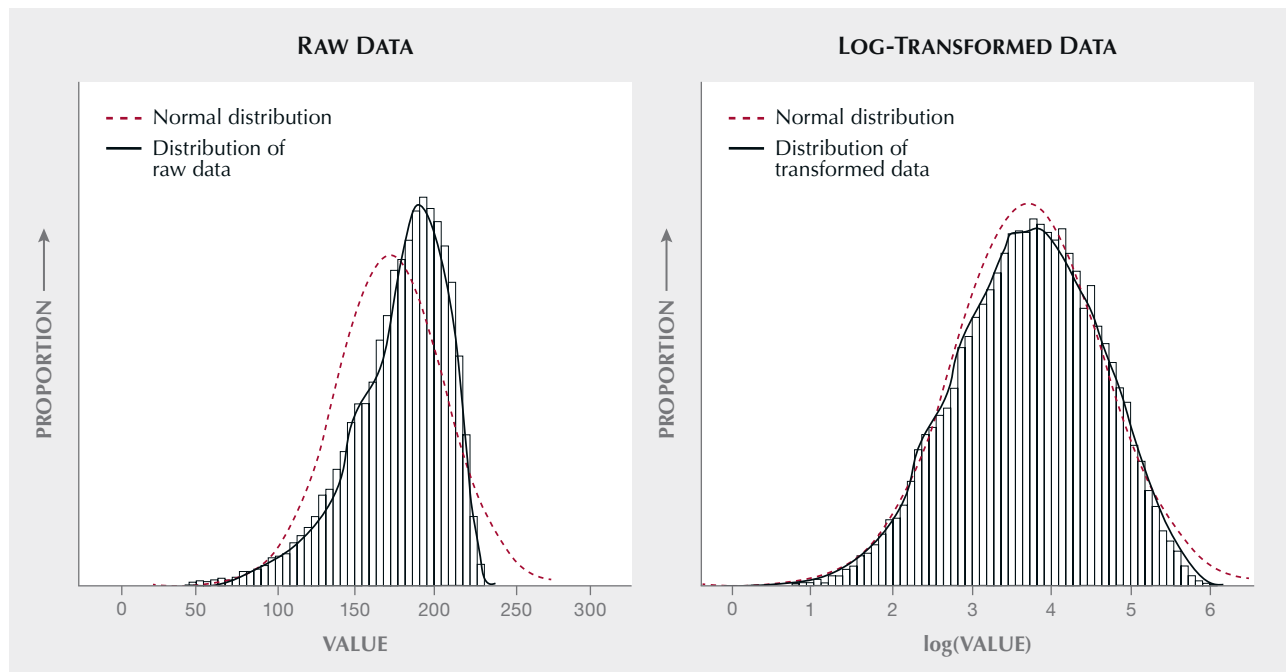
spectra may display bands and/or peaks corresponding to particular defects or features in the material being analyzed. Some spectra may have significant “backgrounds,” appearing as broad bands in certain regions of the spectrum (figure 5). To determine the areas of some peaks, this background must first be removed. To reduce the background effects in this study, a modeled baseline was subtracted from each spectrum (figure 5). Background effects can also introduce noise, or generally weak spectroscopic features unrelated to the sample itself. Under some circumstances, noise can be strong enough to be confused with an actual peak. If noise is inadvertently used to calibrate a statistical model, the model outputs may be unpredictable or meaningless, so it is very important to identify these features. We have screened all spectra in this study to identify actual peaks, but note that distinguishing a weak peak from noise may be difficult or impossible in some cases.

For CVD-grown diamonds, peak heights above background levels can be calculated and compared with other spectra, including those from CVD-grown diamonds previously classified as treated or untreated. However, PL spectra are semiquantitative, and peak heights can vary due to factors such as

changes in laser power during analysis. To compare peak heights across different spectra, the peak height can be normalized by dividing it by the height of the diamond Raman line (with a peak position of 552.4 nm when measured with a 514 nm laser; figure 3). This line corresponds to diamond’s intrinsic Raman peak at 1332 cm^{-1} .

Data can also be transformed into a different structure to better fit the requirements of particular ML models. In reality, data populations can have nonsymmetrical distributions (figure 6, left), skewing toward high or low values rather than a clear central value (as is expected for normal or Gaussian distributions). For example, iron concentrations in an alexandrite database may have a nonnormal distribution similar to that in figure 6, left. However, some statistical models require that calibration data have approximately symmetrical distributions. Log transformation, which replaces all values in a dataset with their logarithms, can shift a population of nonnormal data closer to normality while preserving the relative distribution of their values (figure 6, right). In addition, variables in some datasets may have very different scales. For example, some elements may have concentrations at the parts per thousand level and others at the parts per million

Figure 6. Left: An arbitrary dataset with a nonsymmetrical distribution that has a negative skew (tail to the left) compared to an ideal normal (Gaussian) distribution. Right: The same set of values following log transformation.



level. These data can be rescaled to improve the comparability of all data by preserving their relative order but adjusting their distributions to cover approximately the same range.

To perform statistical analysis in this study, we imported trace element and spectroscopic data files into the R statistical freeware package (v4.3.2; R Core Team, 2024). Using R, all trace element and spectroscopic data were rescaled and log transformed. Example scripts can be accessed in the appendix available online (<https://www.gia.edu/gems-gemology/fall-2024-machine-learning>).

Model Calibration and Validation. At GIA, traditional approaches to colored stone classification include the use of bivariate scatterplots and selective plotting to compare trace element compositions. For stones with compositions that overlap with multiple populations using these methods, statistical models can add more classification variables to enhance discrimination. For trace elements, this may mean considering larger numbers of elements, as compositional overlap in analyzing the concentrations of two elements may be reduced by using three or more elements.

An initial step is dataset *simplification*, particularly if the dataset contains many samples, each with a large number of variables. Principal component analysis (PCA) and linear discriminant analysis (LDA) are techniques that can simplify a dataset by projecting the relationships between many variables onto fewer dimensions for easier interpretation (box A). The dataset can then be visualized using bivariate scatterplots with the transformed variables as axes (see box A), or PCA or LDA can be used to preprocess a dataset for subsequent use in more complex ML models.

The choice of ML model depends on the goals of the user (see box B). Some models are extremely interpretable and intuitive, such as decision trees where the logic of decision-making is provided directly to the user for easy application. Other models are more “black box” in nature, involving complex computational, mathematical, or statistical logic hidden from the user. In these cases, the user inputs data and receives an outcome from an opaque computer model. In general, the more complex models are much more difficult to interpret but also more powerful and capable of producing better results. ML models discussed in this study include random forest (RF), support vector machines (SVM), and artificial neural networks (ANN). Box B provides an overview

of all three. Previous research has documented the advantages and disadvantages of these models when applied to different problems, with RF in particular performing extremely well in many different fields (e.g., Fernández-Delgado et al., 2014; Bolton et al., 2019).

Feature selection is an application of ML that objectively analyzes a dataset to identify variables that have a strong influence on classification. The “Boruta” algorithm, for instance, ranks all statistically significant variables in a dataset by testing whether their exclusion leads to a significant decrease in model performance (box B; Kurasa et al., 2010). This list can then be inspected more closely by a user. For datasets with hundreds or thousands of variables, feature selection saves significant time and effort.

Validation is extremely important. ML models are often reported with *error rates*, values that indicate their success in making correct classifications. Ideally, models with low error rates will produce more accurate results when testing new samples. One approach to accurately evaluate model quality is *k*-fold cross-validation, in which a dataset is partitioned into a random subset for model calibration (e.g., 80%) and the remainder is allocated to testing the model (e.g., 20%). This process is repeated *k* times, with *k* chosen by the user. The error rates of all folds are averaged to provide a measure of model quality. By using a larger *k* value, the average error rate of all folds will generally be less biased toward the results of individual folds (Hastie et al., 2009). ML models can be recalibrated and retested to improve model quality and reduce error rates.

An additional measure of quality is classification *confidence*, which indicates the probability that an individual result produced by the ML model is accurate. Some ML models will output a classification probability with every decision, indicating the model’s confidence in its own decision. When testing an unclassified sample using an ML model, the user can accept or reject the classification based on this probability value. For classification problems with two possible outcomes, a probability value of approximately 50% indicates low certainty for either outcome, possibly due to the unclassified sample having similar characteristics to both groups. The threshold probability that users will set during decision-making is subjective and may be chosen by comparing with the probabilities assigned to stones with known origin. It is important to consider confidence values before accepting ML results.

