# SURFACE ENHANCED RAMAN SPECTROSCOPY FOR DIAGNOSTICS IN OCULAR HEALTH

**Aradhya Agnihotri**

Thesis submitted to the Faculty of Engineering
in partial fulfillment of requirements for the degree of

**Master of Applied Science**

in

Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering
University of Ottawa

# Abstract

This thesis focuses on the development of a Surface-Enhanced Raman Spectroscopy (SERS) assay for detecting biofilms in ocular health. Biofilms are complex communities of microorganisms encased in an extracellular polymeric substance (EPS) that present significant challenges for clinical diagnostics due to their resistance to antimicrobial treatments and their evasion of traditional detection techniques. By amplifying Raman signals through metal nanoparticles, SERS offers a promising solution for overcoming these challenges in detecting biofilm-associated infections in tissues such as the cornea.

Silver nanoparticles (AgNPs) were synthesized and optimized to interact with biofilms using a factorial design approach. Optimal conditions of 1.75x nanoparticle concentration and a 24-hour interaction time between the sample and the nanoparticles were identified. This combination produced the most robust signal, particularly for detecting pyocyanin, a key virulence factor in *Pseudomonas aeruginosa* (PA01) biofilms.

The novelty of this work is the application of this optimized protocol to pig cornea samples, which closely mimic human corneal composition. The SERS spectra from the corneal samples were consistent with earlier validation results we achieved from the collagen samples, with some expected differences due to the distinct structural composition of the cornea. Machine Learning (ML) techniques were integrated into the SERS data analysis to enhance the diagnostic power of the assay. By combining unsupervised methods like K-means clustering with supervised techniques such as Support Vector Machines (SVM), the model effectively differentiated biofilm-present from biofilm-absent samples, achieving an overall accuracy of 92.5%, specificity of 100%, and sensitivity of 80%. This integration of SERS and ML presents a significant advancement for real-time, minimally invasive diagnostics in ocular health.

This research establishes a strong foundation for using SERS in biofilm detection within ocular health, demonstrating the potential of an optimized SERS assay combined with ML techniques to improve early detection and treatment of ocular infections such as bacterial keratitis. However, challenges remain in managing biofilm composition variability and addressing spectral overlap with corneal tissue. Future work will focus on refining diagnostic algorithms to enhance accuracy and sensitivity, expanding the assay's scope to detect a broader range of bacterial infections, and advancing toward human clinical trials.

# Acknowledgements

I would like to thank everyone who has supported me throughout the course of my research and this thesis. First, I would like to thank my supervisor, Dr. Hanan Anis, for her invaluable support, guidance, continuous encouragement, and constructive feedback. Thank you so much for giving me this opportunity and for accepting me into your lab. I am grateful for your patience throughout my research journey. Your motivation pushed me to always give my best, and your guidance helped me navigate the challenges of this process.

I am also very grateful to Robert Hunter, who mentored me throughout this journey. Rob, your dedication to your work has always been a true inspiration for me. Thank you for always being available when I encountered challenges, and for sharing your knowledge and ideas with me. Your passion for research and learning has been a great source of encouragement and I am very thankful for your mentorship.

The work presented in this thesis was done in close collaboration with another research team, the BeATS laboratory, from the Ottawa Heart Institute. Thank you for providing the materials and research facilities I needed to fulfill my research goals. Their support was crucial to the success of this project.

To my colleagues, thank you for creating such a supportive and collaborative environment. I am grateful to have had the opportunity to work with such a great group of people. Thank you, Khaled Atia, Ebin Joseph, and Isaiah Brine, for helping me troubleshoot various problems and for offering invaluable suggestions and ideas. Your kindness and teamwork made this experience rewarding, and I deeply appreciate your friendship.

I am very grateful to my family for their unconditional love and support. Even from afar, I always felt their encouragement. I would also like to thank all my friends who stood by me during this research journey. A special thanks to them for helping me edit this thesis. A very heartfelt thank you goes to my partner, Nithin Chandra, for being a constant source of strength throughout this entire process. Your belief in me, your encouragement during the most difficult moments, and your endless patience and kindness have been my anchor. I cannot thank you enough for standing by my side and supporting me.

Finally, I would like to extend my sincere gratitude to the University of Ottawa's Graduate Studies, the Faculty of Engineering, and the funding agencies that supported this

research. Their financial support made this work possible, and I am deeply appreciative of their contributions.

# Table of Contents

# Table of Figures

# Table of Tables

# Glossary

**Ag:** Silver

**AgNPs:** Silver Nanoparticles

**airPLS:** Adaptive Iteratively Reweighted Penalized Least Squares

**ANOVA:** Analysis of Variance

**Au:** Gold

**AuNP:** Gold Nanoparticles

**CARS**: Coherent Anti-Stokes Raman Scattering

**CLSM:** Confocal Laser Scanning Microscopy

**CNNs:** Convolutional Neural Networks

**DIW:** Deionized Water

**DNA:** Deoxyribonucleic Acid

**ECG:** Electrocardiogram

**EPS:** Extracellular Polymeric Substance

**FN:** False Negative

**FP:** False Positive

**IR:** Infrared

**LB:** Lysogeny Broth

**LSD:** Least Square Difference

**LSPR:** Localized Surface Plasmon Resonance

**LSPs:** Localized Surface Plasmons

**ML:** Machine Learning

**MSE:** Mean Squared

**NA:** Numerical Aperture

**PA01:** Pseudomonas Aeruginosa

**PCA:** Principal Component Analysis

**PLS:** Partial Least Square

**R6G:** Rhodamine 6G

**RBF:** Radial Based Function

**RPi:** Raspberry Pi

**SEM:** Scanning Electron Microscopy

**SERS:** Surface Enhanced Raman Spectroscopy

**SOLIS:** Solution for Imaging and Spectroscopy

**SPR:** Surface Plasmon Resonance

**SVM:** Support Vector Machine

**TERS:** Tip-Enhanced Raman Spectroscopy

**TN:** True Negative

**TP:** True Positive

# Chapter 1

# Introduction

This chapter includes the thesis objectives and my personal contributions to the research work. We also talk about the overall structure of the thesis. This chapter sets the stage for a detailed exploration of Surface-Enhanced Raman Spectroscopy (SERS) as a diagnostic tool for biofilm detection.

## 1.1 Objective

Biofilms are clusters of microorganisms that stick to surfaces and are encased in a matrix to protect themselves from their environment [122]. The main objective of my thesis is to establish a reliable SERS assay for detecting these bacterial biofilms and helping with the diagnosis of ocular infections like bacterial keratitis. This work addresses a significant gap in the research by applying SERS to the detection of biofilms specifically in the context of ocular infections like bacterial keratitis, where existing methods are limited.

My research addresses the significant challenges of biofilms in ophthalmology, where their complex structure and resistance to treatment make detection critical. Traditional diagnostic techniques struggle in detecting biofilms accurately, leading to delayed treatments and poorer patient outcomes.

This gap in current research is addressed by performing two major experimental investigations. The first involved optimizing the existing SERS conditions to ensure consistent detection of biofilms in controlled laboratory settings. The second phase focused on validating these optimized conditions using pig cornea to simulate human ocular tissues. This validation is an important step in translating this technology from the lab to clinical practice.

## 1.2 Original Contributions

The original contributions of this thesis are summarized below:

- After joining the lab and familiarizing myself with the SERS setup, I used it to detect biofilms in ocular environments. I used pig cornea samples both with and without bacteria, to provide a model that closely mimics human corneal composition. This is the first targeted use of SERS for this purpose to my knowledge. My work involved optimizing two key experimental parameters—nanoparticle concentration and interaction time—to enhance the sensitivity and specificity of biofilm detection, particularly for *Pseudomonas Aeruginosa* (PA01) biofilms. My findings show the capabilities of SERS in ophthalmic diagnostics, supporting its future implementation in clinical practice.
- I have also applied a machine learning workflow that preprocesses and clusters the SERS biofilm-cornea data that I collected. After collecting the data, preprocessing techniques like baseline correction, Savitzky-Golay filter, and normalization were applied. Principal Component Analysis (PCA) was then performed for dimensionality reduction. After this, data was then clustered using k-means to account for within-sample variation. This step helped group similar subcultures within each sample. The clustered data was then classified using a SVM model, which differentiated the pig cornea from pig cornea with bacteria with 92.5% accuracy. This is the first time K-means has been used as a preprocessing technique before the classification of corneal samples, according to my literature review.

## 1.3 Thesis Outline

This thesis is organized into the following different chapters:

Chapter 2 introduces Raman spectroscopy and SERS covering the basic theory, limitations, and the role of nanoparticles in enhancing SERS signals. It distinguishes between intrinsic and extrinsic SERS configurations and outlines the SERS setup used in this research, emphasizing the validation processes to ensure reliable results.

Chapter 3 focuses on the optimization of the SERS experimental setup for biofilm detection. It details the experimental design and the selection of variables, including the synthesis of silver nanoparticles. The chapter then presents the results of these optimization experiments and discusses the identification of optimal conditions for effective biofilm detection.

Chapter 4 starts with a literature review on biofilms in biomedical and ocular health settings, highlighting the significant challenges and limitations of current detection methods. The chapter then introduces SERS as a viable solution to overcome these challenges and describes the initial validation experiments conducted to assess whether the optimized SERS conditions can successfully detect *P. aeruginosa* biofilms and collagen. The results from these validation experiments are discussed, confirming the effectiveness of the SERS and setting the stage for its application to corneal samples.

Chapter 5 explores the classification of bacteria absent and present corneal samples using ML techniques integrated with SERS data analysis. Building on the validated SERS detection of biofilms and collagen, this chapter reviews relevant ML applications in biomedical contexts and SERS data interpretation. It then describes the specific algorithms employed for corneal biofilm detection experiments. The results of these analyses are presented, demonstrating the successful classification of corneal samples. The chapter further evaluates the performance metrics of the model, highlighting improvements in accuracy and efficiency achieved through ML integration.

Chapter 6 summarizes the key findings, discusses their contributions to ocular health diagnostics, and outlines potential future research directions.

At the end of each chapter, challenges faced during experiments are discussed, providing a comprehensive understanding of the research process and the strategies employed to overcome obstacles.

# Chapter 2

# Raman Spectroscopy and Surface Enhanced Raman Spectroscopy (SERS)

## 2.1 Introduction

Raman Spectroscopy is a non-invasive technique based on vibrational spectroscopy commonly used to study molecular vibrational characteristics [2]. This technique works by analysing the inelastic scattering of light where a small fraction of the incident light interacts with a molecule and undergoes an energy shift, known as the Raman shift [1] [2]. This shift provides insights into the molecule's vibrational energy levels, offering a unique spectral fingerprint that reflects its chemical and structural properties.

In the early part of the 20th century, Grigorij Samuilovič Landsberg and Nobel laureate Chandrasekhara Venkata Raman independently discovered Raman spectroscopy [3][4][5][6]. Applications for Raman have expanded throughout time in a variety of sectors, including pharmaceuticals, textiles, and food [7]. There are currently around 25 distinct kinds of Raman spectroscopy techniques, including hyper-Raman scattering, tip-enhanced Raman scattering, and coherent anti-Stokes Raman scattering (CARS) [8].

In this chapter, we begin by exploring the fundamental theory of Raman spectroscopy, highlighting its inherent limitations and challenges. We then delve into the principles of SERS, emphasizing the critical role that nanoparticles play in enhancing the Raman signal. We then distinguish between extrinsic and intrinsic SERS approaches, explaining our focus on the intrinsic method for this thesis. Finally, a detailed overview of our experimental SERS setup is provided, including a comparative analysis of spectrometers to select the most suitable equipment. Finally, we address the challenges encountered during calibration and maintenance of the SERS setup, emphasizing the importance of critically evaluating our procedures. This approach sets the foundation for applying these validated techniques to biofilm analysis in the following chapters.

## 2.2 Basic Theory of Raman Spectroscopy

Photons can be transmitted, dispersed, or absorbed when light interacts with matter[9]. In infrared (IR) spectroscopy, absorption occurs when the photon's energy matches the vibrational energy difference between molecular states [10]. If the photon's energy does not match these energy levels, it undergoes scattering, depending on how the light interacts with the molecule.

One type of scattering is Rayleigh scattering, an elastic interaction where the scattered photon retains the same energy as the incident photon, without transferring energy to the molecule's vibrational modes [10]. The molecule acts as a dipole radiator, radiating dispersed light without altering the photon's energy because the incoming light's electric field creates an oscillating dipole moment [10].

In contrast, Raman scattering is an inelastic interaction, where the scattered photon's energy changes due to interactions with the molecule's vibrational modes [11]. The incident radiation creates a temporary induced dipole moment in the molecule. The molecular vibrations then interact with this induced dipole moment, leading to the scattering of photons and generating the Raman signal [11].

This process depends on the molecule's polarizability—the ability of its electron cloud to be distorted by the electric field of the incident light. When the molecule's electron cloud is easily distorted, the resulting induced dipole moment interacts strongly with the molecular vibrations [10][11]. This interaction occurs through a brief virtual state, and the energy exchange between the photon and the molecule results in a frequency shift, referred to as the Raman shift.

Equation ($1$) can be used to determine the Raman shift ($\Delta E_R$) , which is the energy difference between incident and scattered photons [12].

$$\Delta E_R = E_l - E_s \tag{1}$$

where, $E_l$ is the energy from the incident photon and $E_s$ is the energy of the scattered photon. This energy difference reflects the molecule's vibrational modes, creating a molecular fingerprint that can be used to identify specific molecular structures [12].

Figure 1 illustrates the difference between Rayleigh scattering, where no energy transfer occurs, and Raman scattering, where energy is transferred between the photon and the molecule. It also shows Stokes and Anti-Stokes Raman scattering, which result in different energy shifts based on whether the photon gains or loses energy during the interaction.

Once the Raman shift is calculated, the result is plotted on a Raman spectrum, which shows scattered light intensity (in arbitrary units, AU) versus Raman shift (in $cm^{-1}$). The spectrum's peak positions indicate the vibrational frequencies, while the intensity indicates the strength of the Raman scattering. Each peak in the spectrum represents a distinct vibrational mode.

Not all vibrational modes contribute equally to the spectrum. Raman selection rules, which are governed by the molecule's symmetry and changes in its polarizability during vibration, dictate which modes are Raman-active [13]. A molecule's polarizability ($\alpha$), or its capacity to have an electric field distort its electron cloud, is a key factor in predicting whether a vibrational mode would generate a powerful Raman signal. The interaction between the molecule and the light is governed by the equation [13]:

$$P = \alpha E \tag{2}$$

where $P$ is the amplitude of the induced dipole moment by Raman scattering and $E$ is the amplitude of the applied electric field. Molecules with higher polarizability tend to exhibit stronger Raman signals. In normal conditions, where there is no enhancement, the Raman signal intensity can be written as [120]:

$$I_{Raman} \propto \alpha^2 |E|^2 \tag{3}$$

Although Raman scattering is relatively rare—occurring in only about 1 in every $10^6$ to $10^8$ photons—it provides valuable molecular insights by revealing the molecule's vibrational modes [12][15].

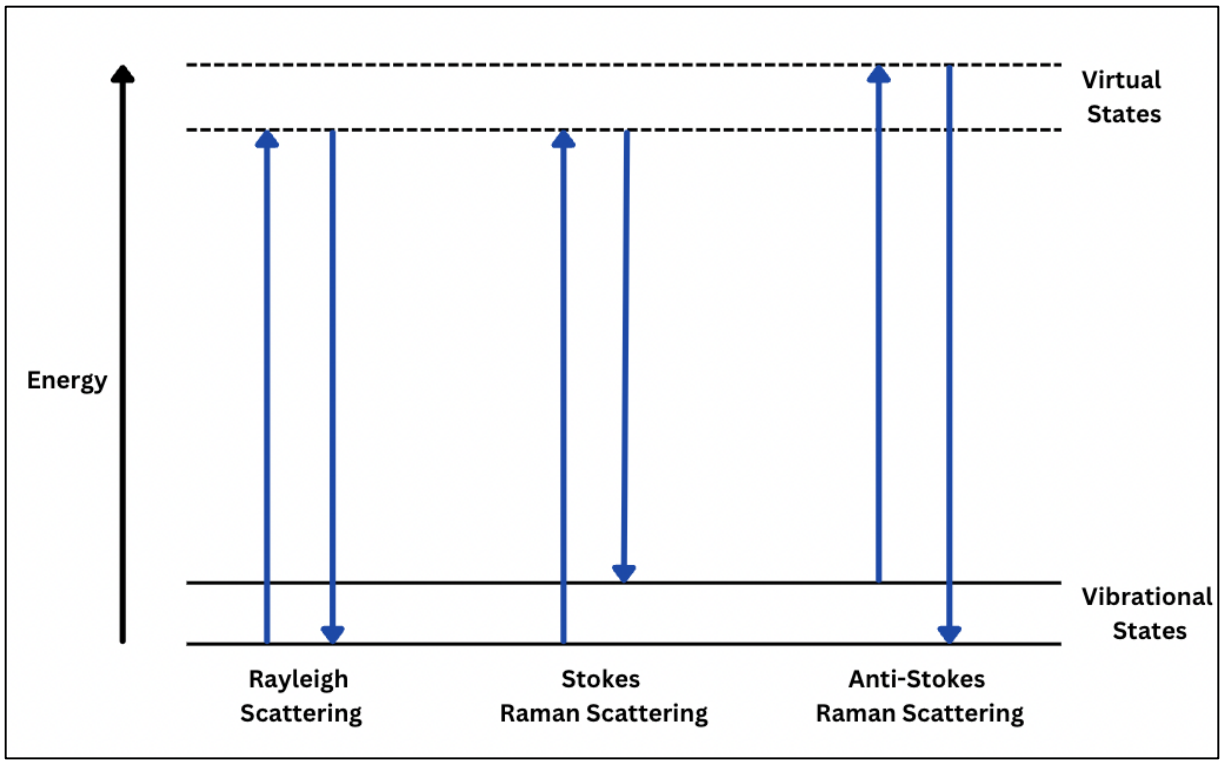


Figure 1: Illustrations of Rayleigh and Raman scattering methods [11,21].

## 2.3 Limitations and Challenges of Raman Spectroscopy

A key limitation of traditional Raman spectroscopy is its signal weakness which makes detecting low-concentration molecules challenging, especially in complex biological environments. Zhang et al. highlight the need for signal enhancement techniques [16]. Methods like optimized excitation conditions can improve detection but remain insufficient in addressing weak signals in such settings [16].

Another challenge is overlapping spectral features, where molecules with similar vibrational frequencies can produce overlapping peaks, complicating the differentiation between them. To address these overlaps, Hedegaard et al. stress the application of multivariate analysis methods such partial least squares (PLS) and principal component analysis (PCA) [17]. However, these computational methods rely on strong signals for effective analysis.

Limited depth penetration further restricts traditional Raman spectroscopy, which primarily probes surface layers [17]. This limits its ability to capture information from deeper regions of multi-layered samples. Stevens et al. propose solutions like CARS and fiber-optic-based Raman probes to improve penetration depth, but balancing depth and signal quality remains a challenge [18].

SERS addresses these limitations by enhancing surface-level signal strength and improving detection sensitivity [120]. The principles and advantages of SERS are discussed in the following section.

## 2.4 Surface Enhanced Raman Spectroscopy (SERS)

### 2.4.1 Introduction to SERS

SERS is a powerful analytical method that enhances the Raman scattering signals of molecules adsorbed on metal nanostructures. Because of this enhancement, SERS can detect individual molecules, making it an effective tool for molecular analysis in various scientific domains, such as biosensing and biomedical applications. The technique was first observed in the mid 1970s when it was discovered that molecules adsorbed on roughened metal surfaces exhibited much stronger Raman signals than those on smooth surfaces [19]. Since then, this discovery has improved the sensitivity and specificity of conventional Raman spectroscopy, hence expanding its possibilities.

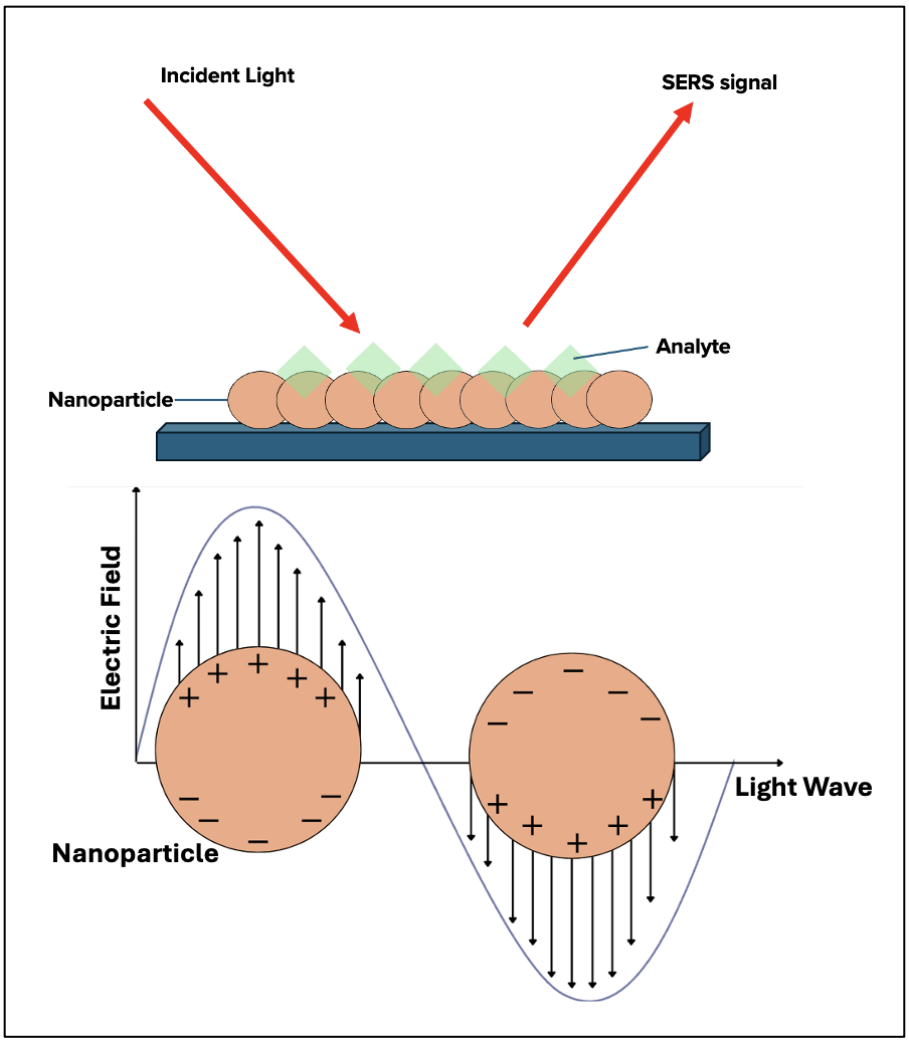


Figure 2: Schematic of SERS enhancement and plasmon oscillation in metal nanospheres [12][21].

SERS uses metal nanostructures to improve Raman signals, commonly silver (Ag) or gold (Au). The localized surface plasmon resonance (LSPR), a mechanism where conduction electrons on the metal surface fluctuate in response to incident light, producing strong electromagnetic fields, is essential to this increase (Figure 2). The Raman signals of molecules close to the metal surface can be amplified by several orders of magnitude by these fields [12][20].

When using SERS, an active nanostructure must be created and fabricated, the analyte must be bound to it, and the SERS signals must then be excited and collected. For a successful application, SERS-active surface design and fabrication are essential. These surfaces are created

using different methods, such as electrochemical roughening, chemical reduction, and nanolithography [22].

SERS's extreme sensitivity, even at the single-molecule level, is one of its main benefits. The electromagnetic enhancement mechanism, in which the LSPR effect produces a powerful localized electric field that intensifies the Raman signal of adsorbed molecules, is responsible for achieving this sensitivity.[22]. Additionally, the narrow Raman peaks allow for multiplex analysis, making SERS a versatile tool for detecting and identifying various analytes simultaneously [22].

SERS is used widely in biosensing and biomedical research [120]. For instance, it is used to detect and identify DNA, proteins, and pathogens with high specificity and sensitivity [120]. The method is beneficial for early disease diagnosis, tracking the course of the disease, and assessing the effectiveness of treatment since it produces comprehensive molecular fingerprints [12][22].

### 2.4.2 Principles of SERS

The enhanced SERS signal is primarily attributed to two mechanisms: electromagnetic enhancement and chemical enhancement [14].

Electromagnetic enhancement, which includes stimulating localized surface plasmons (LSPs) on metal nanostructures, is the primary process of SERS. Strong localized electric fields are produced when light interacts with these nanostructures, causing an overall oscillation of free electrons. The Raman signals of molecules close to the nanostructures can be amplified by several orders of magnitude by these fields. [23].

An incident electric field, $E_0$, first excites the nanoparticle which produces an enhanced localized field, $E_{loc}$, near the particle [120]. The enhancement factor of the excitation field, $M(\lambda_{exc})$ at an incident wavelength $\lambda_{exc}$ can be expressed with the following equation [120]:

$$E_{loc} = M(\lambda_{exc}) \cdot E_0 \tag{4}$$

$M(\lambda_{exc})$ depends on the properties of the nanoparticles, such as shape, size, and material. At this stage, $E_{loc}$ will excite any molecule that is located on the surface of the nanoparticle. This will scatter a Raman signal at shifted wavelength, $\lambda_{sc}$ [120]. The amplitude of the Raman scattering, $P$, is then proportional to the local field, $E_{loc}$ [120]:

$$P = \alpha \cdot E_{loc} = \alpha \cdot M(\lambda_{exc}) \cdot E_0 \quad (5)$$

where, similar to Equation (2), $\alpha$ is the polarizability of the molecule. Because the Raman scattering, $P$, is an electromagnetic field, it can also interact with the nanoparticle. This interaction also induces an enhancement of the Raman scattering [120]:

$$E_{sc} = M(\lambda_{sc}) \cdot P = M(\lambda_{exc}) \cdot M(\lambda_{sc}) \cdot \alpha \cdot E_0 = E_{SERS} \quad (6)$$

where $M(\lambda_{sc})$ is the enhancement factor at the scattered wavelength, $\lambda_{sc}$. Due to both enhancement processes, we account for enhancement at both the excitation $(\lambda_{exc})$ and scattered $(\lambda_{sc})$ wavelengths, making the SERS intensity [120]:

$$I_{SERS} = [M(\lambda_{exc}) \cdot M(\lambda_{sc})]^2 \cdot \alpha^2 \cdot |E_0|^2 \quad (7)$$

When the excitation and scattering wavelengths are very close, both the enhancement factors are estimated to be equal as $M(\lambda_{exc}) \approx M(\lambda_{sc})$. Simplifying Equation (7) and substituting the $I_{Raman}$ from Equation (3)[120]:

$$I_{SERS} \approx [M(\lambda_{exc})]^4 \cdot I_{Raman} \quad (8)$$

Equation (8) emphasizes that SERS is able to significantly amplify the Raman signal. The Raman enhancement is proportional to the fourth power of the enhancement of the electric field near the surface [120]. This amplification makes SERS a highly sensitive tool as it is able to detect very low concentrations of the samples [120].

The second main process, chemical enhancement, occurs by the transfer of charge between the adsorbed molecules and the metal surface. The molecules' polarizability can increase as a result of this interaction, intensifying their Raman signals. [21]. Although it is often thought to contribute a lower enhancement factor than the electromagnetic mechanism, the chemical enhancement mechanism is essential for certain sample types [21].

The total SERS enhancement factor can be approximated as the product of the electromagnetic enhancement factor and the chemical enhancement factor [25]. The electromagnetic enhancement factor can reach values as high as $10^6$ to $10^8$ and the chemical enhancement factor typically contributes to an additional factor of 10 to 100 [25]. This combined

enhancement can lead to an overall increase in the Raman signal by several orders of magnitude, making SERS a highly sensitive analytical technique [25].

