Thesis progression

Thesis title: Comparative feasibility of plasma, serum, and other clinical specimens for Tuberculosis diagnosis using Raman spectroscopy and machine learning

Thesis progression title: Optimization of serum and plasma concentration in sample preparation step using Raman spectroscopy and multivariate analysis

Student: Mr. Patipan Wongboothong **Student ID**: 677070005-9

Advisor: Prof. Dr. Kiatichai Faksri

Date: 24th December 2025

---

1. Introduction

Tuberculosis (TB), caused by the airborne transmission of *Mycobacterium tuberculosis* (Mtb), remains the most persistent and concerning infectious diseases, and is a major global health crisis. With approximately 1.25 million deaths and 8.2 million new cases reported in 2023 alone, its impact is undeniable (World Health Organization, 2023). A primary obstacle to its control is a fundamental screening bottleneck that hinders effective treatment and prevention. This challenge is compounded by the vast reservoir of latent TB infection (LTBI), estimated by the World Health Organization (WHO), about one-quarter of the global population is infected. LTBI are non-contagious, but if left untreated, 5-10% of them will progress to active TB (ATB) during their lifetime and LTBI also requires distinct clinical management from ATB (Comstock et al., 1974; *Tuberculosis (TB)*, n.d.).

However, current diagnostics are full of limitations. Conventional methods for ATB, such as sputum smear microscopy and culture, are slow and require significant laboratory infrastructure, while molecular tests remain costly for widespread use (4). The primary specimen for ATB diagnosis, sputum, can also be difficult to obtain from certain populations, such as children and paucibacillary patients (5). For LTBI, a definitive gold standard is absent; immunological assays cannot reliably differentiate LTBI from ATB or predict which individuals will progress to disease, leaving clinicians without a clear path forward and crippling public health efforts to halt transmission (6). Furthermore, current LTBI diagnostics, such as interferon-gamma release assays (IGRAs), typically rely on venous blood draws, which present logistical challenges for mass screening in community or remote settings, variable sensitivity, poor reproducibility, and unknown prognostic value (7,8). The limitations of these current diagnostic approaches are linked to the specimens they require.

The development of rapid, accurate, and accessible TB tests depends not only on new analytical technology but also on its successful application to more patient-friendly and minimally invasive sample types. The lack of a validated diagnostic platform for such alternative specimens remains a significant barrier to realize Point-of-Care Testing (POCT) and scalable mass screening strategies essential for achieving the WHO's End TB goals.

To overcome these challenges, vibrational spectroscopy has emerged as a powerful candidate technology. Specifically, Surface-Enhanced Raman Spectroscopy (SERS) offers a potential solution by capturing a unique 'biomolecular fingerprint' from a clinical sample. By enhancing the inherently weak Raman signal from matrix complex biological specimens. SERS provides a rapid, label-free analysis of the systemic biochemical changes caused by TB (Chang et al., 2025; Fang et al., 2024). When coupled with machine learning algorithms capable of discerning disease-specific patterns inside the data that are invisible to the human eye, this approach holds the potential for a highly accurate screening tool (Fang et al., 2024; Puravankara et al., 2024). There are studies that have demonstrated the promise of SERS for identifying TB, cancers, and other infectious diseases like typhoid and dengue (Eiamchai et al., 2024; Khalil et al., 2025; Lin et al., 2025). However, a critical knowledge gap persists: the comparative utility of this technology across different, more accessible clinical specimens for TB has not been systematically investigated.

While the combination of SERS and machine learning presents a promise for TB test, its practical implementation depends on answering one key question: which clinical specimen offers the optimal balance of accuracy, accessibility, and patient convenience? This thesis confronts this obstacle directly by systematically evaluating the feasibility of using this platform across a range of biofluids. The successful identification of a suitable alternative specimen would be a critical step toward developing a field-deployable, point-of-care tool for mass screening and effective TB control. Therefore, the primary aim of this research is to systematically evaluate and compare the performance of venous plasma, venous serum, fingertip serum, saliva, and tongue swabs for the differentiation of ATB, LTBI and HC groups using SERS coupled with machine learning.

2. Objective

2.1. To establish the optimal venous blood matrix by comparing the diagnostic performance of serum and plasma with SERS and machine learning to distinguish between TB groups.

2.2. To evaluate the feasibility of point-of-care translation by comparing fingertip serum to venous serum with SERS and machine learning to distinguish between TB groups.

2.3. To investigate the viability of screening by evaluating saliva and tongue swab samples with SERS and machine learning to distinguish between TB groups.

