# Automated Data Processing of Raman Spectra for Supporting Spinel Origin Determination

**Tasnara Sripoonjan[1,2], Montira Seneewong Na Ayutthaya[3], Sirinapa Saengdech[3], Sutas Singbamroong[4], Alongkot Fanka[2]**

[1] G-ID Laboratories, Yan Nawa, Bangkok 10120 Thailand; *tasnara@hotmail.com

[2] Department of Geology, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand

[3] The Gem and Jewelry Institute of Thailand (Public Organization), Bangkok 10500 Thailand

[4] Dubai Central Laboratory Department, Dubai Municipality, Dubai, UAE

**Keywords: Spinel, Raman spectroscopy, provenance, automated data processing, origin determination, machine learning**

## Introduction

Spinel is a highly valued gemstone, with its color and quality influenced by trace-element chemistry and crystal structure. Accurate determination of its geographic origin is crucial for gemmological authentication, scientific research, and maintaining supply-chain integrity. Raman spectroscopy provides a non-destructive analytical approach, offering vibrational fingerprints that reflect both the crystal lattice and impurity-related features. However, manual interpretation is time-consuming and susceptible to operator bias, particularly when dealing with large datasets or subtle spectral variations. To address these challenges, we present an end-to-end automated pipeline that integrates spectral preprocessing, feature extraction, and chemometric analysis. Machine learning (ML) models are then applied to classify spinel samples according to their geographic origin. The performance of this approach is demonstrated using both red and blue spinels from three major deposits.

## Materials and Methods

A total of 315 spinel samples (Figure 1) were sourced from local, trusted dealers and verified by gemmological means. Red spinel ($MgAl_2O_4$; n = 154) originated from Mogok (Myanmar), Luc Yen (Vietnam) and Lukande (Tanzania), while blue samples (n = 161) comprised spinel ($MgAl_2O_4$) from Luc Yen and Lukande and gahnite (ideally $ZnAl_2O_4$; n = 52) from Jemaa (Nigeria). Each sample was polished to present at least one flat surface for Raman analysis. Raman spectra were collected on a confocal Renishaw inVia Raman microscope equipped with a 50× objective (NA 0.50) using a 785 nm excitation laser. Three accumulations of 10s each were acquired over 200–1000 $cm^{-1}$ with 1 $cm^{-1}$ spectral resolution, conditions verified to avoid local heating (no peak drift between successive scans). Raw spectra were exported as CSV files and processed via the automated data processing pipeline.

Figure 1. Representative spinel samples used in this study, with weights ranging from 0.37 to 2.70 carats. (Photo by M. Seneewong Na Ayutthaya)

## Automated Data Processing Pipeline

Our data-processing pipeline was implemented in Python 3.12 and proceeds as follows. First, Raman spectra are smoothed using a Savitzky–Golay filter (window length 7, polynomial order 3) to suppress high-frequency noise while preserving peak shapes. The smoothed spectra are then baseline-corrected via Asymmetric Least Squares (ALS) algorithm with $\lambda = 10^7$ and $p = 0.01$, followed by max-normalization (I/Imax) to remove intensity scale effects. Next, we isolate the four vibrational bands at 310, 405, 665, and 765 cm$^{-1}$ and fit each with a Lorentzian profile (accepting fits only if $R^2 \geq 0.95$) to extract the full width at half maximum (FWHM). The resulting FWHM values, together with two dimensionless ratios (e.g., 405/665 and 405/765), are assembled into feature vectors and visualized in bi-scatter plots to assess natural clustering by geographic origin. Finally, we train both an artificial neural network (ANN) and a random forest (RF) models using 10-fold cross-validation and a 20% held-out test set (red spinel n = 28; blue spinel n = 30), optimize hyperparameters by grid search, and report overall accuracy, macro F1-score, and mean cross-validation accuracy (mean CV).