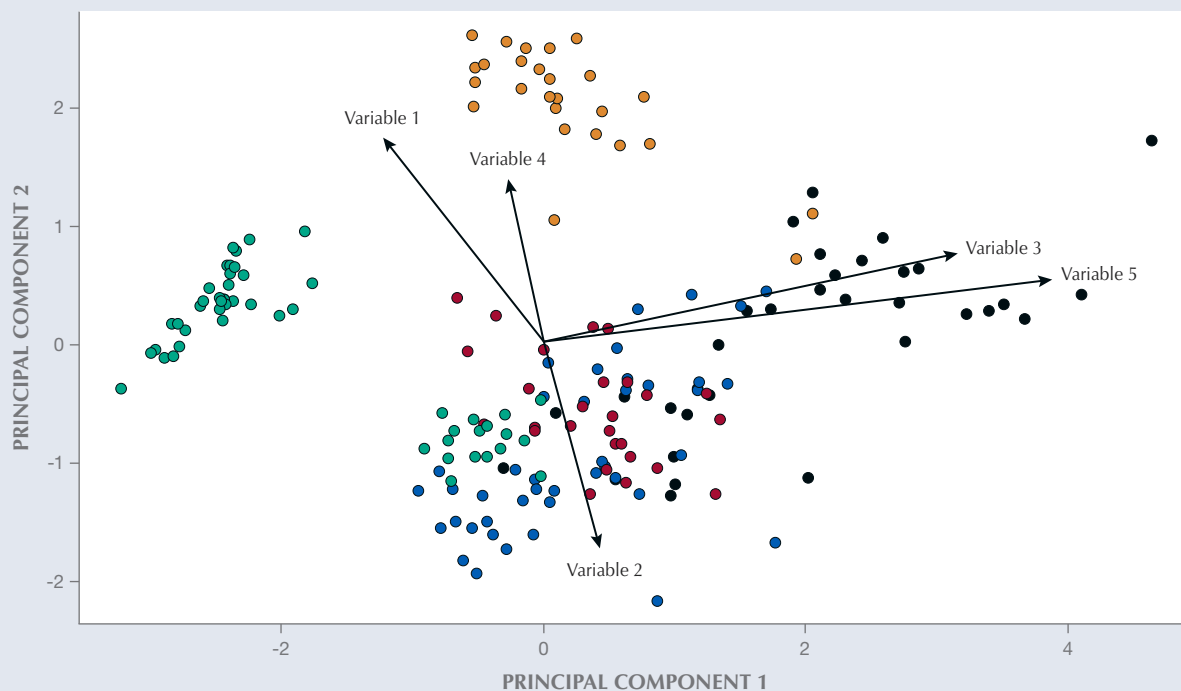
BOX A: DIMENSIONALITY REDUCTION

Analyzing large datasets with many different variables per sample can make it difficult or impossible to uncover complex relationships or subgroups of samples hidden among the data. Principal component analysis and linear discriminant analysis are dimensionality reduction techniques that can be used to simplify large datasets, identifying key relationships and condensing meaningful information into fewer variables.

PCA reduces the number of dimensions in a dataset by transforming the data, producing a new list of variables called principal components. These components retain as much information as possible from the original

data but are much simpler to visualize. Principal components are linear combinations of the starting variables, structured in such a way that the first principal component represents the majority of the variance of the data, with the second principal component representing additional variance, but less than the first. Additional components represent progressively less variance. Therefore, most of the variation in a dataset can be explored or visualized using fewer variables. PCA results shown in a bivariate scatterplot can be a valuable tool (e.g., figure A-1). The relative distribution of data highlights differences and similarities between samples. Vectors superimposed

Figure A-1. Example plot of principal components 1 and 2 in two-dimensional space. Vectors corresponding to five variables used to generate the PCA solution are indicated. Variables 3 and 5 may be positively correlated (as they have similar length and direction), and both strongly influence principal component 1. Variables 2 and 4 may be negatively correlated, given their opposing direction; both variables strongly influence principal component 2 (positively and negatively, respectively). Each principal component is a linear combination of the original variables. The numerical values for each principal component do not have any physical meaning but convey the level of variation between data points.



GIA's research department uses R, Python, and MATLAB software to construct scripts and models that are used to classify gem materials. These software tools are computationally complex, with commands written using scripts, and can be difficult for general users to navigate. Therefore, models cali-

brated using these tools are exported and integrated into existing software already being used for gemological classification. This reduces the complexity of the ML logic and provides laboratory gemologists with the rapid and concise classification outputs they need.

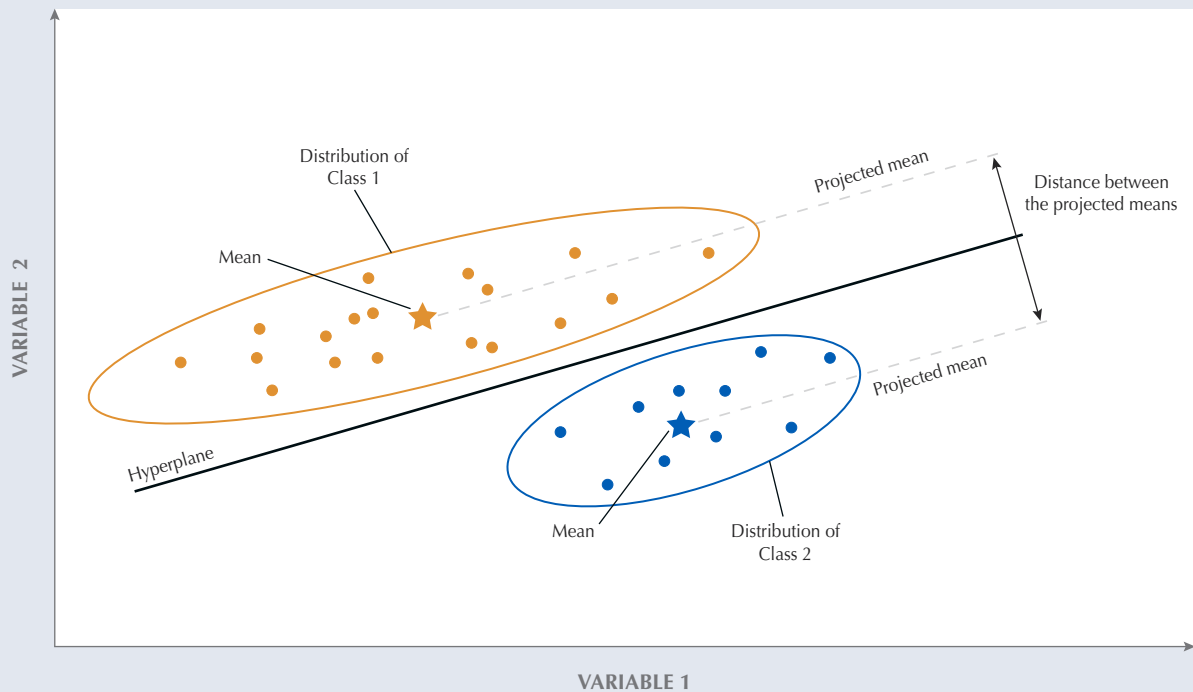


Figure A-2. Example linear discriminant analysis plot. Populations of data can be separated using a hyperplane, a linear equation constructed from variables perpendicular to the maximum separation between the projected means of the two groups. New samples can be classified using the hyperplane equation.

on the data graphically show how each variable or feature influences each principal component. Vector length indicates the intensity of the influence, and vector direction shows whether it is oriented positively or negatively toward one or both of principal components 1 and 2. Broadly speaking, correlated vectors may indicate features that are correlated to one another.

LDA reduces the dimensionality of a dataset by projecting it into a lower-dimensional space, maximizing the separation between labeled groups. It produces a series of linear equations from the initial starting variables (Zhao et al., 2020). Unlike PCA, LDA explicitly determines the best combination of features for classifying data into different groups. These linear boundaries can then be used to clas-

sify new samples. LDA results can also be shown in bivariate scatterplots (e.g., figure A-2).