3. Conceptual Framework

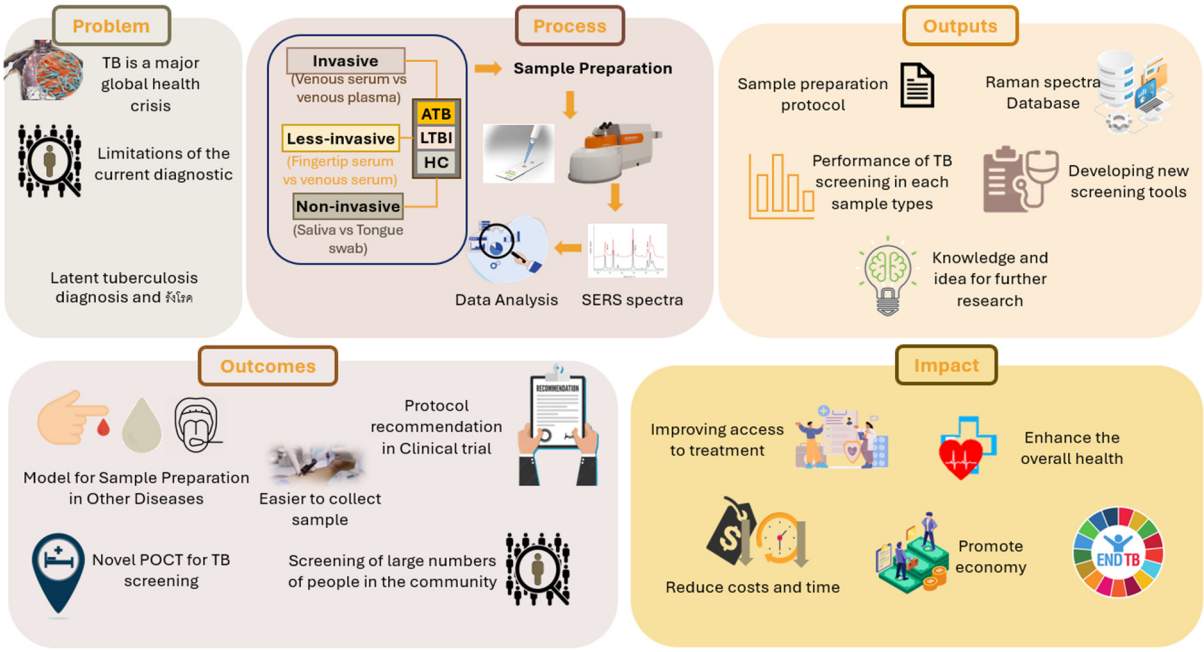

Figure 1. Conceptual framework

From **Error! Reference source not found.** TB is the second most concerning infectious disease globally, posing a significant health problem. The current lack of widespread screening protocols, particularly in communities, and the invasive nature of traditional venous blood collection for diagnosis are major challenges. This project aims to address these issues by evaluating the feasibility of alternative specimens: less-invasive (fingertip serum) and non-invasive (saliva and tongue swab) sample types, alongside comparing the performance of venous serum to venous plasma by using SERS and machine learning to develop improved TB screening protocols and tools. The goal is to enhance access to treatment, improve public health, reduce healthcare costs, and promote economic well-being, contributing to the "END TB" initiative.

