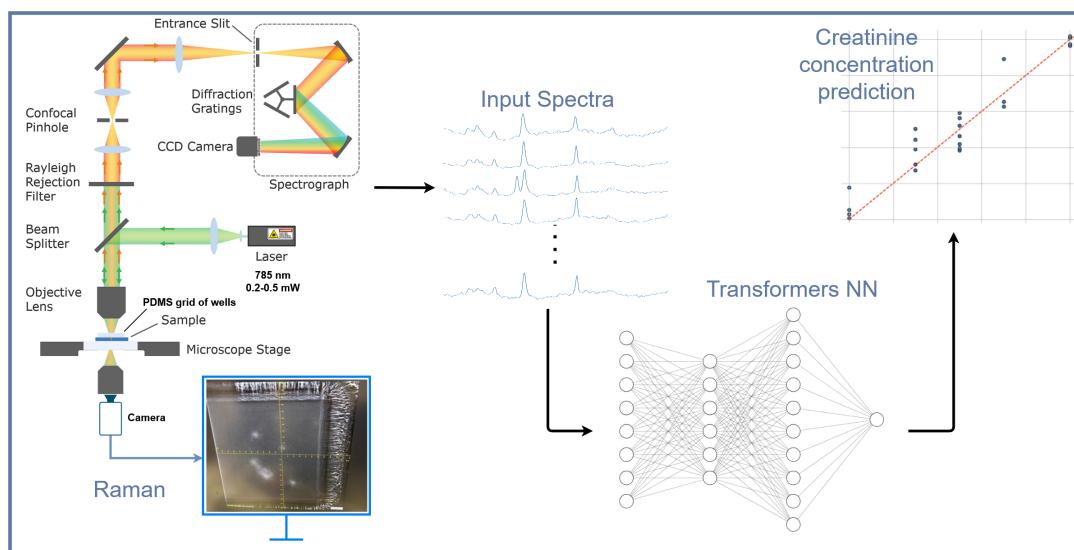


# Surface-Enhanced Raman Spectroscopy for Quantitative, Label-Free Creatinine Sensing Using Transformers Neural Networks



**Daniel Abraham Elmaleh**

Semester Project  
Master's in Microengineering

**EPFL**

**Professor:** Hatice Altug

**Supervisor:** Prof Hatice Altug

**Location:** Bio-Nano-Photonic Systems laboratory

**Date:** June 26, 2025

## Abstract

Continuous, label-free monitoring of creatinine is essential for early diagnosis and management of kidney disease. This project investigates the feasibility of using surface-enhanced Raman spectroscopy (SERS) combined with machine learning to achieve quantitative detection of creatinine in phosphate-buffered saline (PBS). SERS measurements were performed on SERS substrates optimized for 785 nm excitation, with samples spanning concentrations from 30 to 300  $\mu\text{M}$ . A confocal Raman microscope was used to collect spectra from dried micro-spotted samples. The processed data served as an input to a transformer-based neural network architecture. The model was developed to predict creatinine concentrations from Raman spectra using vector regression (SVR) on embeddings from the penultimate layer of the NN model. Yielding strong predictive performance ( $RMSE < 20 \text{ M}$ ,  $R > 0.95$  for the full range;  $RMSE < 10 \text{ M}$ ,  $R > 0.88$  for 0–100  $\text{M}$ ). These results demonstrate the potential of combining SERS with advanced machine learning for sensitive, quantitative detection of creatinine in PBS. Future work will focus on improving reproducibility, extending to serum samples, and developing an automated, portable sensing device with integrated microfluidics for continuous monitoring.

# Contents

<b>Acknowledgments</b>	ii
<b>1 Introduction</b>	1
1.1 Literature Review . . . . .	3
<b>2 Methodology</b>	6
<b>3 Results &amp; Discussion</b>	11
<b>4 Future &amp; Conclusion</b>	18
<b>A Appendix</b>	21
A.0.1 Data ML Stats . . . . .	21
A.0.2 ThorLabs SERS substrate characterization . . . . .	23
A.0.3 T-NN model . . . . .	24

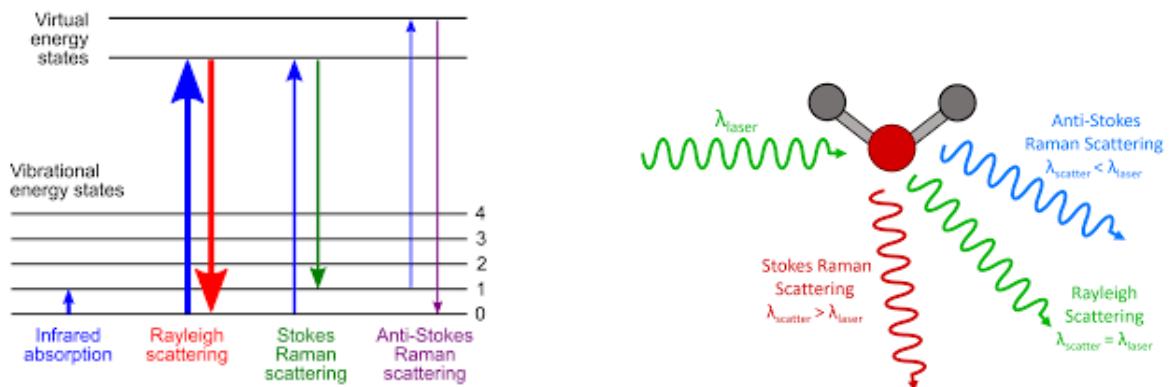
# Acknowledgments

I would like to thank the BIOS Lab for hosting me for this project and for allowing me to use their facilities. I am grateful to Professor **Hatice Altug** for supervision, guidance, and leadership of the lab.