Like all statistical methods, PCA and LDA have particular stipulations. For example, LDA assumes a normal distribution of variables (Hastie et al., 2009). Both are linear techniques and thus appropriate when the data are linearly separable. The outputs from PCA and LDA can be used as inputs to calibrate machine learning models, as the restructuring of the database using PCA and LDA may allow the data to be more effectively processed by machine learning. In this case, the user may select a small number of principal components that collectively describe a large amount of the variance of the original data.

This study examines GIA's application of PCA, LDA, RF, ANN, SVM, and Boruta to gem-related classification problems. The authors used R to construct all of the machine learning models and to validate each one using *k*-fold cross-validation. For alexandrite and natural saltwater pearls, we applied

PCA, LDA, RF, ANN, and SVM to build new models for provenance determination. For CVD-grown diamonds, the authors applied the Boruta algorithm to identify their spectroscopic features and identify those that could enhance the detection of post-growth treatment.

BOX B: MACHINE LEARNING MODELS

Machine learning is the process of training computer algorithms to solve complex problems or identify patterns without explicit programming. ML models are characterized by complex mathematics, and each is suited to particular types of datasets, with specific advantages and disadvantages. Several ML models are summarized here. For each model, the algorithm learns to identify the best combination of features to separate samples into groups by predefined criteria (e.g., species and provenance). These criteria can then be applied to classify new samples.

Decision trees are intuitive ML models consisting of successive binary true or false statements drawn as trees (figure B-1A; Hastie et al., 2009). New samples can be tested and classified rapidly using the logic of the decision tree. However, the method of generating a decision tree can lead to “overfitting,” in which samples are split into groups based on random features in the data, leading to poor results when tested against new samples. The random forest method builds many decision trees (potentially thousands) from the initial dataset; the outcomes from all trees are aggregated, and the majority vote determines the classification (figure B-1B; Breiman, 2001). As an ML model, RF often outperforms the others, being less sensitive to analytical noise than models such as single decision trees (Breiman, 2001; Fernández-Delgado et al., 2014; Bolton et al., 2019).

For datasets where a linear boundary can separate populations of data, linear discriminant analysis can be applied (see box A). LDA produces multiple linear boundaries that separate different groups within a dataset. In cases where data cannot be separated clearly by a linear boundary, other models such as support vector machines can be applied (Boser et al., 1992; Cristianini and Shawe-Taylor, 2000). SVM uses a mathematical function (known as a “kernel function”) to transform the original data and remap it to a new coordinate space where the data can be separated using a linear equation (figure B-1B). Artificial neural networks are another option, organized in a fashion similar to the human brain. Information about variables in the dataset passes through successive layers of “artificial neurons” at each layer, and the neurons transform the dataset and pass it through to other layers (LeCun et al., 2015). After repeated transformation and classification stages, the original data receives a classification outcome (figure B-1C).

Some ML models cannot extrapolate to new data that differ significantly from the data used during the initial

calibration. In these cases, new types of samples tested against ML models—such as RF—may produce unpredictable outcomes (Hengl et al., 2018). But if the new data have a linear relationship with other available samples, linear models such as LDA may be capable of accurately predicting the outcome class. The choice of ML model is critical, and all ML outcomes should be considered carefully. Model performance can be improved by iteratively testing combinations of parameters until finding the combination that produces the model with the best performance.

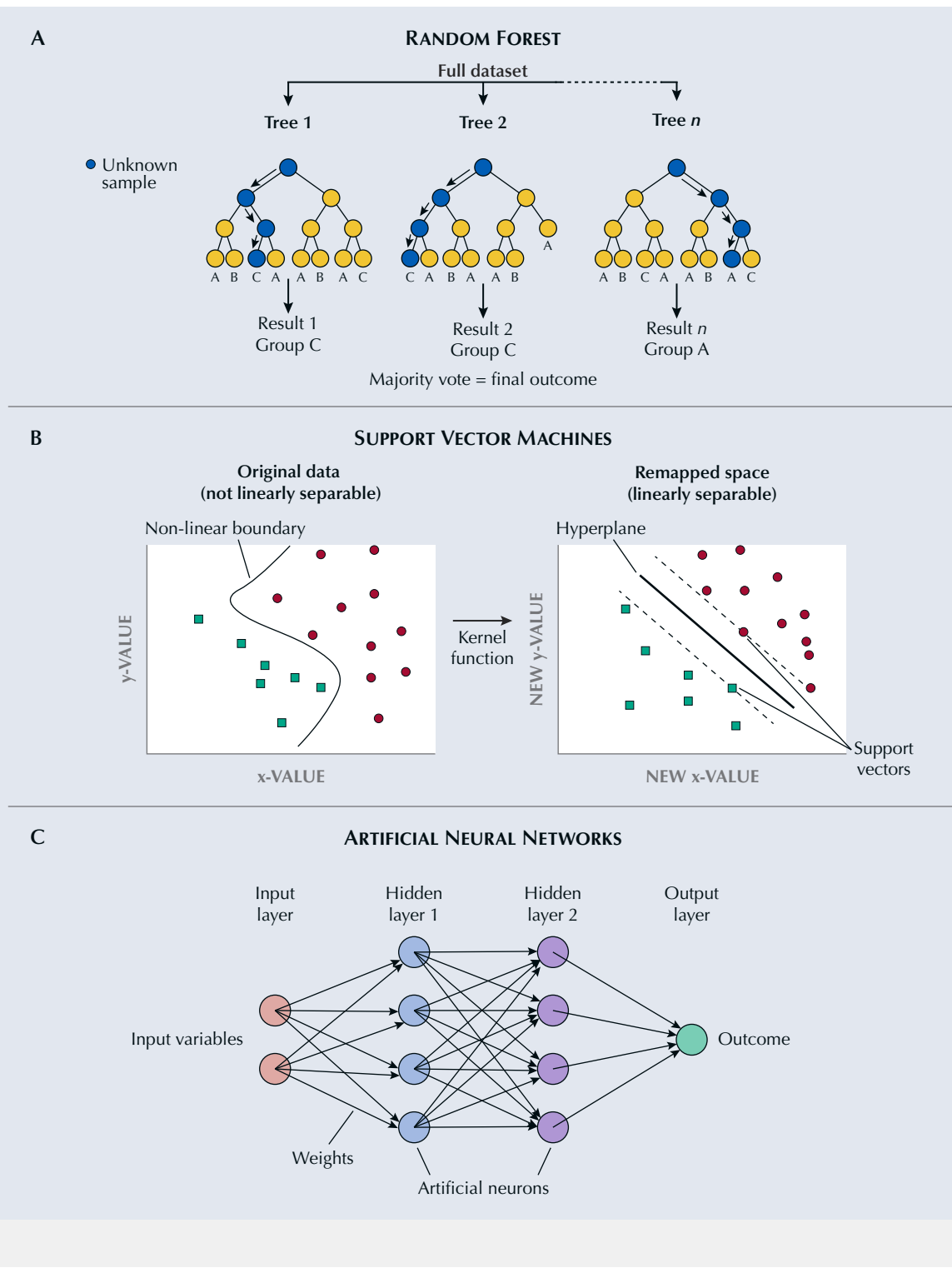
Finally, feature selection is a statistical approach for investigating large databases and identifying the variables that are most important for successful classification of a particular dataset. For spectroscopic data, feature selection can be used to pinpoint the most useful peaks for solving a specific classification problem. The Boruta algorithm, for example, ranks the importance of all variables based on how the error rate of the model changes when each variable is excluded from model calibration (Kursa et al., 2010). Although feature selection algorithms are useful time-saving tools, all “important” variables should be scrutinized to ensure their value. The Boruta approach has been previously applied in determining the importance of geological and environmental processes (e.g., Amiri et al., 2019; Prasad et al., 2021).

Figure B-1. Schematic overviews of three machine learning methods: random forest, support vector machine, and artificial neural network. A: Individual decision trees within RF consist of nodes (circles) representing true or false decisions where unclassified data are tested and then filtered down the tree until assigned a class at the bottom. The outcome for a test dataset is the majority vote from all trees (in this example, Group C). B: In SVM, data that are separable by a nonlinear boundary can be transformed into linearly separable data using a kernel function, a mathematical function that projects the data into a new remapped coordinate space. Support vectors are data points that define the orientation of the boundary (or hyperplane) between groups. C: ANNs consist of input variables and successive hidden layers. Decisions are made at the artificial neurons, where weights indicate the relative importance of different variables at each layer. This is a “feedforward” neural network, where the data are processed in one direction without recursive feedback loops.