4. Materials and methods

4.1. Overview and study design

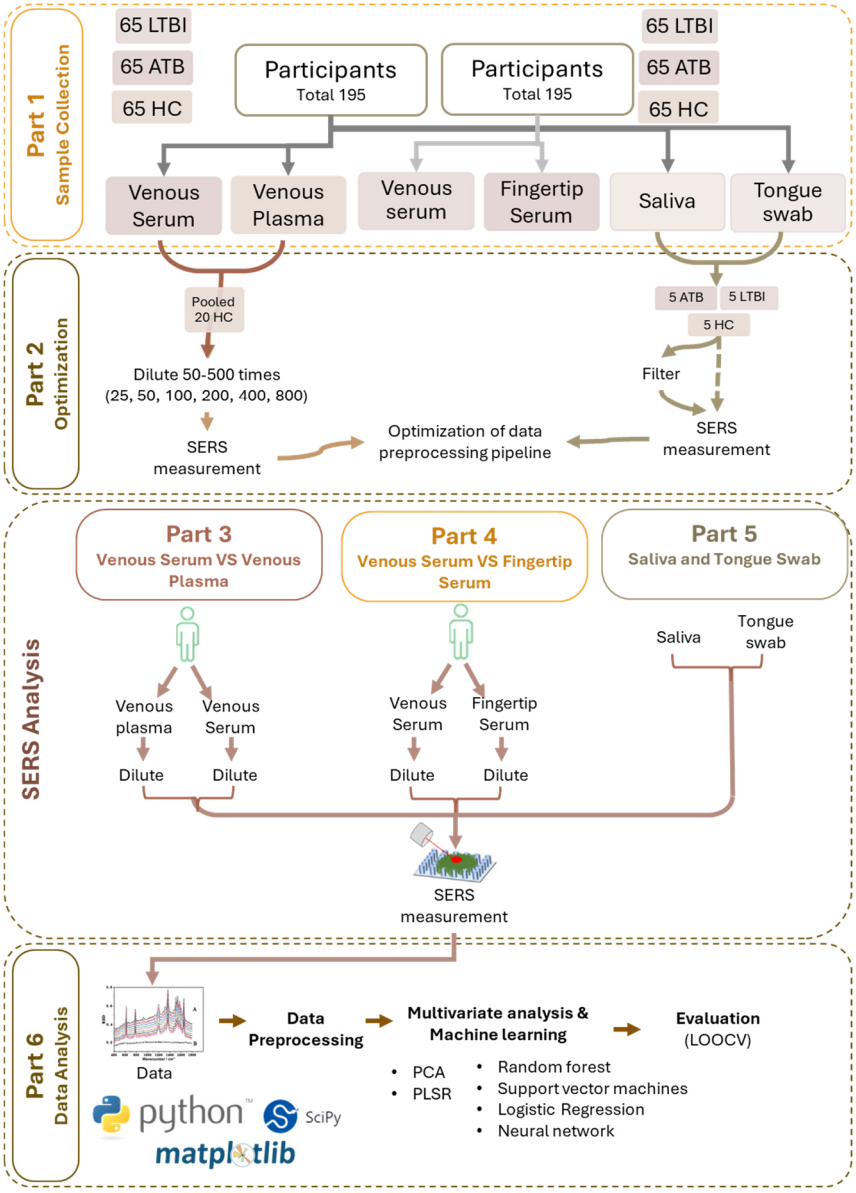

Figure 2. Study Design

This study aimed to evaluate and compare different biological sample types for detecting tuberculosis infection using SERS and machine learning. A total of 195 participants were divided into three groups: 65 active tuberculosis (ATB), 65 latent tuberculosis infection (LTBI), and 65 healthy controls (HC). To achieve this, we will systematically evaluate the efficacy of machine learning models trained on SERS spectra from different sample types: venous serum, venous plasma, fingertip serum, saliva, and tongue. The ultimate objective is to provide clear, data-driven guidance on the optimal specimen, balancing high performance with clinical practicality and patient invasiveness for the potential development of a future rapid SERS-based tool for TB screening.

Multiple biological specimens were collected, including venous serum, plasma, fingertip serum, saliva, and tongue swabs. Serum, plasma, and fingertip serum were stored at -80°C, while saliva and tongue swabs were preserved with formalin before storage.

4.2. Participants and setting

A total of 975 samples, including venous serum, venous plasma, fingertip serum, saliva, and tongue swabs, will be collected from 65 active tuberculosis (ATB), 65 latent tuberculosis infection (LTBI), and 65 healthy control (HC) participants at Srinagarind Hospital, Khon Kean Hospital, Kalasin Hospital, Mahasarakham Hospital, Roiet Hospital, and Chum Phae Hospital, Thailand.

Medical records information will be collected from participants. All participants' blood will be sent to the Office of Disease Prevention and Control, Provincial Agencies 7 for LTBI, and National Healthcare Systems Company Limited (N Health) for ATB and HC to test and obtain IGRA results.

Table 1. The inclusion and exclusion criteria of each group are in the table below.

Criteria	Inclusion	Exclusion
ATB	• Age 18-70 years. • Chronic cough for at least 2 weeks, or hemoptysis, or fever, loss of appetite, and unexplained weight loss. • Positive molecular test for TB (e.g., GeneXpert), regardless of AFB smear result. • Treatment duration within 0-14 days • Abnormal CXR	• Pregnancy. • Confirmed NTM or other lung infections. • Negative molecular test for TB and negative culture for MTB. • HIV infection
LTBI	• Age 18-70 years. • History of household contact or being a healthcare worker. • No history of having ATB. • Living with TB patients >15 hours per week or >180 hours during the 3 months before. • Positive IGRA and/or TST result. • Normal CXR < 1 month before.	• Pregnancy. • Confirmed NTM or other lung infections. • HIV infection
HC	• Age 18-70 years. • A healthy individual with no prior history of ATB.	

	• No history of close contact with TB patients. • Negative IGRA and/or TST result. • Normal CXR < 1 month before.	

4.3. Sample collection and storage

The specimens will be placed on ice and transported to the laboratory within 6 hours.

The heparinized blood and clotted blood will be centrifuged at 3500 rpm for 10 minutes to separate plasma and serum, respectively. They will then be aliquoted and stored at -80°C.