## Results and Discussion

The full dataset of 315 spectra was processed on a MacBook Pro (M1 Pro, 16 GB RAM) in 149 seconds, averaging 0.47 seconds per spectrum. FWHM-based plots reveal well-defined clusters corresponding to geographic provenance. For red spinel, plotting the FWHM of the 405 cm$^{-1}$ band against that of the 665 cm$^{-1}$ band distinguishes samples from Myanmar (MM), Vietnam (VN), and Tanzania (TZ), though some overlap occurs between MM and VN. In blue spinel ($MgAl_2O_4$), the 405 cm$^{-1}$ and 665 cm$^{-1}$ modes are decisive. Nigerian gahnite ($ZnAl_2O_4$) occupies a distinct field due to homologous peak shifts to approximately 415 cm$^{-1}$ and 657 cm$^{-1}$. Applying adaptive fitting windows of ±10 cm$^{-1}$ further isolates gahnite without disturbing the true-spinel clusters (Fig. 2). While gahnite is also distinguishable from blue Mg–Al spinels by RI, SG, and Zn-rich chemistry, Raman provides a complementary method, especially for small or mounted stones. These FWHM shifts likely reflect variations in A–B site inversion and mass-related frequency shifts from $Mg^{2+}$ substitution with heavier $Zn^{2+}$ (Malavasi *et al.*, 2002; Wang *et al.*, 2020). Additionally, the separation—and remaining overlap—between TZ and VN blue spinels may be influenced by Fe/Co variations affecting site distortion (Furuya, 2023). Classification models trained on FWHM features performed strongly, as summarized in Table I. Confusion matrix analysis indicates most misclassifications occurred between MM and VN red spinels, and between TZ and VN blue ones, reflecting partially overlapping geological and chemical characteristics (Chauviré *et al.*, 2015; Chankhantha *et al.*, 2020; Krzemnicki *et al.* 2023, Wu *et al.*, 2023). Overall, these findings demonstrate that deposit-specific FWHM features provide an effective foundation for automated, machine learning–based provenance determination of spinel.

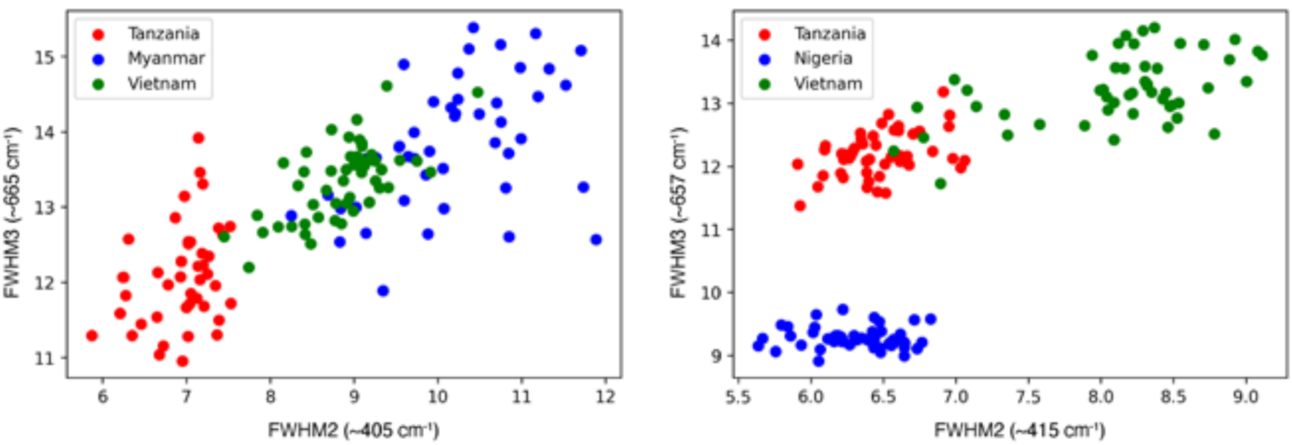


Figure 2. Scatter plots illustrating FWHM-based feature distributions for classification: (left) red spinel samples from Myanmar, Vietnam, and Tanzania; (right) blue spinel (Vietnam, Tanzania) and gahnite (Nigeria).