### 2.4.3 The Role of Nanoparticles

Silver nanoparticles (AgNPs) are effective for SERS due to their excellent plasmonic properties. Silver supports strong plasmon resonances over a wide range of wavelengths, especially in the visible region, which is crucial for achieving high signal enhancement [26]. Compared to other metals like gold, silver provides a higher enhancement factor, making it ideal for applications requiring high sensitivity [26].

The dielectric function of silver allows for minimal energy loss, resulting in more significant amplification of the electromagnetic fields around the nanoparticles. Additionally, AgNPs are relatively easy to synthesize and can be functionalized with various chemical groups, enhancing their compatibility with different analytes [22].

AgNPs can be synthesized through various methods, including chemical reduction, photochemical methods, and green synthesis using biological agents. The choice of synthesis method can affect different properties of the nanoparticles, including the shape, the size, and the surface properties [27]. These can then influence their SERS performance [27]. For example, colloidal AgNPs can be stabilized with capping agents such as citrate, which not only prevent aggregation but also play a role in the interaction with target molecules [27].

In practical applications, the creation of hot spots—regions of extremely high electromagnetic field enhancement—is critical. These hot spots are typically formed at the junctions between closely spaced nanoparticles, and they can provide enhancement factors as high as $10^9$ [22]. The ability to control the formation and distribution of hot spots is a key factor in maximizing the sensitivity and reliability of SERS measurements [22].

### 2.4.4 Mechanistic Approaches in SERS: Extrinsic and Intrinsic

In the context of SERS, it is essential to distinguish between extrinsic and intrinsic SERS, as these two approaches differ in their principles and applications.

***Extrinsic SERS***

Extrinsic SERS involves the use of external labels or tags, Raman-active dyes or nanoparticles, that bind to the target molecules [28]. These labels enhance the Raman signal through the same LSPR mechanism but do not rely on the inherent Raman activity of the analyte [28]. Instead, the tags provide a strong and characteristic Raman signal, which can be used to detect and calculate the presence of the target molecules.

To stop nanoparticle aggregation and Raman reporter detachment, extrinsic SERS immobilizes the reporters on the substrate's surface and then completely encloses them with a protective layer of dielectrics[29]. Finally, ligands or antibodies are applied to these nanoparticles to functionalize them so they can capture the target molecules. [29]. This sandwich structure is applied in situations where intrinsic SERS fails, such as in in-vivo imaging of unique and rare cancer cells (Figure 3(A)) [30]. Extrinsic SERS is advantageous when dealing with target molecules that have weak intrinsic Raman signals or are present in very low concentrations. It allows for highly specific detection and can be used in a variety of applications, including immunoassays and diagnostic applications [30].

***Intrinsic SERS***

The amplification of Raman signals that originate directly from the molecules of interest adsorbed on the metal nanostructures is known as intrinsic SERS [30]. This improvement is accomplished without the need for outside tags or labels. The primary process is the interaction of the analyte molecules with the metal substrate's LSPR, intensifying the analyte's inherent vibrational modes' Raman scattering signals [30].

Intrinsic SERS relies on the inherent vibrational properties of the target molecules. The enhancement occurs due to the electromagnetic and chemical interactions between the analyte and the metal nanostructure [28]. This method is useful for detecting and identifying macro-sized biomolecules and for qualitative analysis (Figure 3(B)) [30]. Intrinsic SERS can speed up the qualitative analysis of related samples by comparing the Raman spectra before and after the molecule is captured. This allows for the collection of structural information about the target molecule [30]. However, intrinsic SERS can have limitations if the target molecules exhibit weak Raman activity or are present in very low concentrations [30].

For this research, we focus on this intrinsic approach. This approach provides us with direct molecular information without altering the sample. By using silver nanoparticles, we were able to

enhance the native Raman signals of the biofilm's molecular components. This approach guarantees sensitivity and specificity in detecting the molecular composition without the need for external labels, making it more suited for our goal of analyzing the molecular composition of biofilms. This is discussed more in detail in Chapter 4.

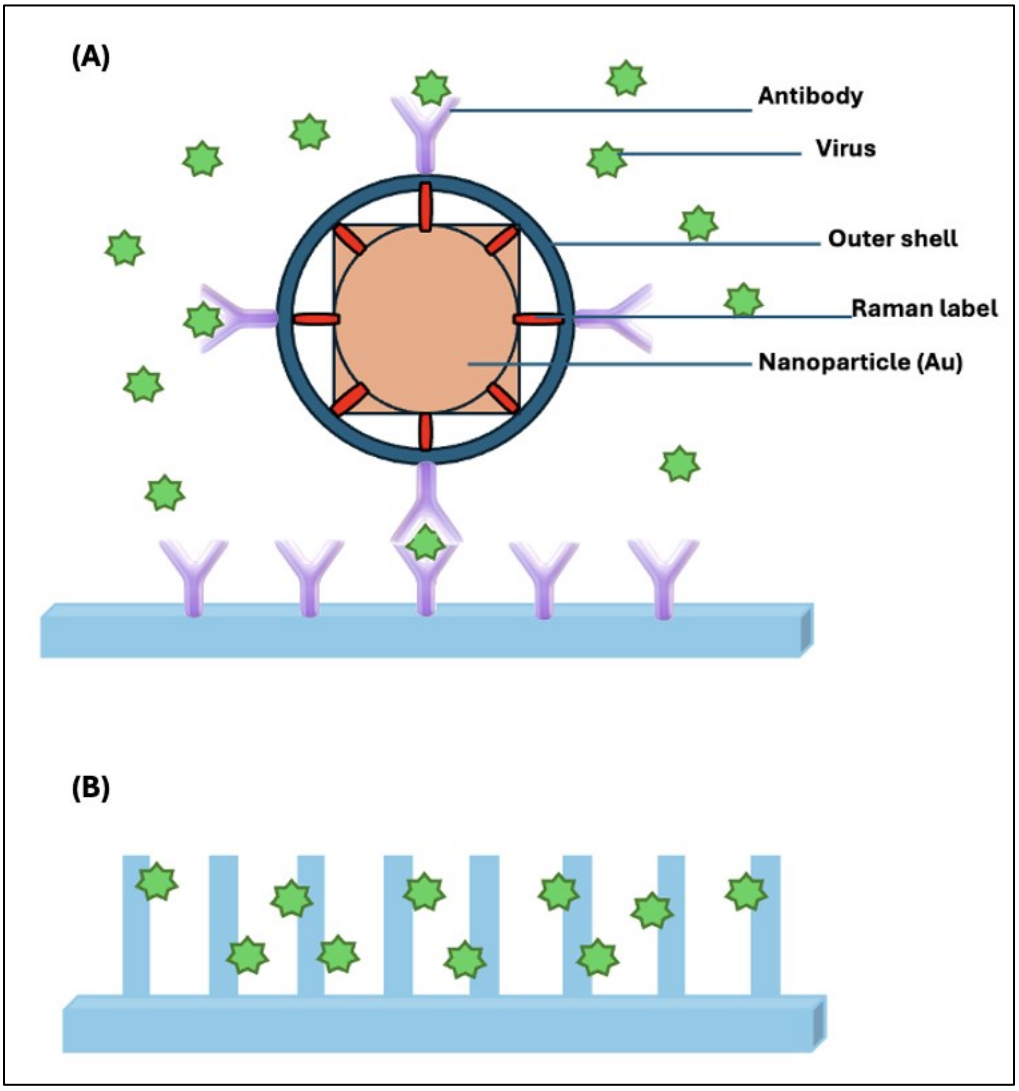


Figure 3: Different SERS configuration. (a) extrinsic approach (b) intrinsic approach [30].

# 2.5 SERS Setup

## 2.5.1 Overview of the experimental setup

For data collection, we used a dark-field microscope SERS setup shown in Figure 4. Previous students from our lab have used this setup for different applications, but in this thesis, we mainly

use it for biofilm detection[107][108][109]. The core component of our system is a continuous wave (CW) distributed Bragg reflector (DBR) laser, the Coherent LM-785-PLR-100-1K, which emits a stable 785 nm beam. This beam was chosen to be the excitation source because of its minimal fluorescence interference. After emission, the beam is collimated and is reflected through an 808 nm notch filter (NF).

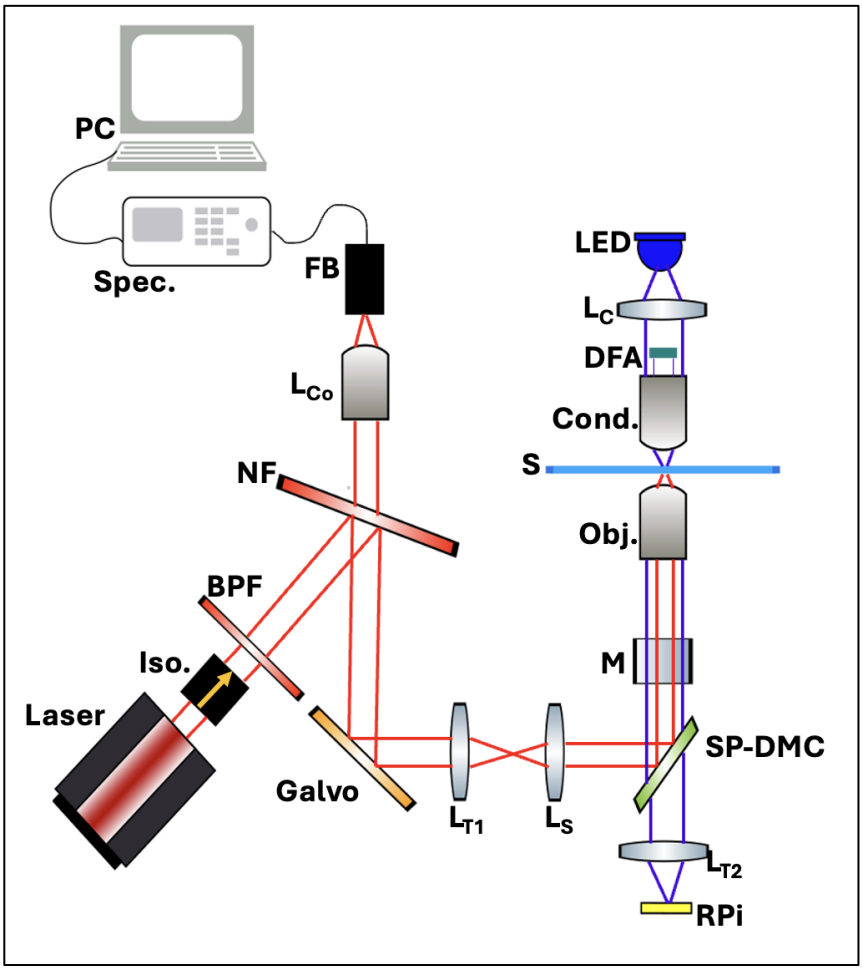


Figure 4: Schematic of the SERS setup used in this research. The setup includes the laser, isolator (Iso.), 785 nm bandpass filter (BPF), 808 nm notch filter (NF), galvo system (Galvo), scan lens (LS), short pass dichroic mirror (SP-DMC), microscope mirror (M), 0.65 NA objective lens (Obj.), sample (S), 0.85 NA condenser lens (Cond.), dark-field aperture (DFA), collector lens (Lc), blue LED, tube lens ($L_{T2}$), fiber coupling lens ($L_{Co}$), multimode fiber bundle (FB), and the spectrometer (Spec.) connected to a PC.

The microscope portion of our setup uses LED illumination and a collector lens ($L_C$), which, along with a water-submerged condenser lens (NA = 0.85), focuses blue light onto the sample. This condenser has a darkfield stop to make sure that the numerical aperture (NA) exceeds the NA

of the objective lens (NA = 0.65). The focused beam then passes through a shortpass dichroic mirror (SP-DMC) which is present in the microscope's infinity space. This improves the clarity and detail of the captured images. These images are projected onto a Raspberry Pi (RPi) camera via a tube lens ($L_{T2}$). The RPi was programmed with an existing Python code, detailed in [110]. It controls the camera and displays the images on a touchscreen, facilitating real-time monitoring and visualization throughout the data collection process [110]. Precision in sample placement was ensured by a Thorlabs XY-directional movable stage, allowing for careful alignment of the sample for optimal Raman scattering.

For spectral analysis, two spectrometers were evaluated: the Kaiser f/1.8i spectrograph with a thermoelectrically cooled Andor CCD camera controlled by the Andor SOLIS software and iHR320 HORIBA spectrograph controlled by the LabSpec6 software. Both systems were assessed under consistent experimental conditions using dried Rhodamine 6G (R6G) as a calibration standard to directly compare their performance. Each spectrometer was tested with a scan acquisition time of 30 seconds, and the R6G was prepared at a concentration of 10mM.

To ensure the reliability of our spectral findings, a calibration procedure is conducted before any experimental runs. This involves applying a 10 mM solution of R6G in deionized water onto a glass slide, which is then allowed to dry, confirming that all system components are aligned correctly. Additionally, a power meter is used to maintain a consistent power output of 14.2mW at the sample before collecting data, ensuring measurement consistency and accuracy. The results from this comparative analysis are further discussed in Section 2.5.2.

### 2.5.2 Instrument Validation

Our lab recently upgraded to a new HORIBA spectrometer, assuming it would enhance the SERS performance because it features newer technology as compared to the older Andor system. To verify whether this upgrade was justified, we tested both systems under identical experimental conditions, using the setup shown in Figure *4*. During the data collection, the only modification made was the switch from one spectrometer (Spec.) to another. For this experiment, we used R6G to determine the SERS performance.

The spectrum from the Andor system (20,000 AU to 28,000 AU) achieved a signal intensity, approximately 19 times higher than the spectrum from the HORIBA system (700 AU to 1,800 AU), as shown in Figure *5*. The resolution of the Andor spectrometer and HORIBA spectrometer

was 0.154nm and 0.125nm, respectively. The comparison between the spectrometers is also supported by the Signal-to-Noise ratio (SNR) calculations detailed in Table *1*.

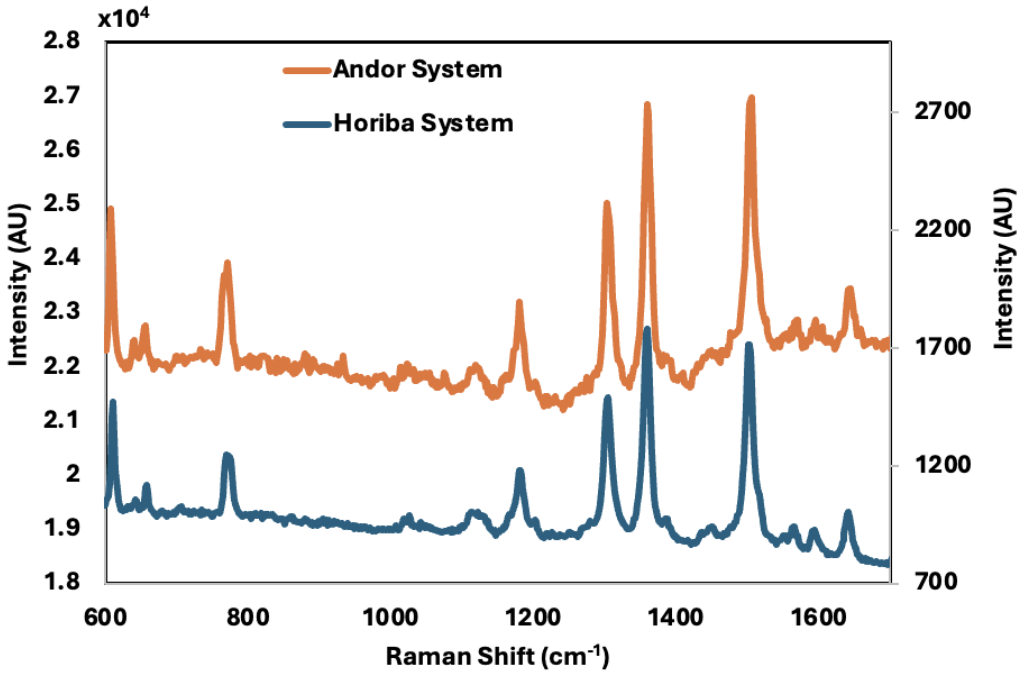


Figure 5: Raman spectra of R6G using Andor (y- axis on the left) and HORIBA (y-axis on the right) systems for setup validation.

The ratio of the signal power to the noise power is known as SNR [31]. SNR provides insight into the spectrometer's effectiveness in distinguishing true signals from background noise [31]. The formula used for calculating SNR is [31]:

$$SNR = \frac{Signal\ Mean}{Standard\ Deviation\ of\ Noise} \tag{9}$$

Table 1: Signal-to-Noise Ratio (SNR) comparison between Andor and HORIBA spectrometers.

| Spectrometer | Signal (AU) | Noise (AU) | SNR |
|---|---|---|---|
| Andor | 22,427.71 | 423.32 | 52.98 |
| HORIBA | 1,018.28 | 115.49 | 8.82 |

Using this formula, the Andor spectrometer achieved an SNR of 52.98, which is approximately 6 times higher than the SNR achieved by the HORIBA spectrometer. This indicates that the Andor system not only captures stronger signals but also delivers much clearer and smoother data with minimal noise interference.

This analysis confirmed that the Andor spectrometer was the better choice for our research, based on its higher signal intensity and better SNR. Choosing the Andor system for our investigation guarantees that we are using the most dependable equipment accessible to us. This choice ensures the precision of our results.

### 2.5.3 Challenges and Reflection

We experienced significant obstacles beginning this research project. Because there were multiple components in the setup, one major setback we encountered was laser misalignment. We overcame this by manually shifting the angles of the mirrors and lenses. Additionally, we slightly repositioned some components to ensure the laser would reach the sample. The SERS setup also required constant attention to calibration and maintenance. For example, the power of our laser might vary and cause the sample to burn. This meant that we would have to go back and realign the components to collect accurate and reliable results. This fine-tuning process took us a couple of months before we could start collecting usable data.

## 2.6 Conclusion

SERS offers a powerful solution to the limitations of traditional Raman spectroscopy by significantly enhancing molecular Raman signals. This enhancement is achieved using nanoparticles which amplify the Raman signal via localized surface plasmon resonance. For this research, we chose to go with the intrinsic SERS approach, as it avoids the need for external labels and probes. Additionally, the careful calibration of the setup is vital for ensuring accurate and reliable results. By integrating these factors, SERS not only proves to be an effective tool for detailed molecular characterization but also lays a solid foundation for its application in analyzing biofilm samples in the later chapters.

# Chapter 3

# Optimization Experiment for SERS Biofilm detection

## 3.1 Introduction

Optimizing SERS conditions is crucial for enhancing the detection sensitivity of biofilms [32]. These are organized communities of microorganisms enclosed in a protective matrix that makes them resistant to traditional methods [36]. More details on biofilm characteristics and their relation to healthcare are discussed in Chapter 4.

This chapter focuses on experimentally optimizing key variables that influence SERS signal enhancement for detecting *Pseudomonas Aeruginosa* (PA01) biofilms. We start by outlining the experimental design considerations, and then detailing the synthesis and characterization of AgNPs used to amplify the signal in this research [33].

Using a factorial design approach, which allows for the systematic investigation of multiple variables and their interactions [34], we examine how interaction time and nanoparticle concentration affect SERS signal intensity. Through controlled experiments and statistical analyses, we identify the optimal conditions that maximize signal enhancement without compromising biofilm integrity, establishing a robust framework for subsequent analyses. Additionally, we reflect on the challenges encountered during nanoparticle synthesis and experimental optimization, underscoring the importance of protocol development and validation to achieve reliable and reproducible results.

### 3.1.1 Experimental Design Considerations

In the early stages of experiment planning for SERS biofilm detection, we conducted a detailed analysis to define the key factors influencing the experimental design. These factors were classified into constants, variables, and potential noise sources. To ensure consistency and reproducibility, constants such as laser power, laser wavelength, nanoparticle synthesis protocol, sample

preparation techniques, spectrometer settings, slide material, and environmental conditions were tightly controlled throughout all experiments.

The primary focus, however, was on the variables — interaction time and nanoparticle concentration — which were systematically varied to assess their impact on SERS signal intensity.

While potential noise factors, such as sample variability and environmental conditions, were acknowledged, they were not directly investigated in this study. These early considerations provided a robust experimental framework and laid the groundwork for future optimization efforts.

The following sections describe the methods used in the optimization experiments, as well as the results from the factorial design and subsequent analyses, aimed at identifying the optimal conditions for SERS signal enhancement.

## 3.2 Materials and Methods

Silver nitrate, hydroxylamine hydrochloride, R6G were all purchased from Sigma-Aldrich [111]. Sodium hydroxide pellets were purchased from Fisher Scientific [112]. Glass microscope slides were purchased from UltiDent Scientific Inc [113].

### 3.2.1 Synthesis and Analysis of Silver Nanoparticles (AgNPs)

AgNPs were synthesized using the Leopold and Lendl method [33]. First, a 750 mL silver nitrate stock solution ($AgNO_3$) was prepared. Deionized water (DIW) was deoxygenated by bubbling nitrogen gas through it for 2 hours in a covered beaker. After this, 148 mg of silver nitrate was dissolved in the DIW. To keep the solution safe from light, it was kept at room temperature in an amber bottle.

Two stock solutions were prepared separately - hydroxylamine (final concentration $1.5 \times 10^{3}$ M) and sodium hydroxide (final concentration $3 \times 10^{-3}$ M). These were mixed to form a solution. The solution was rapidly added to the silver nitrate stock solution ($10^{-3}$ M), inducing the formation of silver nanoparticles. The immediate color change to brownish grey confirmed the successful synthesis of the nanoparticles (Figure 6(A)). The synthesized nanoparticles were then concentrated through centrifugation at 4000 rpm for 30 minutes, repeated twice. This solution was stored away from ambient light at 4°C and sonicated for 3 minutes prior to each use.

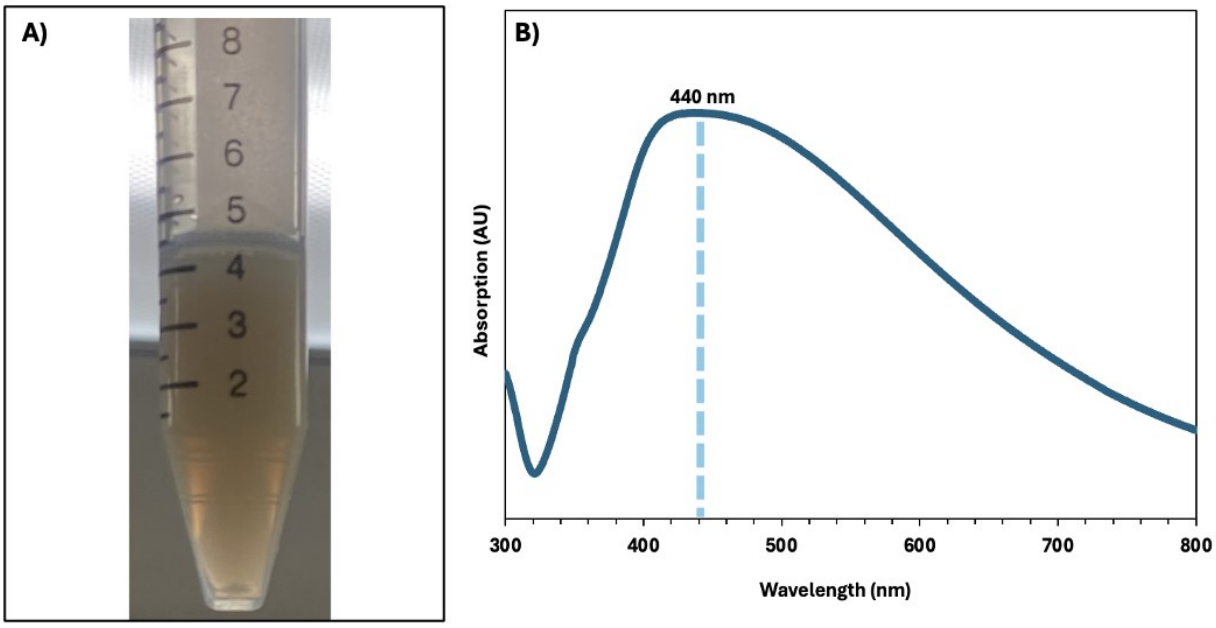


Figure 6: (A) Colloidal solution of the synthesized silver nanoparticles (B) UV-Vis absorption spectrum of the silver nanoparticles.

The synthesized AgNPs were characterized using UV-Vis spectroscopy to assess their optical properties and confirm their formation. The UV-Vis spectrum was characterized by a pronounced surface plasmon resonance (SPR) peak around 440 nm, indicative of nanoparticle formation. As depicted in Figure 6(B), this peak confirms the presence of nanoparticles and suggests a polydisperse distribution. The wide range of plasmon resonance modes causes the broad absorption band, which is advantageous for a number of uses, including Raman scattering [115].

Bulk SERS enhancement was calculated using the following equation [25]:

$$A_{SERS} = \frac{I_{SERS}}{I_{Raman}} \frac{C_{Raman}}{C_{SERS}} \tag{10}$$

where $I_{SERS}$ is the measured intensity for SERS and $I_{Raman}$ is the measure intensity for normal Raman scattering. Intensity measurements are in AU. $C_{SERS}$ and $C_{Raman}$ are the concentrations used in the measurement. $C_{SERS}$ used here was 50μM of R6G and $C_{Raman}$ was 10mM of R6G.

### 3.2.2 Optimization Experiment: Factorial Design

The optimization experiment aimed to determine the ideal combination of interaction time and nanoparticle concentration for maximizing SERS signal intensity in PA01 biofilms. A $2^2$ factorial design with center runs was implemented to generate a first-order response surface, with the main variables being the interaction time between the biofilm samples and the AgNPs, and the concentration of these nanoparticles. Detailed methodologies on biofilm and collagen sample preparation are discussed in Section 4.6.

The baseline condition was set at a 48-hour interaction with a 1x concentration ($10^{-3}$ M) of nanoparticles, while variable conditions tested included shorter (24 hours) and longer (72 hours) interaction times, paired with lower (0.25x) and higher (1.75x) concentrations. The 1x concentration AgNPs were freshly synthesized and adjusted—either concentrated or diluted—to achieve the desired concentrations for each experimental setup. We concentrated the nanoparticles through centrifugation for 30 minutes at 4000rpm and then pipetted out the excess water carefully. To dilute the nanoparticles, we added DIW. The volume of the water removed or added depended on the desired concentration we were trying to get.

Six different batches of nanoparticles were synthesized to account for variability in particle properties, and the bulk enhancement (calculated using Equation (*10*) for these batches is illustrated in Figure 7 (Batches 1-6).