For saliva, participants will first be instructed to rinse their mouths. After 15 minutes, they will passively drool saliva in their mouths for 1 minute before allowing 3-5 mL of saliva to flow into a designated sterile collection container. To prevent bacterial growth after collection, 5 mL of 0.5% formalin will be added. Finally, the sample will be transferred into a cryotube and stored at -80°C.

For tongue swabs, use a sterile flocked swab to firmly press the dorsal surface of the tongue for 5 seconds, then swab from left to right 3 times. The swab will be placed into a sterile tube containing 3 ml of 0.1% formalin to prevent bacteria from growing. The preservative 3 mL of 0.1% formalin will be added to the solution; the swab will be discarded after mixing, and the solution will then be stored at -80°C.

4.4. Optimization of preprocessing step of serum and plasma dilution

To identify the ideal analyte concentration for SERS analysis, venous serum and venous plasma of healthy control group from 20 pooled sample were prepared across a range of dilutions, with dilution factors varying from 1:25 to 1:800. This optimization is critical as SERS intensity is highly dependent on analyte concentration at the nanoparticle surface, where overly concentrated samples can lead to signal quenching and overly dilute samples yield insufficient signals. Following this comprehensive analysis, the dilution factor that provided the best balance of high intensity, consistency, and reliability was selected. This optimized dilution factor, determined for venous serum, was subsequently applied to fingertip serum samples, assuming comparable proteomic and metabolomic concentrations.

4.5. Data visualization

All SERS spectra obtained from the confocal Raman spectroscopic microscope was conducted in a structured pipeline designed to extract meaningful biological insights and build robust classification models. The workflow was executed by using the Python

programming language (version 3.9 or higher). Key scientific libraries that will be employed include Scikit-learn for machine learning and model validation, Pandas and NumPy for data manipulation, SciPy for statistical analyses, Raman Spy for preprocessing, and Matplotlib and Seaborn for data visualization.

Upon the importation, all SERS spectra was preprocessed by using the optimized pipeline established in Section 3.2.5.3. Following preprocessing, the reproducibility of the data will be quantitatively confirmed by calculating RSD across spectra within the sample and across all samples within the same group. PCA and UMAP were used for visualization of the clustering and differentiating of the inherent structure of the data.

5. Results and Conclusion

The optimization of sample concentration is a critical prerequisite for quantitative Surface-Enhanced Raman Scattering (SERS) analysis, as it directly influences signal stability and measurement repeatability. As illustrated in Figures 3 and 4, the SERS spectra of serum and plasma derived from healthy controls were acquired across a gradient of concentrations. The spectral profiles of serum and plasma appeared remarkably similar, reflecting their overlapping biochemical compositions, primarily differing in the presence of clotting factors such as fibrinogen in plasma (Baker et al., 2014).

To quantitatively assess the reproducibility of the SERS signal across different dilution factors, the peak intensity at 1003 cm^{-1} was selected as the reference marker. This vibrational mode, predominantly attributed to the ring-breathing of phenylalanine, is consistently reported as the most intense and distinct feature in biofluid SERS spectra, making it an ideal candidate for normalization and stability calculations (Bonnier & Byrne, 2011; Feng et al., 2010).

The Relative Standard Deviation (RSD) of the peak intensity at 1003 cm^{-1} was calculated to measure precision. As depicted in Figure 5, the data reveals a non-linear relationship between concentration and signal stability. Specifically, samples prepared at a 100-fold dilution (1:100) exhibited the lowest %RSD values for both serum and plasma compared to other dilution ratios. This suggests that at this specific concentration, the adsorption kinetics of bio-analytes onto the silver SERS substrate is suitable for dispersion and thickness of biofluid (Lu et al., 2023).

This finding is further corroborated by statistical variance analysis. A box plot distribution of average spectral intensities across the full wavelength range demonstrates that

the 1:100 dilution factor yields the narrowest interquartile range and the lowest standard deviation. This reduction in variance indicates a significant enhancement in the intra-assay reproducibility of the SERS spectra at this operational concentration.

Furthermore, multivariate analysis via Principal Component Analysis (PCA) provides a visualization of spectral consistency. As shown in Figures 7 and 8, the PCA score plots reveal distinct clustering behaviors associated with different dilution factors. The scores corresponding to the 100-fold dilution are the most tightly clustered, exhibiting minimal dispersion along the principal component axes. In chemometrics, tight clustering in PCA space is definitive evidence of high spectral homogeneity and low experimental variance(Jolliffe & Cadima, 2016; Stiles et al., 2008).

Collectively, the convergence of evidence from specific peak kinetics (1003 cm^{-1} RSD), global intensity variance (box plots), and multivariate clustering (PCA) establishes the 1:100 dilution as the optimal operational parameter. High spectral reproducibility and consistency are paramount for diagnostic reliability; ensuring that spectral differences are attributable to biological variation rather than technical artifacts is essential for the valid clinical application of SERS technology.

