EXPLAINABLE AI-ASSISTED RAMAN MICROSPECTROSCOPY FOR RELIABLE CLASSIFICATION OF BREAST CANCER CELLS

Hemanth Noothalapati[1,2], Ajinkya Anjikar[3] and Tatsuyuki Yamamoto[1]

[1] Faculty of Life and Environmental Sciences, Shimane University, Japan
[2] Department of Chemical Engineering, Indian Institute of Technology Hyderabad, India
[3] The United Graduate School of Agricultural Sciences, Tottori University, Japan
E-mail: *nvhnag@life.shimane-u.ac.jp*

Reliable and early diagnosis of cancer remains a cornerstone challenge in healthcare. While histopathology is considered the gold standard, it is invasive, time-consuming and requires highly skilled pathologists. In contrast, Raman microspectroscopy provides a label-free, non-invasive alternative capable of detecting subtle biomolecular changes at the single-cell level. When combined with artificial intelligence (AI), Raman spectroscopy has shown promising results in automated cell classification tasks. However, despite high accuracy, most AI models lack interpretability making them unsuitable for real-world diagnostics where decisions must be explainable and clinically meaningful.

In this study, we propose a novel framework combining Raman spectroscopy with explainable AI (XAI) to address this limitation. Raman spectra were collected from normal human mammary epithelial cells (HMEpC) and breast cancer cells (MCF-7) using a laboratory built confocal Raman microscope with 632.8 nm excitation. Averaged Raman spectra obtained from 30 independent MCF-7 and HMEpC cells are shown in Figure 1A where marked difference in ratios of DNA/Lipids (Figure 1B) and Protein/Lipids (Figure 1C) were observed. Further, we employed advanced multivariate statistics and machine learning models such as Principal Component Analysis, Linear Discriminant Analysis, Neural Networks and Support Vector Machines (SVM) to discriminate cancer cells successfully. However, none of the popular methods give insights into the molecular basis for discrimination.

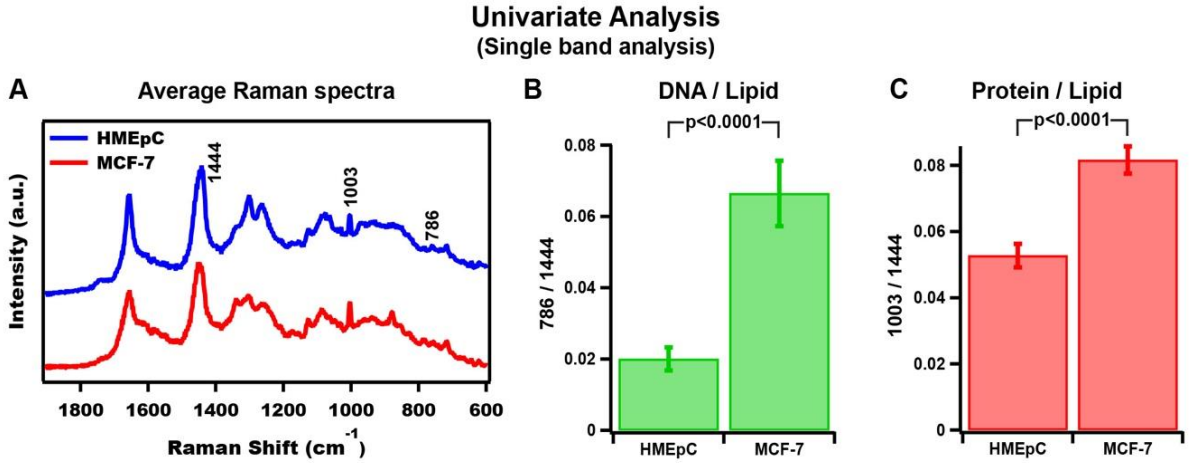

Figure 1. *Average Raman spectra of HMEpC and MCF-7 cells (A), intensity ratios of DNA/Lipid (B) and Protein/Lipid (C) with corresponding 'p' values.*

To address this, first a Feed Forward Neural Network (FFNN) was constructed to improve model robustness. We achieved over 98% classification accuracy. More importantly, we applied SHapley Additive exPlanations (SHAP), a game-theory-based XAI approach, to interpret the contribution of each Raman wavenumber to the model's predictions. SHAP analysis provided an important features highlighting the biomolecular origins of classification decisions. Key vibrational modes related to DNA, proteins and lipids were consistently identified as discriminative features. Interestingly, the analysis also revealed how certain models, despite high accuracy, relied on non-chemical background signals. This emphasizes that accuracy alone is not sufficient and model interpretability with chemical validity must also be assessed to ensure reliability.

We further demonstrated the generalizability of our XAI framework by applying the same SHAP-based interpretability to other machine learning models such as SVM classifier. Both FFNN and SVM models showed agreement in identified key spectral regions, affirming the robustness of the method. Our framework also enabled post hoc analysis of misclassifications, revealing molecular-level reasons such as heterogeneous expression of key biomolecules within individual cells. We believe, our work marks a significant step toward demystifying the "black-box" nature of AI in spectral diagnostics. Unlike prior approaches that rely on dimensionality reduction or handcrafted feature zones, our method considers the entire Raman spectrum and provides explicit, per-feature explanations that align with known biochemical pathways involved in cancer development. By incorporating domain knowledge into AI model interpretation, we enhance both scientific understanding and clinical trust.

In summary, we introduce a feature-aware model selection strategy that combines high-performance AI with full spectral interpretability. This strategy not only improves classification robustness but also serves as a new paradigm for building trustworthy AI-driven diagnostic tools in vibrational spectroscopy. The approach is scalable to other spectral data types and disease models, supporting the broader goal of making AI models safe, transparent and effective in healthcare applications.

References

1) K. Iwasaki, A. Araki, C.M. Krishna, R. Maruyama, T. Yamamoto and H. Noothalapati, Int. J. Mol. Sci. 22, 800, 1-14 (2021).