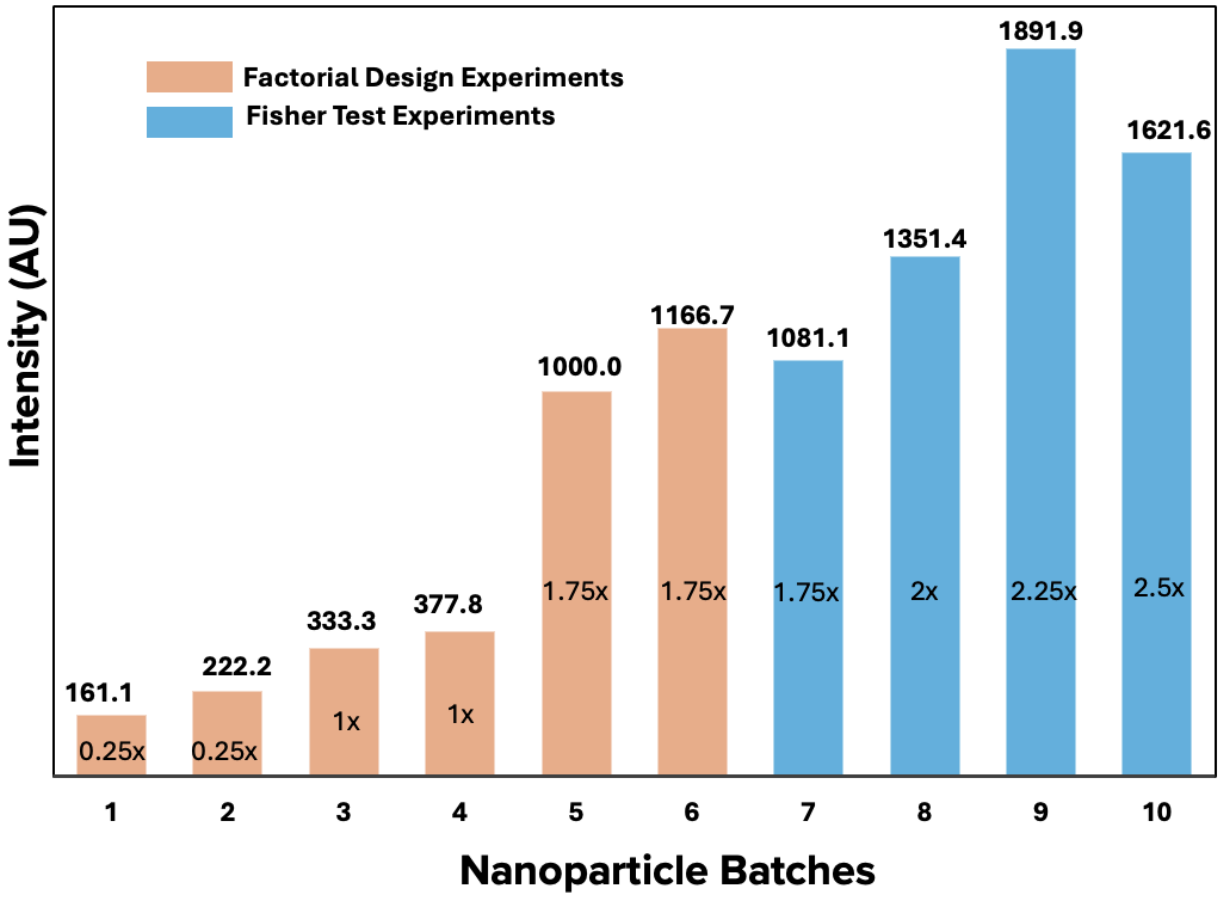


Figure 7: Bar chart showing bulk enhancement calculations across all experimental batches.

After the initial factorial design analysis, the second phase of optimization was performed using a Fisher test. Here we further explored the influence of shorter interaction times and higher nanoparticle concentrations on SERS enhancement. The Fisher test is a statistical method used to see whether the differences observed between the variables are significant or due to random chance [106]. In these experiments, four different batches of nanoparticles were synthesized: 2x concentration with an 18-hour interaction, 2.25x concentration with 12 hours, 2.5x concentration with 6 hours, and 1.75x concentration with 24 hours. Bulk enhancement calculations (calculated using Equation(10) for these batches are also provided in Figure 7 (Batches 7-10).

As shown across all batches, the introduction of the AgNPs consistently resulted in a substantial enhancement of the SERS signal. This consistent increase in signal intensity, as detailed in Figure 7, confirms the efficacy of the nanoparticles in amplifying the Raman scattering. Especially, even at the lowest nanoparticle concentration, a significant enhancement was observed.

However, at the highest concentration (2.5x), there was a noticeable decline in the enhancement value, possibly due to the aggregation of nanoparticles, which can interfere with optimal signal enhancement by reducing the number of effective scattering sites [32][35].

The details of how we prepared each sample are given in Section 4.6.1 and illustrated in Figure 10. To ensure data integrity, slides washed and prepared on the same days, were processed on the same day, minimizing any variations that arise from storage or extended delays between preparation and analysis. This approach helped maintain consistency across experimental runs.

### 3.2.3 Data Collection and Processing

Before the data collection and processing, the SERS setup was calibrated as outlined in Section 2.5.1. The integration time for each measurement was set at 15 seconds across 10 accumulation cycles. Each experimental trial was averaged over 5 spectra, resulting in $n = 50$ spectra per experimental condition.

For the data analysis, we used Python. The code was designed to handle factorial and center runs focusing on interaction times and nanoparticle concentrations, crucial for optimizing the SERS signal. All spectra were preprocessed using techniques detailed in Section 5.4.3.1 before any analysis.

Regression analysis is a common statistical method used to analyse the relationship between two or more variables, while Analysis of Variance (ANOVA) is used to confirm whether this relationship is due to random chance[116][117]. We use both regression analysis and ANOVA to determine the effects of our variables on the SERS signal, particularly focusing on the pyocyanin peak. The detailed analysis of this peak is further elaborated in Chapter 4. We used ANOVA to test the null hypothesis ($H_0$) that differences in SERS signal across experimental conditions are due to random variation, as opposed to a significant influence of variables such as nanoparticle concentration and interaction time.

For pairwise comparisons and to evaluate potential additional optima in experimental conditions, we employed Fisher's Least Significant Difference (LSD) test. The Fisher test uses the Mean Squared Errors (MSEs) derived from the center runs to calculate the Least Significant Difference (LSD), facilitating a comparison of mean intensity differences at the pyocyanin peak across varying conditions. A two-tailed test with an alpha threshold of 0.05 was used to establish statistical significance with the null hypothesis ($H_0$) posting no significant difference in mean

intensities across the different conditions. This approach allowed us to determine if the identified 24-hour interaction with 1.75x nanoparticle concentration represented the true optimum, or if additional conditions could yield comparable or superior results, thus ensuring a thorough exploration of potential optima.

## 3.3 Results and Discussion

As illustrated in Figure 8, the SERS spectra obtained from the different experimental conditions show that the combination of a 24-hour interaction time and 1.75x nanoparticle concentration produced the highest signal intensity. Figure 8 also shows how this trial gave us some more information about the sample, by giving us more peaks overall, as compared to the other trails.

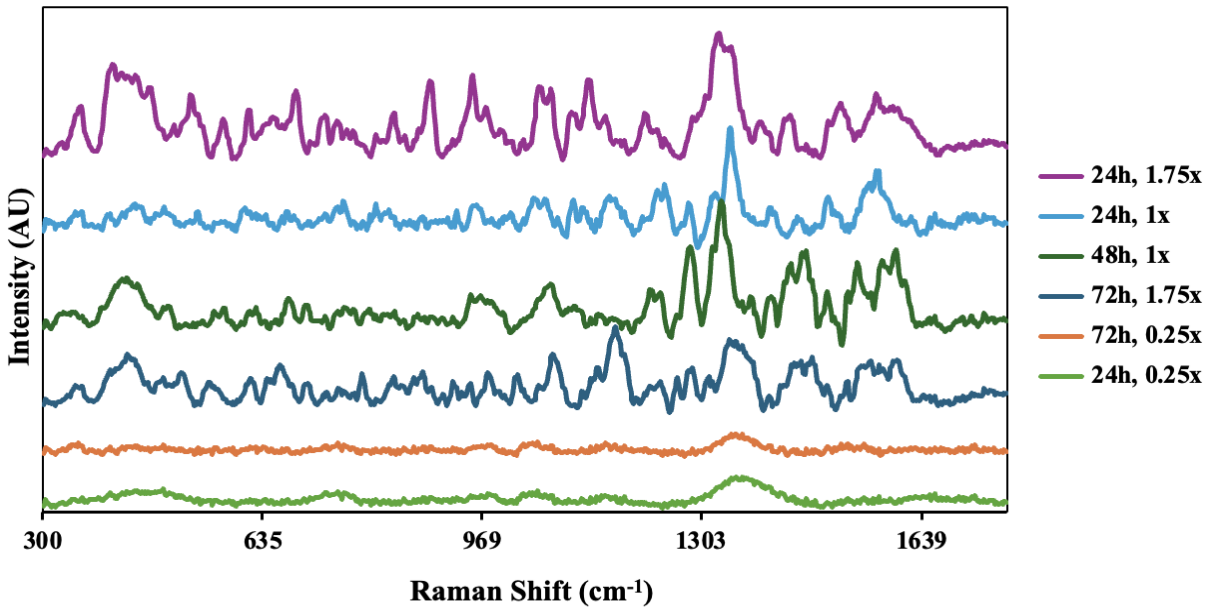


Figure 8: Raman spectra from optimization experiments at varying interaction times and nanoparticle concentrations.

Table *2* shows the statistical analysis's findings. ANOVA revealed that the interaction time and nanoparticle concentration had a significant impact on the SERS signal intensities. Therefore, ANOVA rejected the null hypothesis ($H_0$) confirming a statistically significant effect of both

interaction time and nanoparticle concentration on the SERS signal ($p < 0.05$). The nanoparticle concentration showed a positive coefficient of 4441.66, indicating that higher concentrations promote the formation of electromagnetic 'hot spots' essential for enhancing the SERS signal. This observation aligns with studies by Schultz et al., which emphasized the role of nanoparticle density in amplifying Raman signals [35].

Table 2: ANOVA results detailing the effects of interaction times and nanoparticle concentrations.

| Source | Coefficient/Slope | F-Statistic | p-value |
|---|---|---|---|
| x1 : Interaction Time | -2375.87 | 161.94 | 0.0061 |
| x2: Nanoparticle Concentration | 4441.66 | 565.99 | 0.0018 |
| I(x1 ** 2) | -3950.86 | 191.92 | 0.0051 |
| x1∗x2 | -2607.83 | 195.11 | 0.00508 |

Conversely, the negative coefficient for interaction time suggests that prolonging the interaction beyond 24 hours may lead to a reduction in signal intensity, possibly due to oxidative stress affecting the biofilms, as noted by Keleştemur et al. [36]. This finding is supported by Choo-Smith et al., who reported that optimal adsorption of biofilm components onto nanoparticles occurs within 24 hours, with longer durations potentially causing biofilm disruption or analyte desorption [36].

The interaction term between interaction time and nanoparticle concentration ($x1 * x2$) underscores the need for a balanced approach in optimizing these parameters. This interaction could potentially be attributed to reaction kinetics, where higher concentrations of nanoparticles might accelerate reaction processes up to a saturation point. Beyond this point, any additional increase could lead to suboptimal SERS signals, a phenomenon that Sharma et al. suggest occurs due to the complex dynamics of nanoparticle aggregation and biofilm interaction [36]. These dynamics underscore the importance of not only the concentration but also the timing of interactions.

Further validation of these conditions was achieved through a Fisher LSD test, which confirmed that a 24-hour interaction time with 1.75x nanoparticle concentration was most effective, outperforming other tested conditions (Figure 9). These additional experiments, did not yield better results, reaffirming the effectiveness of the established optimal condition. Notably,

the experiments utilizing the lowest interaction time with the highest nanoparticle concentration demonstrated minimal improvement. Despite the presence of enhancement, the signal captured was insufficiently robust, indicating that a balance between interaction time and nanoparticle concentration is crucial for optimal signal generation.

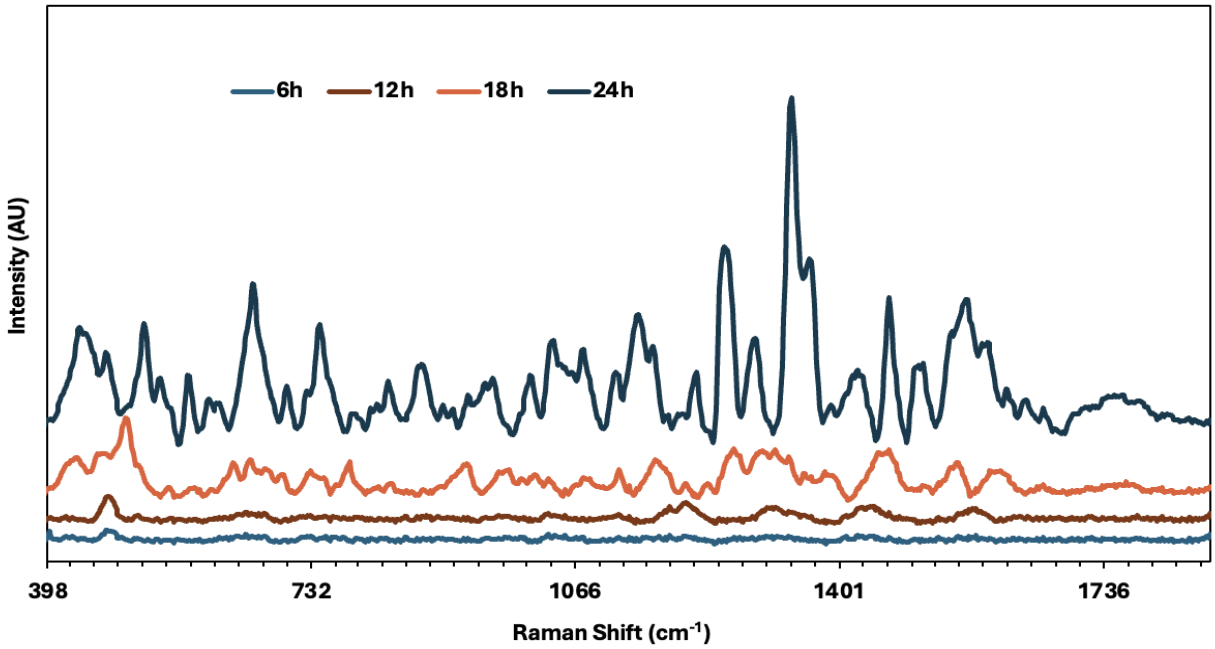


Figure 9: Raman spectra across time illustrating effects of shorter interaction times and increased nanoparticle concentrations.

Statistical analysis, presented in Table 3, quantifies the impact of these variables on the SERS signal. The analysis highlights significant differences in signal intensity among the tested conditions, confirming that a 24-hour interaction with 1.75x nanoparticle concentration yields superior results. For example, the mean difference in signal intensity between the 24-hour, 1.75x condition and the 12-hour, 2.25x condition was significant, with a calculated t-statistic of 21.754 and a highly significant p-value ($p = 2.10\text{e-}08$), as detailed in the table. The t-statistic measures how large the difference in variability is between any two conditions[118]. Therefore, higher the t-statistic, the difference between the two conditions gets larger as compared to the variability in the data[118]. The p-value ($p < 0.05$) confirms that this difference is not due to random chance for all our trials, further rejecting the null hypothesis ($H_0$).

This demonstrates not only the effectiveness of the 24-hour, 1.75x condition but also the robustness of these findings across various experimental setups.

Table 3: Fisher LSD test results for Raman signal intensity variations across different interaction times and nanoparticle concentrations.

| Condition 1 | Condition 2 | Mean Difference | t-statistic | p-value | Significant |
|---|---|---|---|---|---|
| 24-hour, 1.75x AgNP | 12-hour, 2.25x AgNP | 9966.86 | 21.754 | 2.10e-08 | Yes |
| 24-hour, 1.75x AgNP | 6-hour, 2.5x AgNP | 9045.90 | 19.744 | 4.51e-08 | Yes |
| 24-hour, 1.75x AgNP | 18-hour, 2x AgNP | 5625.81 | 12.279 | 1.80e-06 | Yes |
| 12-hour, 2.25x AgNP | 18-hour, 2x AgNP | 4341.05 | 9.475 | 1.27e-05 | Yes |

## 3.4 Challenges and Reflections

The synthesis of silver nanoparticles for SERS presented significant challenges. Initial attempts frequently resulted in nanoparticles that did not meet the necessary criteria for effective signal enhancement. These early setbacks led to delays, as it became necessary to revisit and refine the synthesis protocols. This portion of our research required a deeper comprehension of the underlying chemical processes in addition to mastering the technical aspects of nanoparticle synthesis. Even after a successful synthesis protocol was established, maintaining consistency between batches was a continual challenge, as any slight deviation could result in variations in the SERS signals, complicating data analysis. To overcome this, after each synthesis batch, we obtained SERS spectra to assess the enhancement efficiency. If the results were unsatisfactory, we repeated the synthesis process until the desired level of signal enhancement was achieved. This iterative approach was crucial for establishing a robust and reproducible nanoparticle synthesis protocol.

The optimization of experimental parameters, such as interaction time and nanoparticle concentration, was another area marked by extensive trial and error. This process was labour-intensive and required careful analysis of the data after each experiment to recognise trends and determine the next steps. There were many instances where initial hypotheses had to be rejected based on unexpected results, requiring an adaptive approach to experimental design.

## 3.5 Conclusion

Through this optimization, it was determined that a 24-hour interaction time combined with a 1.75x concentration of silver nanoparticles yielded the highest signal intensity. The increased nanoparticle concentration significantly amplified the Raman signal. However, interaction times longer than 24 hours led to a reduction in signal intensity. Hence, the interaction between nanoparticle concentration and interaction time underscored the necessity of balancing these parameters to optimize SERS enhancement effectively.

# Chapter 4

# Biofilms and Bacterial Detection

## 4.1 Introduction

In this thesis, we focus on biofilms in ocular health because of the unique challenges these microbial communities present in this specific area of healthcare. The presence of biofilms forming in contact lenses, intraocular lenses and other surfaces of the eye can result in severe infections like bacterial keratitis that pose a substantial risk to vision [37]. These infections are difficult to treat, as biofilms are shielded from both host immune responses and standard antimicrobial therapies, complicating diagnosis and treatment [37].

We start this chapter with the structural and functional characteristics of biofilms, highlighting the limitations of traditional diagnostic methods in effectively detecting these complex structures. Then, we focus on the application of SERS within the context of detecting ocular infections. This chapter shows how this technology can enhance diagnostic precision and improve clinical outcomes in managing biofilm-associated infections.

After this foundational discussion, we apply these concepts to our research and discuss the results obtained when using our experimental setup to detect biofilms. We first conducted initial experiments under baseline conditions. These preliminary investigations were important for determining whether SERS could effectively detect these microbial communities and for providing a foundation for optimization experiments, already detailed in Chapter 3. We then analyze and discuss detection experiments using three sample types: biofilms, collagen, and a combination of biofilm and collagen under optimized conditions. Through these results, we lay the groundwork for our main goal of whether SERS can detect biofilms in corneal samples, which is later detailed in Chapter 5.

## 4.2 Definition and Structure of Biofilms

Biofilms are organized within the 'biofilm matrix' made of polysaccharides, proteins, lipids, and extracellular DNA [38]. Inside this matrix, we can find small clusters of cells mixed with channels that help the flow of water, nourishment and communication, important for distributing the resources essential for their survival [39][40].

Because the outer layers of a biofilm are exposed to the environment, they have easier access to nutrients and oxygen. This results in a higher metabolic activity among the microorganisms in these regions [41]. The deeper layers, on the other hand, where oxygen and nutrients are limited, contain cells in a slow-growing state [41]. This layering contributes to their overall resilience, helping the community to resist any adverse condition like antibiotic exposure.

The structural organization of biofilms also plays a critical role in their presence on surfaces, especially in clinical settings [41]. The sticky EPS matrix attaches the biofilm to these surfaces making it challenging to remove through mechanical or chemical methods [41].

## 4.3 Biofilms in Biomedical Settings and Ocular Health

In healthcare, biofilms are commonly associated with infections related to implanted devices such, as prosthetic joints, heart valves and catheters [42]. These infections are difficult to treat, requiring prolonged antibiotic therapy and, in some cases, surgical removal of the infected device [42].

### 4.3.1 Biofilms in Biomedical Settings

Devices like urinary catheters, intravenous lines, and endotracheal tubes provide ideal surfaces for biofilm development [42]. The interchange of genetic material, including genes that provide resistance to antibiotics, is facilitated by the close alignment of bacteria within a biofilm, making treatment even more challenging [41]. Biofilm development on these surfaces also lead to chronic infections that need extensive medical care [43][44]. For example, biofilms on prosthetic joints can lead to infections that may require the removal and replacement of the joint, significantly increasing patient morbidity and healthcare costs [43]. Similarly, biofilms on heart valves can result in endocarditis, a life-threatening condition that demands a thorough treatment [44].

### 4.3.2 Biofilms in Ocular Health

The transparent, dome-shaped cornea at the front of the eye serves as crucial for concentrating light and protecting the eye from contaminants. [45]. It consists of multiple layers, including the epithelium, which serves as a protective barrier, and the stroma, which provides structural integrity and transparency [46]. Any disruption to these layers from microbial agents, can significantly impair vision, making the cornea highly susceptible to infection [46].

Bacterial keratitis is a leading cause of vision impairment [48]. It is estimated that microbial keratitis affects over 1.5 million people worldwide each year, with biofilm formation on contact lenses being a major contributing factor [47]. The warm, moist environment between the lens and the corneal surface creates an ideal breeding ground for these microorganisms [37]. This environment makes treating such infections challenging, requiring more aggressive and prolonged treatment strategies. Additionally, biofilms can form on other ocular surfaces, such as lenses used in cataract surgery [49]. These devices can cause post-operative infections that are hard to treat and can result in serious side effects like endophthalmitis, which can cause blindness [49].

Current diagnostic methods fail to identify biofilms in their early stages, leading to delays in treatment and worsening of the infection. Advances in detection techniques, such as SERS, offer promising solutions for identifying biofilms more rapidly and accurately, which could improve clinical outcomes.

## 4.4 Limitations and Challenges in detecting biofilms

One significant challenge in biofilm detection is the difficulty in visualizing and identifying biofilms using standard staining and imaging techniques [39]. Traditional methods like Gram staining often fail to penetrate the thick EPS matrix, resulting in under-detection of biofilm-associated bacteria [42]. Standard imaging techniques, such as light microscopy, may lack the resolution required to differentiate the complex, layered structures of biofilms in the early stages of their development [42].

Advanced imaging methods, such scanning electron microscopy (SEM) and confocal laser scanning microscopy (CLSM), provide better resolution but are impractical for normal clinical usage because they take a lot of time and require specialist equipment. [50]. CLSM requires staining of the samples and SEM needs conductive coating on samples to function effectively,

which are procedural complications that can be challenging for long term use [50]. Additionally, these techniques may still struggle to detect biofilms on intricate surfaces or within deep tissue layers, limiting their effectiveness in clinical diagnostics [51]. For example, Relucenti et al. noted that SEM, despite its high resolution, failed to detect early-stage biofilm formation on certain medical implants due to the biofilm's deep embedding within the material [51].

Another complication arises from the presence of dormant or metabolically inactive cells within biofilms, known as persister cells, as explored by Conlon et. al [52]. These cells can survive antimicrobial treatments that would typically eliminate active bacteria, making infections difficult to remove. They found that since many diagnostic techniques rely on detecting metabolic activity, they fail to identify these dormant cells, leading to incomplete diagnoses and ineffective treatments [52]. The small size and early-stage microcolonies within biofilms can evade detection, allowing infections to progress unnoticed until they become more severe and harder to treat [52].

Furthermore, the variability in biofilm formation across different surfaces and environments adds another layer of complexity to their detection [42]. This variability complicates the development of standardized detection methods that are effective across different clinical settings.

## 4.5 Use of SERS in Biofilm Detection

Given the challenges discussed earlier, there is a need for more advanced and reliable methods for biofilm detection. SERS enhances sensitivity and specificity, enabling the detection of biofilms at molecular levels and even in their early stages [53]. Unlike traditional methods, SERS can penetrate the EPS matrix and identify microbial signatures even bypassing the protective barriers [53]. Silva et al. found that conventional techniques fail to detect early-stage biofilms due to low microbial concentrations [54]. However, SERS offers a significant advantage, enabling molecular-level sensitivity even at low concentrations of biofilm-associated molecules. Chao et al. demonstrated that SERS could detect biofilm components at nanomolar concentrations, highlighting its potential in clinical diagnostics [55]. Lee et al. showed that SERS amplifies the Raman scattering signals of molecules within microbial cells, allowing for detailed characterization of the biochemical composition of biofilms [56]. This provides clinicians with the information needed to intervene before the infection progresses to more severe stages, such as corneal ulcers or vision loss [56].

The minimally invasive nature of SERS is another advantage relevant in ocular applications. The delicate tissues of the eye require diagnostic methods that do not cause additional harm or discomfort. Kong et al. highlighted that SERS could perform detailed molecular analysis without damaging ocular tissues, making it ideal for ophthalmology where preserving the integrity of ocular tissues is fundamental [57].

SERS can be integrated with other imaging techniques, such as confocal microscopy, to enhance the overall diagnostic process in ocular health. Chao and Zhang noted that combining SERS with confocal microscopy allows for simultaneous visualization and molecular analysis of biofilms at the cellular level [55]. This integration offers a multi-dimensional understanding of biofilm structure and function, which is valuable in complex clinical cases where understanding the full scope of biofilm involvement is necessary for successful intervention.

# 4.6 Methods

## 4.6.1 Sample Preparation

PA01 bacterial biofilms and rat tail collagen, provided by The Ottawa Heart Institute, were used in this thesis. Each sample came on a circular glass slide arranged within a 24-well plate. We then introduced 500 μL of freshly synthesized AgNPs into each well, ensuring that the samples were uniformly covered. The well plate was then covered with foil to protect them from light and left to interact at 4°C for 24 hours. This step was crucial for promoting the interaction between the nanoparticles and the samples. After the interaction period, the wells were rinsed twice with normal saline to remove any unbound nanoparticles, ensuring that only nanoparticles that had interacted with the biofilm or collagen remained attached to the slides.

After the rinsing process, the slides were carefully removed from the wells and transferred onto larger rectangular microscopic glass slides. They were then secured with coverslips to stabilize the samples for SERS analysis. Figure 10 shows the workflow of this sample preparation method.

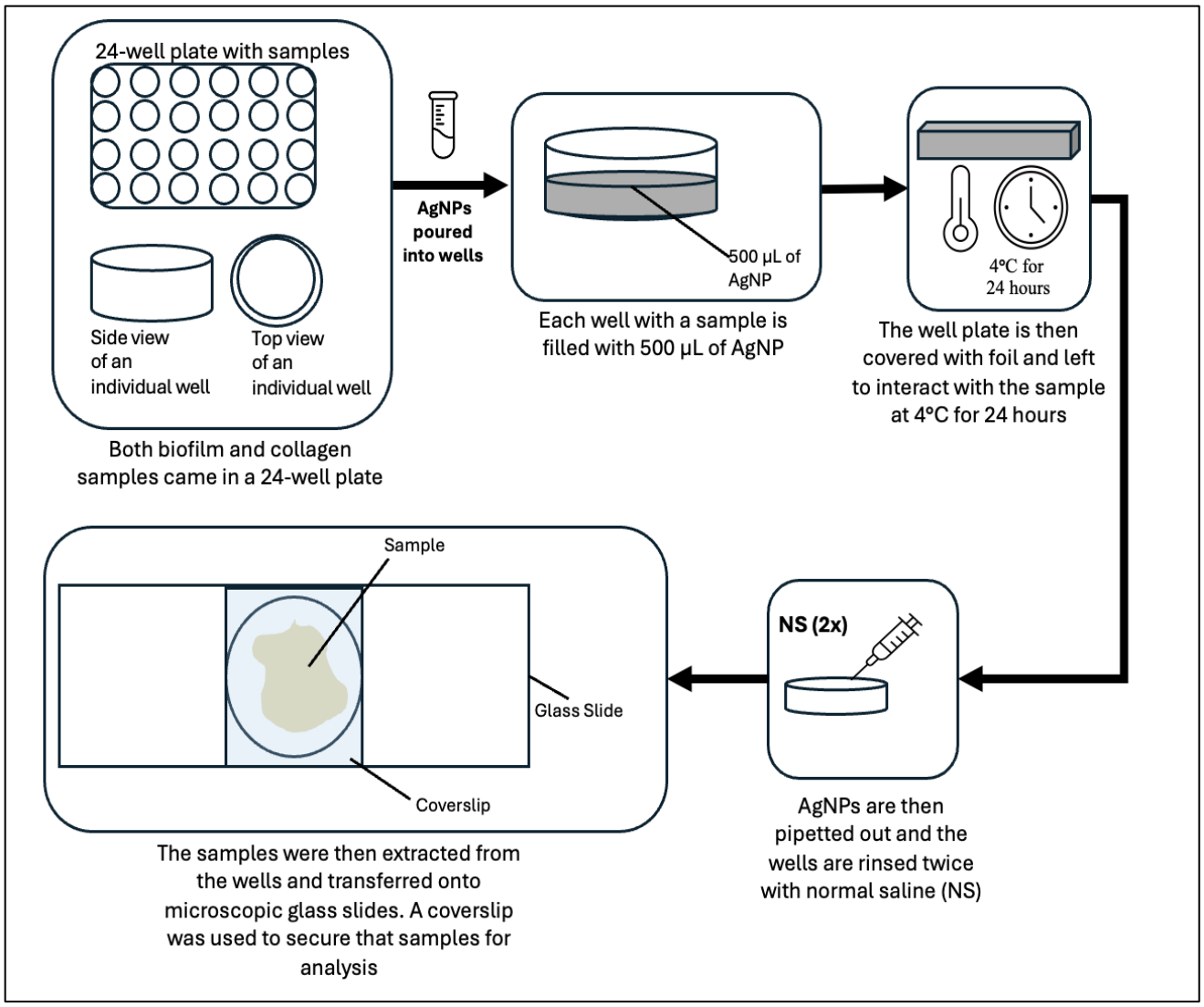


Figure 10: Illustration for the sample preparation of biofilm and collagen samples.