RESULTS

Alexandrite. PCA was applied to the trace element concentrations for alexandrite in this study; the results are shown in a bivariate scatterplot with prin-

cipal components 1 and 2 as axes (figure 7). Principal components are linear combinations of the initial variables in a database, serving to summarize and simplify complex datasets; the vectors in figure 7 in-



dicate the relative concentrations of each element and highlight samples with elevated concentrations of each (see box A). The scatterplot reveals separation between several populations of alexandrite from dif-

ferent localities—such as India, Brazil, and Russia—though alexandrite from other localities overlaps to different degrees. Alexandrite from some localities have different compositions than those from other

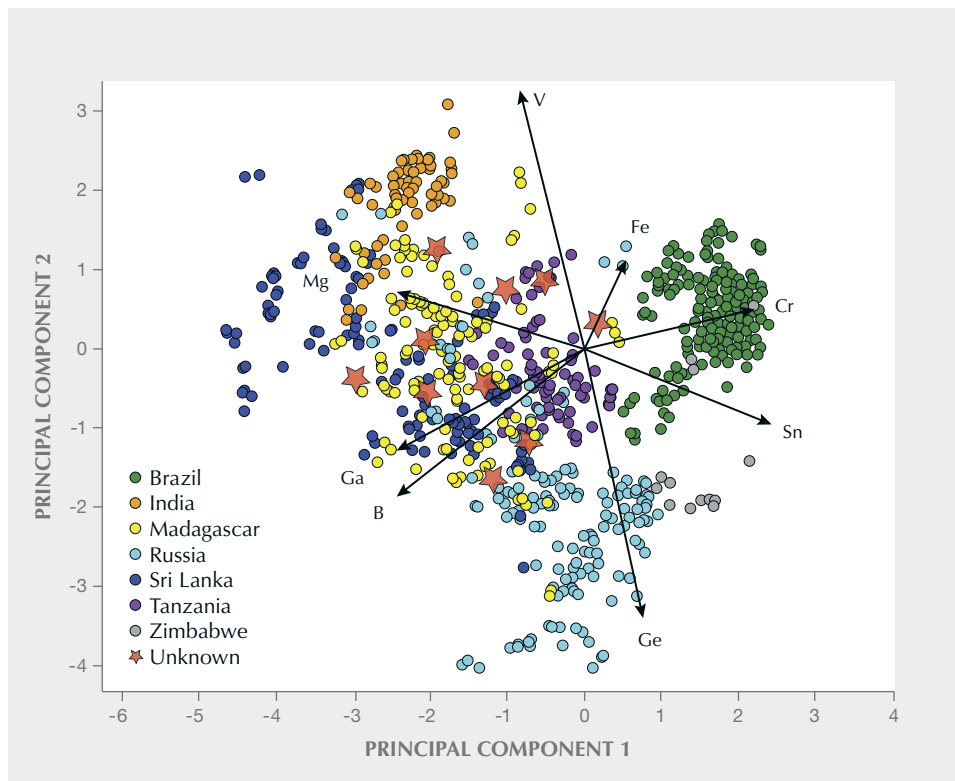
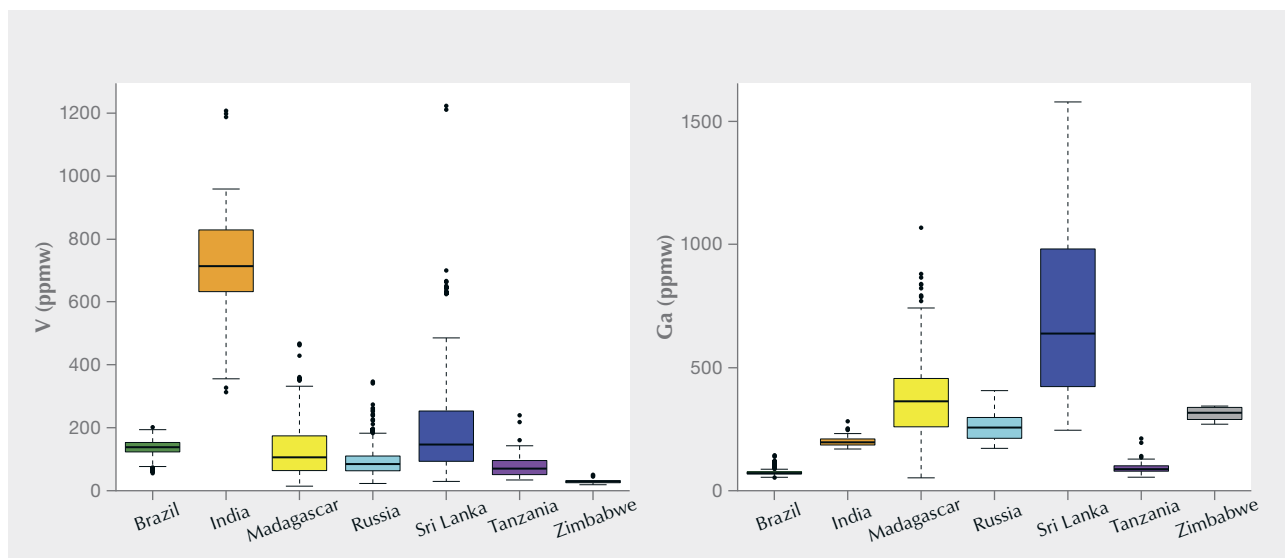


Figure 7. Plot of principal components 1 and 2 for alexandrite samples in this study. The stars denote alexandrite with an “undetermined” provenance using traditional approaches. The values for principal components 1 and 2 convey the relative variation between data points, but the values themselves do not have physical meaning.

localities; for example, samples from India have relatively elevated vanadium concentrations, and those from Sri Lanka have elevated gallium concentrations (figure 8). Calibrated models using LDA, RF, SVM, and ANN achieved cross-validated error rates <4%

for alexandrite provenance determination, within error of one another (table 4). A set of 10 alexandrite samples with elemental compositions previously considered inconclusive using traditional approaches is plotted relative to alexandrite with known prove-

Figure 8. Box and whisker diagrams for the concentrations of vanadium (left) and gallium (right) in alexandrite from the seven localities in this study. The colors match the groups in figure 7.



nance using PCA (figure 7) and classified using the four calibrated models (table 5). The inconclusive alexandrite samples fall within areas of the plot where multiple localities overlap. When tested using the four ML models, half of the samples yielded consistent classifications and the remainder produced contradictory results.

CVD-Grown Diamonds. A previous study (Hardman et al., 2022) applied machine learning to the detection of treatment in CVD-grown diamonds using PL spectra with the Raman-normalized peak intensities of 13 different features as variables. The diamond spectra are complex and many other peaks occur, but their value as indicators of treatment has not been assessed. When the Boruta algorithm is applied to feature selection for CVD-grown diamond PL spectra, it produces a ranked list of peaks. Each peak is evaluated for its ability to improve detection of treated or untreated stones when included in the statistical model. We have tabulated the peaks considered most important by Boruta, and these are presented in figure 9 and table 6. For treated CVD-grown diamonds, the majority

TABLE 4. Error rates^a for statistical models applied to alexandrite and natural saltwater pearl provenance determination.

Machine learning model	Alexandrite (346 samples, 1,157 analyses)	Natural saltwater pearl (604 samples, 1,875 analyses)
Linear discriminant analysis	3.7 ± 3.0%	24.2 ± 4.0%
Random forest	1.3 ± 3.2%	12.1 ± 3.6%
Artificial neural networks	2.0 ± 2.5%	13.2 ± 3.0%
Support vector machines	3.4 ± 3.1%	17.3 ± 3.3%

^aError rates represent mean ± 2σ determined by tenfold cross-validation.

of these features occur in the range of 520–550 nm in PL spectra.