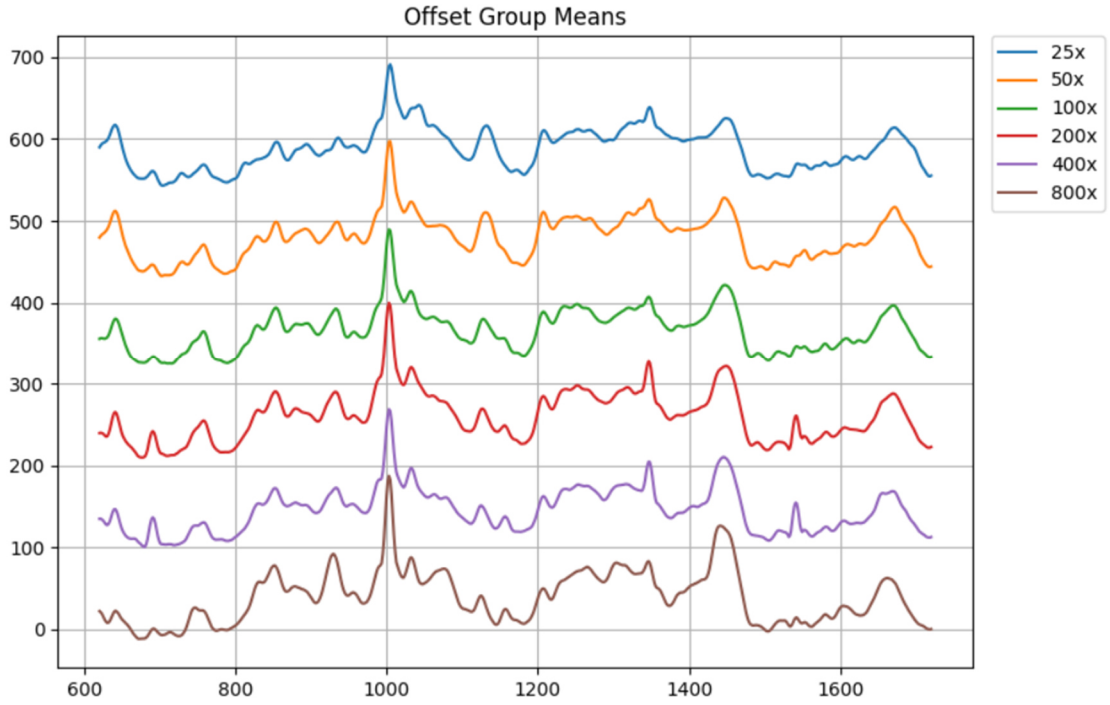

Figure 3 SERS spectra of serum samples with different dilutions (25, 50, 100, 200, 400, 800) obtained under 785 nm laser wavelength excitation.

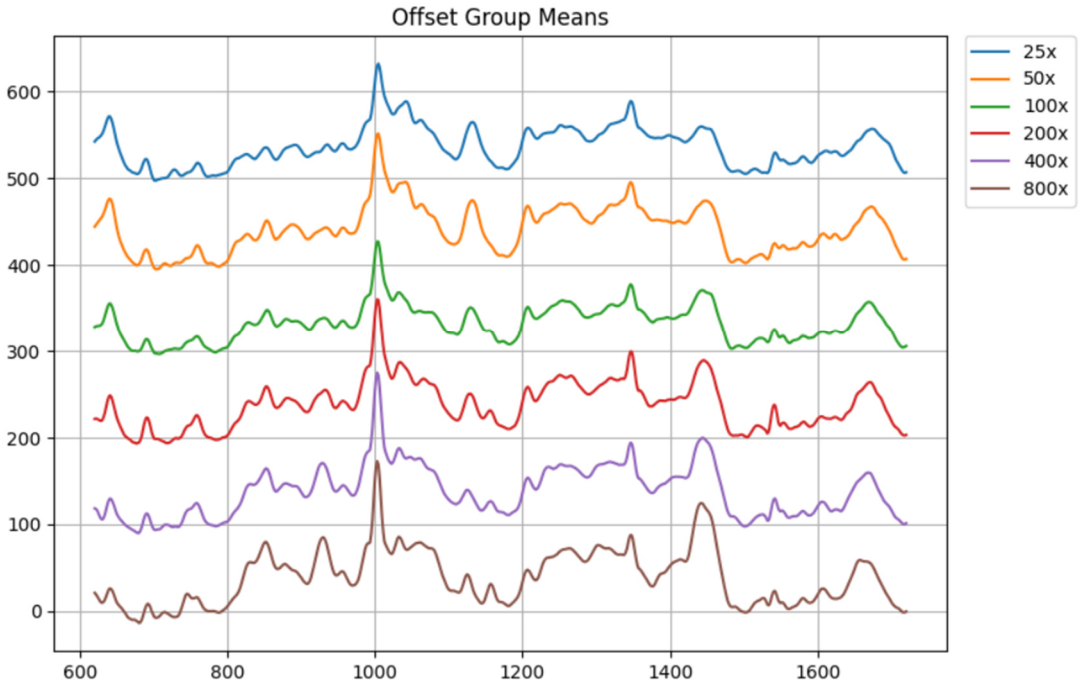

Figure 4. SERS spectra of plasma samples with different dilutions (25, 50, 100, 200, 400, 800) obtained under 785 nm laser wavelength excitation.