### 4.6.2 Data Collection and Processing

As outlined in Section 3.2.3, the integration time for each spectroscopic measurement was maintained at 15 seconds, over 10 accumulation cycles. This duration was chosen to prevent the saturation of the raw spectra, given the Andor spectrometer's signal limit, which is near 650000 AU. To obtain the SERS spectra, 20 spectra, with 10 accumulation cycles each, were averaged to get a spectrum of each sample type, biofilm, collagen and combined. Therefore, the total number of samples for each sample type was 200. All raw data was preprocessed according to the procedures outlined in Section 5.4.3. Representative microscopic images of biofilm-only,

collagen-only, and combined biofilm-collagen samples are displayed in Appendix A. These images provide a visual overview of each sample type, illustrating their distinct appearances.

## 4.7 Results and Discussion

### 4.7.1 Initial Findings: Spectral Analysis of PA01 Biofilms

In the initial stages of this research, the primary objective was to determine whether SERS could effectively detect the presence of *Pseudomonas aeruginosa* (PA01) biofilms under baseline conditions of nanoparticle concentration and interaction time. This foundational investigation was important for establishing the parameters needed to enhance the method's sensitivity and specificity, as detailed in Chapter 3.

In preparation for the detection of the biofilm, we ensured the accuracy of our spectroscopic setup through calibration with R6G. This step, detailed in Section 2.5, was essential to validate our instruments and methodologies before applying them to biofilm detection. The SERS spectra obtained from the *P. aeruginosa* (PA01) biofilms revealed several identifiable Raman bands, with a prominent peak at approximately 1360 $cm^{-1}$, as shown in Figure 11. This peak is attributed to pyocyanin, a virulence factor unique to PA01[58].

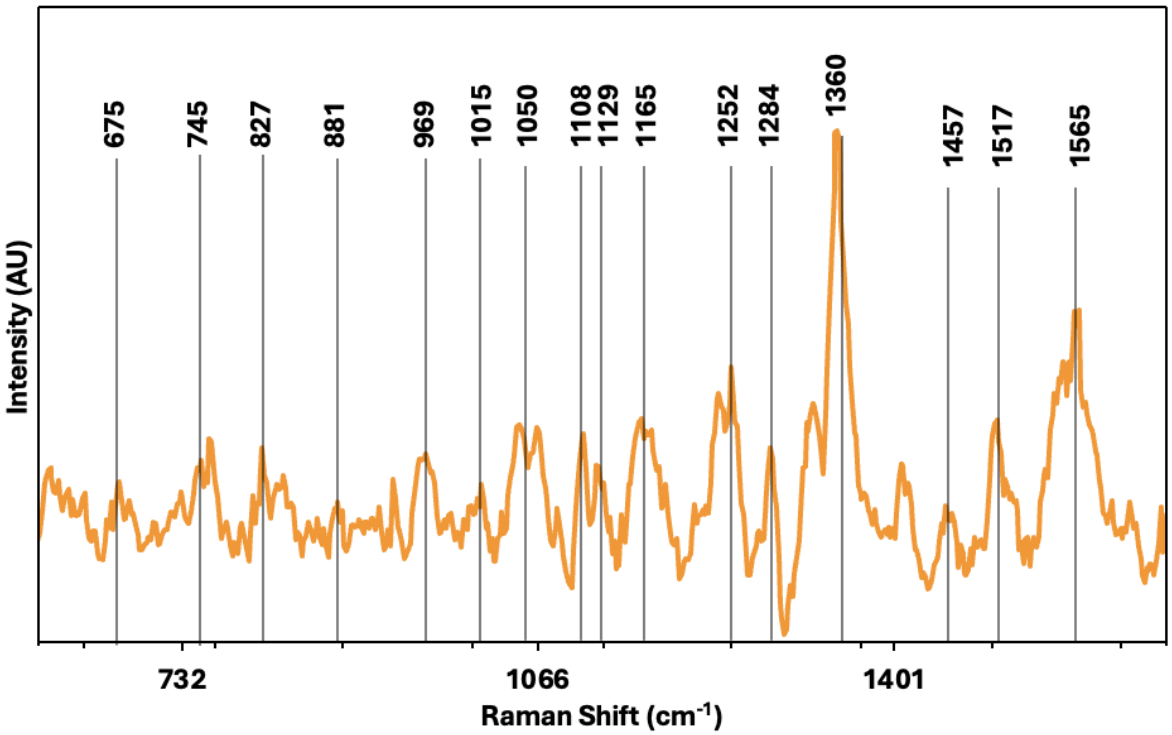


Figure 11: Initial Raman spectrum of PA01 biofilm.

Pyocyanin is a virulence factor that plays a critical role in the pathogenicity of PA01 [114]. It helps with bacterial colonization and enhancing biofilm resilience against host defenses and antimicrobial treatments [59]. As a chromophore— a molecule that absorbs light and produces a strong Raman signal—pyocyanin has a high Raman cross section, allowing it to stand out clearly in the spectrum and providing a strong indication of PA01 presence when detected using SERS [59]. The detection of pyocyanin through SERS highlights this technique's effectiveness for identifying PA01 biofilms, which can advise therapeutic strategies targeting biofilm-related infections.

Further spectral analysis highlighted bands at 1252 $cm^{-1}$ and 1284 $cm^{-1}$, corresponding to the Amide III band, indicative of the biofilm's protein content [58][60]. Peaks at 1015 $cm^{-1}$ and 1050 $cm^{-1}$ associated with phenylalanine and C-H bending were observed, highlighting the biochemical diversity within the biofilms and supporting the findings related to the protein-heavy nature of the biofilm [58][60]. A detailed tentative band assignment of Figure 11 can be found in the Appendix

B. While the initial spectral analysis confirmed that SERS could generate Raman signals from PA01 biofilms, the quality and intensity of these signals suggested room for improvement.

Our initial experimental conditions (1x nanoparticle concentration, 24-hour interaction time) were based on heuristic approaches commonly used in preliminary SERS studies [119]. However, as noted in the literature, these initial parameters are rarely optimal when dealing with complex biological samples like biofilms [61] [119]. The preliminary findings provided proof-of-concept for using SERS to detect PA01 biofilms under baseline conditions. However, the need for improved spectral clarity and intensity emphasized the importance of optimizing the experimental setup. The specific methodologies and the results of the optimization experiment have been discussed in Chapter 3.

### 4.7.2 Evaluation of SERS Spectra for PA01 and Collagen under Optimized Conditions

With the optimal experimental conditions identified in Chapter 3, this section examines the resulting SERS spectra of PA01 biofilms and collagen. We start with a band assignment for PA01 to identify any spectral shifts or enhancements. A similar examination is performed for pure collagen, which serves as an analog for corneal tissue. Finally, the combined spectra of PA01 and pure collagen are evaluated to explore any potential spectral overlaps.

***SERS Spectrum for PA01***

The optimization of experimental conditions resulted in noticeable shifts in peak intensities and the emergence of new peaks, which suggest a stronger interaction between the biofilm components and the nanoparticles under the optimized conditions.

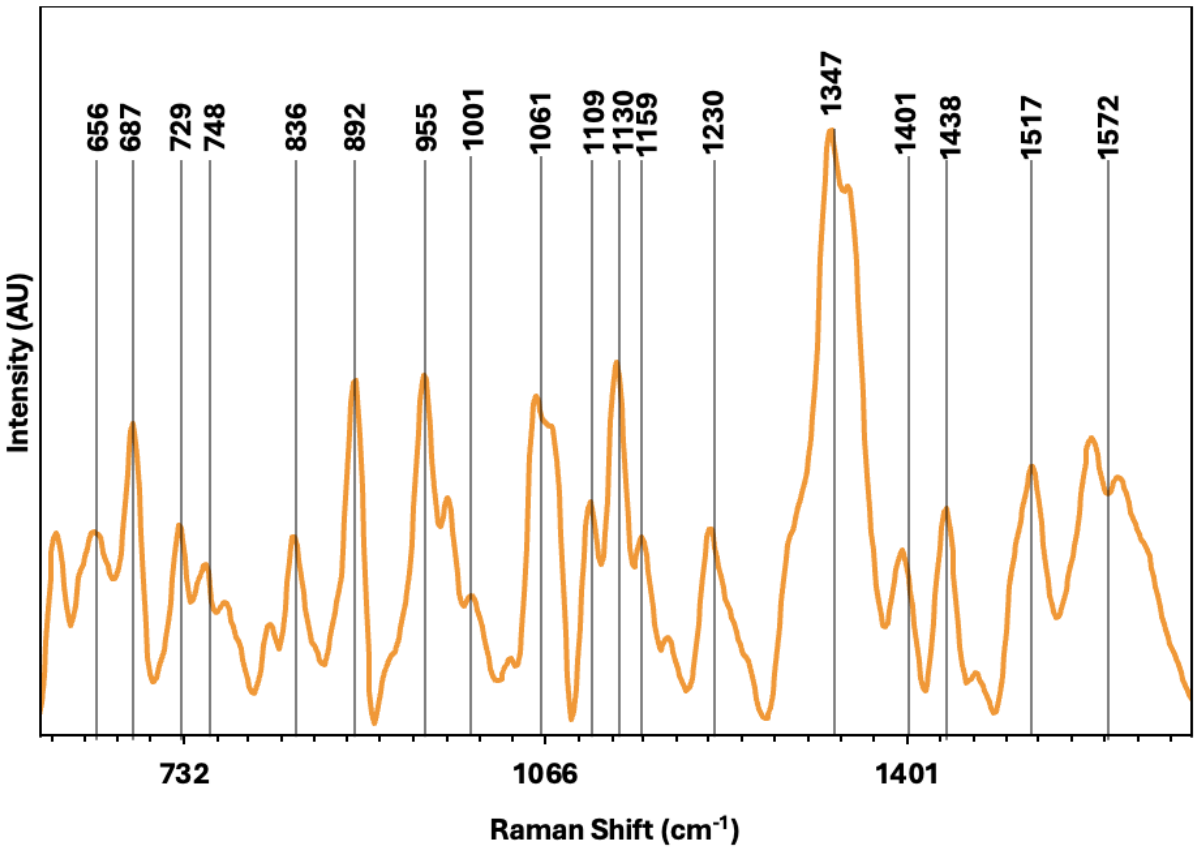


Figure 12: Raman spectrum of PA01 biofilm.

Figure 12 shows that the pyocyanin peak is now at 1347 $cm^{-1}$. As compared to the baseline conditions, it is now sharper and more pronounced [58]. This suggests that the increased concentration created a more favorable environment for pyocyanin adsorption and resulted in a stronger Raman signal intensity. This improvement aligns with studies where higher concentrations of nanoparticles were associated with better SERS signals because of increased electromagnetic enhancement effects that help in enhanced Raman scattering [55].

In terms of signal quality, the optimized conditions produced a clearer and more distinct spectral profile, with enhanced peak intensities crucial for reliable biofilm detection and analysis. The improved spectral clarity reduces any potential misinterpretation, making the SERS technique more robust for diagnostic applications. The detailed band assignments, showing these shifts and enhancements, are presented in Appendix C, Table 6.

### *SERS Spectrum for Pure Collagen*

The SERS spectra obtained from pure collagen samples help verify the system's ability to accurately capture the spectral characteristics of collagen. Establishing this baseline is important for differentiating between healthy tissue and biofilm signals in clinical simulations, which is elaborated in Chapter 5.

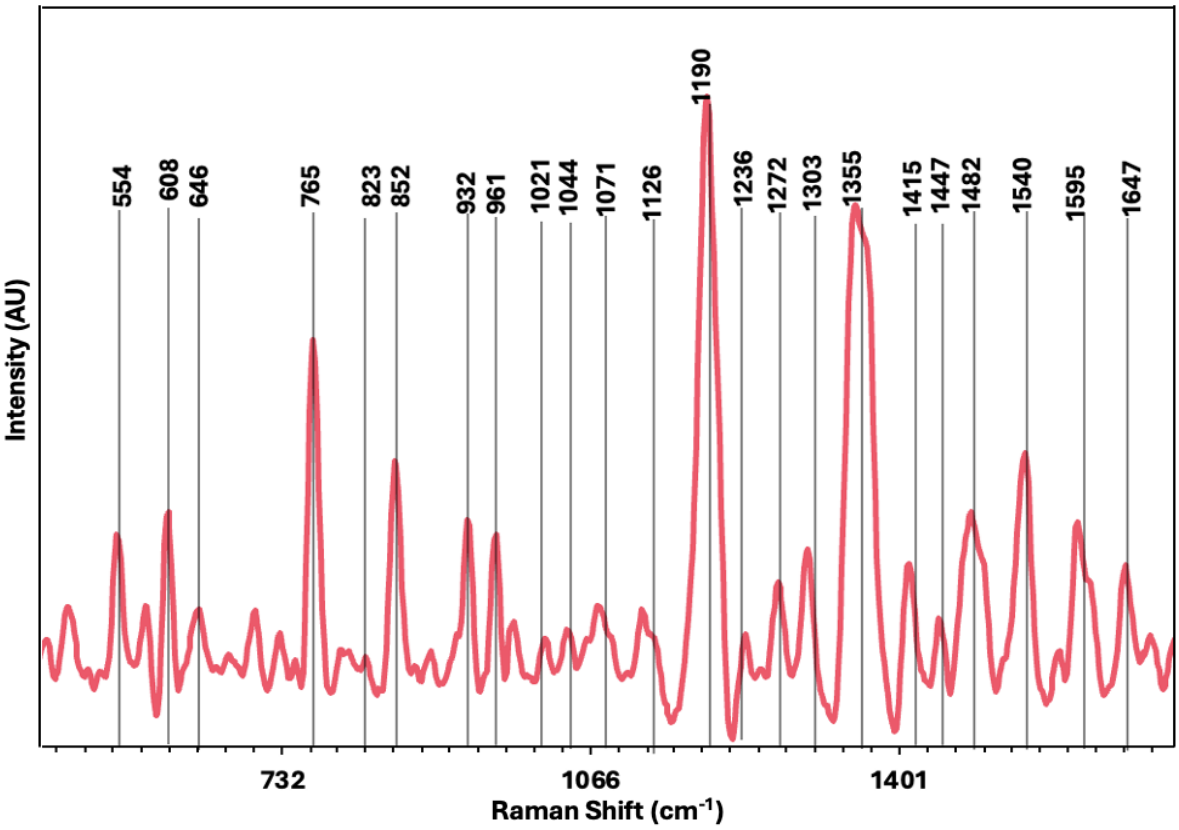


Figure 13: Raman spectrum of collagen.

The Raman spectrum of collagen, shown in Figure 13, present distinctive bands including the Amide I band at 1647 $cm^{-1}$ and the Amide III peaks at 1236 $cm^{-1}$ and 1272 $cm^{-1}$, associated with collagen's structural components [62][63]. Peaks at 852 $cm^{-1}$ and 932 $cm^{-1}$ indicate the presence of proline and hydroxyproline. These peaks are significant in identifying the presence of collagen in tissue and are critical for understanding the structural integrity within the extracellular matrix [63][64].

A notable overlap between collagen and PA01 spectra is observed at 1355 $cm^{-1}$. In collagen, this peak corresponds to the Amide III band, involving C-N stretching and N-H bending [64], while in PA01, it is linked to pyocyanin. This overlap stresses the complexity of spectral differentiation in clinical settings, where accurate interpretation of overlapping peaks is essential for reliable diagnostics. The detailed band assignments for pure collagen are presented in Appendix C, Table 7.

***SERS Spectrum for Collagen with PA01 Biofilm***

The SERS spectrum obtained from samples combining PA01 biofilm and collagen present a complex profile of molecular interactions between these two materials. The spectrum revealed overlapping peaks that are characteristic of both biofilm and collagen components, making it a challenge to assign bands to specific molecular vibrations without ambiguity.

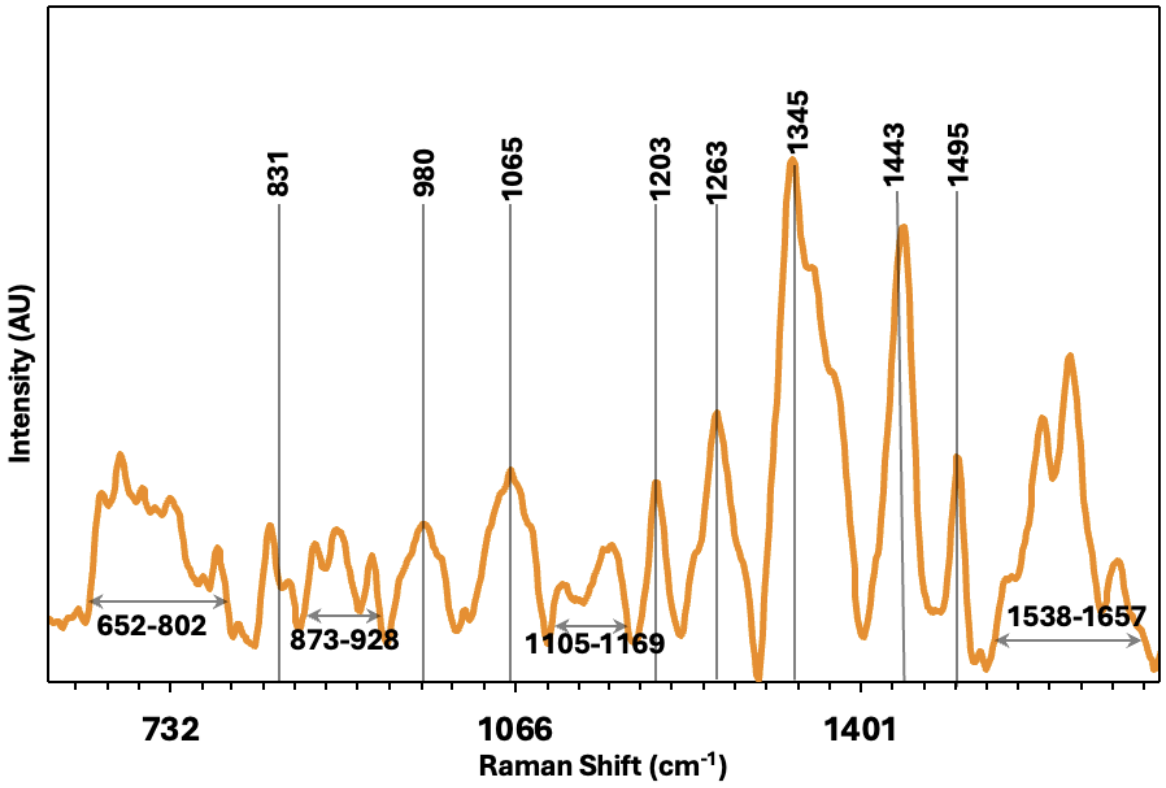


Figure 14: Raman spectrum of collagen with PA01 biofilm.

As shown in Figure 14, the spectrum shows peaks for both PA01 biofilm and collagen, emphasising the complex nature of the combined sample. The band assignment for this spectrum, detailed in Appendix C, Table 8 highlights the most significant features. The spectrum showed a broad peak range between 652-802 $cm^{-1}$, indicating overlapping molecular vibrations from both the biofilm's polysaccharide matrix and the collagen fibrils, suggesting extensive interaction [65].

The peak at 1345 $cm^{-1}$ prominent in both PA01 and collagen spectra [58]. The overlap of these peaks in the combined sample demonstrates the interaction between pyocyanin and collagen, which alters the spectral signature typically observed in separate analyses.

Additionally, the Amide bands between 1105-1169 $cm^{-1}$ and 1538-1657 $cm^{-1}$, linked to protein structures, reflect significant interactions between the collagen matrix and the biofilm's extracellular components [66]. The broadening of these peaks indicates a heterogeneous environment at the biofilm-tissue interface, complicating the assignment of specific molecular vibrations.

The SERS spectra from the combined samples of PA01 and collagen show a complex pattern of molecular interactions. These overlapping signals complicate the ability to differentiate between biofilm and tissue, even though SERS effectively captures both. The complexity highlighted by these results points to the need for advanced analytical methods and computational models as described in Chapter 5.

## 4.8 Challenges and Reflection

Acquiring the samples from the team at the Ottawa Heart Institute was one of the earliest challenges we faced. This required proper coordination and collaboration with the other team, often resulting in delays. We had to follow a specific protocol to carefully transfer the samples. It was important to ensure that the samples were not shaken and were kept at a stable temperature of 4°C during transit.

Obtaining a clear and consistent SERS signal from the biofilm samples was an additional setback. Despite following established protocols, the initial spectra proved to be weak and noisy, lacking the necessary detail for reliable analysis. This issue stayed even after multiple adjustments to the experimental setup, such as varying the laser intensity, acquisition times, and excitation wavelengths. It soon became evident that achieving the desired signal required a deeper

understanding of the factors influencing the SERS effect. This realization led to extensive troubleshooting and multiple iterations of the experimental setup. Ultimately, a clear signal was achieved, but this was only after considerable effort.

Another major challenge was the inherent variability in biofilm composition, which led to inconsistent SERS spectra. Because of their heterogeneous structure even small differences in composition or thickness could result in significant variations in the spectra obtained. This variability made it difficult to consistently replicate results and required a careful approach to both sample preparation and data collection.

In some cases, different spectra were obtained from different points on the same biofilm sample, depending on where the laser was focused. To address this issue, various strategies were explored, including averaging spectra from multiple points. We wanted to make sure that our results are reproducible with the conditions we set. Although, these adjustments ultimately led to more reliable results, they were lengthy and involved a process of trial and error.

## 4.9 Conclusion

The initial baseline experiments provided a foundation for refining the experimental conditions. Optimized interaction times and nanoparticle concentrations led to enhanced Raman signal intensities, with pyocyanin serving as a strong biomarker for PA01 detection. The spectral analysis of biofilm, collagen, and their combined samples revealed key molecular interactions and highlighted challenges, such as spectral overlaps between biofilm components and collagen. With these findings, we provide the foundation for exploring the use of machine learning in the following chapter, where we use this experimental setup to categorize pig corneal samples and assess the classification performance of SERS and ML.

# Chapter 5

# Machine Learning in SERS Data Analysis

## 5.1 Introduction

In many scientific fields, machine learning (ML) is being used as a vital tool for analyzing large, complicated datasets [67][68]. ML is a group of techniques that let computers learn from data, spot trends, and make judgments or predictions without needing to be specifically programmed for each activity [67][68]. This data-driven approach has made ML valuable in the biomedical sciences, where the increasing complexity and volume of data pose significant challenges [68].

Biomedical data exhibits high dimensionality, variability, and noise, which can limit traditional statistical methods [69]. ML addresses these challenges by offering robust techniques to identify meaningful patterns in such complex datasets, making it a powerful asset for applications such as disease diagnosis, treatment prediction, and personalized medicine [70][71].

ML methodologies can be divided into two main categories: supervised and unsupervised learning [67][68]. In supervised learning, every input is associated with a known outcome, and models are trained using labeled datasets [67][68]. The goal is for the model to learn and make accurate predictions on new, unseen data—useful in tasks such as classification and regression. In biomedical contexts, supervised learning is frequently employed to classify patient data or detect disease patterns [68].

Unsupervised learning, on the other hand, is applied to unlabeled data, where the model seeks to uncover hidden patterns or structures without explicit guidance [67]. This is commonly used in clustering, grouping similar data points, or dimensionality reduction, which simplifies datasets by retaining essential features [67]. Unsupervised learning is useful for exploratory data analysis, where insights or hypotheses are generated directly from the data [68].

In this thesis, ML is applied to enhance the analysis of the SERS data that is collected. The goal of this project is to increase the accuracy and reliability of SERS-based diagnostics by combining supervised and unsupervised machine learning approaches. Unsupervised methods are employed for data preprocessing and clustering, while supervised approaches are used for classification tasks.

This chapter begins by exploring the broader role of ML in biomedical research and then transitions to the specific applications of ML in analyzing SERS data. After this, the algorithms implemented in this study are discussed, and the chapter concludes with an evaluation of the results, focusing on the performance of ML models in detecting biofilms in corneal tissue samples.

## 5.2 ML in Biomedical Applications

One key area where ML has made an impact is in the processing of biomedical signals. For example, Hannun et al. developed a neural network model capable of classifying 12 types of cardiac arrhythmias from single-lead ECG recordings, achieving an overall accuracy of 94%. This model outperformed average cardiologists in diagnostic tasks, demonstrating the potential of ML to augment clinical decision-making [72].

Another study where ML is used for signal processing is to analyse electroencephalogram (EEG) data. Acharya et al. applied ML algorithms to predict epileptic seizures based on EEG signals, achieving sensitivity rates as high as 97% [73]. The study showed how ML could improve patient outcomes by providing real-time seizure alerts, thereby enabling timely medical intervention [73].

Additionally, ML applications extend into microbiology, in combating bacterial infections and antibiotic resistance. A study by Nguyen et al. used different supervised learning techniques to predict antibiotic resistance from different pathogens like *Escherichia coli* [74]. By analyzing whole-genome sequencing data, they predict antibiotic susceptibility with up to 94% accuracy, helping to enhance clinical decision-making in treating bacterial infections [74].

## 5.3 ML in Biomedical SERS Applications

In biomedical research, machine learning approaches are being used extensively to interpret the high-dimensional and complicated data produced by SERS. Among these techniques, PCA, SVM, and K-means clustering are widely used for their ability to efficiently handle intricate spectral datasets, making them ideal for diagnostic applications.

For example, Li et al. demonstrated the effectiveness of PCA combined with SVM in diagnosing Crohn's disease through non-invasive analysis of SERS spectra from urine samples, achieving a classification accuracy of 82.5% [75]. This approach reduces data dimensionality and improves classification performance, allowing SVM to distinguish between disease and control groups.

K-means clustering has also been integrated with SVM to enhance classification in cancer diagnostics. Wang et al. applied a weighted K-means SVM model for cancer prediction, emphasizing the reduction in computational complexity by clustering the data prior to classification [76]. Although specific accuracy metrics aren't always highlighted, the technique has proven to improve performance in imbalanced datasets common to cancer studies.

In other biomedical applications, Random Forests and Decision Trees have been employed for SERS-based cancer diagnostics [77]. For example, Random Forests have been used effectively by Liu et al. to identify prostate and bladder cancers from SERS data by analyzing serum samples, achieving a classification accuracy of 96.5% [77]. These models excel at managing in handling high-dimensional data and preventing overfitting, which is beneficial given the complexity of SERS spectra.

While PCA, K-means, and SVM have proven highly effective in handling the types of SERS data used in this thesis, newer techniques such as Convolutional Neural Networks (CNNs) have also shown promise for more complex datasets. For example, Li et al. applied CNNs to mRNA biomarker detection in cancer diagnostics outperforming other methods like support vector regression [78]. However, the choice of ML technique depends on the nature of the dataset. For moderate-sized, complex datasets—such as those commonly encountered in biomedical SERS studies—methods like PCA, K-means, and SVM remain highly effective, offering a balance between computational efficiency and the interpretability of results [79].