Special thanks go to **Jiayi Tan** and **Xinyi Huang** for their help with sample spotting, Raman measurements, logistics, Raman microscope & software setup, and general support. I would also like to thank **Daniil Riabov** for performing the spectral characterization and SEM measurements of the Thorlabs SERS substrates.

# 1 Introduction

Raman scattering is the inelastic scattering of light by a material in response to laser excitation. When the system absorbs incident photons, it is promoted from a low-energy vibrational state to a high-energy *virtual state*—a transient state that is not a true molecular eigenstate. The scattered light can then relax to a different vibrational level. If the scattered photon has a longer wavelength (lower energy) than the incident light, the process is called *Stokes scattering*. If it has a shorter wavelength (higher energy), it is called *anti-Stokes scattering* (Figure 1.1b and Figure 1.1a).



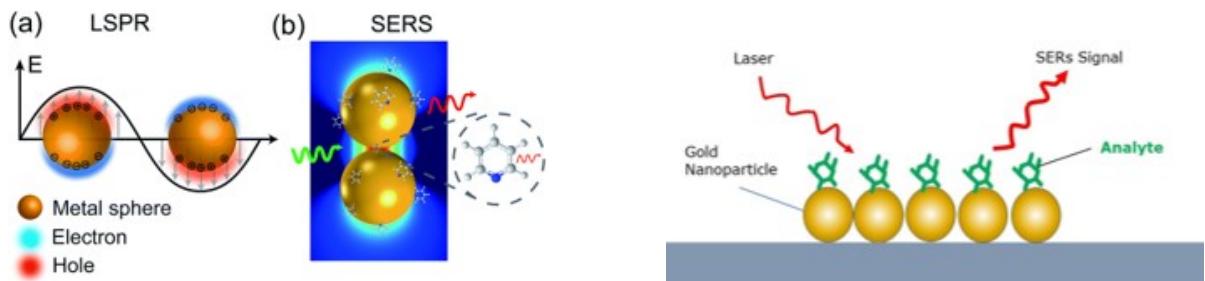
(a) Energy-level diagram illustrating Stokes and anti-Stokes Raman scattering via virtual states. (*Wikipedia*)

(b) Illustration of Raman scattering with incident photon excitation and inelastic scattering to different vibrational levels. (*Wikipedia*)

**Figure 1.1:** Raman scattering process: (a) energy-level diagram showing transitions between vibrational states via virtual states; (b) schematic illustrating photon excitation and inelastic scattering paths.

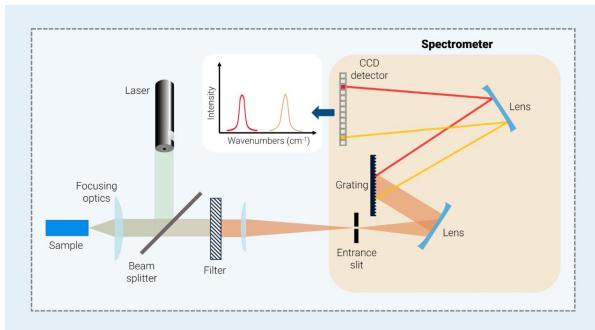
In this project, we focus specifically on *Stokes Raman scattering*, in which molecules in their ground vibrational state are excited to a virtual state and then relax to a higher vibrational energy level, emitting a photon of lower energy (longer wavelength) than the incident light.

*Surface-Enhanced Raman Scattering (SERS)* further enhances this signal using metallic nanoparticles that exploit *localized surface plasmon resonance (LSPR)*. When incoming light excites the surface plasmons on the nanoparticles, it induces strong local polarization. At the resonance frequency, these nanoparticles generate highly enhanced local electric fields (Figure 1.2a). These local fields amplify the Raman signal from nearby analyte molecules, resulting in much stronger signals even at low laser power and analyte concentrations (Figure 1.2b).



**Figure 1.2:** Illustrations of the SERS phenomenon: (a) LSPR effects leading to Raman signal enhancement; (b) conceptual mechanism of SERS for analyte detection.

The experimental setup used for Raman measurements (Figure 1.3) consists of a laser source that illuminates the sample. The scattered light from the sample is collected and directed through a beam splitter, optical filters, slits, and lenses. The light is then dispersed by a diffraction grating to separate different wavelengths before being recorded on a CCD detector. This produces the so-called *fingerprint spectrum* that is characteristic of the analyte in solution.