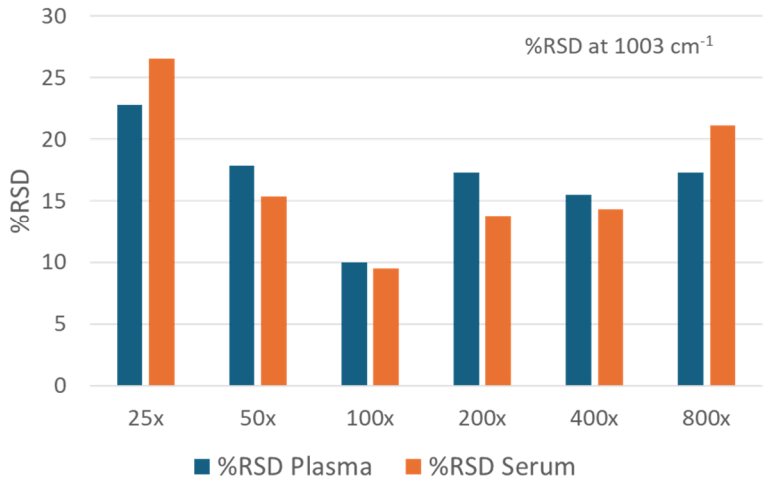

Figure 5. %RSD of plasma and serum at different concentrations.

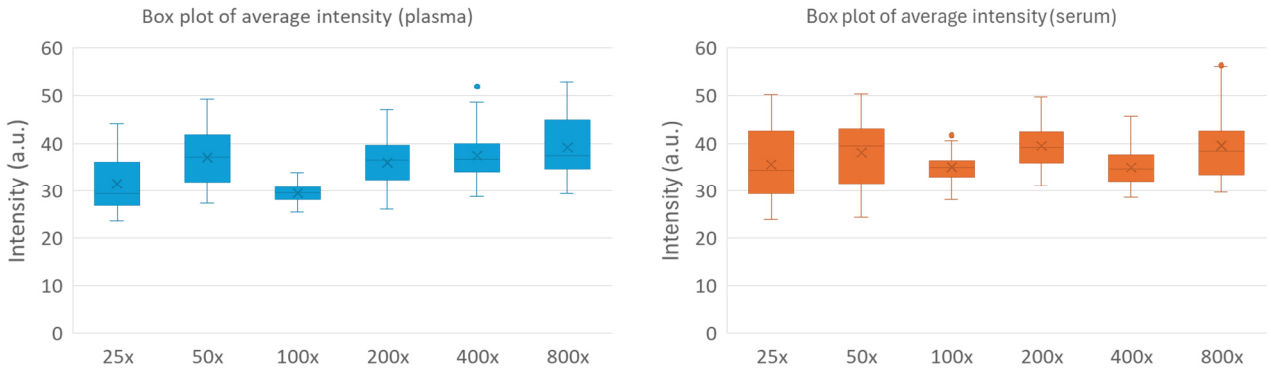

Figure 6. Box plot of average intensity of plasma (left) and serum (right) at different concentrations.