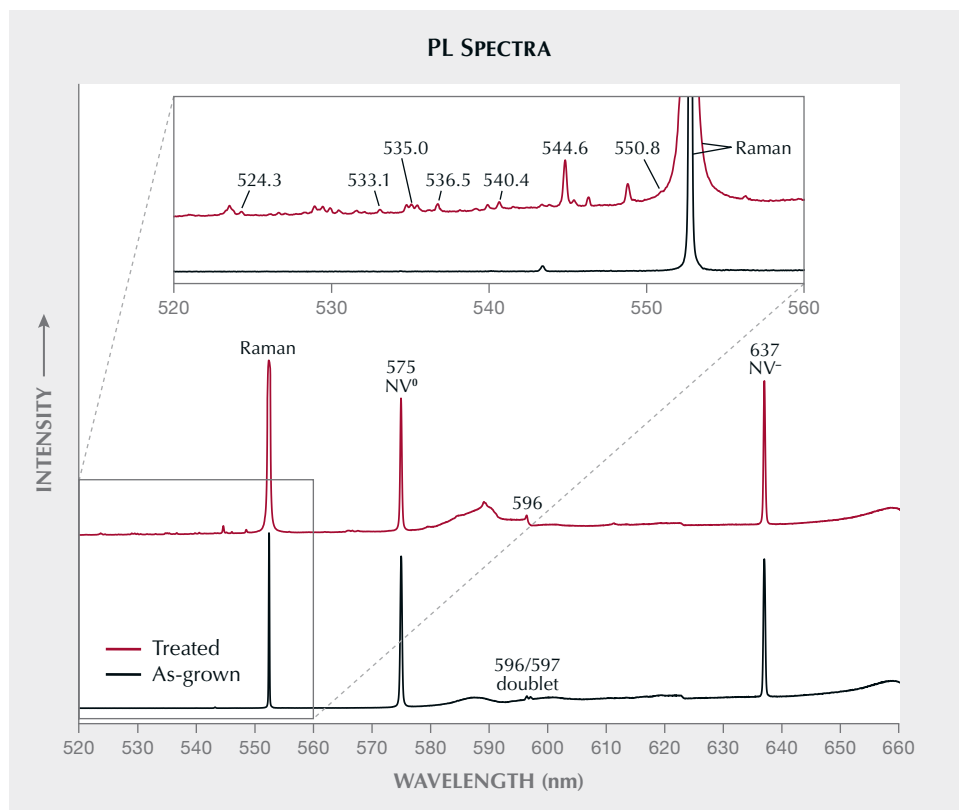


Figure 9. Example PL spectra for as-grown and HPHT-treated CVD-grown diamonds. A subset of 10 peaks whose Raman-normalized intensities are identified as important by the Boruta feature selection algorithm are labeled. The spectra are offset vertically for clarity and scaled so that the Raman peaks have equal intensity. Note that not all peaks appear in every spectrum. For example, the 596/597 nm peak doublet is present in some but not all as-grown CVD diamonds (whereas only 596 nm is present in some treated CVD-grown diamonds).

TABLE 5. Classification outcomes for ten alexandrite samples with previously inconclusive provenance, determined using four calibrated ML models.

Sample	Linear discriminant analysis	Random forest	Artificial neural networks	Support vector machines
1	Madagascar	Madagascar	Madagascar	Russia
2	Madagascar	Madagascar	Madagascar	Madagascar
3	Tanzania	Madagascar	Madagascar	Tanzania
4	Tanzania	Madagascar	Madagascar	Madagascar
5	Sri Lanka	Madagascar	Madagascar	Madagascar or Sri Lanka ^a
6	Madagascar	Madagascar	Madagascar	Madagascar
7	Madagascar	Madagascar	Madagascar	Madagascar
8	Madagascar	Madagascar	Madagascar	Madagascar
9	Tanzania	Tanzania	Tanzania	Tanzania
10	Madagascar or Brazil	Madagascar	Madagascar	Madagascar or Russia

^aSamples with multiple outcomes are those with subequal probabilities for each.

Natural Saltwater Pearls. There are varying levels of population separation when PCA results are shown using bivariate scatterplots and when pearls are color-coded based on different classification variables (figure 10, A and B). Many pearls reportedly fished from Kuwait and Bahrain (with relatively close spatial proximity) and produced by the same *Pinctada radiata* mollusk species have similar compositions and overlap; Oman pearls, which were reportedly produced by *Pinctada margaritifera* mollusks, are separated from Kuwait and Bahrain pearls (figure 10A). The population of Bahrain pearls is broadly bimodal, with one subgroup having strong overlap with Kuwait pearls and the other showing elevated magnesium and manganese concentrations (figure 10A). Moreover, nacreous and non-nacreous pearls differ in composition, with non-nacreous pearls (mostly from Bahrain) having higher magnesium concentrations (figure 10, B and C). The magnesium concentrations of non-nacreous pearls from Bahrain, Kuwait, and Oman largely overlap with each other (figure 10C).

Using models based on LDA, RF, SVM, and ANN, geographic provenance can be determined with model error rates <25%, with RF performing the best (<13% error rate; table 4). For outcomes produced using the RF model, the misclassification of natural saltwater pearls is tabulated in table 7. The most sig-

nificant misclassifications are between Kuwait and Bahrain pearls: 13% of Bahrain and 18% of Kuwait

TABLE 6. Ranking of top 10 peaks in CVD-grown diamond spectroscopic data collected using 514 nm PL, identified by the Boruta algorithm.

Rank ^a	Approximate peak position (nm)
1	524.3
2	575
3	540.4
4	535.0
5	544.6
6	596/597 doublet
7	536.5
8	533.1
9	637
10	550.8

^aRanked importance of the Raman-normalized intensities of each peak when used as a classification variable in machine learning models. These ranks are produced by examining PL spectra for 300 as-grown and 1,795 post-growth treated CVD diamonds.