**Figure 1.3:** General schematic of a Raman spectroscopy setup, showing laser excitation, sample interaction, and spectral detection components. ([Bruker.com](http://Bruker.com))

Quantitative, continuous, label-free sensing of creatinine is highly desirable because creatinine levels provide a direct measure of kidney function. Elevated creatinine levels indicate impaired kidney function, making it one of the most widely used biomarkers for early diagnosis of kidney disease. Creatinine is a non-protein nitrogenous compound produced by the breakdown of creatine in muscle. Continuous monitoring of creatinine levels allows real-time assessment of kidney health and enables early detection of disease progression.

The motivation for this project is to develop a SERS-based approach for continuous, quantitative sensing of creatinine in PBS solution. This is partly driven by participation in the SensUs student competition, which challenges teams to design devices capable of continuously monitoring clinically relevant analyte concentrations in serum and PBS. The goal of this work is to provide a SERS-based feasibility assessment for quantitative,

quasi-continuous sensing of creatinine in PBS, as a first step toward developing such a device for the competition and broader clinical applications.

## 1.1 Literature Review

Creatinine is an endogenous degradation product of muscle metabolism and serves as a crucial biomarker for assessing kidney health and monitoring chronic kidney disease (CKD) progression [1]. It is also widely used as an internal standard in forensic toxicology for quantitative analysis of drugs and xenobiotics [1]. Given the high and rapidly increasing global incidence of CKD, there is significant demand for rapid, quantitative, and low-cost techniques for detecting creatinine concentrations in human urine and blood [1].

Traditional methods for determining urinary and serum creatinine include the Jaffe reaction and enzymatic assays [1]. The Jaffe reaction is commonly used but suffers from low specificity due to interference from other molecules in bodily fluids [1]. Enzymatic methods, such as the “creatinine kit (CK),” offer improved specificity but come with increased detection costs due to the required enzymes [1, 2]. High-Performance Liquid Chromatography (HPLC) has also been explored as an alternative [1]. Consequently, there is a clear need for alternative, efficient methods, particularly optical spectroscopy techniques offering high-throughput screening capabilities [2].

Surface-Enhanced Raman Spectroscopy (SERS) has emerged as a powerful optical technique for rapid, trace-level detection of analytes in complex media, offering unique fingerprint recognition and high sensitivity [1, 3, 4, 2]. SERS combines the intrinsic advantages of Raman spectroscopy—such as non-destructive analysis and compatibility with aqueous environments—with significantly enhanced sensitivity, even achieving single-molecule detection in some cases [3, 2, 5].

Despite its potential, applying SERS for routine quantitative chemical analysis in biological samples remains challenging [1, 6]. Biological matrices introduce interferents such as small molecules, organic salts, nucleotides, and proteins that generate significant background noise or compete for binding on the SERS substrate, reducing detection specificity and sensitivity [1]. Moreover, the field enhancement of SERS signals is highly dependent on the morphology of nanostructures—including nanoparticle size, shape, and interparticle spacing—leading to variability in signal intensities [2].

Several strategies have been developed to address these challenges in quantitative creatinine detection:

### Isotope Dilution SERS

Stosch et al. [6] developed a SERS-based method for quantitative creatinine determination in human serum using isotopically labeled creatinine as an internal standard. This approach leverages the identical chemical properties of isotopomers to correct for variations in sample properties and experimental conditions, similar to isotope dilution mass spectrometry (IDMS) [6]. After thorough serum conditioning involving ultrafiltration and ion exchange to remove proteins and other interferents, they achieved quantification at clinically relevant levels with high accuracy (within 3% of GC/IDMS values). Their method demonstrated a limit of detection (LOD) below 0.1 µg/mL, compared to approximately 100 µg/mL for conventional Raman spectroscopy [6].

## Solvent Extraction and Portable SERS

Zhu et al. [1] developed a portable SERS method for rapid, cost-effective creatinine quantification in human urine using a gold nanoparticle solution (Au sol) as the substrate. They optimized experimental conditions, including the type and concentration of agglomerating salts (with NaCl producing the strongest signals) [1]. To mitigate interference from the complex urine matrix, they employed a simple liquid-liquid solvent extraction procedure (adjusting pH to 10 or 11 with NaHCO<sub>3</sub>–NaOH buffer, using n-butanol as the solvent) [1]. This approach achieved an LOD of 1.45 mg/L, outperforming the 3.4 mg/L of a clinically validated enzymatic CK method [1]. The average difference between their SERS method and the CK method was 5.84%, with the entire detection process completed in 2 minutes (vs. 11 minutes for CK) [1]. The cost was also lower (¥0.2 per SERS sample vs. ¥5 for CK) [1].