### 5.3.1 Integration of Supervised and Unsupervised Methods

In this thesis, unsupervised and supervised learning methods were integrated to address the challenges of variability and classification in the data. Preprocessing steps were employed to clean the raw spectra and enhance their comparability. These initial steps are crucial for reducing noise and standardizing the data, making it suitable for further analysis, as demonstrated by Byrne et al., who noted the significance of preprocessing in ensuring clearer spectral signals for SERS analysis [80].

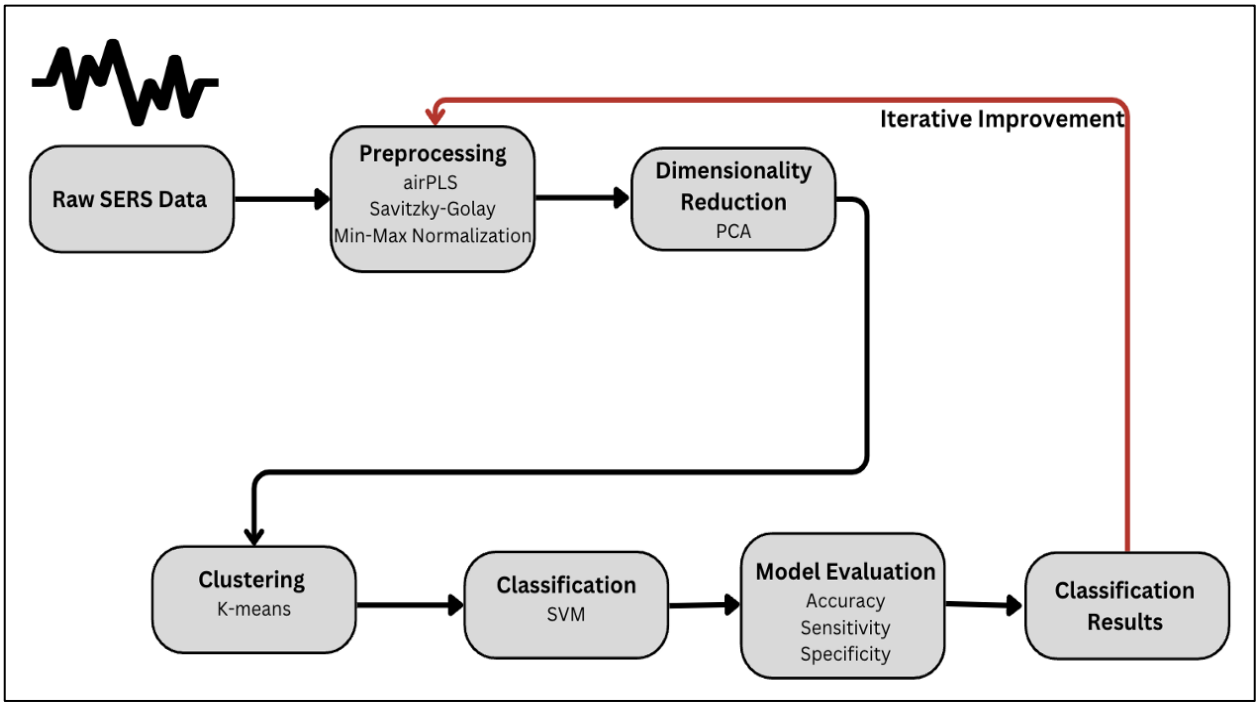


Figure 15: Flowchart for data analysis and classification.

PCA was then used to reduce the dimensionality of the SERS spectra. This step is critical in handling the vast data without losing essential chemical information, a technique underscored by Cao et al. as pivotal in SERS data analysis. PCA simplifies the dataset, maintaining the most informative features which are crucial for the efficient clustering and classification that follows [81].

After the PCA, K-means clustering grouped the data based on spectral similarities. This unsupervised approach is very helpful in capturing within-sample variance, critical for managing tissue variability in biological samples like the cornea or biofilms. Clustering not only enhances

data clarity but also isolates distinct groups, making it easier for subsequent classification stages. Hedegaard et al. have highlighted the utility of clustering in distinguishing meaningful spectral patterns from noise, in the analysis of Raman spectra from single cells [17].

The SVM classification was chosen for its robustness in handling complex and high-dimensional data spaces typical of SERS spectra. Dong et al. have noted that SVM performs well even with limited training data, making it suitable for biomedical applications where the feature space may be extensive despite smaller sample sizes [82]. Its application is justified by its resistance to overfitting, ensuring reliable classifications of the spectral data, as further discussed by Cao et al.[81].

This integrated approach effectively uses the strengths of unsupervised methods to address within-sample variability and enhance spectral clarity, while the supervised SVM method classifies the distinct biological conditions with high accuracy. The final pipeline, depicted in Figure 15, shows an iterative feedback loop where model evaluation metrics like accuracy, precision, and recall guide the continuous refinement and optimization of the classification process. Details on implementing each part of this model are further elaborated in Section 5.4.3.

# 5.4 Methods

## 5.4.1 Sample Preparation

The BeATS laboratory from the Ottawa Heart Institute provided both infected and non-infected pig cornea samples. They were provided in centrifuge tubes mixed with or without bacterial cultures in LB Broth. A 1:1 ratio of synthesized silver nanoparticles (AgNPs) at a 1.75x concentration in relation to the sample volume was applied to each sample. The samples were placed on slides and maintained in a humidity-controlled chamber for 24 hours at 4°C to simulate clinical diagnostic conditions closely.

After the incubation, the cornea samples were air-dried under dust-free conditions to prepare them for subsequent SERS analysis. Four replicates were prepared for each set of conditions—infected and non-infected cornea—to ensure the reliability and reproducibility of the results. This experimental procedure is illustrated in Figure 16.

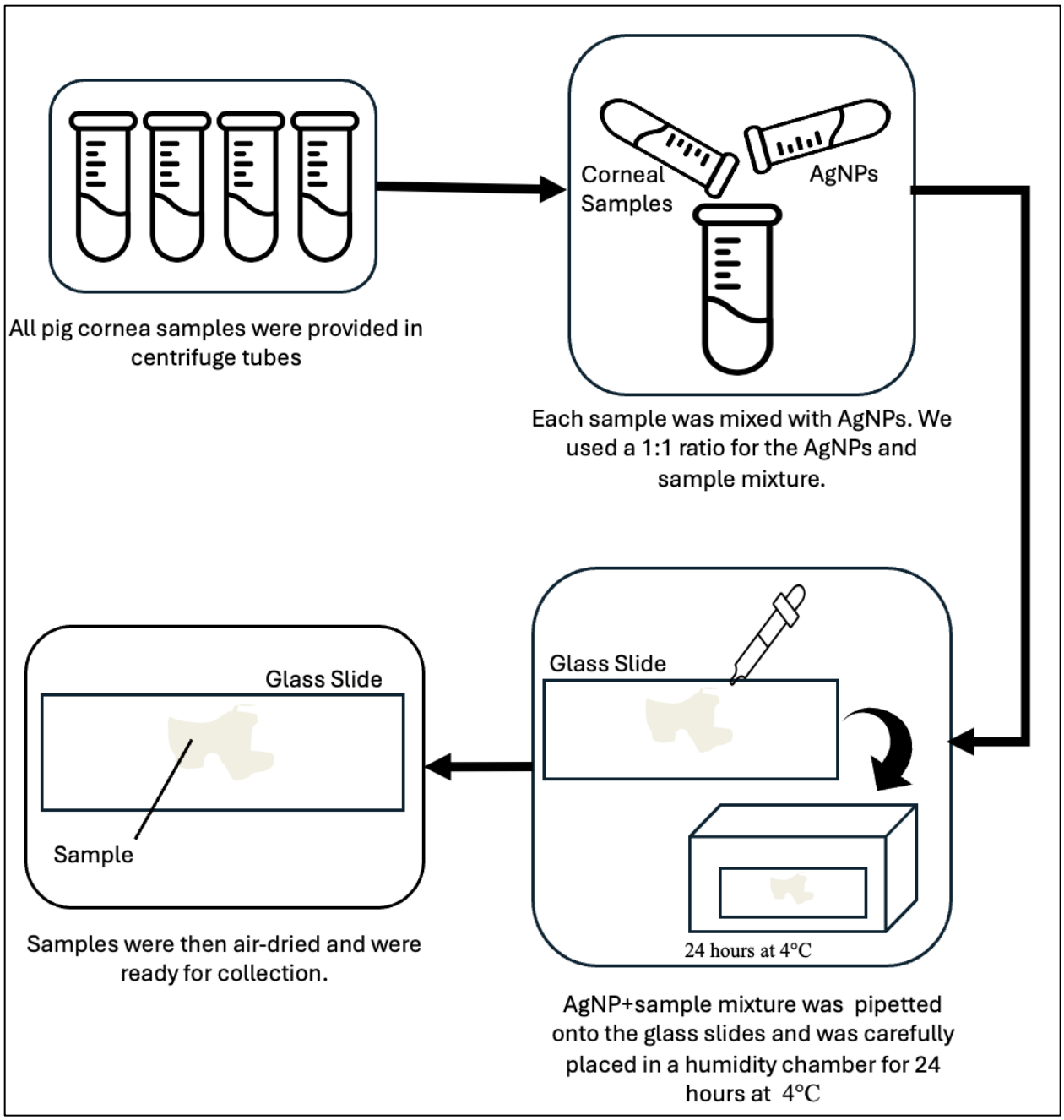


Figure 16: Illustration of sample preparation steps for the pig cornea samples.

### 5.4.2 Data Collection and SERS measurement

Data collection procedures were like the procedure detailed in Section 4.6.2. To obtain the SERS spectra, 27 spectra, with 10 accumulation cycles each, were averaged to get a spectrum of non-infected cornea samples. 24 spectra, with 10 accumulation cycles each, were averaged to get

an average spectra of infected cornea samples. Each of these spectra can be visualized in the images shown in Appendix A Figure 23. So, in total, 270 spectra were used for the non-infected cornea samples and 240 for the infected cornea samples. Representative microscopic images of infected and non-infected cornea samples are shown in Appendix A Figure 24. These images provide a visual overview of each sample type and show their differences. The preprocessing steps and the analysis of all raw data is detailed in the next section.

### 5.4.3 Algorithm Implementation

A previous student in our lab developed a MATLAB code for bacterial classification, which used airPLS for preprocessing the SERS spectra [107]. Building on this idea, I developed a new code in Python. Some elements of the MATLAB code were recreated, such as the method to load the files, the background subtraction using airPLS, and the normalization steps. However, some improvements were also introduced.

These include Savitzky-Golay filtering for spectral smoothing and PCA for dimensionality reduction. Additionally, k-means clustering and an updated SVM classifier were incorporated to handle the pig cornea classification. All the data processing and classification were done using this newly developed Python code. The dataset was split between 30% testing and 70% training sets in order to assess SVM classification performance. The entire code can be found in the GitHub repository linked in Appendix G.

#### *5.4.3.1 Preprocessing of SERS Spectra*

Preprocessing involved baseline correction using the Adaptive Iteratively Reweighted Penalized Least Squares (airPLS) [83], noise reduction through Savitzky-Golay (SG) filtering [84], and normalization via Min-Max scaling [85]. This ensured the data's readiness for advanced analysis, which is crucial for maintaining chemical integrity within the SERS spectra.

- ***Baseline Correction:***

Baseline drift arises from instrumental fluctuations or sample heterogeneity. airPLS iteratively adjusts and subtracts the slow-varying background noise, enhancing the true signal and preserving spectral peaks [83]. This preservation is crucial for retaining all chemical information for accurate analysis. For the purposes of our data, we set the lambda ($\lambda$) parameter of the technique to 10,000 for optimal smoothness and the polynomial order to 2 to capture the moderate baseline curvature

of our biofilm spectra. Additionally, we allowed up to 100 iterations to ensure convergence. airPLS works by iteratively adjusting weights for each data point and down-weighting those above the baseline in successive cycles. When background noise is reduced, the algorithm stops, producing a clean baseline highlighting important spectral patterns for additional examination.

- ***Smoothing:***

After baseline correction, the SG filter is applied to reduce high-frequency noise and enhance signal clarity. The filter parameters, such as window length and polynomial order, are carefully chosen based on the specific characteristics of the SERS data. In our case, we set the window length to 13, meaning the filter considers 13 neighboring points around each data point to compute the smoothed value. A higher window length creates a smoother curve but can risk over smoothing if set too high. The polynomial order was chosen as 2, balancing noise reduction with the preservation of spectral peaks and details.

- ***Normalization:***

By adjusting the spectra to a common scale (range [0,1]), we minimize discrepancies caused by variations in sample concentration or measurement conditions [85]. This step is important in SERS analysis, as it ensures that differences in overall intensity between spectra do not confuse the analysis, allowing for a more accurate comparison of spectral features.

Examples of these techniques being applied to an example of SERS PA01 Biofilm sample spectra can be found in Appendix D.

#### *5.4.3.2 Unsupervised Learning: Clustering Techniques*

- ***Principal Component Analysis (PCA) [86]:***

PCA is used for dimensionality reduction. PCA creates a new set of uncorrelated variables called principal components by projecting the high-dimensional data onto directions that maximize variance. First, PCA calculates the covariance matrix of the data, which helps identify the direction in which the data varies the most [124]. This variance is calculated using [127]:

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \tag{11}$$

where Σ represents the covariance matrix, $x_i$ is each data point, $\bar{x}$ is the mean of the data points and $n$ is the total number of data points. This measures how much each dimension varies from the mean with respect to each other [124]. After getting this matrix, the eigenvectors and eigenvalues are calculated using [127]:

$$\Sigma v = \lambda v \tag{12}$$

where $v$ represents an eigenvector and $\lambda$ is its corresponding eigenvalue. These eigenvectors define the principal components, and their corresponding eigenvalues determine their significance [124]. Higher the eigenvalue, the more variance that component explains [125]. The highest eigenvalue is the direction along which the data varies the most.

In Figure 17, we can visualize this transformation for our SERS data. The first two principal components, PC1 and PC2, representing the most significant data variance, are used to project the data from its original high-dimensional space into a new plane. The blue and red arrows in the plot represent the directions of maximum variance (PC1) and the second-highest variance (PC2), respectively. This transformation shifts the data from a potentially correlated space to one where principal components are uncorrelated, making it more suitable for further analysis [125].

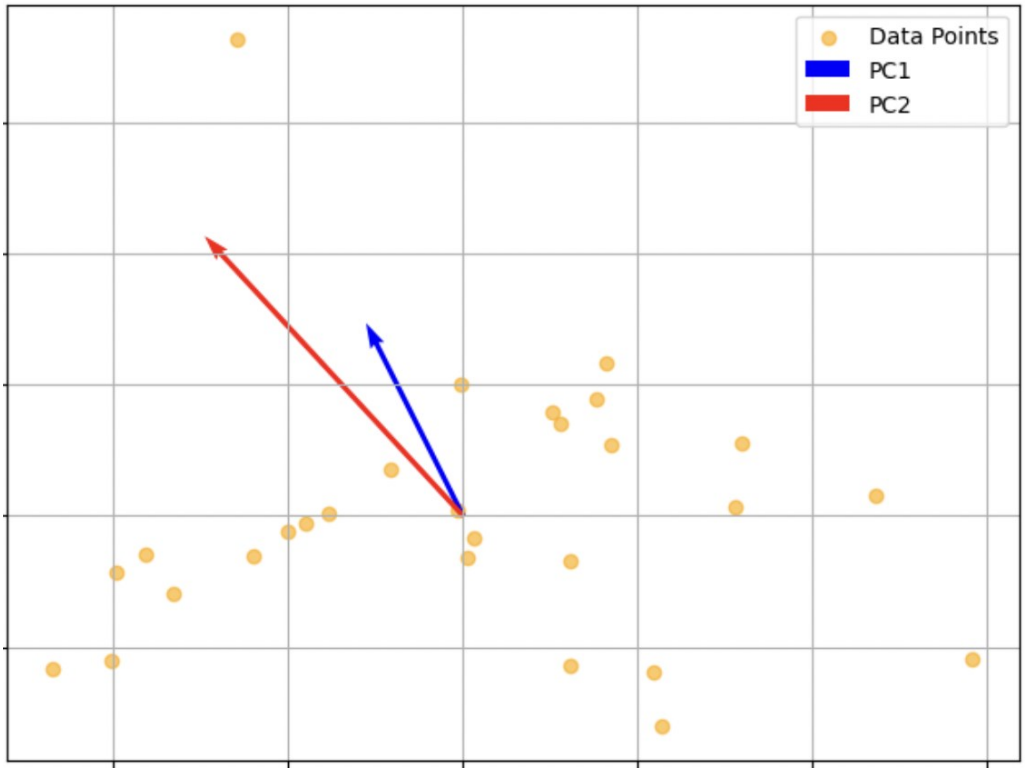


Figure 17: Visualization of Principal Component Analysis (PCA) on our SERS data.

The explained variance plot (Appendix F Figure 27) illustrates that while the first two components capture a significant portion of the variance, our analysis utilizes approximately 30 components, cumulatively accounting for nearly 95% of the total variance. PCA preserves vital variance while simplifying data complexity by keeping only the most essential components. The number of primary components for additional analysis is chosen based on this cumulative variance contribution.

- ***K-means Clustering [87]:***

After dimensionality reduction, K-Means clustering organizes the data into distinct groups based on spectral similarities. After the data is divided into clusters, every data point is allocated to the cluster with the closest centroid. Figure 18 illustrates this process as applied to our SERS dataset, showing the state of the data both before and after applying K-means. In the left panel, the data is shown with ground truth labels, while the right panel demonstrates how our data is clustered after K-means, with each point assigned to the cluster closest to its centroid (indicated by the red 'X'). This step is crucial for enhancing the clarity of the spectral features, thereby aiding in noise

reduction and improving the interpretability for further classification. K-means is mathematically expressed as [88]:

$$\min_{C} \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - \mu_k\|^2 \tag{13}$$

where $C_k$ represents the k-th cluster, $x_i$ is a data point, and $\mu_k$ is the centroid of the k-th cluster. The iterative process of updating centroids and reassigning data points continues until the clustering converges, minimizing the within-cluster variance.

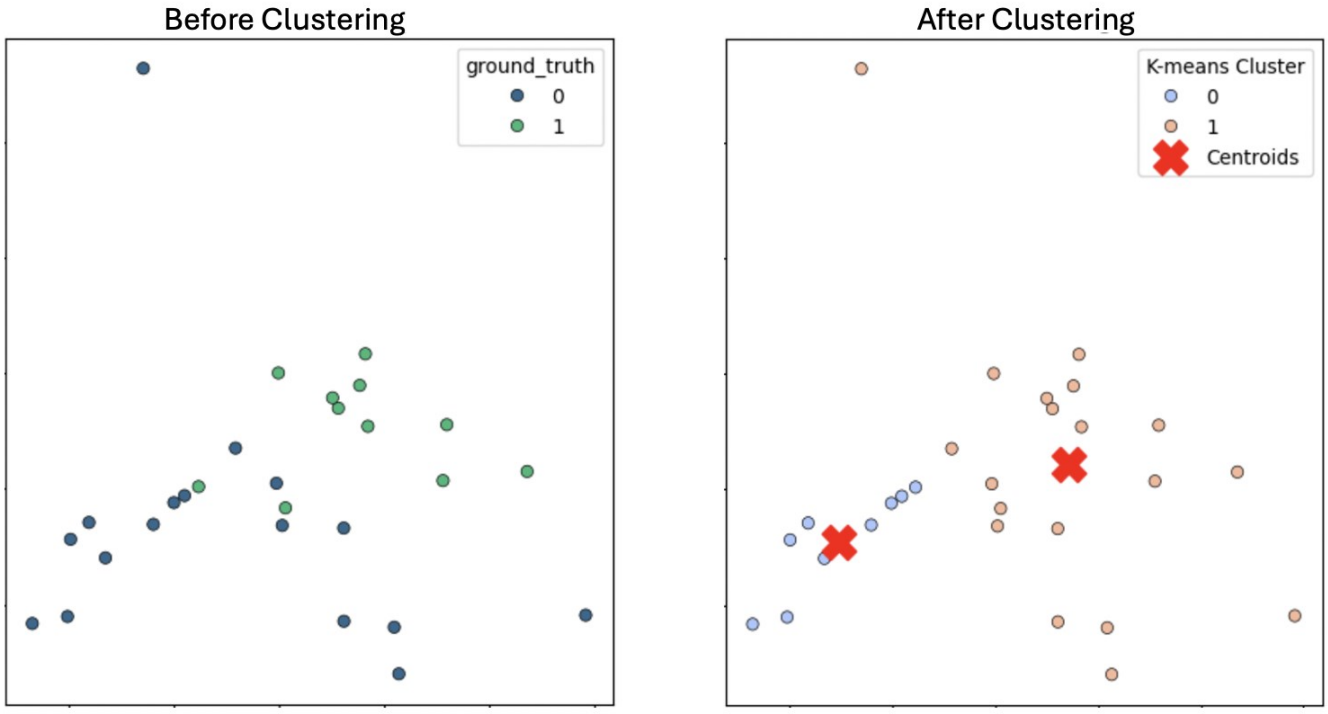


Figure 18: Visualization of the effect of K-means clustering on our corneal SERS data.

The combination of PCA and K-Means clustering simplifies the complexity of our SERS data while retaining important features and organizing them into meaningful groups as visualized in Figure 18. These steps effectively prepare the data for the subsequent classification tasks.

Once the data has been preprocessed and clustered, it is ready for classification using supervised learning techniques, which are discussed in detail in the next section.

#### *5.4.3.3 Supervised Learning for Classification*

After categorizing the SERS data into distinct clusters, we moved on to the supervised learning phase. This phase classified the spectral data into defined categories that indicate the presence or absence of PA01. SVM was chosen due to its proficiency in managing high-dimensional spaces and robustness in achieving distinct class separations.

- ***SVM Implementation***

The core objective of an SVM is to establish a hyperplane that optimally separates the classes with the maximum margin [89]. This is formally represented by the equation [89]:

$$f(x) = w^T x + b \tag{14}$$

where $w$ is the weight vector perpendicular to the hyperplane, $x$ is the input feature vector, and $b$ is the bias term. The objective of SVM is to find the values of $w$ and $b$ that maximize the margin, while ensuring that the data points are correctly classified [89]. Support vectors are the data points that are closest to the hyperplane and have a direct impact on its position. These points are crucial for defining the margin, which separates each class's closest data point and the hyperplane. By maximizing this margin, SVM aims to increase the classification accuracy. This can be seen in Figure 19 (A). The blue and the orange points represent two different classes.

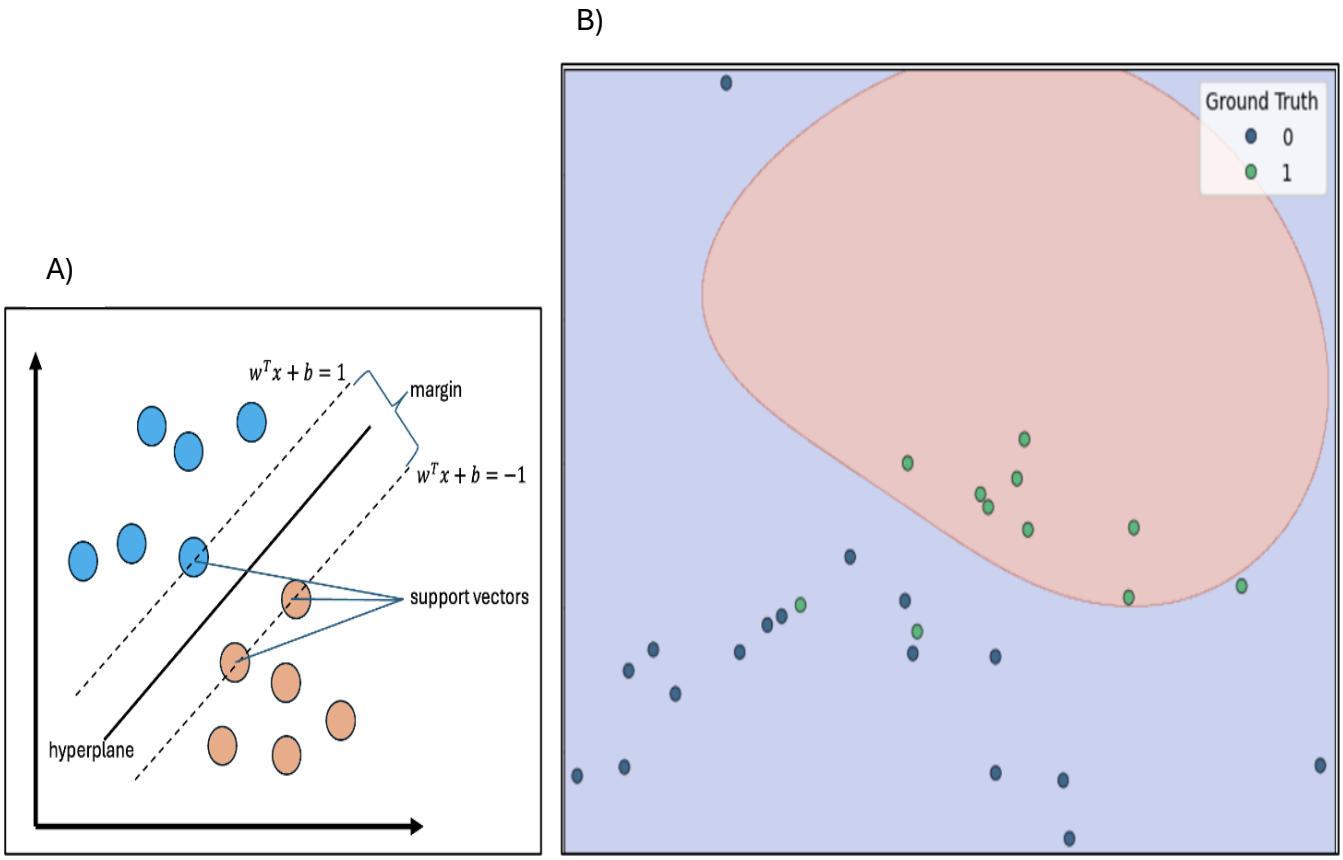


Figure 19: A) Visualization of support vector machines on linearly separable data. B) SVM RBF kernel decision boundary for corneal samples SERS data.

For data that is not linearly separable, SVMs employ kernel functions to enable separations in higher-dimensional spaces without explicit transformation computations [90]. The kernel trick allows the SVM to operate in this transformed space without explicitly computing the transformation, thus maintaining computational efficiency [90]. Common kernels include the radial basis function (RBF) and polynomial kernels [91]. The input data is mapped into a higher-dimensional space using the RBF kernel, making a linear separation possible [91]:

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{15}$$

where $\gamma$ is a parameter that defines the influence of a single training example [129]. The exponential term resembles an infinite series expansion like a Taylor series, which expands functions into infinite sums of polynomial terms [128]. This helps illustrate how the RBF kernel

is helpful in capturing the non-linear relationships by approximating infinite mappings [128]. This capability makes the RBF kernel an effective choice for classifying SERS data in biofilm detection, where the relationships between data points are not straightforwardly linear. Figure 19 (B) shows the SVM RBF kernel decision boundary applied to our PCA-reduced SERS data. This plot illustrates how the SVM model separates the two classes—representing the presence and absence of PA01 biofilms—using a non-linear boundary enabled by the RBF kernel. The shaded areas on the plot depict the regions the SVM classifier associates with each class. The red region indicates the area classified as biofilm-present, while the blue area represents biofilm-absent samples. The points in green and blue represent individual spectral data points after PCA reduction, making it possible to see how the SVM decision boundary aligns with the data distribution in this transformed, lower-dimensional space.

The RBF kernel's ability to transform the data into a higher-dimensional space shapes the boundary itself, enabling the model to establish a separation that is not linearly achievable in the original space. By doing so, the SVM model can effectively distinguish between classes that exhibit complex, non-linear relationships in the SERS spectra. This approach helps achieve a robust classification, essential for distinguishing subtle spectral differences in biofilm detection.

