An implementation of machine learning in ruby and sapphire origin determination

Montira Seneewong-Na-Ayutthaya[1], Bhuwadol Wanthanachaisaeng[1], Waratchanok Suwanmanee, Thanapong Lhuaamporn[1], Pimlapat Kamkeaw[1], Sarun Phibanchon[2] Tasnara Sripoonjan[3], and Thanong Leelawatanasuk[1]

[1]The Gem and Jewelry Institute of Thailand (Public Organization), ITF-Tower Building,
Silom Road, Suriyawong, Bangrak, Bangkok, 10500 Thailand
[2]Faculty of Education, Burapha University, Chonburi, 20131 Thailand
[3]G-ID Laboratories in Yannawa, Bangkok, 10120 Thailand
smontira@git.or.th

Introduction

Country-of-origin is one of the most important value factors of corundum (ruby and sapphire) due mainly to its historical or legendary popularity in the trade. Generally, gemmological laboratories in the early days determined ruby and sapphire origins simply by relying mostly on diagnostic inclusions scenes because those stones were supplied globally from only a few well-known important sources (Groat et al., 2019). However, as more and more stones have been sourced from many more geographical origins, frontline laboratories have to apply additional scientific approaches, such as spectroscopic and chemical data, to be able to well distinguish the different geological and geographical origins based on such data. (Sutherland et al., 2015; Harlow et al., 2013). Among those approaches, chemical fingerprinting or trace-element data discrimination is the most well-known technique and has become increasingly more reliable for determining the origin of the stones (Guillong & Günther 2001; Abduriyim and Kitawaki, 2006; Karampelas et al. 2019; Palke et al. 2019; Seneewong-Na-Ayutthaya et al., 2021). In addition, machine learning algorithms, a branch of artificial intelligence (AI), have been initially applied to classify the chemical database and assist the country-of-origin determination of stones (Wang & Krzemnicki 2021) For this study, well-documented samples of rubies and sapphires from selected gem-deposits were measured for their trace element contents by both EDXRF and LA-ICP-MS techniques and grouped in the form of a database of each gem deposit. The discrimination of the trace element data to evaluate the country-of-origin was carried out by 3D scattering plots and our self-developed machine learning program using our preselected samples in this work.
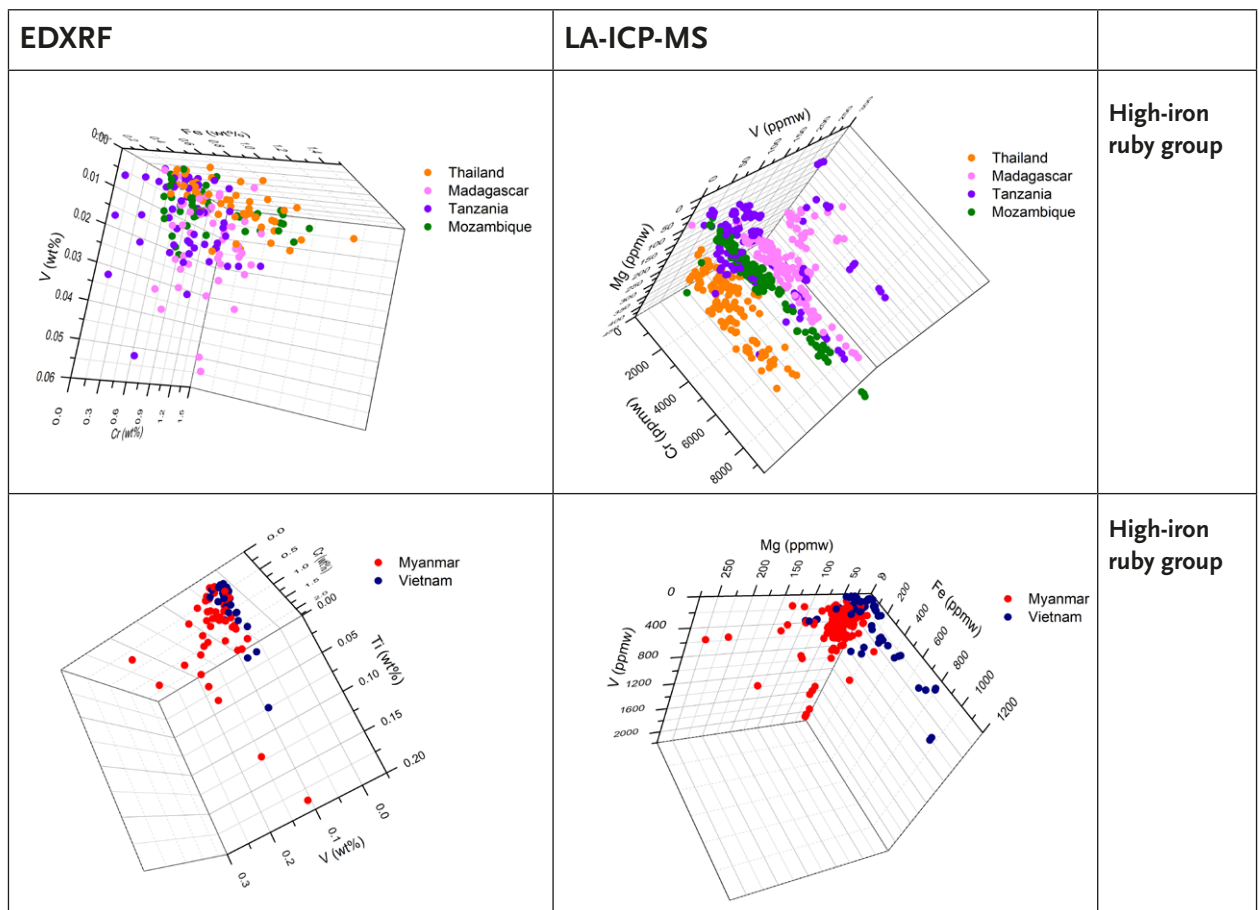
Materials and Methods