## Evaporation-Induced Hotspots on Au Nanocubes

Zhang et al. [2] demonstrated a quantitative SERS approach in serum using evaporation-induced optimal hotspots of gold nanocubes (Au NCs). Time-dependent SERS measurements showed optimal hotspot formation just before complete drying, driven by capillary forces [2]. This yielded reproducible results with a relative standard deviation (RSD) of approximately 4.16% across droplet replicates [2]. They observed a clear linear relationship between log-transformed SERS intensity and log-transformed creatinine concentration ( $R^2 = 0.987$ ) [2], with an LOD of 11 nM in aqueous solutions [2]. When applied to human serum samples (diluted 10,000-fold), their results showed an average difference of less than 5% compared to the Jaffe reaction method, with detection completed in 5 minutes (vs. 25 minutes for Jaffe) [2].

## Integrated Micro-Optical System (MOS) SERS Chip

Yang et al. [4] proposed a high-performance SERS chip integrating a micro-optical system (MOS) with the SERS substrate for serum analysis. The MOS, combining a micro-reflecting cavity and micro-lens, optimizes optical matching between the substrate and Raman system, significantly improving collection efficiency [4]. They prepared a uniform monolayer of silver nanoparticles on gold film using liquid-liquid interface self-assembly [4]. This chip achieved an analytical enhancement factor (AEF) of  $1.46 \times 10^8$ , over 22 times higher than a Si-based substrate [4]. It demonstrated detection limits of 1 μM in aqueous solutions and 5 μM in serum [4]. Using PCA and LDA on serum SERS spectra, they could distinguish healthy individuals from CKD patients with 90% accuracy [4].

## SERS Experimental Parameters

The choice of experimental parameters—excitation wavelength, laser power, and integration time—is crucial for optimizing signal-to-noise ratio (SNR), sensitivity, and avoiding sample damage.

### Usual Excitation Wavelength and Rationale

- **Near-Infrared (NIR)** regions (NIR-I: 650–950 nm, NIR-II: 1000–1350 nm, NIR-III: 1500–1800 nm) are preferred for biomedical applications due to lower tissue extinction [5].
- **785 nm**: Widely used for reduced fluorescence interference and strong SERS enhancement [5, 1, 4].
- **633 nm**: Common as well, offering higher Raman cross-section ( $\propto \tilde{\nu}^4$ ), but with increased risk of sample damage [5, 4, 2].

## Laser Power and Rationale

- SERS signal scales linearly with laser power [5].
- Excessive power can induce plasmonic heating and damage samples [4, 5].
- Reported powers vary widely (e.g., 500 mW [1]; 5–17 mW in microscope systems [4]; 0.326–1.63 mW/cm<sup>2</sup> for solution/serum measurements [2]).

## Integration Time and Rationale

- Longer times improve SNR but increase degradation risk [5].
- Reported times range from 1–3 s for rapid detection [4, 1] to 200 s for high-accuracy analysis [6].

## Relevance to This Project

These literature findings highlight the demand for rapid, low-cost, and accurate creatinine detection, underscoring the clinical relevance of this work [1]. Although this project focuses on PBS—a simpler matrix than urine or serum—the challenges and solutions developed for complex matrices (e.g., solvent extraction [1], conditioning procedures [6], advanced substrates [2, 4]) remain highly informative.

Neural Networks (NN) offer a natural progression in data analysis for SERS, capable of modeling subtle, overlapping spectral features without needing internal standards [3]. Studies such as Koyun et al. [3] demonstrate NN's effectiveness in quantifying chemical mixtures from Raman spectra, suggesting that integrating NN can further improve accuracy and robustness in detecting and quantifying creatinine concentrations from PBS-based SERS measurements.

## 2 Methodology

The plan for the experiment began with a review of the existing literature and an assessment of the constraints imposed by the available instrumentation. Based on these factors, I developed a strategy for a short-term project of approximately four months aimed at producing sufficiently robust results.

The next steps involved preparing the samples precisely, conducting measurements and data acquisition, and then performing systematic data processing. Afterward, a neural network model was developed, trained, validated, and tested on held-out samples.

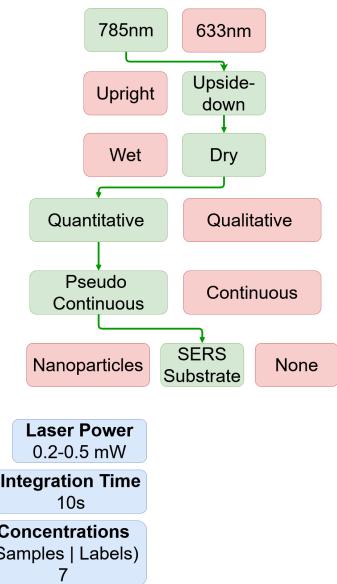
Several key experimental design choices were carefully considered. These included the selection of the laser wavelength, the configuration of the Raman setup (upright or inverted), and the decision to use dry versus wet sample environments. The approach to detection—whether quantitative or pseudo-continuous sensing—was also evaluated. Finally, the choice of SERS enhancement method (substrate-based or colloidal nanoparticles in solution), as well as the laser power and integration time, were important parameters optimized for this project.