- ***Model Training and Validation***

In order to assess the SVM model's performance, the dataset was divided into both testing and training subsets for training and validation. We applied a stratified train-test split with a test size of 30%, ensuring that the class distribution remained balanced across both sets. The data was split using stratified $k$ -fold cross-validation with $k = 10$ folds to maintain the proportion of samples from each class in the training and validation sets.

Cross-validation is crucial for assessing the model's generalization ability [92]. The model uses $k$ -1 folds for training, and the remaining fold is used for validation [130]. This procedure is repeated $k$ times, with each fold being the validation set once [131]. The overall model performance is then averaged across all folds, providing a more reliable estimate of its effectiveness.

In addition to cross-validation, hyperparameter tuning is performed to optimize the SVM model's performance. The two primary hyperparameters tuned are the penalty parameter $C$ and the

kernel parameter $\gamma$ in the RBF kernel. The penalty parameter $C$ controls the trade-off between maximizing the margin and minimizing classification errors, while $\gamma$ establishes how much each training example influences the decision boundary [93]. Usually, a grid search is utilized to methodically investigate various combinations of these hyperparameters, and the model's performance is assessed at each setting using cross-validation. The $C$ and $\gamma$ combination that yields the greatest cross-validated performance is chosen. We selected an RBF kernel with $\gamma = scale$ and set the penalty parameter $C = 100$ balancing between maximizing the margin and minimizing classification errors. These parameters were chosen to enhance model stability and accuracy based on the characteristics of our dataset.

- ***Performance Metrics***

We calculated a number of statistical measures using the confusion matrix, which divides predictions into false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN), in order to assess the classification model's performance.

The model's overall performance is measured by accuracy, which is determined by dividing the number of correct predictions (TP and TN) by the total number of cases [94] [132]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

However, more than accuracy is needed to fully capture the model's performance, especially in cases where the data is imbalanced. To address this, we look at sensitivity and specificity. Sensitivity (also referred to as recall or true positive rate) evaluates the model's capacity to correctly identify positive instances, calculated as [94] [132]:

$$Sensitivity = \frac{TP}{TP + FN} \tag{17}$$

This metric is crucial in biomedical applications, where missing positive cases can be costly. Specificity, or the true negative rate, measures the model's ability to correctly identify negative cases, calculated as [94]:

$$Specificity = \frac{TN}{TN + FP} \tag{18}$$

Specificity is important when the cost of false positives is high, as it ensures that negative cases are correctly identified. These metrics, derived from the confusion matrix, form the basis for understanding the model's performance, which is discussed in Section 5.5.2.

# 5.5 Results and Discussion

## 5.5.1 Clinical Simulation: SERS Analysis of Pig Cornea Samples

### *SERS Spectrum of Pig Cornea Samples Without PA01*

In the analysis of pig cornea samples using SERS under optimized conditions, distinct Raman bands were observed that correspond to the molecular structure of corneal tissue. The spectra, illustrated in Figure 20, reveal key features that provide a deeper understanding of the biochemical composition of the cornea, which is rich in collagen and other structural proteins [95].

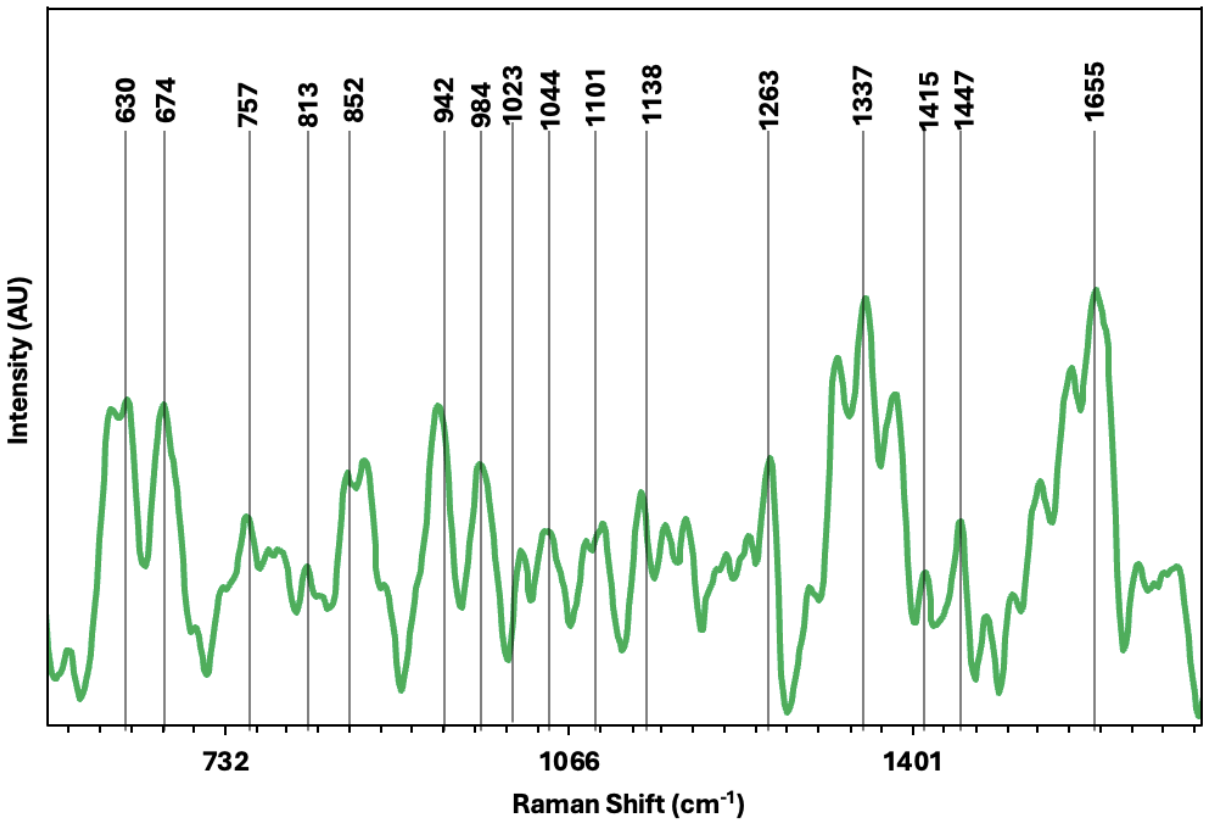


Figure 20: Raman spectrum of cornea without PA01 biofilm.

One of the most prominent features observed is the Amide I band at 1655 $cm^{-1}$, associated with proteins. This band is a crucial marker reflecting the collagen structure within the corneal matrix [96].

The presence of a peak at 852 $cm^{-1}$, associated with proline and hydroxyproline ring breathing modes, provides additional confirmation of the collagen-rich nature of the cornea [97]. Collagen is made up of these vital amino acids, which support the tissue's structural integrity. Additionally, the detection of nucleic acid-associated peaks, such as the one at 1101 $cm^{-1}$, indicates the presence of cellular components within the cornea [98]. Appendix E, Table 9 presents the detailed band assignments, highlighting significant molecular vibrations observed during the analysis.

### *SERS Spectrum of Pig Cornea Samples with PA01 Biofilm*

The SERS spectra obtained from pig cornea samples combined with PA01 biofilm is shown in Figure 21.

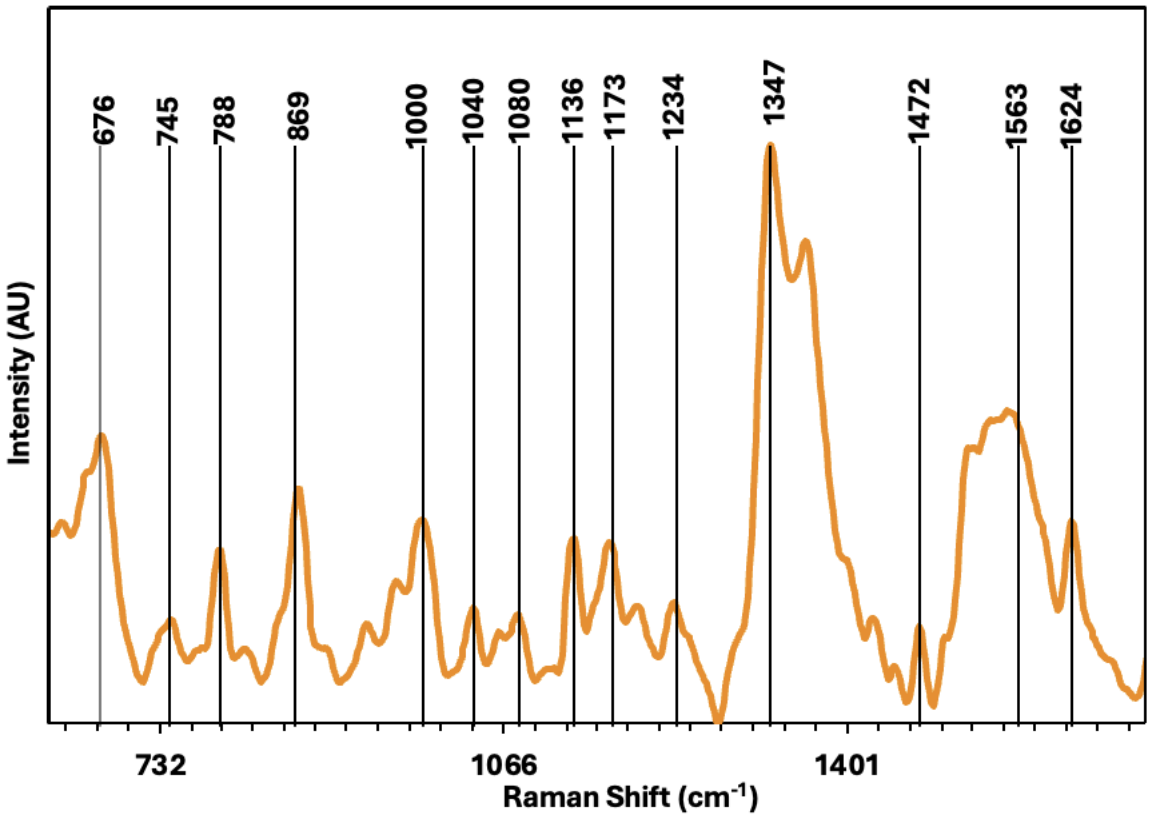


Figure 21: Raman spectrum of cornea with PA01 biofilm.

One of the key observations in the combined spectra is the appearance of a distinct peak at 1347 $cm^{-1}$, which is attributed to the presence of pyocyanin [58]. This peak is of particular significance as it overlaps with the $CH_2$ bending vibrations typically found in lipids and proteins. The presence of this peak in the spectra not only confirms the presence of the PA01 biofilm but also illustrates the intricate interaction between the biofilm and corneal tissues. This overlap suggests that the biofilm components are closely associated with the lipid and protein matrix of the cornea, possibly altering the structural integrity and biochemical environment of the tissue.

Another critical feature is the band at 788 $cm^{-1}$, associated with nucleic acids such as adenine and cytosine. This peak relates to the biofilm's cellular material [99]. The presence of nucleic acids in the spectrum reinforces the notion of active bacterial colonization and the potential implications for biofilm resilience and pathogenicity.

Another important peak is 745 $cm^{-1}$, associated with cytochrome c [100]. In cellular respiration, cytochrome c is a crucial part of the electron transport chain. Because we can identify

it in the spectra, the SERS approach is sensitive enough to pick up signals from specific proteins associated with biofilms [100].

The Amide I band at 1624 $cm^{-1}$, associated with C=O stretching in proteins, remains prominent in the combined spectra, highlighting the structural proteins of the corneal tissue, even in the presence of the biofilm [98] [99]. The detailed band assignments for these spectra are provided in Appendix E, Table 10.

***Comparative Analysis and Discussion***

The comparison of spectra from no biofilm-integrated and biofilm-integrated cornea samples highlighted overlapping peaks at 1347 $cm^{-1}$, associated with both protein structures and the pyocyanin virulence factor, suggesting complex interactions at the molecular level, possibly involving chemical modifications like hydrogen bonding [101].

Additionally, broad overlapping peaks between 1100 and 1200 $cm^{-1}$ were observed, indicating interactions between biofilm components and the corneal tissue, which complicates specific molecular vibration assignments. The appearance of new or intensified peaks in the combined spectra, such as at 1040 $cm^{-1}$, suggests potential new molecular formations or enhanced Raman signals due to biofilm interactions [102][103].

This analysis highlights the challenge of overlapping and broad spectral features in SERS, necessitating advanced techniques for precise molecular identification. The intricate nature of these interactions at the biofilm-corneal interface, as well as the necessity for robust spectral deconvolution methods, is critical for advancing SERS applications in clinical diagnostics.

### 5.5.2 Analysis and Discussion of Computational Model

The computational model developed for bacterial detection using SERS spectral data was evaluated to ensure effectiveness in distinguishing between bacteria-present and bacteria-absent samples. The overall results demonstrate the model's robustness, achieving an accuracy of 92.5% with a standard deviation of 6.12%, statistically significant at $p = 0.0352$ (using a significance level of 0.05) (Figure 22).

Traditional methods like smear microscopy (Gram and Giemsa stains) and culture techniques vary in effectiveness [104]. While smear microscopy can achieve high sensitivities for detecting

specific pathogens, sometimes up to 100% for certain fungi, the specificity can vary [104]. Culture methods often considered a definitive method for identifying bacterial infections; however, cultures can have relatively low sensitivity, ranging from 30% to 50% due to the low bioburden in infectious keratitis [104]. The turnaround time for culture results can also be quite long, up to three days, which may delay appropriate treatment and cause patient distress [105].

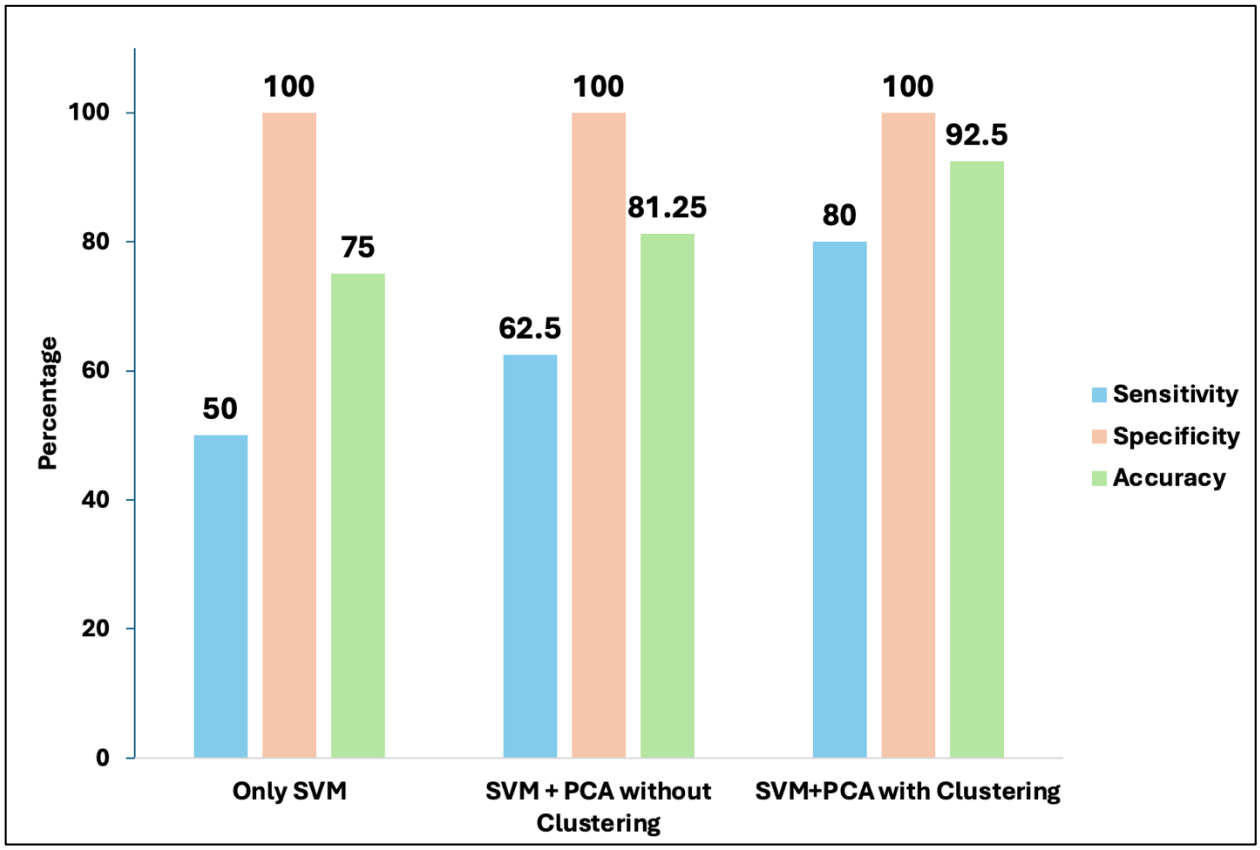


Figure 22: Performance metrics of the bacteria classification model.

We also varied model configurations to identify their impact on the performance metrics, as can be seen in Figure 22. The SVM-only setup provided a baseline with a sensitivity of 50%, suggesting challenges in accurately detecting all positive cases those near the decision boundaries where spectral features may overlap. The incorporation of PCA improved sensitivity to 62.5% and accuracy to 81.25%, demonstrating PCA's role in enhancing feature separation. This also highlights PCA's role in better identifying true positives. The addition of k-means clustering to SVM and PCA maximized the model's performance, achieving 80% sensitivity and 92.5%

accuracy, highlighting the significant impact of clustering in managing within-sample variance and enhancing overall model robustness.

The model relies on subtle spectral differences to distinguish between bacteria-present and bacteria-absent states, making overlapping peaks and broad spectral features challenging. Preprocessing steps such as baseline correction, smoothing, and normalization is crucial in enhancing spectral clarity. However, they also risk confusing critical spectral differences needed for accurate classification—a key trade-off in SERS data preprocessing.

Dimensionality reduction through PCA and k-means clustering helps manage the SERS data's complexity. PCA identifies informative components, simplifying the dataset while preserving essential features. Despite this, small spectral shifts in complex biological samples, like those mixing corneal tissue and bacterial biofilm, pose ongoing classification challenges. To better understand the spectral data's driving features, PCA loading vectors have been also analyzed. Detailed plots of these loading vectors, along with the cumulative variance plot to determine the optimal number of principal components, are provided in Appendix F. We primarily focused on the first two principal components. These components were selected because they capture the majority of the variance within the data, providing a clear and concise overview of the most impactful spectral features.

Table 4: Confusion matrix for classification of corneal samples into bacteria present and absent.

| | Predicted Negative | Predicted Positive |
|---|---|---|
| True Negative | **100%(**TN) | 0% (FP) |
| True Positive | 20% (FN) | **80%** (TP) |

The analysis of the confusion matrix reveals the results of the model's performance (Table 4). The model's high specificity of 100% effectively rules out false positives, critical in clinical settings to avoid unnecessary treatments and patient distress. However, the presence of false negatives indicates a need for further refinement to enhance sensitivity. This indicates that while the model is highly sensitive, certain samples with spectral features close to the decision boundary,

pose a classification challenge. These samples likely exhibit spectral characteristics that overlap with those of bacteria-absent samples, making them more difficult to classify correctly. Improvements include advanced machine learning techniques and integrating more diverse data types, such as clinical metadata, to improve the model's understanding of sample characteristics.

The results suggest that while the model is highly promising in its sensitivity, there are areas where further refinement is necessary. The existence of false negatives in our models emphasizes the necessity for extra techniques to improve the model's ability to distinguish between challenging samples. While SERS is widely used for detecting bacterial and viral infections, its application in diagnosing ocular infections is still emerging. But the high accuracy of 92.5% is indicative of our model's overall reliability and aligns well with other studies done to prove the effectiveness of ML in diagnostics. For example, Jammula et. al, used advanced machine learning models for ocular disease diagnosis and achieved 90% accuracy, 94% sensitivity and 86% specificity [123]. This benchmark was drawn from computer-aided diagnosis of funduscopic images for conditions like glaucoma or diabetic retinopathy[123].

## 5.6 Challenges and Reflection

Implementing machine learning algorithms to classify the SERS spectra and detect biofilms accurately was a significant challenge. Through dedicated independent study and practice, I developed a strong foundation of ML and the different techniques. Various parameters had to be tested and fine-tuned to achieve the desired performance for this research. The coding aspect of this project required not only selecting the right algorithms but also adapting them to work with the specific characteristics of biofilm SERS data.

Issues such as overfitting, computational efficiency, and data preprocessing had to be addressed through a series of iterations, each requiring debugging and refinement. This aspect of the research was demanding due to the need for the right combination of tools to achieve the highest accuracy. This iterative process of coding, testing, and refining was essential in developing a reliable framework for data analysis.

## 5.7 Conclusion

This chapter presented the integration of ML techniques with SERS data analysis to effectively detect biofilms in corneal samples. By combining PCA for dimensionality reduction, k-means clustering for handling spectral variability, and SVM for classification, we were able to achieve high diagnostic accuracy. The resulting model demonstrated an overall accuracy of 92.5%, with a specificity of 100% and a sensitivity of 80%, reflecting the robust performance of the integrated approach.

The SERS spectra for corneal tissue were consistent with earlier validation results from collagen samples, though some spectral differences were noted—expected due to the distinct structural composition of the cornea compared to collagen. The ML model's performance highlighted its ability to accurately differentiate between biofilm-present and biofilm-absent samples, providing an effective diagnostic framework.

While the model performed well in avoiding false positives, further improvements could enhance sensitivity, addressing the challenge of false negatives. The application of these techniques establishes a baseline for future exploration in clinical diagnostics, with the potential for real-world implementation in biofilm detection.

# Chapter 6

# Summary and Future Work

## 6.1 Summary

SERS is a powerful analytical technique known for its ability to provide highly sensitive and specific detection of molecular species, especially at low concentrations. Bacterial biofilms are difficult to detect with traditional techniques, so this ability is useful. This thesis aimed to establish a baseline for detecting biofilms using SERS, specifically focusing on PA01 biofilms, which are significant contributors to ocular infections. The goal was to demonstrate that the optimized SERS technique could reliably detect biofilms under controlled experimental conditions.

Optimizing important experimental parameters, such as interaction time and silver nanoparticle concentration, was the main focus of this study's first phase. Based on our experiments, the most reliable SERS signals were produced by combining a 1.75x concentration of silver nanoparticles with a 24-hour interaction period. This optimization was essential in detecting specific Raman peaks, particularly the pyocyanin peak at 1345-1347 $cm^{-1}$, which serves as a marker for the presence of PA01 biofilms. This finding was supported by preliminary experiments and existing literature, which suggested that this combination of parameters would enhance the detection sensitivity.

After the optimization, the next phase of the research aimed to validate these SERS conditions using more complex biological samples, including collagen and pig cornea, which are representative models for human ocular tissues. The results from these validation studies demonstrated that the optimized SERS technique could effectively detect biofilms in these models, reinforcing the potential of SERS as a diagnostic tool for ocular health. This validation is a critical step in confirming that the SERS method can be applied in real-world clinical settings for the rapid and accurate detection of biofilm-related infections.

### 6.1.1 Limitations

While this thesis has successfully established a baseline for the detection of biofilms using SERS, we encountered several limitations that highlight areas for further refinement. One of the

primary challenges was the variability in biofilm sample preparation, which occasionally led to inconsistencies in the SERS signal. Specifically, variations in the uniformity of the biofilm layer, the concentration of bacteria, and the extent of nanoparticle interaction were identified as key factors contributing to this variability. These challenges emphasize the inherent complexity of working with biological samples and the need for more standardized and refined preparation protocols. For future research to produce consistent and trustworthy results, these problems need to be resolved.

Additionally, the focus on PA01 biofilms, while necessary for establishing the baseline, limits the generalizability of the findings. The ML models developed in this study were specifically tailored to detect PA01, and their applicability to other bacterial strains still needs to be explored. Although this research provides a solid foundation, expanding these models to include a broader range of bacterial species and biofilm types is important for extending the diagnostic utility of this technique.

Moreover, while pig cornea samples provided some results, there are inherent differences between pig and human corneas that could impact the direct translation of the findings into clinical practice. These differences, which include variations in corneal thickness, biochemical composition, and immune response, could influence the sensitivity and specificity of SERS when applied to human tissues.

## 6.2 Future Works

### 6.2.1 Optimization of Experimental Conditions

This thesis successfully established a baseline for detecting biofilms using silver nanoparticles with a 24-hour interaction time and a 1.75x concentration. Going forward, an important area for future research is to explore other types of nanoparticles. For example, gold nanoparticles, which are known for their strong plasmonic properties, could be investigated to see if they offer any advantages in signal strength or consistency. Testing these could enhance the sensitivity and specificity of SERS. Future research can also focus on the use of positively charged capping agents, such as chitosan or polyethyleneimine, to improve bacterial interaction with nanoparticles. Understanding the surface charge of the cornea, and how it changes with infection, could further optimize nanoparticle adherence and biofilm detection.

### 6.2.2 Expansion of Bacterial Classification

While this thesis focused on detecting PA01 biofilms, expanding the scope of SERS to detect a broader range of bacterial species is the next step. Future research should explore how this technique can be adapted to identify other bacteria that are commonly associated with biofilm formation – pathogens such as *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Escherichia coli*, which can be further investigated because of their prevalence in ocular infections. This could involve developing new classification models tailored to different bacterial species, potentially expanding the diagnostic capabilities of SERS. Additionally, it would be valuable to test SERS in more complex environments, such as mixed biofilm communities where multiple bacterial species coexist.

### 6.2.3 Transition to Human Cornea Testing

Building on the success of using SERS to detect biofilms in pig cornea models, the logical progression is to begin trials involving human corneas. This shift is important for determining whether the SERS parameters, optimized in earlier experiments, can be applied effectively in clinical scenarios that closely resemble actual medical environments. Future work should prioritize evaluating the sensitivity and specificity of these conditions when applied to human tissues. Such efforts will not only confirm the practical significance of the current findings but also set the foundation for eventual clinical trials. As we aim toward practical applications, it is critical to create safe and efficient techniques for incorporating SERS into diagnostic

Demonstrating the capability of SERS to detect biofilms in human corneas would mark a major step forward, potentially leading to its adoption as a diagnostic tool for biofilm-associated infections in ocular health.