The first group of stones consists of 291 ruby samples from Myanmar (50), Vietnam (27), Thailand (50), Mozambique (51), Madagascar (Votomandry) (53), Tanzania (Winza, Mahenge, and Morogoro) (50), and test-samples (10), ranging from 0.11 to 7.21 ct. The second group comprises 290 sapphire samples from Sri Lanka (50), Myanmar (50), Madagascar (54), Thailand (51), Nigeria (50), Australia (25), and test-samples (10), ranging from 0.10 to 4.92 ct. For chemical analyses we used EDXRF (Eagle III XPL model, EDAX) and LA-ICP-MS (ThermoScientific iCAP RQ series ICP-MS coupled with a New Wave NWR-213 Nd:YAG laser) to analyze major, minor, and trace element concentrations. To evaluate the consistency of the data and to manually discriminate the country-of-origin, 3D scattering plots were constructed to depict the clustered concentration of the trace elements. Furthermore, we used a self-developed supervised machine learning program (named "AI for Gem Origin Determination"), which was pioneered and developed by Wanthanachaisaeng et al. (2020) as part of the research mission of the Gem and Jewelry Institute of Thailand (GIT). Learning algorithms consisted of K Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neuron Networks (ANN) chosen to assess the prediction accuracy of stone's country-of-origin through the program.