There were two excitation wavelengths to choose from: 633 nm and 785 nm. 785 nm was selected for two main reasons. First, the substrate that was planned to be used was optimized for 785 nm excitation. Second, the 633 nm laser generally produces more luminescence from gold due to its higher energy, which increases background luminescence and Stokes Raman scattering from the gold itself.

The experimental Raman microscope available in the lab was an inverted system (sample has to be upside-down), which influenced the decision to use dry samples instead of wet samples. Dry samples were preferable for several reasons: the inverted setup complicates liquid-phase measurements, and our experimental design required depositing very small volumes (tiny drops) using a spotting machine that dries samples quickly.

The aim was for quantitative, pseudo-continuous sensing. Pseudo-continuous sensing because each measurement area on the substrate is used only once (i.e., single-use), it is not truly continuous in the same location, but allows for repeated, quasi-continuous sampling across the substrate.

The SERS substrates were chosen instead of pure Raman or colloidal nanoparticles. Pure Raman spectroscopy was unsuitable because it yielded very weak signals from the analyte at relevant concentrations. Colloidal nanoparticles were also rejected because



**Figure 2.1:** Flow chart of project plan and strategy

they tend to form “hot spots”—aggregates of particles that create highly localized signal enhancement. The number, location, and intensity of these hot spots are unpredictable, making quantitative analysis challenging.

For these reasons, SERS substrates were used featuring a homogeneous, random arrangement of gold-like nanoparticle “blobs” approximately 50–300 nm in size. This random and uniform distribution is crucial for obtaining consistent, reliable, and quantitative measurements across different sample areas.

A laser power in the range of 0.2 to 0.5 mW was used. This choice is justified for two main reasons. First, using low-power laser excitation is a good foundation for developing a continuous monitoring device, as it reduces power consumption and heat generation. Second, lower power minimizes heating of the substrate and the analyte, reducing the risk of damaging or chemically altering the analyte during measurement.

An integration time of 10 seconds was selected. This duration was determined experimentally to balance sufficient signal strength with noise suppression. Ensuring good signal-to-noise ratio in the Raman peaks is essential for consistently quantifying creatinine levels from the spectra.

Seven samples corresponding to seven concentrations were prepared (i.e., seven labels), ranging from 0  $\mu\text{M}$  (pure PBS) to 300  $\mu\text{M}$ , which matches the upper concentration limit required for the SensUs competition [[sensus.org](#)]. The selected concentrations were: 0, 30, 50, 70, 100, 200, and 300  $\mu\text{M}$ . This range was designed to provide good resolution at lower concentrations while covering the full required range with a manageable number of samples. Note: the PBS reference was also used as the 0 concentration for a better generalization of the neural network model and was not considered as a reference in it.

The substrates used were SERS substrates from ThorLabs [[www.thorlabs.com](#)], with dimensions of  $4.3 \times 4.3$  mm. In [Figure A.4a](#) a comparative plot from ThorLabs between bare substrate, gold-coated substrate, and the gold-textured SERS substrate can be shown. As seen in [Figure Placeholder], the Raman peaks are significantly enhanced with the textured substrate due to localized surface plasmon resonance and the resulting electromagnetic field enhancement.

It is important to note that these substrates are specified by the manufacturer as being optimized for 785 nm excitation.

The samples were prepared using materials from Sigma-Aldrich [[www.sigmaaldrich.com](#)], specifically PBS and creatinine, which were mixed with deionized water to achieve the desired concentrations. (Note: the creatinine-PBS-water solution was approximately one month old and stored at 5 °C at time of experiment.)

The Spotting Machine from Cellenion [[www.cellenion.com](#)] was used to deposit very precise volumes onto the SERS substrate, as well as a manual pipette to place small drops on the glass slip for measurements with nanoparticles and for bare substrate tests. This setup is shown at the top of [Figure 3.1](#).

These nanoparticles, also referred to as nanostars (they have a star shape), were purchased from Cytodiagnostics [<https://www.cytodiagnostics.com>] and have dimensions of 60 nm and 70 nm.