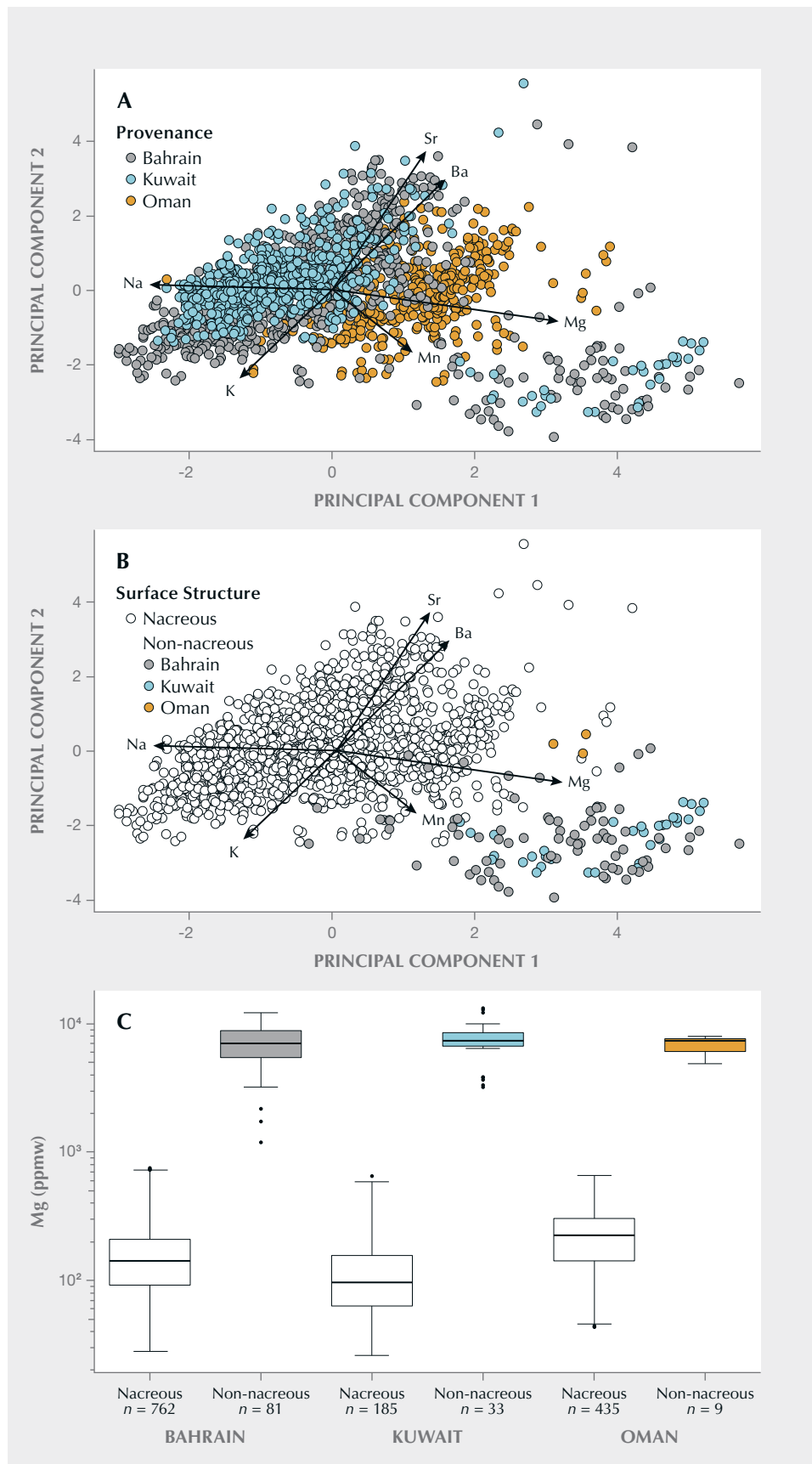


Figure 10. Plot of principal components 1 and 2 for natural saltwater pearls, based on (A) geographic provenance (Bahrain, Kuwait, and Oman) and (B) surface structure (nacreous and non-nacreous). C: Box and whisker plot of magnesium concentrations for all analyses of nacreous and non-nacreous pearls in this study, divided by locality. Solid lines correspond to median values, boxes represent the range from the 1st to 3rd quartile (interquartile range; IQR), the vertical lines are whiskers that range from $IQR \pm (1.5 \times IQR)$, and the black dots correspond to pearls with values beyond the whiskers, considered to be outliers.

pearls misclassify, primarily as one another, with very few misclassifications as “Oman.” Conversely, only 4% of pearls from Oman receive an incorrect provenance determination, all as “Bahrain.”

DISCUSSION

Alexandrite Provenance Determination. When alexandrite trace element compositions are transformed using PCA, separation between samples from different localities becomes apparent, enabling the broad classification of several alexandrite samples with previously inconclusive provenance (figure 7). PCA indicates several elements that may discriminate alexandrite from different localities, such as high vanadium in Indian alexandrite and gallium concentrations that are elevated in Sri Lankan alexandrite and low in Brazilian samples. These differences are also resolved in box and whisker plots for these elements (figure 8). Alexandrite from Sri Lanka and Madagascar display overlapping compositions that may be differentiated using additional principal components. PCA also indicates that Brazilian alexandrites are compositionally diverse, with some having higher germanium despite reportedly originating from the same mine (Sun et al., 2019). This underscores the importance of compiling representative databases for accurate provenance determination of unclassified stones. The compositional variability of alexandrite within a particular country may increase as stones from other mines are collected.

A moderately strong positive correlation between the elements boron and gallium for the full dataset is inferred from the vectors with similar length and direction shown in figure 7. The primary advantage

of applying PCA to alexandrite is its ability to consider each of these aspects simultaneously, enabling efficient visualization and interpretation of alexandrite compositional data. If alexandrite samples from different localities are formed through distinct geological processes, PCA plots may reveal detectable differences in their compositions.

Machine learning models produced using the LDA, RF, SVM, and ANN methods can determine alexandrite provenance with error rates <4% (table 4). Of the 10 alexandrite samples with inconclusive provenance determinations using traditional approaches, 50% received a consistent classification using all four methods, while others were classified similarly using three different models (table 5). However, some samples received conflicting outcomes or classification probabilities (not shown) that were intermediate. These samples should be referred to a trained specialist before the ML outcome is accepted. Discrepancies between models calibrated using the same dataset may arise from the underlying mathematical properties of each model, some of which are more suitable for particular datasets, combined with the fact that some alexandrite samples have elemental compositions that are transitional between stones from different localities. This may be further complicated by changes in quality and consistency of data collection methods over time, including changes in standard reference materials. Previous studies have shown RF to be a particularly powerful method (Fernández-Delgado et al., 2014; Bolton et al., 2019), a finding consistent with model error rates in this study (table 4). In cases of conflicting outcomes, the RF model results may carry more weight with trained specialists than LDA, for example.

TABLE 7. Summary of classification results for natural saltwater pearls in this study using the random forest model. Samples with matching known and predicted provenance are considered to be correctly classified.

Known provenance	Predicted provenance			Error rate
	Bahrain	Kuwait	Oman	
Bahrain (268 samples)	233	28	7	13.1%
Kuwait (188 samples)	33	155	0	17.6%
Oman (148 samples)	6	0	142	4.1%

For alexandrite from different localities with similar compositions, it can be difficult to accurately determine provenance even with very robust ML models. To reduce model error rates and improve provenance determination for more stones, new compositional variables can be added. While the addition of ML to traditional approaches failed to definitively classify all samples in this study, it reduced the number of “undetermined” results by 50%, using existing data routinely acquired during gemological assessment.

CVD-Grown Diamond Treatment Detection. Machine learning approaches previously applied to detecting treatment in CVD-grown diamonds achieved error rates <5%, using the intensities of 13 peaks in PL spectra as variables in RF models (Hardman et al., 2022). The models described by Hardman et al. (2022) are currently used at GIA to aid in treatment detection for CVD-grown diamonds. For each stone analyzed, a confidence value is provided to the user, reflecting the model’s level of certainty in the classification outcome. This is very important for accurate detection of treatment: Accepting ML outcomes without careful consideration, even with error rates of 1%, can scale to a large number of misclassifications as more stones are tested.

Laboratory-grown diamond producers are continually exploring new approaches to post-growth treatment (Eaton-Magaña et al., 2024). These developments may introduce new defect distributions, rendering existing classification criteria obsolete and necessitating the development of new criteria. Beyond the 13 peaks used by Hardman et al. (2022), the spectroscopic signatures of CVD-grown diamonds contain many other peaks, not all of which have been fully characterized. As many peaks occur in different combinations in thousands of available spectra, machine learning offers a powerful tool to mine these data. By applying ML to diamonds with known treatment histories, researchers can identify potential new classification criteria significantly faster.

Applying Boruta to the CVD-grown diamond database in this study produced a ranking of the PL spectra peaks considered statistically useful for treatment detection. For brevity, we have listed only the top 10 important peaks as identified by the feature selection algorithm (table 6), shown graphically in figure 9. The majority of these peaks (e.g., 524.3 and 575 nm) were previously used for treat-

ment detection by Hardman et al. (2022). The peak doublet at 596/597 nm has a high ranking and was previously reported as an important indicator of untreated CVD-grown diamonds (Wang et al., 2003; Martineau et al., 2004). Other peaks, such as 536.5 nm, are assigned high importance using Boruta but were not applied directly by Hardman et al. (2022). This feature occurs in the 520–550 nm spectral range, which contains many previously documented peaks associated with post-growth treatment of CVD-grown diamonds (Wang et al., 2012; Barrie, 2020; Hardman et al., 2022). With further validation and testing, this peak—and others identified using feature selection—may become new criteria for detection, complementing existing variables. Boruta analysis also provides a sense of comparative importance that is useful for classification. For example, a 524.3 nm peak may be very important for identifying HPHT treatment. In its absence, the relative intensity of the 575 nm peak—which is present in nearly all CVD-grown diamonds and ranked as the second most important variable (table 6)—becomes an even more important variable. New variables identified by Boruta must be validated before they are added into new ML models for treatment detection: Future studies can evaluate how strongly new variables improve classification error rates. Regardless, the feature selection procedure as a whole represents a significant time savings for objectively parsing large datasets and identifying subtle variations in available data that may be difficult for a human to quantify without computer software.

Natural Saltwater Pearl Provenance Determination. PCA applied to trace element compositions of natural saltwater pearls reveals that many pearls from Oman are compositionally distinct from those of Kuwait and Bahrain, driven in part by elevated magnesium concentrations (figure 10, A and B). By comparison, many pearls from Kuwait and Bahrain are compositionally similar to each other. The elemental compositions of pearls are the result of complex biological processes related to mollusk species, water chemistry, and nutrient sources (Homkrajac et al., 2019). Therefore, it is difficult to specify which biogenic processes cause the compositional similarity between pearls from Kuwait and Bahrain. For example, while the Kuwait and Bahrain pearls are derived from *Pinctada radiata* mollusks, their nutrient sources might differ. There is also no guarantee that

all mollusks from a single locality and from the same species had similar nutrient sources. This may account for the compositional spread in pearls from each locality (figure 10).