3D scattering plots of trace element contents

For this study, we divided ruby and blue sapphire samples into two groups based on their iron content, i.e., high-iron and low-iron groups (Figure 1). 3D scatter plots allowed us to gain a more detailed insight into data overlapping. For our selected samples, we found that Mg, Cr, Fe, Ga, Ti, and V were good discriminators for distinguishing corundum's country-of-origin as the selected features. In comparison, however, LA-ICP-MS data enabled a better origin discrimination than by using EDXRF data. We assume that this is due to the difference in the detection limit and quantitative accuracy of these two analytical methods. The manual rotation of plotting axes and the overlapping of some origin-related data groups proved to be a weakness in determining the origin using 3D scatter plots. To address this, machine learning methods were applied in this study.
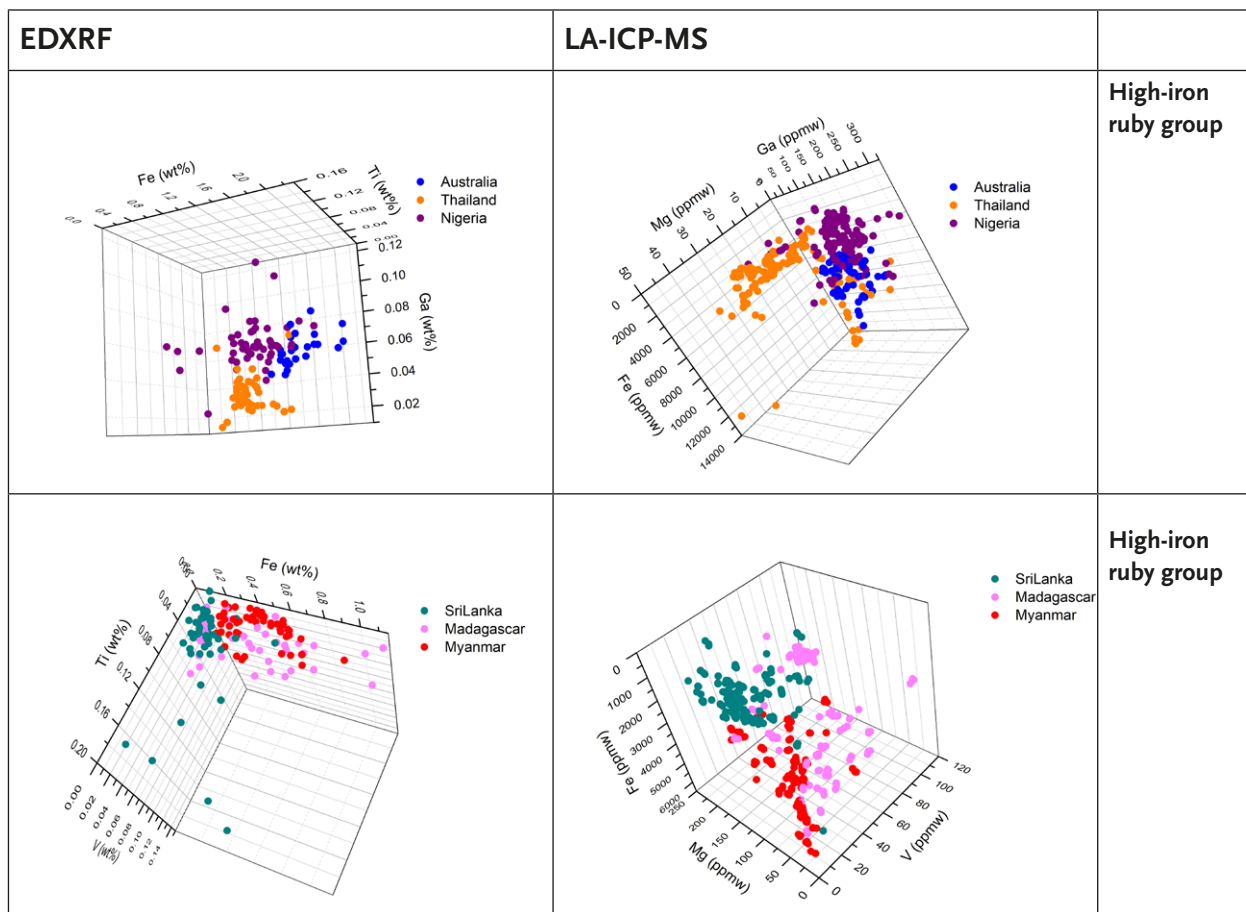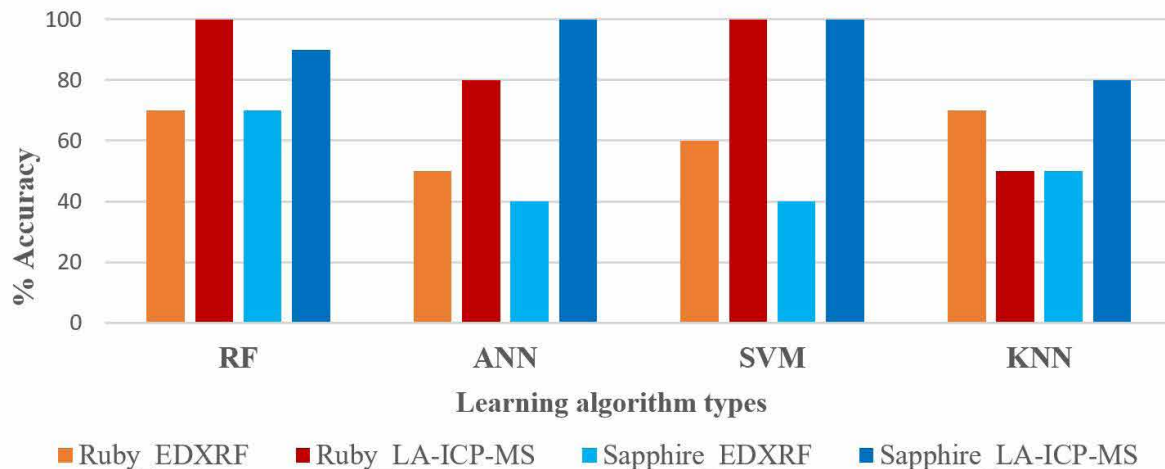
EDXRF	LA-ICP-MS	
		High-iron ruby group
		High-iron ruby group