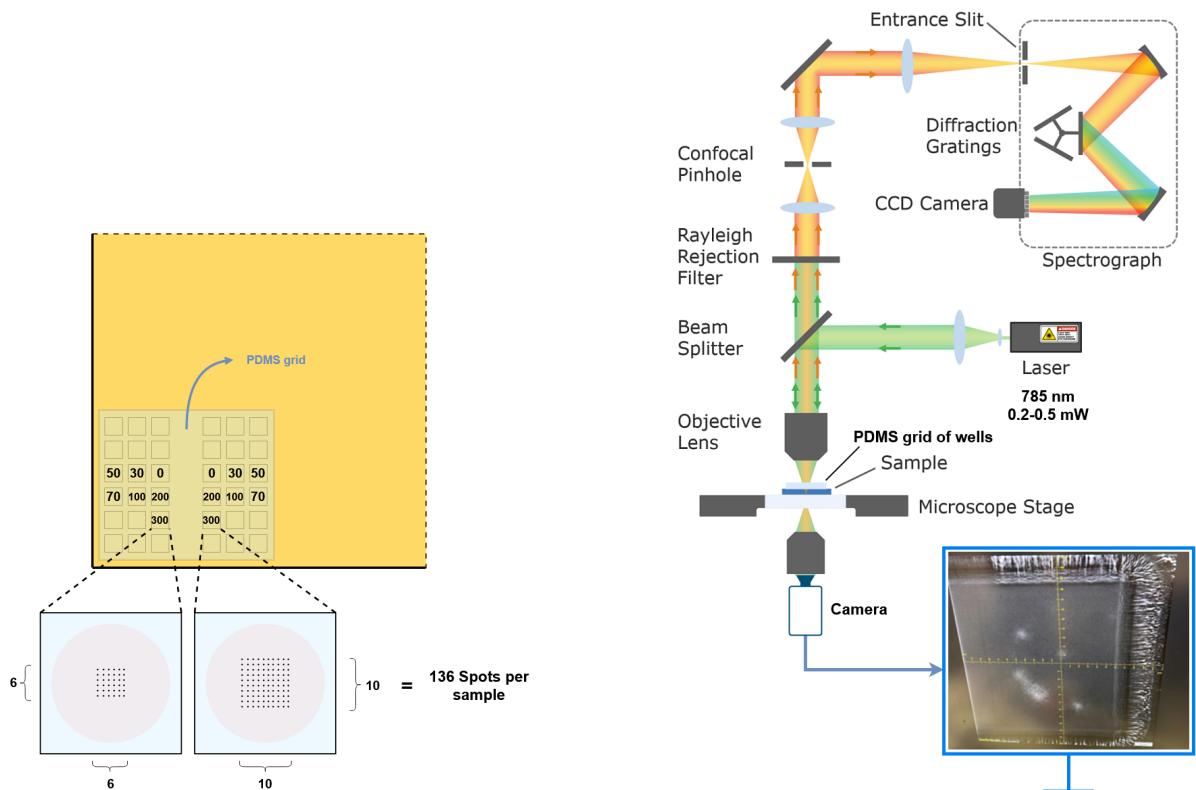
Each drop from the spotting machine was 600 picoliters in volume, and we deposited two drops of each sample in each target well. In the bottom left corner of the substrate (as seen in [Figure 2.2a](#)), a PDMS grid was manually placed to localize and contain the different drops, preventing them from mixing. Accurate localization was essential for ensuring that Raman measurements could later be taken from the correct spots.

As seen in this figure, replicates of each sample were prepared, meaning there are two wells for each concentration on the substrate. Each well in the PDMS grid contained two 600 picoliter drops, which were then dried before measurement.

Regarding the measurement design, the wells on the right side of the substrate were measured with a  $10 \times 10$  grid of Raman laser spots, while those on the left used a  $6 \times 6$  grid. Each “spot” refers to a focused Raman acquisition point. In total, 136 measurements were performed for each sample—that is, for each concentration (hence - label).

In [Figure 2.2b](#) below, the Raman setup is shown, which is a confocal Raman microscopy system using 785 nm laser excitation at a power of 0.2–0.5 mW. The laser is focused onto the sample, and the scattered light is collected, directed through a diffraction grating, and finally detected by a CCD detector to acquire the Raman spectrum.

In the same figure, we can see that an optical camera is mounted in transmission mode. This configuration means that, to visualize the wells, we observed them by imaging through the entire substrate in transmission. For example, the image on the screen in this figure shows a single well in the PDMS grid imaged in transmission through the substrate. The laser excitation is incident on the bottom surface (hence - inverted setup).