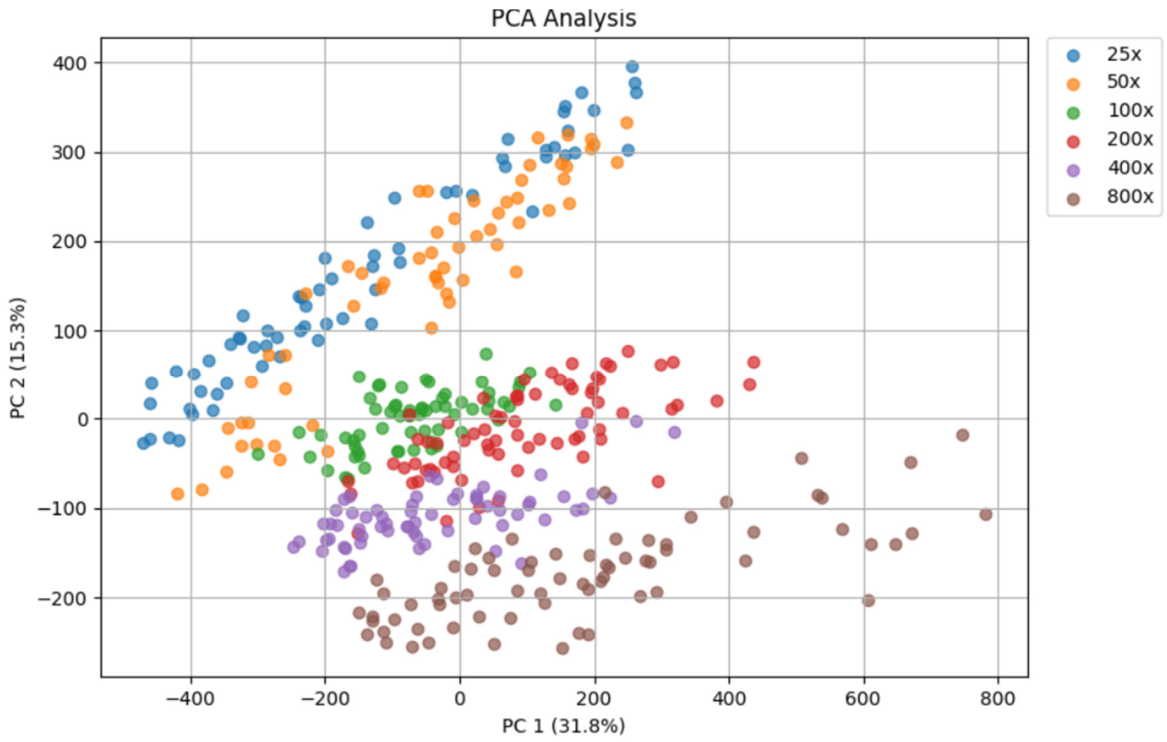

Figure 7 PCA of serum samples with different dilutions (25, 50, 100, 200, 400, 800) obtained under 785 nm laser wavelength excitation.

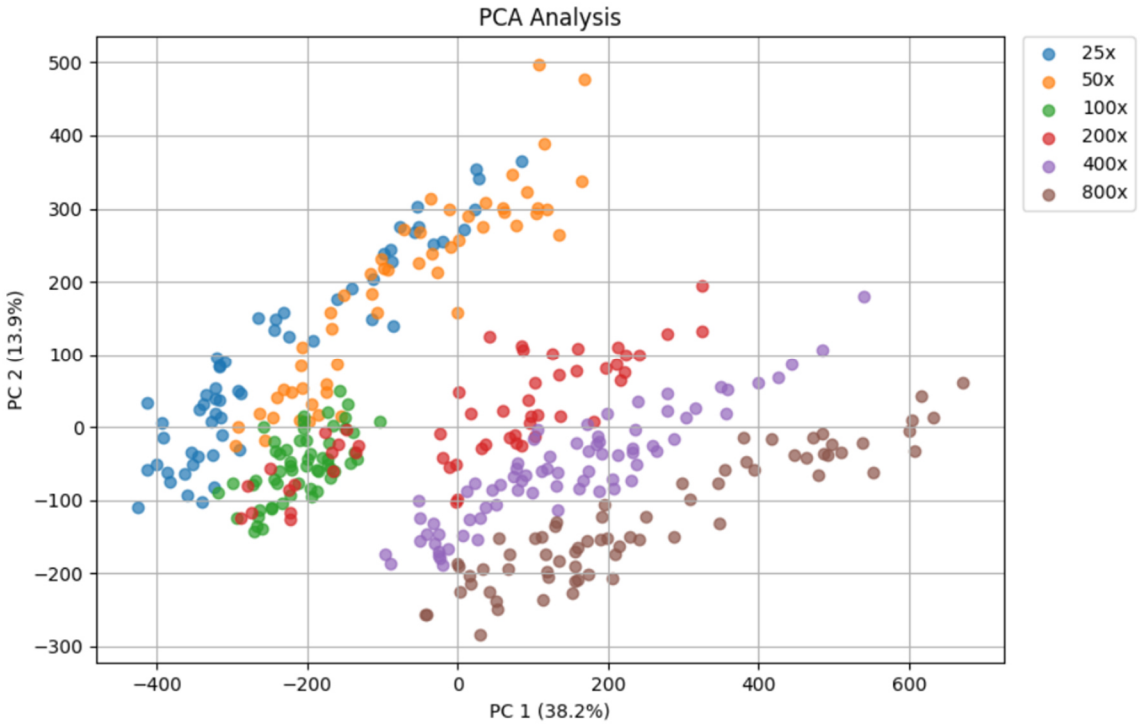

Figure 8 PCA of plasma samples with different dilutions (25, 50, 100, 200, 400, 800) obtained under 785 nm laser wavelength excitation.