| Dataset | Model | Accuracy | Macro F1 | Mean CV | Misclassified samples |
|---|---|---|---|---|---|
| **Red spinel** | ANN | 93 | 93 | 89% ± 6.6% | 2/28: (1MM→VZ, 1VN→TZ) |
| | RF | 93 | 93 | 88% ± 7.1% | 2/28: (1MM→VZ, 1VN→TZ) |
| **Blue spinel & gahnite** | ANN | 97 | 97 | 95% ± 4.0% | 1/30: (1VN→TZ) |
| | RF | 93 | 94 | 96% ± 4.1% | 2/30: (2VN→TZ) |

Table I Performance summary of ML models applied to red and blue spinel classification

## Conclusion

We present an automated, high-throughput workflow for Raman spectral processing and provenance classification of blue and red spinel. By using FWHM of key vibrational bands as features, the pipeline captures subtle lattice variations associated with geographic origin. This approach achieves over 90 % classification accuracy on the current dataset and significantly reduces analysis time, enabling scalable and reproducible gemmological assessments. The strong performance of FWHM-based features highlight their sensitivity to crystallographic disorder and trace-element chemistry—factors that are critical for provenance differentiation. Building on these results, future work will expand the geographic diversity of the spinel dataset and extend the pipeline to a broader range of gemstone species.

## Limitations and Outlook

The current model is calibrated using a limited set of deposits; incorporating additional localities will likely introduce greater spectral variability, necessitating model retraining or, at minimum, site-specific recalibration. Broader datasets may also increase class overlap, which could complicate cluster separation and reduce classification accuracy. Furthermore, heating above ~800 °C has been shown to broaden the same Raman bands (Saeseaw *et al*., 2009) used for provenance discrimination, potentially obscuring geographic signals. As a result, Raman-based FWHM classification should be integrated with complementary techniques—such as trace-element analysis—to disentangle provenance from heat treatment effects and establish a more robust, multimodal framework for origin determination.

## References:


- Chankhantha, C., Amphon, R., Rehman, H.U., and Shen, A.H. 2020. Characterisation of Pink-to-Red Spinel from Four Important Localities. Journal of Gemmology, 37(4).
- Chauviré, B., Rondeau, B., Fritsch, E., Ressigeac, P., and Devidal, J.L. 2015. Blue spinel from the Luc Yen District of Vietnam. Gems & Gemology, 51(1).
- Furuya, M. 2023. Cobalt spinel from Bai Buoi mine, Yen Bai, Vietnam. 37th IGC2023, Tokyo, Japan
- Krzemnicki, M.S., Leuenberger, A. and Walter, A.B. 2023. Cobalt-bearing Blue Spinel from Lukande, near Mahenge, Tanzania. Journal of Gemmology, 38(5).
- Malavasi, L., Galinetto, P., Mozzati, M.C., Azzoni, C.B., & Flor, G. 2002. Raman spectroscopy of AMn2O4 (A= Mn, Mg and Zn) spinels. Physical Chemistry Chemical Physics, 4(15).
- Saeseaw, S., Wang, W., Scarratt, K., Emmett, J.L., and Douthit, T.R. 2009. Distinguishing heated spinels from unheated natural spinels and from synthetic spinels. GIA Res. News, 13.
- Wang, C., Shen, A.H., and Liu, Y. 2020. Characterization of order-disorder transition in MgAl2O4: $Cr^{3+}$ spinel using photoluminescence. Journal of Luminescence, 227, 117552.
- Wu, J., Sun, X., Ma, H., Ning, P., Tang, N., Ding, T., Huihuang, L., Zhang, T., and Ma, Y. 2023. Purple-Violet Gem Spinel from Tanzania and Myanmar: Inclusion, Spectroscopy, Chemistry, and Color. Minerals, 13(2), 226.


## Acknowledgments

The authors sincerely acknowledge the support provided by the Gem and Jewelry Institute of Thailand (GIT) and its staff throughout the course of this research. Hub of Talents in Gem and Jewelry funded from National Research Council of Thailand is gratefully acknowledged.