For example, non-nacreous pearls have higher magnesium concentrations than nacreous pearls from the same locality (figure 10C). When analyzing pearls using LA-ICP-MS, it is important to accurately record the position of the analysis within the pearl, along with other useful descriptive information about the material (such as the presence or absence of nacre). These observations provide important context for compositional variation displayed in PCA plots.

Despite the compositional overlap of pearls from Bahrain and Kuwait using PCA (figure 10, A and B), the LDA, RF, SVM, and ANN models can still determine the provenance of many pearls (table 4). Each model has a moderate error rate for the classification of pearl provenance (>10%), indicating compositional differences between many pearls from different localities. However, some compositional overlap remains unresolved using the current set of elemental variables, even when using ML techniques. These error rates are higher than those observed for the classification of alexandrite (table 4), reflecting the nature of biological and growth environment conditions—some problems cannot be fully resolved even with powerful statistical methods such as RF. Nevertheless, the majority of Oman pearls from this study received a correct classification using RF, whereas classifications for Bahrain and Kuwait were less conclusive (table 7). When finalizing a provenance determination, pearls classified by ML as “Oman” tend to be more accurate. The similar compositions of some pearls from Bahrain and Kuwait may be attributed to the fact that they are produced by the same mollusk species (*Pinctada radiata*) located within the Persian (Arabian) Gulf, which provides similar biological and growth environment conditions. Conversely, Oman pearls form in *Pinctada margaritifera* mollusks found in the Gulf of Oman region.

Trace element composition also provides insight into the biogenic processes that formed the pearls. Freshwater pearls have much higher manganese content than saltwater pearls (Homkrajae et al., 2019). Among the natural saltwater pearls in this study, we noted several other variations in elemental compositions that may reflect local environmental conditions, such as varying concentrations of strontium, sodium, potassium, and barium (figure 10, A and B). Pearls harvested over time from a variety of environ-

ments may also be sensitive to small- and large-scale climate changes. Consequently, analyzing pearl compositional data using ML may be complicated by temporal changes in environment.

Pearl research is an active area of study at GIA, one that relies on acquiring pearls for which the provenance and culturing history are known and well characterized. Findings from ML will help shape future research projects by identifying important elemental variables and highlighting populations of pearls that would benefit from expanded sampling. Adding more elements into the ML models in this study—if these elements are shown to be free of analytical interferences (compare to Homkrajae et al., 2019)—may further improve the determination of pearl provenance.

CONCLUSIONS

Traditional approaches to gem classification at GIA encompass microscopic examination and the acquisition of trace element or spectroscopic data. Trace element data can be compared using bivariate scatterplots and selective plotting, while spectroscopic data are assessed through visual inspection to pinpoint key classification features. Unclassified materials are compared to curated datasets of samples for which the classification has been independently verified (figure 11). Machine learning complements these approaches, enhancing the determination of provenance and detection of color treatment in gem materials. The ML models in this study can successfully determine the provenance of alexandrite and reduce the number that may otherwise receive an inconclusive outcome. For CVD-grown diamonds, ML identifies peaks in spectroscopic data that offer valuable indicators of post-growth treatment. For natural saltwater pearls, principal component analysis helps to assess biogenic processes related to growth environment, enabling preliminary provenance determination.

Despite high success rates for some classification problems (>95%; table 4), ML models still provide incorrect answers sometimes. For certain natural gemstones, this could be associated with geology—similar geological processes occurring at different localities may result in samples with very similar properties and elemental compositions, complicating provenance determination (e.g., Smith et al., 2022). In these cases, ML does not always provide a confident provenance determination, no matter how robust the approach. If a decision cannot be made after classification by ML



Figure 11. Double-strand natural saltwater pearl necklace. Photo by Robert Weldon; courtesy of Karmshil Enterprise. Alexandrite exposed to incandescent illumination (left, red) and daylight-equivalent illumination (right, green). Photos by Robert Weldon; courtesy of GIA's Dr. Edward J. Gübelin collection, no. 33805. Ring containing a CVD-grown diamond. Photo by Johnny Leung.

and review by a trained specialist, the sample may receive a final decision of “undetermined.” Nevertheless, ML successfully minimizes the small number of “undetermined” outcomes by leveraging data that is

already part of routine gemological practices. These techniques are powerful complements to existing laboratory practices, offering substantial time and cost savings.

ABOUT THE AUTHORS

Dr. Matthew F. Hardman is a research scientist, Artitaya Homkrajee is supervisor of pearl identification, Dr. Sally Eaton-Magaña is senior manager of diamond identification, Dr. Christopher M. Breeding is senior manager of analytics, Dr. Aaron C. Palke is senior manager of research, and Ziyin Sun is a senior research associate, at GIA in Carlsbad, California.


ACKNOWLEDGMENTS

We thank Troy Ardon for software development, as well as Chunhui Zhou, Abeer Al-Alawi, Mei Yan Lai, Ulrika D'Haenens-Johansson, James Shigley, and three anonymous reviewers for useful comments and suggestions that improved the quality of this manuscript.