EDXRF	LA-ICP-MS	
		High-iron ruby group
		High-iron ruby group

Figure 1: Representative 3D scatter plots of trace elements in ruby and blue sapphires from important sources

Machine learning algorithms for Gem Origin Determination

Machine learning algorithms were used in our study to attribute corundum origins, specifically in cases in which the 3D scatter plots revealed overlapping data groups. We found that our trained model was able to produce a rather high accuracy for origin determination if using LA-ICP-MS data. For example, it was able to achieve an average accuracy of 82.5% for ruby and 92.5% for blue sapphires, whereas the EDXRF only yielded 62.5% and 52.5% for rubies and blue sapphires, respectively (Figure 2). Briefly, the results revealed that the Random Forest Algorithm (RF) demonstrated the best overall learning algorithm for predicting the origin of the selected stones used in the study. The ANN and SVM algorithms, nonetheless, could give reasonable results when the input data were obtained from LA-ICP-MS analyses.

Concluding remarks

Trace element data fingerprinting is an essential method for studying diagnostic chemical features, especially when combined with machine learning methods and suitable algorithms and selected features. They can effectively help to determine the geographical origin of ruby and sapphire with a low level of error, especially when using LA-ICP-MS data. However, their prediction accuracy and determination success depend on various factors, such as instrumentation performance and limitation, data preparation and processing, model optimization, and validation. Importantly, a sufficient number, reliability, and homogeneity of gemstone samples are the most important key factors to maintaining the accuracy of origin determination by using the chemical database with machine learning methods, consistent application of analytical protocols during sample measure-

ment, and cautious handling of any outliers present in the dataset. Nonetheless, a gemologist is still the one who will decide the final result from considering various analytical data as well, including internal features and others spectral analyses.

Figure 2: The accuracy of origin determination was assessed by applying our program to analyze 20 test-samples of rubies and sapphires.

References

• Abduriyim, A., Kitawaki, H., 2006. Determination of the origin of blue sapphire using Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS), The Journal of Gemmology, 30(1/2), 23-36

• Groat, L.A., Giuliani, G., Stone-Sundberg, J., Renfro, N.D., Sun, Z., 2019. A review of analytical methods used in geographic origin determination of gemstones, Gems & Gemology, 55(4), 512–535.

• Guillong, M. & Detlef, G. 2001. Quasi 'non-destructive' laser ablation-inductively coupled plasma-mass spectrometry fingerprinting of sapphires, Spectrochimica Acta Part B Atomic Spectroscopy, 56(7), 1219-1231.

• Gubelin Gem Lab. 2020. Gemtelligence. Source: https://www.gubelingemlab.com/ en/news/detail/gemtelligence [05 January 2021]

• Harlow, G., Bender, W.M., 2013. A study of ruby (corundum) compositions from the Mogok Belt, Myanmar: Searching for chemical fingerprints, American Mineralogist, 98(7), 1120-1132

• Karampelas, S., Al-Shaybani, B., Mohamed, F., Sangsawong, S., Al-Alawi, A., 2019. Emeralds from the most important occurrences: chemical and spectroscopic data, Minerals, 9(9), 561.

• Palk A.C., Saeseaw S., Renfro N.D., Sun Z., Mcclure S.F., 2019. Geographic origin determination of ruby, Gems & Gemology, 55(4), 580–612.

• Sutherland, F.L., Zaw, K., Meffre, S., Yui, T.F., Thu, K., 2015. Advances in trace element "fingerprinting" of gem corundum, ruby and sapphire, Mogok area, Myanmar, Minerals, 5(1), 61-79

• Seneewong-Na-Ayutthaya, M., Chongraktrakul, W., Sripoonjan, T., 2021. Gemological characterization of peridot from Pyaung-Gaung in Mogok, Myanmar, Gems & Gemology, 57(4), 318–337

• Wanthanachaisaeng, B., Sripoonjan, T., Rungruengphol, A. 2020. Smart database of world gemstone origins (in Thai). The Gem and Jewelry Institute of Thailand (GIT), Bangkok, 105pp.

Acknowledgements