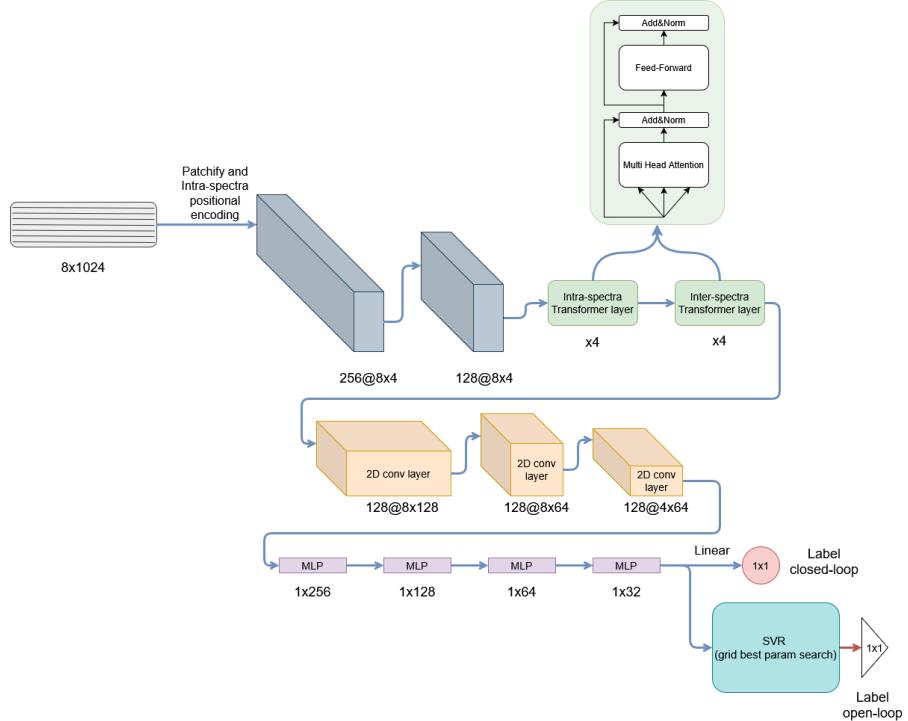


**Figure 2.2:** Experimental setup illustrations: (a) SERS chip conceptual design; (b) Raman system schematic.

## Neural Network Architecture

The model input consists of an  $8 \times 1024$  matrix, representing eight different spots from the same sample (i.e., same label/ concentration), each with 1024 intensity values along the Raman shift axis. Including multiple spots from the same label helps ensure consistent and robust prediction by capturing local variability on the substrate.

The first stage of the network is based on a Transformer architecture. Four Transformer layers were implemented with *intra-spectra attention* to learn dependencies within individual spectra, and *inter-spectra Transformer layers* to capture correlations and shared patterns across different spectra. This design allows the model to learn both local features within a spectrum and relationships among spectra from the same sample. (First row in Figure 2.3).



**Figure 2.3:** Architecture of the neural network with transformer layers, convolutional blocks, mlp and regression output (with the SVR performed on the enultimate-layer embeddings).

Next, the architecture includes 2D convolutional layers. We treat the eight spots as one dimension and the intensity vector as the second dimension, enabling the model to learn spatial and spectral patterns jointly. These convolutional layers feed into a series of multi-layer perceptrons (MLPs), also known as feedforward layers, which perform the final regression step to predict the creatinine concentration. (Second and third rows in Figure 2.3).

This setup constitutes a *closed-loop model*, in which predicted concentrations are compared to ground truth labels during training, the loss is computed, and the error is backpropagated through the entire network to update the weights.

For the Transformer layers, the *patch size* was optimized experimentally and set to 256, meaning the 1024-point spectra are divided into patches of 256 Raman shift ranges [Figure A.5]. Attention layers and feedforward blocks operate on these patches both

within a single spectrum and across spectra from different spots. Positional encoding was applied to maintain the sequential structure of spectral data, preserving critical wavelength-dependent information.

It was found that the final predictions of the initial neural network model produced poor results, as will be shown later. These disappointing results are not yet fully understood, but the bottom line is that direct one-by-one regression of concentration values using the model’s final layer did not perform well.

To improve performance, a workaround was introduced by extracting 32-dimensional embeddings from the penultimate layer of the trained neural network. These compressed, informative features of the input spectra were used as inputs to a Support Vector Regression (SVR) model. The approach included performing a grid search with cross-validation to optimize SVR hyperparameters, training multiple SVRs on slightly varied datasets (bagging) for robustness, and evaluating predictive performance using RMSE and  $R^2$  metrics. The final SVR ensemble model was then saved for later use to predict continuous creatinine concentration labels from new data.

The output in both cases remains a single scalar value per input sample. Since the underlying problem is a regression task, unlike classification problems where spectra would be assigned to discrete categories, here the goal is to predict continuous concentration values across the full range.

To optimize the pipeline, iterative hyperparameter tuning was conducted, partly through trial and error and partly informed by intuition. Parameters explored included the label normalization method, learning rate, patch size, and others.

## Data Processing

The data processing pipeline [Figure A.6] began with the raw spectra acquired from the Raman spectrometer. The first step was formatting the data: saving it in .npy format, converting wavelengths to Raman shifts, and cropping the spectral range to 450–2000 cm<sup>-1</sup>, which is relevant for the model’s focus. The spectra were then interpolated to 1024 points to produce a consistent vector length that is a power of two, aligning with the neural network’s architectural requirements.

Calibration was already applied by the instrument itself, specifically aligning the silicon peak at 520 cm<sup>-1</sup>.

Subsequent pre-processing included baseline correction, outlier removal using several techniques, spectral smoothing, and standard deviation normalization to make the data consistent and model-ready.

Next, the dataset was split into training, validation, and test sets. From the 136 spots per label, the split was 72 for training, 24 for validation, and 40 for testing. Given the relatively small size of these sets, data augmentation was applied to the training and validation data. This augmentation included Gaussian noise addition, spectral shifting, scaling, and random peak dropout. The test set was left untouched to ensure reliable and unbiased evaluation of model performance.