REFERENCES

- Amiri M., Pourghasemi H.R., Ghanbarian G.A., Afzali S.F. (2019) Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma*, Vol. 340, pp. 55–69, <http://dx.doi.org/10.1016/j.geoderma.2018.12.042>
- Arnault J.-C., Saada S., Ralchenko V. (2022) Chemical vapor deposition single-crystal diamond: A review. *Physica Status Solidi Rapid Research Letters*, Vol. 16, No. 1, article no. 2100354, <http://dx.doi.org/10.1002/pssr.202100354>
- Barrie E. (2020) Lab Notes: HPHT-processed CVD laboratory-grown diamonds with low color grades. *G&G*, Vol. 56, No. 2, pp. 289–290.
- Bassoo R., Eames D., Hardman M.F., Befus K., Sun Z. (2023) Topaz from Mason County, Texas. *G&G*, Vol. 59, No. 4, pp. 414–431, <http://dx.doi.org/10.5741/GEMS.59.4.414>
- Bendinelli T., Biggio L., Nyfeler D., Ghosh A., Tollan P., Kirschmann M.A., Fink O. (2024) Gemtelligence: Accelerating gemstone classification with deep learning. *Communications Engineering*, Vol. 3, article no. 110, <http://dx.doi.org/10.1038/s44172-024-00252-x>
- Blodgett T., Shen A.H. (2011) Application of discriminant analysis in gemology: Country-of-origin separation in colored stones and distinguishing HPHT-treated diamonds. *G&G*, Vol. 47, No. 2, p. 145.
- Bolton M.S.M., Jensen B.J.L., Wallace K., Praet N., Fortin D., Kaufman D., Batist M.D. (2019) Machine learning classifiers for attributing tephra to source volcanoes: An evaluation of methods for Alaska tephra. *Journal of Quaternary Science*, Vol. 35, No. 1–2, pp. 81–92, <http://dx.doi.org/10.1002/jqs.3170>
- Boser B.E., Guyon I.M., Vapnik V.N. (1992) A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152.
- Breiman L. (2001) Random forests. *Machine Learning*, Vol. 45, pp. 5–32, <http://dx.doi.org/10.1023/A:1010933404324>
- Cristianini N., Shawe-Taylor J. (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge, UK.
- Dawson J.B., Stephens W.E. (1975) Statistical classification of garnets from kimberlite and associated xenoliths. *Journal of Geology*, Vol. 83, No. 5, pp. 589–607, <http://dx.doi.org/10.1086/628143>
- D'Haenens-Johansson U.F.S., Butler J.E., Katrusha A.N. (2022) Synthesis of diamonds and their identification. *Reviews in Mineralogy & Geochemistry*, Vol. 88, No. 1, pp. 689–754, <http://dx.doi.org/10.2138/rmg.2022.88.13>
- Eaton-Magaña S., Hardman M.F., Odake S. (2024) Laboratory-grown diamonds: An update on identification and products evaluated at GIA. *G&G*, Vol. 60, No. 2, pp. 146–167, <http://dx.doi.org/10.5741/GEMS.60.2.146>
- Fernández-Delgado M., Cernadas E., Barro S., Amorim D. (2014) Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 3133–3181.
- Fuge R., Palmer T.J., Pearce N.J.G., Perkins W.T. (1993) Minor and trace element chemistry of modern shells: A laser ablation inductively coupled plasma mass spectrometry study. *Applied Geochemistry*, Vol. 8, Supplement 2, pp. 111–116, [http://dx.doi.org/10.1016/S0883-2927\(09\)80020-6](http://dx.doi.org/10.1016/S0883-2927(09)80020-6)
- Griffin W.L., Fisher N.L., Friedman J.H., O'Reilly S.Y., Ryan C.G. (2002) Cr-pyrope garnets in the lithospheric mantle 2. Compositional populations and their distribution in time and space. *Geochemistry Geophysics Geosystems*, Vol. 3, No. 12, pp. 1–35, <http://dx.doi.org/10.1029/2002GC000298>
- Groat L.A., Giuliani G., Stone-Sundberg J., Sun Z., Renfro N.D., Palke A.C. (2019) A review of analytical methods used in geographic origin determination of gemstones. *G&G*, Vol. 55, No. 4, pp. 512–535, <http://dx.doi.org/10.5741/GEMS.55.4.512>
- Gübelin E., Schmetzer K. (1982) Gemstones with alexandrite effect. *G&G*, Vol. 18, No. 4, pp. 197–203, <http://dx.doi.org/10.5741/GEMS.18.4.197>
- Hardman M.F., Pearson D.G., Stachel T., Sweeney R.J. (2018a) Statistical approaches to the discrimination of crust- and mantle-derived low-Cr garnet – Major-element-based methods and their application in diamond exploration. *Journal of Geochemical Exploration*, Vol. 186, pp. 24–35, <http://dx.doi.org/10.1016/j.gexplo.2017.11.012>
- (2018b) Statistical approaches to the discrimination of mantle- and crust-derived low-Cr garnets using major and trace element data. *Mineralogy and Petrology*, Vol. 112, No. S2, pp. 697–706, <http://dx.doi.org/10.1007/s00710-018-0622-7>
- Hardman M.F., Eaton-Magaña S.C., Breeding C.M., Ardon T., D'Haenens-Johansson U.F.S. (2022) Evaluating the defects in CVD diamonds: A statistical approach to spectroscopy. *Diamond and Related Materials*, Vol. 130, article no. 109508, <http://dx.doi.org/10.1016/j.diamond.2022.109508>
- Hastie T., Tibshirani R., Friedman J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer-Verlag, New York.
- Hazen R.M., Morrison S.M. (2022) On the paragenetic modes of minerals: A mineral evolution perspective. *American Mineralogist*, Vol. 107, No. 7, pp. 1262–1287, <http://dx.doi.org/10.2138/am-2022-8099>
- Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M., Graler B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, Vol. 6, article no. e5518, <http://dx.doi.org/10.7717/peerj.5518>
- Homkrajac A., Sun Z., Blodgett T., Zhou C. (2019) Provenance discrimination of freshwater pearls by LA-ICP-MS and linear discriminant analysis (LDA). *G&G*, Vol. 55, No. 1, pp. 47–60, <http://dx.doi.org/10.5741/GEMS.55.1.47>
- Homkrajac A., Manustrong A., Nilpetploy N., Sturman N., Lawanwong K., Kessrapong P. (2021) Internal structures of known *Pinctada maxima* pearls: Natural pearls from wild marine mollusks. *G&G*, Vol. 57, No. 1, pp. 2–21, <http://dx.doi.org/10.5741/GEMS.57.1.2>
- Krebs M.Y., Hardman M.F., Pearson D.G., Luo Y., Fagan A.J. (2020) An evaluation of the potential for determination of the geographic origin of ruby and sapphire using an expanded trace element suite plus Sr-Pb isotope compositions. *Minerals*, Vol. 10, No. 5, article no. 447, <http://dx.doi.org/10.3390/min10050447>
- Kursa M.B., Jankowski A., Rudnicki W.R. (2010) Boruta – a system for feature selection. *Fundamenta Informaticae*, Vol. 101, No. 4, pp. 271–285, <http://dx.doi.org/10.3233/FI-2010-288>
- LeCun Y., Bengio Y., Hinton G. (2015) Deep learning. *Nature*, Vol. 521, No. 7553, pp. 436–444, <http://dx.doi.org/10.1038/nature14539>
- Luo Z., Yang M., Shen A.H. (2015) Origin determination of dolomite-related white nephrite through IB-LDA. *G&G*, Vol. 51, No. 3, pp. 300–311, <http://dx.doi.org/10.5741/GEMS.51.3.300>
- Martineau P.M., Lawson S.C., Taylor A.J., Quinn S.J., Evans D.J.F., Crowder M.J. (2004) Identification of synthetic diamond grown using chemical vapor deposition (CVD). *G&G*, Vol. 40, No. 1, pp. 2–25, <http://dx.doi.org/10.5741/GEMS.40.1.2>
- Palke A.C., Saeseaw S., Renfro N.D., Sun Z., McClure S.F. (2019a) Geographic origin determination of blue sapphire. *G&G*, Vol. 55, No. 4, pp. 536–579, <http://dx.doi.org/10.5741/GEMS.55.4.536>
- (2019b) Geographic origin determination of ruby. *G&G*, Vol. 55, No. 4, pp. 580–612, <http://dx.doi.org/10.5741/GEMS.55.4.580>
- Prasad P., Loveson V.J., Das S., Chandra P. (2021) Artificial intelligence approaches for spatial prediction of landslides in mountainous regions of western India. *Environmental Earth*

- Sciences*, Vol. 80, No. 21, article no. 720, <http://dx.doi.org/10.1007/s12665-021-10033-w>
- R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <https://www.R-project.org>
- Schönig J., von Eynatten H., Tolosana-Delgado R., Meinhold G. (2021) Garnet major-element composition as an indicator of host-rock type: A machine learning approach using the random forest classifier. *Contributions to Mineralogy and Petrology*, Vol. 176, No. 12, article no. 98, <http://dx.doi.org/10.1007/s00410-021-01854-w>
- Smith E.M., Smit K.V., Shirey S.B. (2022) Methods and challenges of establishing the geographic origin of diamonds. *G&G*, Vol. 58, No. 3, pp. 270–288, <http://dx.doi.org/10.5741/GEMS.58.3.270>
- Sun Z., Palke A.C., Muyal J., DeGhionno D., McClure S.F. (2019) Geographic origin determination of alexandrite. *G&G*, Vol. 55, No. 4, pp. 660–681, <http://dx.doi.org/10.5741/GEMS.55.4.660>
- Wang W., Moses T., Linares R.C., Shigley J.E., Hall M., Butler J.E. (2003) Gem-quality synthetic diamonds grown by a chemical vapor deposition (CVD) method. *G&G*, Vol. 39, No. 4, pp. 268–283, <http://dx.doi.org/10.5741/GEMS.39.4.268>
- Wang W., D'Haenens-Johansson U.F.S., Johnson P., Moe K.S., Emerson E., Newton M.E., Moses T.M. (2012) CVD synthetic diamonds from Gemesis Corp. *G&G*, Vol. 48, No. 2, pp. 80–97, <http://dx.doi.org/10.5741/GEMS.48.2.80>
- Xu Y., Liu X., Cao X., Huang C., Liu E., Qian S., Liu X., Wu Y., Dong F., Qiu C.-W., Qiu J., Hua K., Su W., Wu J., Xu H., Han Y., Fu C., Yin Z., Liu M., Roepman R., Dietmann S., Virta M., Kengara F., Zhang Z., Zhang L., Zhao T., Dai J., Yang J., Lan L., Luo M., Liu Z., An T., Zhang B., He X., Cong S., Liu X., Zhang W., Lewis J.P., Tiedje J.M., Wang Q., An Z., Wang F., Zhang L., Huang T., Lu C., Cai Z., Wang F., Zhang J. (2021) Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, Vol. 2, No. 4, article no. 100179, <http://dx.doi.org/10.1016/j.xinn.2021.100179>
- Zhao H., Lai Z., Leung H., Zhang X. (2020) Linear discriminant analysis. In *Feature Learning and Understanding: Algorithms and Applications*. Information Fusion and Data Science Series. Springer Nature, Cham, Switzerland, https://doi.org/10.1007/978-3-030-40794-0_5



Join our growing G&G Facebook group of nearly 45,000 members, connecting gem enthusiasts from all over the world!