6. References

Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., Fielden, P. R., Fogarty, S. W., Fullwood, N. J., Heys, K. A., Hughes, C., Lasch, P., Martin-Hirsch, P. L., Obinaju, B., Sockalingum, G. D., Sulé-Suso, J., Strong, R. J., Walsh, M. J., Wood, B. R., … Martin, F. L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols*, *9*(8), 1771–1791. https://doi.org/10.1038/nprot.2014.110

Bonnier, F., & Byrne, H. J. (2011). Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst*, *137*(2), 322–332. https://doi.org/10.1039/C1AN15821J

Chang, H., Hur, W., Kang, H., & Jun, B.-H. (2025). In vivo surface-enhanced Raman scattering techniques: Nanoprobes, instrumentation, and applications. *Light: Science & Applications*, *14*(1), 79. https://doi.org/10.1038/s41377-024-01718-5

Comstock, G. W., Livesay, V. T., & Woolpert, S. F. (1974). THE PROGNOSIS OF A POSITIVE TUBERCULIN REACTION IN CHILDHOOD AND ADOLESCENCE.

American Journal of Epidemiology, *99*(2), 131–138. https://doi.org/10.1093/oxfordjournals.aje.a121593

Eiamchai, P., Juntagran, C., Somboonsaksri, P., Waiwijit, U., Eisiri, J., Samarnjit, J., Kaewseekhao, B., Limwichean, S., Horprathum, M., Reechaipichitkul, W., Nuntawong, N., & Faksri, K. (2024). Determination of latent tuberculosis infection from plasma samples via label-free SERS sensors and machine learning. *Biosensors and Bioelectronics*, *250*, 116063. https://doi.org/10.1016/j.bios.2024.116063

Fang, S., Wu, S., Chen, Z., He, C., Lin, L. L., & Ye, J. (2024). Recent progress and applications of Raman spectrum denoising algorithms in chemical and biological analyses: A review. *TrAC Trends in Analytical Chemistry*, *172*, 117578. https://doi.org/10.1016/j.trac.2024.117578

Feng, S., Chen, R., Lin, J., Pan, J., Chen, G., Li, Y., Cheng, M., Huang, Z., Chen, J., & Zeng, H. (2010). Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis. *Biosensors & Bioelectronics*, *25*(11), 2414–2419. https://doi.org/10.1016/j.bios.2010.03.033

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Khalil, M. Z., Tariq, R., Majeed, M. I., Nawaz, H., Alwadie, N., Shafique, H., Ali, M., Liaqat, I., Khalid, F., Tahir, R., & Imran, M. (2025). SERS-Based Serological Profiling of Dengue and Typhoid Fever Advancing Infectious Disease Diagnostics Using PCA and PLS-DA. *Plasmonics*. https://doi.org/10.1007/s11468-025-02854-3

Lin, Y., Zhang, Q., Chen, H., Liu, S., Peng, K., Wang, X., Zhang, L., Huang, J., Yan, X., Lin, X., Hasan, U. M. D., Sarwara, M., Fu, F., Feng, S., & Wang, C. (2025). Multi-cancer early detection based on serum surface-enhanced Raman spectroscopy with deep learning: A large-scale case–control study. *BMC Medicine*, *23*(1), 97. https://doi.org/10.1186/s12916-025-03887-5

Lu, Y., Lei, B., Zhao, Q., Yang, X., Wei, Y., Xiao, T., Zhu, S., Ouyang, Y., Zhang, H., & Cai, W. (2023). Solid-State Au Nanocone Arrays Substrate for Reliable SERS Profiling of Serum for Disease Diagnosis. *ACS Omega*, *8*(32), 29836–29846. https://doi.org/10.1021/acsomega.3c04910

Puravankara, V., Manjeri, A., Ho Kim, Y., Kitahama, Y., Goda, K., Dwivedi, P. K., & George, S. D. (2024). Surface-Enhanced Raman spectroscopy for Point-of-Care

Bioanalysis: From lab to field. *Chemical Engineering Journal*, *498*, 155163. https://doi.org/10.1016/j.cej.2024.155163

Stiles, P. L., Dieringer, J. A., Shah, N. C., & Duyne, R. P. V. (2008). Surface-Enhanced Raman Spectroscopy. *Annual Review of Analytical Chemistry*, *1*(Volume 1, 2008), 601–626. https://doi.org/10.1146/annurev.anchem.1.031207.112814

Tuberculosis (TB). (n.d.). Retrieved July 31, 2025, from https://www.who.int/news-room/fact-sheets/detail/tuberculosis

World Health Organization. (2023). *Global tuberculosis report 2023*. World Health Organization; WHO IRIS. https://iris.who.int/handle/10665/373828