# Appendix A

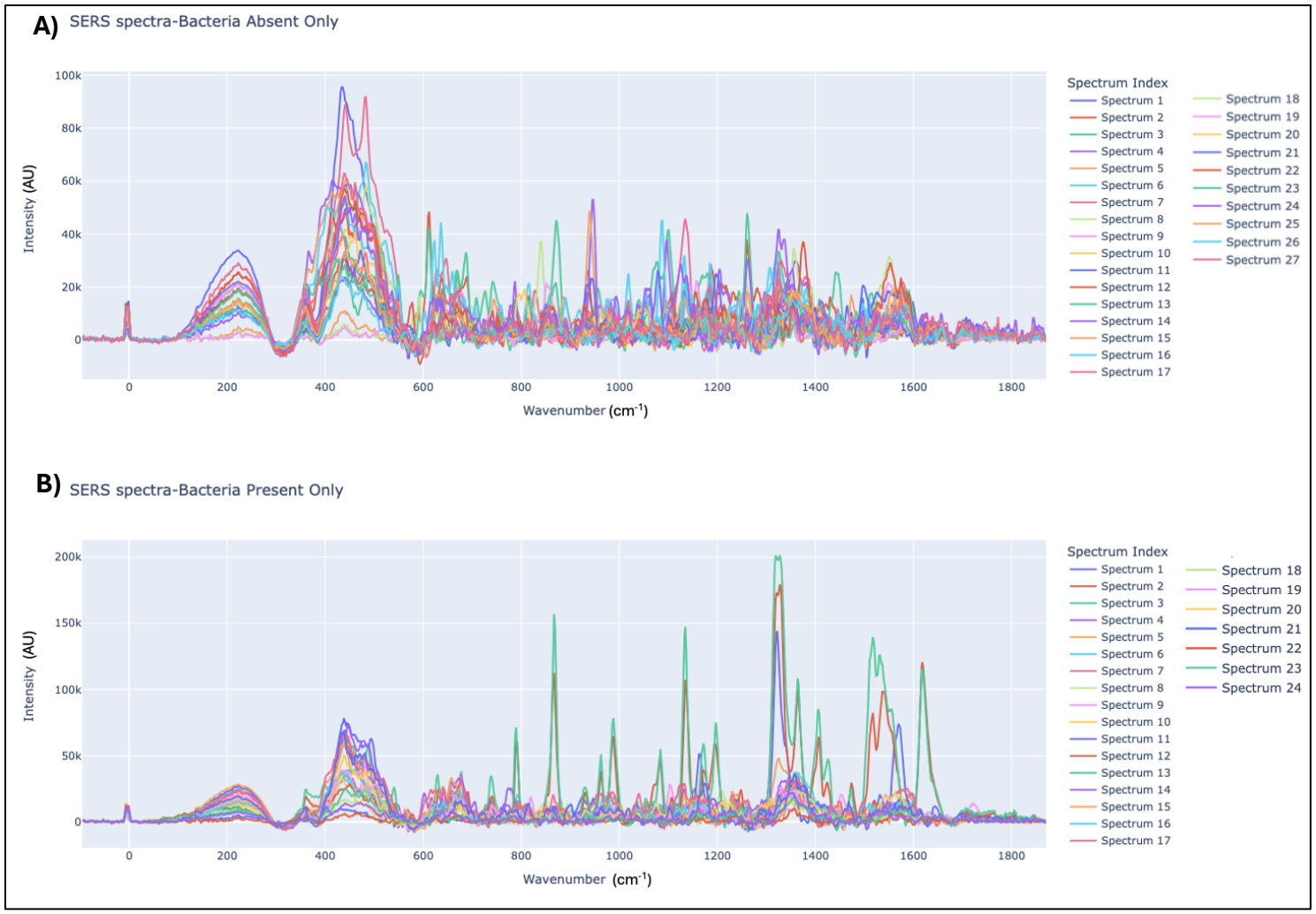


Figure 23: Individual Raman spectra of non-infected cornea samples (A) and the infected cornea samples (B).

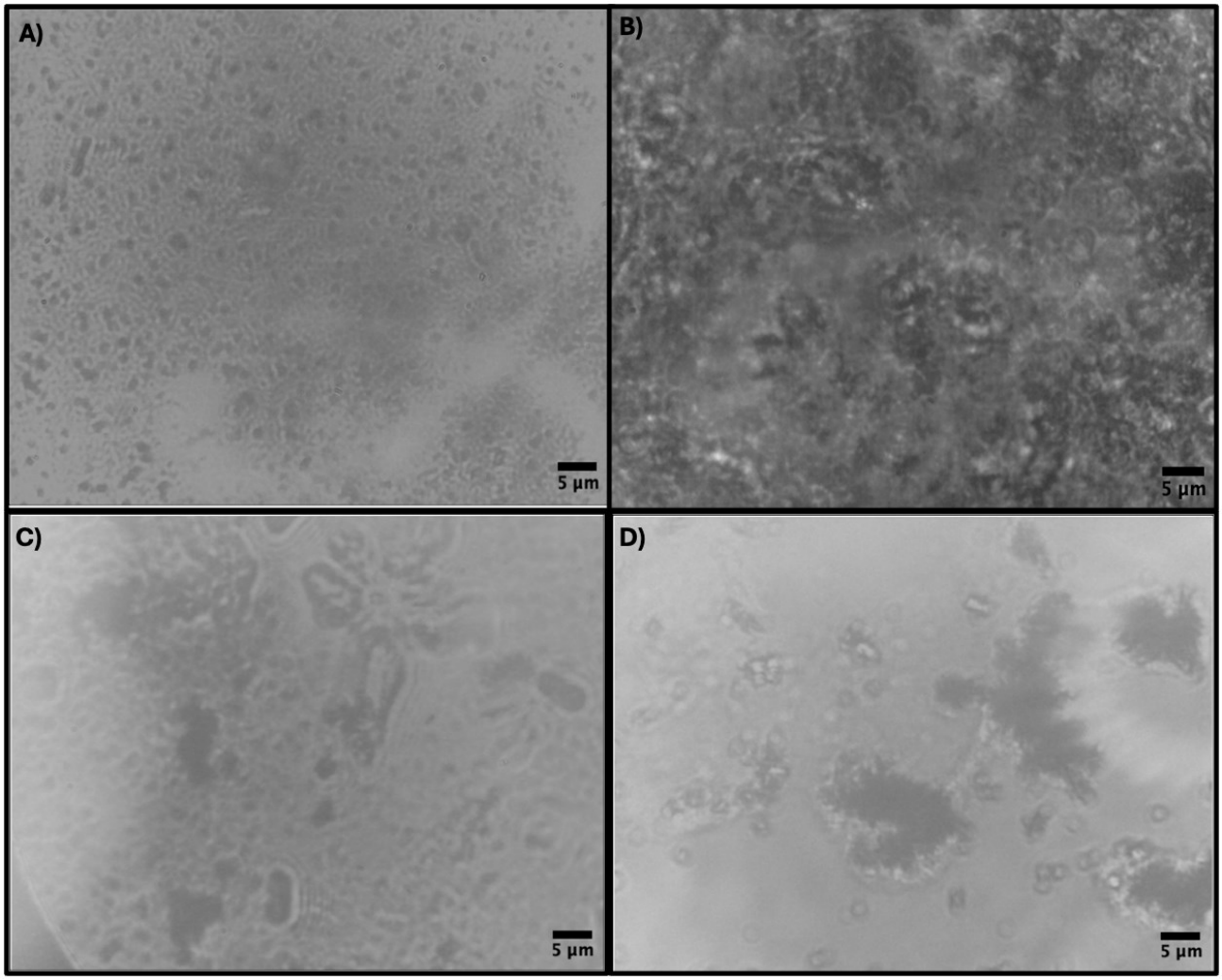

Figure 24: Microscopic images of A) PA01 with AgNPs, B) collagen with AgNPs, C) cornea with PA01 biofilm and AgNPs, and D) cornea with AgNPs, but without PA01 biofilm.

# Appendix B

Table 5: Detailed Raman band assignment for PA01 biofilm under baseline conditions as shown in Figure 11.

| **Raman Band ($cm^{-1}$)** | **Tentative Band Assignment [58][59][60]** |
|---|---|
| 675 | Valine (proteins), $CN^{+}(CH_3)_3$ stretching (lipids) |
| 733 | Adenine (DNA/RNA), COO- deformation (carbohydrates), $CH_2$ rocking (carbohydrates) |
| 745 | Ring breathing Tryptophan (proteins) |
| 827 | O-P-O asymmetric stretching (DNA/RNA), Ring breathing Tyrosine (proteins) |
| 881 | C-C-$N^{+}$ symmetric stretching (lipids), C-O-C ring (carbohydrates) |
| 969 | C-C backbone stretching (proteins) |
| 1015 | Symmetric Ring Breathing Phenylalanine (proteins) |
| 1050 | C-H in plane phenylalanine (proteins) |
| 1108 | $PO_{2-}$ symmetric stretching (DNA/RNA), C-C stretching, C-O-C skeletal stretching, -glycosidic linkage of saccharides (carbohydrates) |
| 1129 | C-C stretching, C-O-C skeletal stretching, -glycosidic linkage of saccharides (carbohydrates) |
| 1165 | Guanine (DNA/RNA), cytosine (DNA/RNA), C-H Tyrosine (proteins), Phenylalanine (proteins) |
| 1252 | Amide III band (C-N stretching and N-H bending in proteins) |
| 1284 | Amide III (proteins), CH deformation (carbohydrates) |
| 1360 | Pyocyanin (PA01-specific metabolite) |
| 1457 | Guanine (DNA/RNA), Adenine (DNA/RNA), CH deformation (carbohydrates) |
| 1517 | C=C stretching |
| 1565 | Guanine (DNA/RNA), Adenine (DNA/RNA), Amide II (proteins), COO- asymmetric stretching (carbohydrates) |

# Appendix C

Table 6: Detailed Raman band assignment for PA01 Biofilm as shown in Figure 12.

| **Raman Band ($cm^{-1}$)** | **Tentative Band Assignment [58][59][60]** |
|---|---|
| 656 | C-S stretching (proteins), CO stretching (carbohydrates) |
| 687 | C-C skeletal stretching (lipids), C-H deformation (lipids) |
| 729 | Adenine (DNA/RNA), $COO^-$ deformation (carbohydrates) |
| 748 | Ring breathing mode of tryptophan (proteins) |
| 836 | O-P-O symmetric stretching (DNA/RNA), C-O-C ring (carbohydrates) |
| 892 | C-C stretching (proteins), C-H deformation (lipids) |
| 952 | C-C backbone stretching (proteins) |
| 1001 | Symmetric ring breathing of phenylalanine (proteins) |
| 1061 | C-N stretching (proteins), C-O-C skeletal stretching (carbohydrates) |
| 1109 | $PO_2$ symmetric stretching (DNA/RNA), C-C stretching, glycosidic linkage of saccharides |
| 1130 | C-N stretching (proteins), C-H deformation (lipids) |
| 1159 | C-H deformation (lipids), C-N stretching (proteins) |
| 1230-1271 | Amide III band (C-N stretching and N-H bending in proteins) |
| 1347 | Pyocyanin (PA01-specific metabolite), CH deformation (carbohydrates) |
| 1401 | $COO^-$ symmetric stretching (carbohydrates), C-N stretching (proteins) |
| 1438 | $CH_2/CH_3$ deformation (lipids/proteins) |
| 1516 | C=C stretching (lipids), C-H deformation (lipids) |
| 1572 | Guanine (DNA/RNA), Adenine (DNA/RNA), Amide II (proteins) |

Table 7: Detailed Raman band assignment for pure collagen samples shown in Figure 13.

| **Raman Band ($cm^{-1}$)** | **Tentative Band Assignment [62][63][64]** |
|---|---|

| | |
|---|---|
| 554 | C-C twisting (collagen), S-S stretching vibrations (disulfide bridges) |
| 608 | C-C stretching (proteins), Amide V (out-of-plane C=O bending) |
| 646 | Out-of-plane C-H bending (aromatic side chains) |
| 765 | Amide I overtone or combination band (C=O and N-H stretching) |
| 823 | C-C stretching in proline (collagen) |
| 852 | C-H deformation (collagen, proteins), Proline ring breathing |
| 932 | C-C stretching (collagen), Hydroxyproline ring breathing modes |
| 961 | C-C backbone stretching (collagen), C-C stretching vibrations of proline |
| 1021 | $CH_3$ rocking (collagen), C-C backbone stretching (collagen) |
| 1044 | C-H in-plane bending (collagen), C-N stretching and N-H bending (Amide III) |
| 1071 | C-C skeletal stretching (collagen), C-N stretching and N-H bending (Amide III) |
| 1126 | C-N stretching (collagen) |
| 1190 | C-H stretching, C-H bending, and C-C stretching |
| 1236 | Amide III (collagen, proteins) |
| 1272 | Amide III (collagen, proteins) |
| 1303 | $CH_2$ twisting and bending (collagen, lipids) |
| 1355 | Amide III, C-N stretching and N-H bending, C-C stretching, C-H bending |
| 1415 | $CH_3$ bending (collagen, proteins) |
| 1447 | $CH_2$ bending (collagen, lipids) |
| 1482 | C-H deformation (collagen, proteins) |
| 1540 | Amide II (collagen, proteins) |
| 1595 | C=C stretching (collagen, aromatic rings) |
| 1647 | Amide I (collagen, proteins), C=O stretching |

Table 8: Detailed Raman band assignment for the combined PA01 and collagen samples shown in Figure 14.

| **Raman Band ($cm^{-1}$)** | **Tentative Band Assignment [58][64][65][66]** |
|---|---|
| 652-802 | C-C stretching, amide modes (proteins), C-H deformation (lipids), contributions from both PA01 and collagen |
| 831 | C-H deformation (collagen, lipids) |
| 873-928 | C-C backbone stretching, amide modes (proteins), C-H bending (lipids), contributions from both PA01 and collagen |
| 980 | Symmetric C-C stretching (proteins), contributions likely from both sources |
| 1065 | C-N stretching, amide modes (proteins), contributions from both PA01 and collagen |
| 1105-1169 | C-N stretching, $CH_2$ bending (collagen, proteins), contributions from both sources |
| 1203 | Amide III band (proteins), contributions likely from collagen |
| 1263 | Amide III band (proteins), CH deformation (lipids), contributions likely from collagen |
| 1345 | Pyocyanin (PA01-specific metabolite) overlap with $CH_2$ bending (lipids/proteins) from collagen |
| 1443 | $CH_2$ bending (lipids), contributions likely from collagen |
| 1495 | C=C stretching (lipids, aromatic rings), overlap likely from both sources |
| 1538-1657 | Amide I band, C=O stretching (proteins), contributions likely from collagen |

# Appendix D

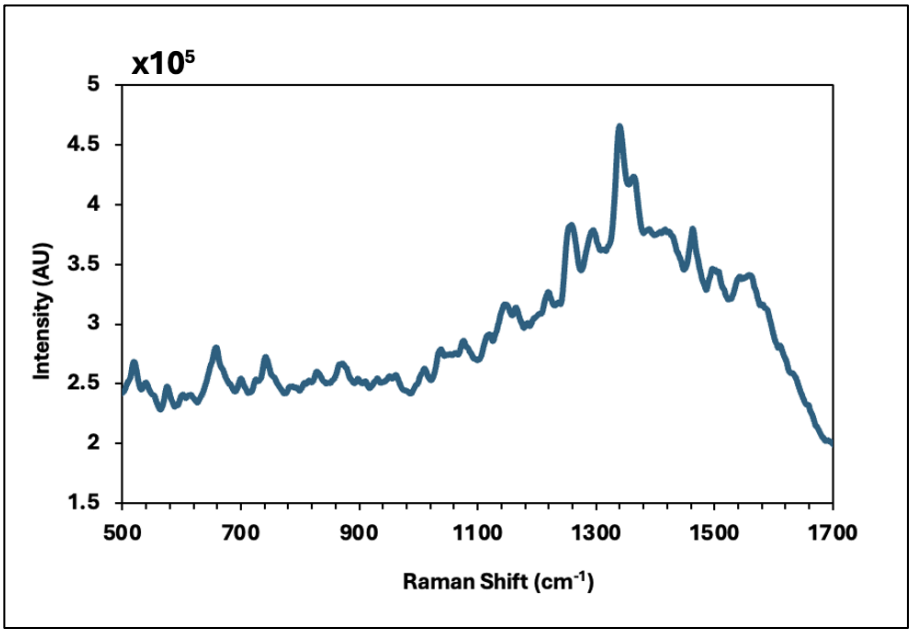


Figure 25: Example raw spectrum of PA01 biofilm.

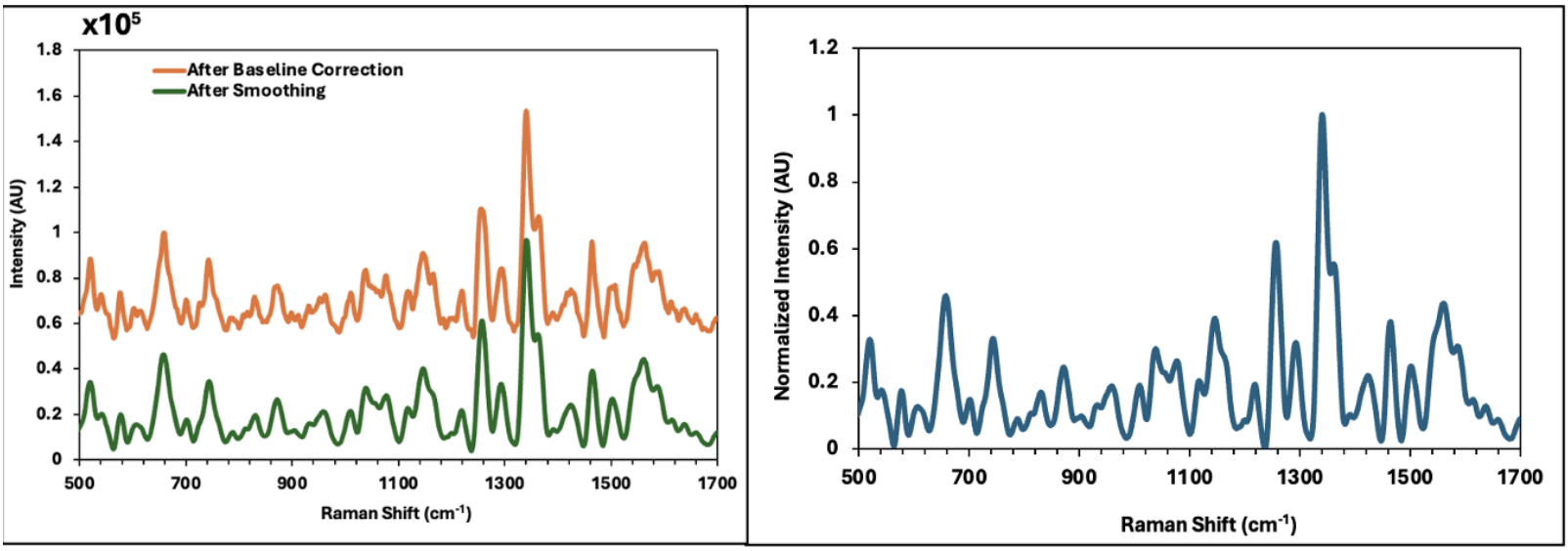


Figure 26: Example spectra of PA01 biofilm after baseline correction and smoothing (left) and after normalization (right).

# Appendix E

Table 9: Detailed Raman band assignment for the bacteria absent cornea samples shown in Figure 20.

| Raman Band ($cm^{-1}$) | Tentative Band Assignment [96][97][98] |
|---|---|
| 630 | C-S stretching (proteins), C-H deformation (lipids) |
| 673 | C-C skeletal stretching, Amide V (out-of-plane C=O bending) |
| 757 | Symmetric ring breathing of tryptophan (proteins) |
| 813 | C-C stretching (collagen), C-H bending (lipids) |
| 852 | Proline ring breathing (collagen) |
| 942 | C-C stretching (collagen), Hydroxyproline ring breathing modes |
| 984 | C-H in-plane bending (proteins), C-C skeletal stretching (collagen) |
| 1023 | C-N stretching (collagen), C-C skeletal stretching (proteins) |
| 1044 | C-H in-plane bending (collagen), C-N stretching and N-H bending (Amide III) |
| 1101 | $PO_2$ symmetric stretching (nucleic acids), C-C stretching (collagen) |
| 1138 | C-H deformation (lipids), C-N stretching (proteins) |
| 1161 | C-H bending, Amide III (collagen, proteins) |
| 1263 | Amide III band (C-N stretching and N-H bending) |
| 1337 | $CH_3$ deformation (proteins), C-H bending (lipids) |
| 1415 | C=C stretching (lipids), $CH_3$ bending (collagen) |
| 1447 | $CH_2$ bending (lipids), C-H deformation |
| 1655 | Amide I band (C=O stretching vibrations) |

Table 10: Detailed Raman band assignment for the bacteria present cornea samples shown in Figure 21.

| Raman Band ($cm^{-1}$) | Tentative Band Assignment [58][96][97][98][99][100] |
|---|---|

| | |
|---|---|
| 676 | C-C twisting vibrations, S-S stretching vibrations (disulfide bridges), Amide V (out-of-plane C=O bending), |
| 745 | Cytochrome c |
| 788 | Nucleic acids (DNA/RNA), Adenine-Cytosine |
| 869 | Proline - Hydroxyproline |
| 1000 | Phenylalanine |
| 1040 | C-H in-plane bending, Amide III |
| 1080 | C-N stretching, CH deformation |
| 1136 | Glycosidic bonds |
| 1173 | C-H deformation (lipids) |
| 1234 | Amide III band (C-N stretching and N-H bending in proteins) |
| 1347 | Pyocyanin (PA01-specific metabolite) overlap with $CH_2$ bending (lipids/proteins) |
| 1472 | CH2 bending vibrations |
| 1563 | C=C stretching |
| 1624 | Amide I band (C=O stretching) |

# Appendix F

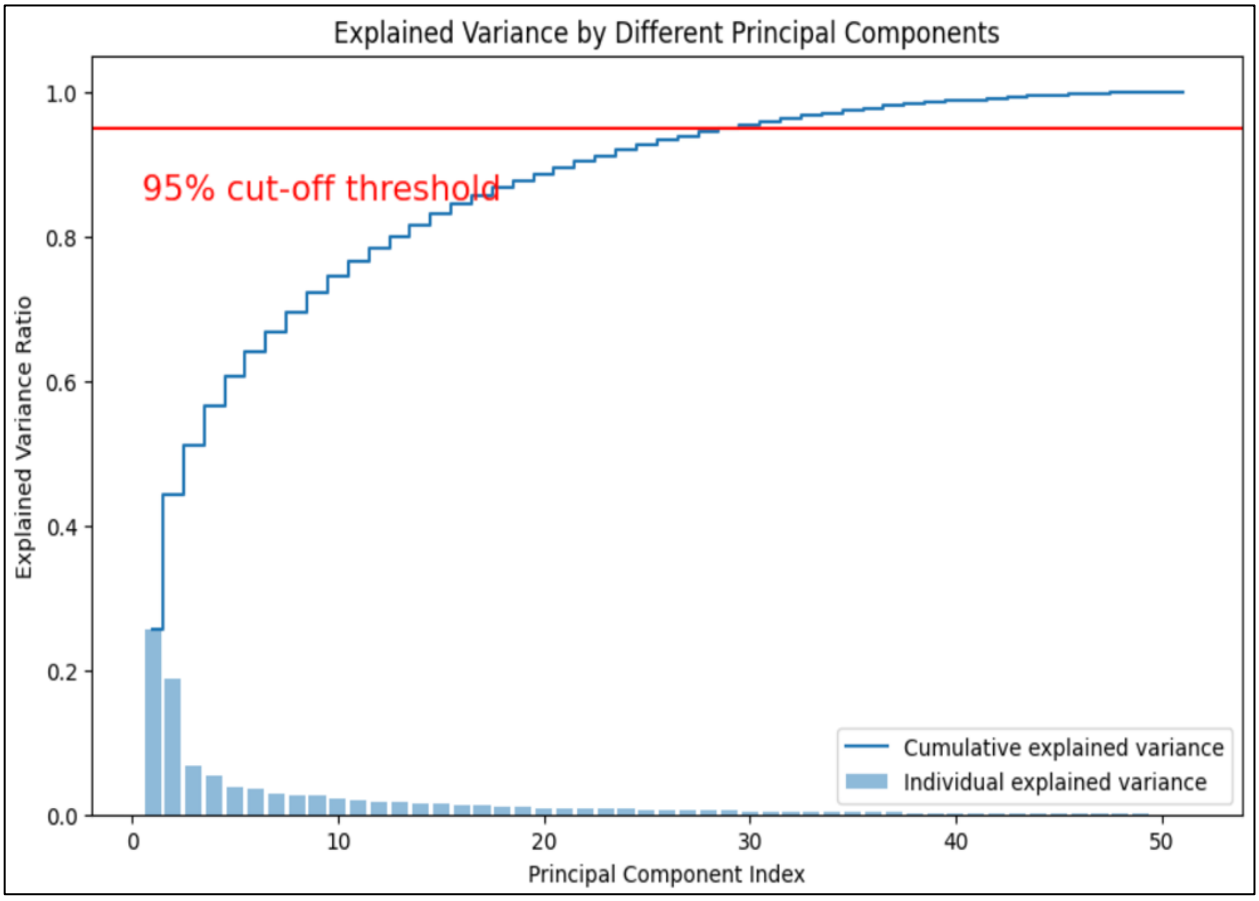


Figure 27: Cumulative and individual explained variance of principal components.

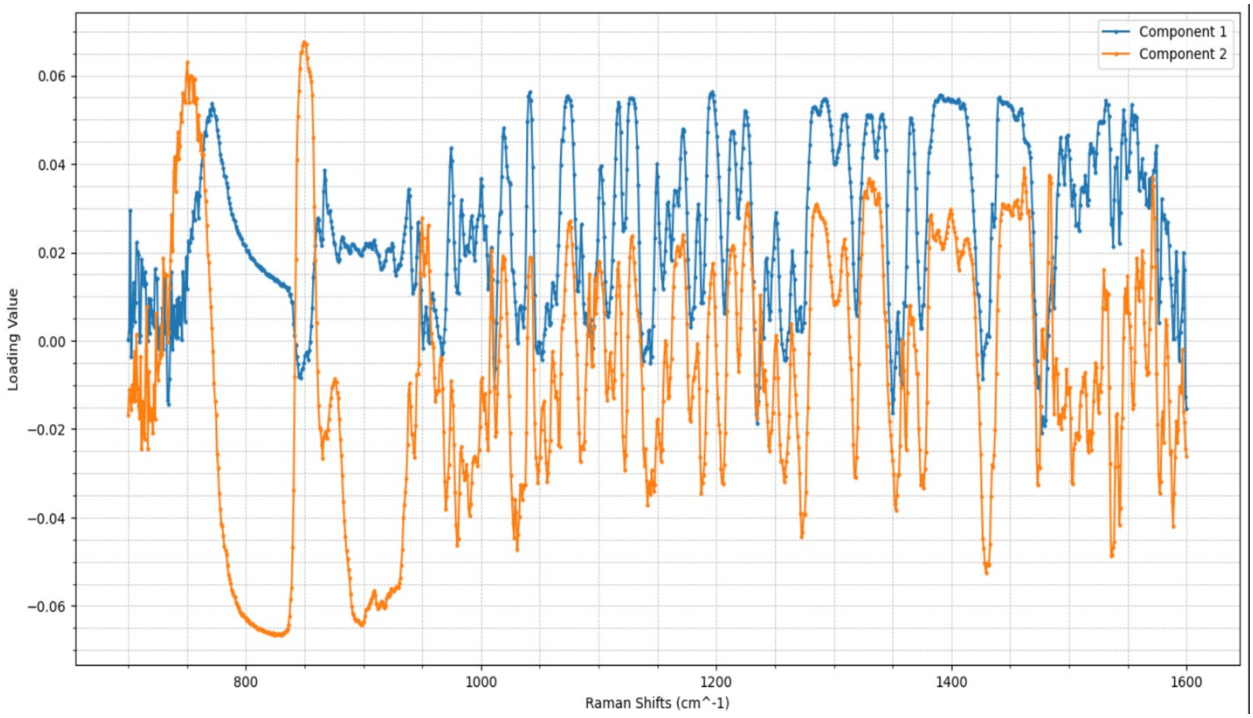


Figure 28: Loadings of the first two principal components for the Raman spectra of corneal samples.

# Appendix G

In order to ensure reproducibility and transparency of the computational processes described in this thesis, all code developed for data preprocessing, analysis, and visualization is available on GitHub:

https://github.com/araagnih/SERS-Corneal-Sample-Classification

# References


[1] A. K. Adya, E. Canetta, "Nanotechnology and its applications to animal biotechnology," in *Animal Biotechnology*, Second edition (Academic Press/Elsevier, 2020), pp. 309–326.

[2] A. Orlando, F. Franceschini, C. Muscas, S. Pidkova, M. Bartoli, M. Rovere, A. Tagliaferro, "A Comprehensive Review on Raman Spectroscopy Applications," *Chemosensors*, 9 (2021) 262. doi:10.3390/chemosensors9090262.

[3] C.V. Raman, "A new radiation" [Reproduced from Indian J. Phys., 1928, 2, 387–398], Current Science 74, 382–386 (1998).

[4] G. Landsberg, L. Mandelstam, "Eine neue Erscheinung bei der Lichtzerstreuung in Krystallen" *Naturwissenschaften 16*(1928), 557–558.

[5] P.J. Hendra, P.M. Stratton, "Laser-raman spectroscopy" *Chem. Rev*. 69(1969) 325–344.

[6] V.T. Platonenko, R.V. Khokhlov, "On the mechanism of operation of a raman laser" *J. Exptl. Theor. Phys*. 19(1966) 555–559.

[7] P. Parvathi, P. V. Parimaladevi, V. Sathe, U. Mahalingam, "Graphene boosted silver nanoparticles as surface enhanced Raman spectroscopic sensors and photocatalysts for removal of standard and industrial dye contaminants," *Sens. Actuators B Chem.*, 281 (2019) 679–688. doi:10.1016/j.snb.2018.10.137.

[8] R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson, V. K. Valev, "Raman Techniques: Fundamentals and Frontiers," *Nanoscale Research Letters*, 14 (2019) 231. doi:10.1186/s11671-019-3039-2.

[9] L. O. Björn (2008). The Nature of Light and Its Interaction with Matter. In L. O. Björn (Ed.), Photobiology (pp. 1–39). Springer. https://doi.org/10.1007/978-0-387-72655-7_1

[10] B. E. A. Saleh & M. C. Teich (2019). *Fundamentals of Photonics* (3rd ed.). John Wiley & Sons. (pp. 38-89)

[11] R. W. Boyd (2020). *Nonlinear Optics*. Elsevier. (pp. 381-493)

[12] E. L. Ru & P. Etchegoin (2009). *Principles of Surface-Enhanced Raman Spectroscopy and related plasmonic effects*. Elsevier.

[13] Traces Centre. (April 2021). Understanding Raman Spectroscopy Principles and Theory. Version: 5. University of Toronto. Retrieved from https://www.utsc.utoronto.ca/~traceslab/PDFs/raman_understanding.pdf

[14] S. Y. Ding, E. M. You, Z. Q. Tian, & M. Moskovits (2017). Electromagnetic theories of surface-enhanced Raman spectroscopy. *Chemical Society Reviews*, 46, 4042-4076.

[15] P.M.V. Raja & A.R. Barron (Ed.). (2022). Chemistry LibreTexts. LibreTexts. [https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Physical_Methods_in_Chemistry_and_Nano_Science_(Barron)/04%3A_Chemical_Speciation/4.03%3A_Raman_Spectroscopy#:~:text=Due%20to%20the%20lack%20of,be%20seen%20in%20Figure%204.3] (Chapter: 4.3: Raman Spectroscopy)

[16] Z. Zhang, T. Peng, X. Nie, G. S. Agarwal, & M. O. Scully (2022). Entangled photons enabled time-frequency-resolved coherent Raman spectroscopy and applications to electronic coherences at femtosecond scale. *Light: Science & Applications*, 11(1), 274.

[17] M. Hedegaard, C. Matthäus, S. Hassing, C. Krafft, M. Diem, & J. Popp (2011). Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging. Theoretical Chemistry Accounts, 130, 1249-1260.
[18] O. Stevens, I. E. I. Petterson, J. Day, & N. Stone (2016). Developing fibre optic Raman probes for applications in clinical spectroscopy. Chemical Society Reviews, 45(7), 1919-1934.
[19] M. Fleischman, P. Hendra, and A. McQuillan, "Raman spectra of pyridine adsorbed at a silver electrode," Chemical Physics Letters 26, 163-166, (1974).
[20] R. L. McCreery (2000). Raman Spectroscopy for Chemical Analysis. John Wiley & Sons. (Print ISBN: 9780471252870 | Online ISBN: 9780471721642 | DOI: 10.1002/0471721646)
[21] Z. Huang, J. Peng, L. Xu, & P. Liu (2024). Development and Application of Surface-Enhanced Raman Scattering (SERS). *Nanomaterials*, 14(17), 1417. https://doi.org/10.3390/nano14171417.
[22] K. Kneipp, M. Moskovits, & H. Kneipp (Eds.). (2006). *Surface-Enhanced Raman Scattering: Principles, Techniques, and Applications*.
[23] A. Sadrolhosseini, A. Noor, and M. Moksin, "Application of surface plasmon resonance based on a metal nanoparticle," in Plasmonics — Principles and Applications, K. Y. Kim, ed. (IntechOpen, 2012), pp. 253–282.
[24] E. C. Le Ru & P. G. Etchegoin (2006). Rigorous justification of the $|E|^4$ enhancement factor in Surface Enhanced Raman Spectroscopy. Chemical Physics Letters, 423(1-3), 63-67.
[25] E. C. Le Ru & P. G. Etchegoin (2013). Quantifying SERS enhancements. MRS Bulletin, 38, 631-640.
[26] Y. Hang, A. Wang, & N. Wu (2022). Plasmonic silver and gold nanoparticles: shape- and structure-modulated plasmonic functionality for point-of-caring sensing, bio-imaging and medical therapy. Materials Horizons, 9(10), 2477-2499.
[27] N. T. T. Phuong, V. Q. Dang, L. V. Hieu, T. N. Bach, B. X. Khuyen, H. K. T. Ta, H. Ju, B. T. Phan, & N. H. T. Tran (2023). Functionalized silver nanoparticles for SERS amplification with enhanced reproducibility and for ultrasensitive optical fiber sensing in environmental and biochemical assays. *Journal of Materials Chemistry C*, 11(39), 12787-12800.
[28] N. D. Israelsen, C. Hanson, and E. Vargis, "Nanoparticle properties and synthesis effects on surface-enhanced Raman scattering enhancement factor: An introduction," The Scientific World Journal 2015, 1–12 (2015).
[29] K. C. Bantz, A. F. Meyer, N. J. Wittenburg, H. Im, Ö. Kurtulus, S. H. Lee, N. C. Lindquist, S.-H. Oh, and C. L. Haynes, "Recent Progress in SERS Biosensing", Physical chemistry chemical physics 13, 11551–11567 (2011).
[30] R. A. Tripp, R. A. Dluhy, and Y. Zhao, "Novel nanostructures for SERS biosensing," NanoToday 3, 31–37 (2008).
[31] D. Harvey. LibreTexts. (2024). The Signal-to-Noise Ratio. [Online]. chem.libretexts.org
[32] B. Sharma, R. R. Frontiera, A.-I. Henry, E. Ringe, & R. P. Van Duyne (2012). SERS: Materials, applications, and the future. *Materials Today*, 15(1-2), 16-25.
[33] N. Leopold & B. Lendl (2003). A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. The Journal of Physical Chemistry B, 107(24), 5723–5727.
[34] D. C. Montgomery (2017). Design and Analysis of Experiments (9th ed.). John Wiley & Sons.

[35] Z. D. Schultz, Y. Ozaki, K. Kneipp, & R. Aroca (Eds.). (2014). Frontiers of surface-enhanced Raman scattering: single nanoparticles to single cells. Analytical and Bioanalytical Chemistry, 406(28), 6083-6084. https://doi.org/10.1007/s00216-014-8072-5
[36] S. Keleştemur, E. Avci, & M. Çulha (2018). Raman and Surface-Enhanced Raman Scattering for Biofilm Characterization. Chemosensors, 6(1), 5. https://doi.org/10.3390/chemosensors6010005.
[37] A. Voinescu, M. Licker, D. Muntean, C. Musuroi, S. I. Musuroi, O. Izmendi, S. Vulpie, R. Jumanca, M. Munteanu, & A. Cosnita (2023). A Comprehensive Review of Microbial Biofilms on Contact Lenses: Challenges and Solutions. Journal of Contact Lens Research, 45(2), 112-125.
[38] L. Hall-Stoodley, J. W. Costerton, & P. Stoodley (2004). Bacterial biofilms: from the natural environment to infectious diseases. Nature Reviews Microbiology, 2(2), 95-108. doi:10.1038/nrmicro821
[39] H. C. Flemming & J. Wingender (2010). The biofilm matrix. Nature Reviews Microbiology, 8(9), 623-633. doi:10.1038/nrmicro2415
[40] P. S. Stewart & M. J. Franklin (2008). Physiological heterogeneity in biofilms. *Nature Reviews Microbiology*, 6(3), 199-210. doi:10.1038/nrmicro1838
[41] M. Kostakioti, M. Hadjifrangiskou, & S. J. Hultgren (2013). Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the post-antibiotic era. Cold Spring Harbor Perspectives in Medicine, 3(4), a010306. doi:10.1101/cshperspect.a010306
[42] J. W. Costerton, P. S. Stewart, & E. P. Greenberg (1999). Bacterial biofilms: a common cause of persistent infections. Science, 284(5418), 1318-1322. doi:10.1126/science.284.5418.1318
[43] W. Zimmerli, A. Trampuz, & P. E. Ochsner (2004). Prosthetic-joint infections. New England Journal of Medicine, 351(16), 1645-1654. https://doi.org/10.1056/NEJMra040181
[44] L. M. Baddour, W. R. Wilson, A. S. Bayer, et al. (2015). Infective Endocarditis in Adults: Diagnosis, Antimicrobial Therapy, and Management of Complications: A Scientific Statement for Healthcare Professionals From the American Heart Association. Circulation, 132(15), 1435-1486. https://doi.org/10.1161/cir.0000000000000296
[45] American Academy of Ophthalmology (2016). Cornea. Retrieved from https://www.aao.org/eye-health/anatomy/cornea-103
[46] Verywell Health (2022). The Anatomy of the Cornea. Retrieved from https://www.verywellhealth.com/cornea-anatomy-4800509
[47] D. S. J. Ting, C. S. Ho, R. Deshmukh, D. G. Said, & H. S. Dua (2021). Infectious keratitis: an update on epidemiology, causative microorganisms, risk factors, and antimicrobial resistance. *Eye.* 35, 1084-1101. https://doi.org/10.1038/s41433-020-01339-3
[48] K. McLaughlin-Borlace, F. Stapleton, M. Matheson, & J. Dart (1998). Bacterial biofilm on contact lenses and lens storage cases in wearers with microbial keratitis. *Journal of Applied Microbiology*, 84(5), 827–838. https://doi.org/10.1046/j.1365-2672.1998.00418.x
[49] P. Garg, R. Garg, & P. Raj (2019). Biofilm-Mediated Diseases of the Eye. In: S. Kumar, N. Chandra, L. Singh, M. Hashmi, & A. Varma (eds), *Biofilms in Human Diseases: Treatment and Control*. Springer, Cham. https://doi.org/10.1007/978-3-030-30757-8_8
[50] M. Pan, L. Zhu, L. Chen, Y. Qiu, & J. Wang (2016). Detection techniques for extracellular polymeric substances in biofilms: A review. BioResources, 11(3), 8092-8115.
[51] M. Relucenti, G. Familiari, O. Donfrancesco, M. Taurino, X. Li, R. Chen, M. Artini, R. Papa, & L. Selan (2021). Microscopy Methods for Biofilm Imaging: Focus on SEM and VP-SEM Pros and Cons. Biology, 10(1), 51. https://doi.org/10.3390/biology10010051

[52] B. P. Conlon, S. E. Rowe, & K. Lewis (2015). Persister Cells in Biofilm Associated Infections. In: G. Donelli (eds), Biofilm-based Healthcare-associated Infections. Advances in Experimental Medicine and Biology, vol 831. Springer, Cham. https://doi.org/10.1007/978-3-319-09782-4_1

[53] Y. Liu, H. Yu, Y. Cheng, et al. (2020). Non-destructive Monitoring of *Staphylococcus aureus* Biofilm by Surface-Enhanced Raman Scattering Spectroscopy. *Food Analytical Methods*, 13, 1710–1716. https://doi.org/10.1007/s12161-020-01792-6

[54] N. B. S. Silva, L. A. Marques, & D. D. B. Röder (2021). Diagnosis of biofilm infections: current methods used, challenges and perspectives for the future. *Journal of Applied Microbiology*, 131(5), 2148–2160. https://doi.org/10.1111/jam.15049

[55] Y. Chao, & T. Zhang (2012). Surface-enhanced Raman scattering (SERS) revealing chemical variation during biofilm formation: from initial attachment to mature biofilm. *Analytical and Bioanalytical Chemistry*, 404, 1465–1475. https://doi.org/10.1007/s00216-012-6225-y

[56] K. S. Lee, M. Palatinszky, F. C. Pereira, J. Nguyen, V. I. Fernandez, A. J. Mueller, F. Menolascina, H. Daims, D. Berry, M. Wagner, & R. Stocker (2019). An automated Raman-based platform for the sorting of live cells by functional properties. *Nature Microbiology*, 4(6), 1035-1048.

[57] K. Kong, C. Kendall, N. Stone, & I. Notingher (2015). Raman spectroscopy for medical diagnostics--From in-vitro biofluid assays to in-vivo cancer detection. *Advanced Drug Delivery Reviews*, 89, 121-134. https://doi.org/10.1016/j.addr.2015.03.009

[58] W. Aljuhani, Y. Zhang, M. P. Wylie, Y. Xu, C. P. McCoy, & S. E. J. Bell (2023). Probing the interaction of ex situ biofilms with plasmonic metal nanoparticles using surface-enhanced Raman spectroscopy. *Analyst*, 148(14), 2842-2850. https://doi.org/10.1039/d3an00301a"

[59] T. M. T. Thi, D. Wibowo, & B. H. A. Rehm (2020). *Pseudomonas aeruginosa* biofilms. *International Journal of Molecular Sciences*, 21(22), 8671. https://doi.org/10.3390/ijms21228671

[60] De Gelder, K. De Gussem, P. Vandenabeele, & L. Moens (2007). Reference database of Raman spectra of biological molecules. *Journal of Raman Spectroscopy*, 38(9), 1133-1147.

[61] I. P. Molobela, & F. M. Ilunga (2012). Impact of bacterial biofilms: the importance of quantitative biofilm studies. *Annals of Microbiology*, 62, 461–467. https://doi.org/10.1007/s13213-011-0344-0

[62] J. Xiao (2024). Raman detection of collagen biomarkers. In: *Collagen Mimetic Peptides and Their Biophysical Characterization*. Springer Series in Biomaterials Science and Engineering, vol 20. Springer, Singapore. https://doi.org/10.1007/978-981-10-1941-8_10

[63] L. Van Gulick, C. Saby, H. Morjani, & A. Beljebbar (2019). Age-related changes in molecular organization of type I collagen in tendon as probed by polarized SHG and Raman microspectroscopy. *Scientific Reports*, 9, 7280

[64] L. Galvis, J. W. C. Dunlop, G. Duda, P. Fratzl, & A. Masic (2013). Polarized Raman anisotropic response of collagen in tendon: Towards 3D orientation mapping of collagen in tissues. *PLOS ONE*, 8(5), e63518. https://doi.org/10.1371/journal.pone.0063518"

[65] F. Adar (2022). Interpretation of Raman spectrum of proteins. *Spectroscopy*, 37(2), 9–13, 25. https://doi.org/10.56530/spectroscopy.lo2270l5

[66] T. T. Nguyen, C. Gobinet, J. Feru, S. Brassart-Pasco, M. Manfait, & O. Piot (2012). Characterization of Type I and IV collagens by Raman microspectroscopy: Identification of spectral markers of the dermo-epidermal junction. *Spectroscopy: An International Journal*, 27(5-6), 421–427. https://doi.org/10.1155/2012/686183

[67] MIT Sloan School of Management (2021). Machine learning explained. *MIT Sloan: Ideas Made to Matter*. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained
[68] I. H. Sarker (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, 160. https://doi.org/10.1007/s42979-021-00592-x
[69] J. Rahnenführer, R. De Bin, A. Benner, et al. (2023). Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. *BMC Medicine*, 21, 182. https://doi.org/10.1186/s12916-023-02858-y
[70] M. M. Ahsan, S. A. Luna, & Z. Siddique (2022). Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare*, 10(3), 541. https://doi.org/10.3390/healthcare10030541
[71] V. S. Diogo, H. A. Ferreira, D. Prata, et al. (2022). Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alzheimer's Research & Therapy*, 14, 107. https://doi.org/10.1186/s13195-022-01047-y
[72] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, et al. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25, 65–69. https://doi.org/10.1038/s41591-018-0268-3
[73] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, & D. P. Subha (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Computer Methods and Programs in Biomedicine*, 161, 103-113. https://doi.org/10.1016/j.cmpb.2018.04.012
[74] Y. Ren, T. Chakraborty, S. Doijad, L. Falgenhauer, J. Falgenhauer, A. Goesmann, A.-C. Hauschild, O. Schwengers, & D. Heider (2022). Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics*, 38(2), 325–334. https://doi.org/10.1093/bioinformatics/btab681
[75] B. Li, Y. Wu, Z. Wang, M. Xing, W. Xu, Y. Zhu, P. Du, X. Wang, & H. Yang (2021). Non-invasive diagnosis of Crohn's disease based on SERS combined with PCA-SVM. *Analytical Methods*, 13, 5264-5273. https://doi.org/10.1039/d1ay01046f
[76] S. Kim (2016). Weighted K-means support vector machine for cancer prediction. *SpringerPlus*, 5, 1162. https://doi.org/10.1186/s40064-016-2677-4
[77] Y. Liu, M. Li, H. Liu, C. Kang, & C. Wang (2024). Cancer diagnosis using label-free SERS-based exosome analysis. *Theranostics*, 14(5), 1966-1981. https://doi.org/10.7150/thno.92621
[78] J. Q. Li, P. V. Dukes, W. Lee, M. Sarkis, & T. Vo-Dinh (2022). Machine learning using convolutional neural networks for SERS analysis of biomarkers in medical diagnostics. *Journal of Raman Spectroscopy*, 53(12), 2044–2057.
[79] D. P. dos Santos, M. M. Sena, M. R. Almeida, et al. (2023). Unraveling surface-enhanced Raman spectroscopy results through chemometrics and machine learning: principles, progress, and trends. *Analytical and Bioanalytical Chemistry*, 415, 3945–3966. https://doi.org/10.1007/s00216-023-04620-y
[80] H. J. Byrne, P. Knief, M. E. Keating, & F. Bonnier (2016). Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chemical Society Reviews*, 45, 1865-1878. https://doi.org/10.1039/C5CS00846H
[81] D. Cao, F. Shi, J. Sheng, et al. (2024). Machine learning–driven SERS analysis platform for rapid and accurate detection of precancerous lesions of gastric cancer. *Microchimica Acta*, 191, 415. https://doi.org/10.1007/s00604-024-06508-9

[82] S. Dong, D. He, Q. Zhang, et al. (2023). Early cancer detection by serum biomolecular fingerprinting spectroscopy with machine learning. *eLight*, 3, 17. https://doi.org/10.1186/s43593-023-00051-5
[83] Z.-M. Zhang, S. Chen, & Y.-Z. Liang (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135, 1138-1146. https://doi.org/10.1039/c0an00021d"
[84] M. Člupek, P. Matějka, & K. Volka (2007). Noise reduction in Raman spectra: Finite impulse response filtration versus Savitzky-Golay smoothing. *Journal of Raman Spectroscopy*, 38(9), 1174–1179
[85] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, & A. J. Tallón-Ballesteros (2023). The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In: García Bringas, P., et al. (Eds.), *18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023)*. Lecture Notes in Networks and Systems, vol 750. Springer, Cham. https://doi.org/10.1007/978-3-031-42536-3_33
[86] Shlens (2014). A tutorial on principal component analysis. *ArXiv*. https://arxiv.org/abs/1404.1100
[87] M. Capó, A. Pérez, & J. A. Lozano (2018). An efficient K-means clustering algorithm for massive data. *ArXiv*. https://arxiv.org/abs/1801.02949
[88] S. Dasgupta. (2013). *K-means*. Lecture notes, University of California, San Diego. Retrieved from https://cseweb.ucsd.edu/~dasgupta/291-geom/kmeans.pdf
[89] C. Cortes, & V. Vapnik (1995). Support-vector networks. *Machine Learning*, 20, 273–297. https://doi.org/10.1007/BF00994018
[90] J. Shawe-Taylor, & N. Cristianini (2004). *Kernel methods for pattern analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809682
[91] C. W. Su, C. C. Chang, & C. J. Lin (2010). *A practical guide to support vector classification*. National Taiwan University. Retrieved from https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf
[92] R. Kohavi (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (IJCAI), 1137-1143.
[93] J. Bergstra, & Y. Bengio (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb), 281-305
[94] S. Mallett, S. Halligan, M. Thompson, G. S. Collins, & D. G. Altman (2012). Interpreting diagnostic accuracy studies for patient care. *BMJ*, 345, e3999. https://doi.org/10.1136/bmj.e3999
[95] M. Xuan, S. Wang, X. Liu, et al. (2016). Proteins of the corneal stroma: importance in visual function. *Cell and Tissue Research*, 364, 9–16. https://doi.org/10.1007/s00441-016-2372-3
[96] E. Smith & G. Dent (2005). *Modern Raman Spectroscopy: A Practical Approach*. John Wiley & Sons, Ltd
[97] A. R. Guerrero & R. F. Aroca (2010). Surface-enhanced Raman scattering of hydroxyproline in gold colloids. *AIP Conference Proceedings*, 1267, 922. https://doi.org/10.1063/1.3482891
[98] E. Garcia-Rico, R. A. Alvarez-Puebla, & L. Guerrini (2018). Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: From fundamental studies to real-life applications. *Chemical Society Reviews*, 47, 4909-4923. https://doi.org/10.1039/C7CS00809K

[99] D. Cialla, S. Pollok, C. Steinbrücker, K. Weber, & J. Popp (2013). SERS-based detection of biomolecules. *Nanophotonics*. https://doi.org/10.1515/nanoph-2013-0024
[100] N. A. Brazhe, E. I. Nikelshparg, A. A. Baizhumanov, V. G. Grivennikova, A. A. Semenova, S. M. Novikov, V. S. Volkov, A. V. Arsenin, D. I. Yakubovsky, A. B. Evlyukhin, Z. V. Bochkova, E. A. Goodilin, G. V. Maksimov, O. Sosnovtseva, & A. B. Rubin (2021). SERS uncovers the link between conformation of cytochrome c heme and mitochondrial membrane potential. *bioRxiv*. https://doi.org/10.1101/2021.01.03.425119
[101] J. R. W. Govan, & V. Deretic (1996). Microbial pathogenesis in cystic fibrosis: Mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiological Reviews*, 60(3), 539–574.
[102] B. Sharma, R. R. Frontiera, A. I. Henry, E. Ringe, & R. P. Van Duyne (2012). SERS: Materials, applications, and the future. *Materials Today*, 15(1-2), 16-25.
[103] E. C. Le Ru, & P. G. Etchegoin (2009). *Principles of Surface-Enhanced Raman Spectroscopy and Related Plasmonic Effects*. Elsevier.
[104] S. P. Sharma, S. Dwivedi, S. Kumar, et al. (2023). Bacterial and fungal keratitis: Current trends in its diagnosis and management. Current Clinical Microbiology Reports, 10, 266–278. https://doi.org/10.1007/s40588-023-00210-9
[105] Corneal Scrape Bacteriology. (2020). *Manchester University NHS Foundation Trust*. Retrieved from https://mft.nhs.uk/app/uploads/2020/10/Corneal-Scrape-bacteriology.pdf
[106] W. Kanyongo, A.E. Ezugwu, "Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives" [Reproduced from *Informatics in Medicine Unlocked*, 2023], Retrieved from https://doi.org/10.1016/j.imu.2023.100746.
[107] R. Hunter, "Medical Diagnostics with Surface Enhanced Raman Scattering," Ph.D. thesis, Ottawa-Carleton Institute of Biomedical Engineering, Faculty of Engineering, University of Ottawa, 2022.
[108] M. Deb, " Surface Enhanced Raman Spectroscopy for Bacteria Detection", M.Sc. thesis, Université d'Ottawa / University of Ottawa, 2022.
[109] A. Mandour, "Improvement of Bacteria Detection Accuracy and Speed Using Raman Scattering and Machine Learning", M.Sc. thesis, Université d'Ottawa / University of Ottawa, 2022.
[110] Lazarus IDE, "Lazarus Integrated Development Environment," Retrieved from https://www.lazarus-ide.org/.
[111] Sigma-Aldrich. (n.d.). Retrieved from https://www.sigmaaldrich.com/CA/en
[112] Fisher Scientific. (n.d.). Retrieved from https://www.fishersci.com/us/en/home.html
[113] Ultident Scientific. (n.d.). Retrieved from https://www.ultident.com/index.php/
[114] Ding, L., Wang, J., Cai, S., Smyth, H., & Cui, Z. (2021). Pulmonary biofilm-based chronic infections and inhaled treatment strategies. *International Journal of Pharmaceutics*. https://doi.org/10.1016/j.ijpharm.2021.120768
[115] Wang, J., Lin, W., Cao, E., Xu, X., Liang, W., & Zhang, X. (2017). Surface Plasmon Resonance Sensors on Raman and Fluorescence Spectroscopy. *Sensors*, 17, 2719. https://doi.org/10.3390/s17122719
[116] Investopedia. (n.d.) What is Analysis of Variance? (ANOVA). Retrieved from https://www.investopedia.com/terms/a/anova.asp.
[117] Mohr, D. L., Wilson, W. J., & Freund, R. J. (n.d.). Chapter 7 - Linear Regression. *In* B978-0-12-823043-5. https://doi.org/10.1016/B978-0-12-823043-5.00007-2

[118] Bevans, R. (2020). T-Test: Definition, Formula and Examples. *Scribbr*. Published on January 31, 2020, revised on June 22, 2023. Retrieved from https://www.scribbr.com/statistics/t-test/
[119] A.I. Pérez-Jiménez, D. Lyu, Z. Lu, G. Liu, B. Ren, "Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments," *Chem. Sci.*, 2020, 11, 4563–4577. doi:10.1039/D0SC00809E.
[120] S. Boujday, M.L. de la Chapelle, J. Srajer, W. Knoll, "Enhanced Vibrational Spectroscopies as Tools for Small Molecule Biosensing," *Sensors*, 2015, 15, 21239–21264. doi:10.3390/s150921239
[121] Understanding Raman Spectroscopy," Traces Lab, University of Toronto Scarborough. Retrieved from https://www.utsc.utoronto.ca/~traceslab/PDFs/raman_understanding.pdf.
[122] M.P. Herrling, S. Lackner, H. Nirschl, H. Horn, G. Guthausen, "Chapter Four - Recent NMR/MRI studies of biofilm structures and dynamics," in *Advances in Chemical Engineering*, 2019, vol. 55, pp. 101–145. doi:10.1016/bs.ache.2019.03.002.
[123] R. Jammula, V. R. Tejus, S. Shankar, "Optimal Transfer Learning Model for Binary Classification of Funduscopic Images through Simple Heuristics," 2020, arXiv preprint arXiv:2002.04189.Retrieved from https://doi.org/10.48550/arXiv.2002.04189.
[124] Built In, "Step-by-Step Explanation of Principal Component Analysis," Built In. Retrieved from https://builtin.com/data-science/step-step-explanation-principal-component-analysis.
[125] NumXL, "Principal Component Analysis (PCA) 101," Retrieved from https://numxl.com/blogs/principal-component-analysis-pca-101.
[126] V. Powell & L. Lehe, "Principal Component Analysis (PCA)," Retrieved from https://setosa.io/ev/principal-component-analysis/.
[127] S. Raschka, "Principal Component Analysis in 3 Simple Steps," Retrieved from https://sebastianraschka.com/Articles/2015_pca_in_3_steps.html.
[128] M. Berggren, "RBF Kernel in Support Vector Machines," Retrieved from https://pages.cs.wisc.edu/~matthewb/pages/notes/pdf/svms/RBFKernel.pdf.
[129] Support Vector Machines (SVM) - Analytics Vidhya, https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/.
[130] S. Sindhia. K-Folds and Grey Matter: The Neuroscience Behind Our Built ... - LinkedIn, https://www.linkedin.com/pulse/k-folds-grey-matter-neuroscience-behind-our-built-in-data-sindhia-s9mcc.
[131] H. Smolic. *K-Fold Cross Validation*, https://graphite-note.com/k-fold-cross-validation/.
[132] Z. Li, J. Jin, W. Long, L. Wei, "PLPMpro: Enhancing promoter sequence prediction with prompt-learning based pre-trained language model," Comput. Biol. Med., 2023, 107260. doi:10.1016/j.compbiomed.2023.107260.