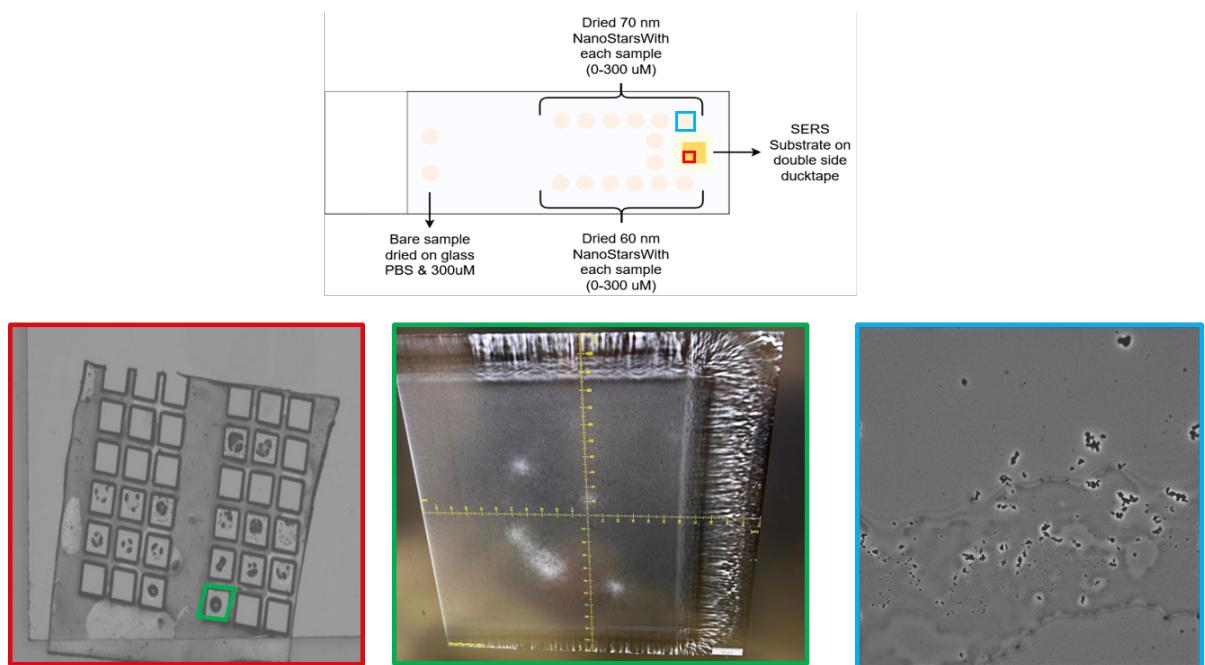
Finally, the data were batched in tensors of shape (8 spots × 1024 intensity points), defining the input shape used during model training. The code can be found on GitHub [7].

### 3 Results & Discussion

In [Figure 3.1](#), the glass cover slip is shown, with hand-pipetted drops of 0.1 mL and the SERS substrate secured to double-sided tape, allowing to flip the glass slip for measurement in the inverted Raman setup.

In the three images in [Figure 3.1](#), we highlight different features:

- **Blue:** the nanoparticle sample, where aggregates of nanoparticles (hot spots) can be clearly observed.
- **Green:** one PDMS wells viewed in transmission through the substrate and the underlying gold layer, captured from the camera under the sample.
- **Red:** the PDMS grid placed on the corner of the SERS substrate, containing the spotted samples deposited by the machine.

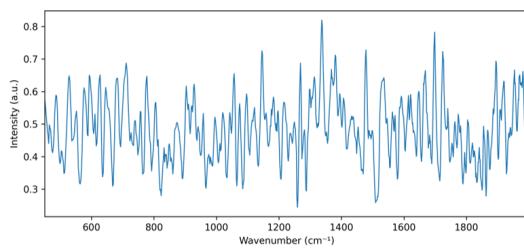


**Figure 3.1:** Illustration of glass slip substrate with: manually spotted samples, SERS substrate and corresponding real microscope images of PDMS grid (red), well and Au nanostars (green) + Creatinine sample (blue).

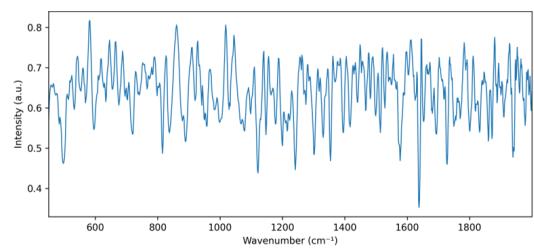
In [Figure 3.2a](#) and [Figure 3.2b](#), the Raman spectra for PBS and 30  $\mu$ M creatinine concentrations measured on bare glass (i.e., without nanoparticles or SERS substrate) are shown. These measurements represent pure Raman (non-SERS) conditions.

In Figure 3.3, a microscope image of the  $70\text{nm}$  nanostars,  $30\mu\text{M}$  creatinine sample is shown. In this image, the dried solution of this drop is visible, a hot spot of nanostars aggregate is circled in green and a regular (non hot-spot) spot in orange. The accompanying plots compares the Raman response for  $30\mu\text{M}$  creatinine in two conditions (a plot of the PBS sample in same conditions is shown below the image):

- **Without hot spots** (minimal SERS enhancement), showing a low signal and noisy.
- **With hot spots**, showing an increase in intensity around  $607\text{ cm}^{-1}$  but still noisy.
- **PBS**, Showing a low signal and noisy.

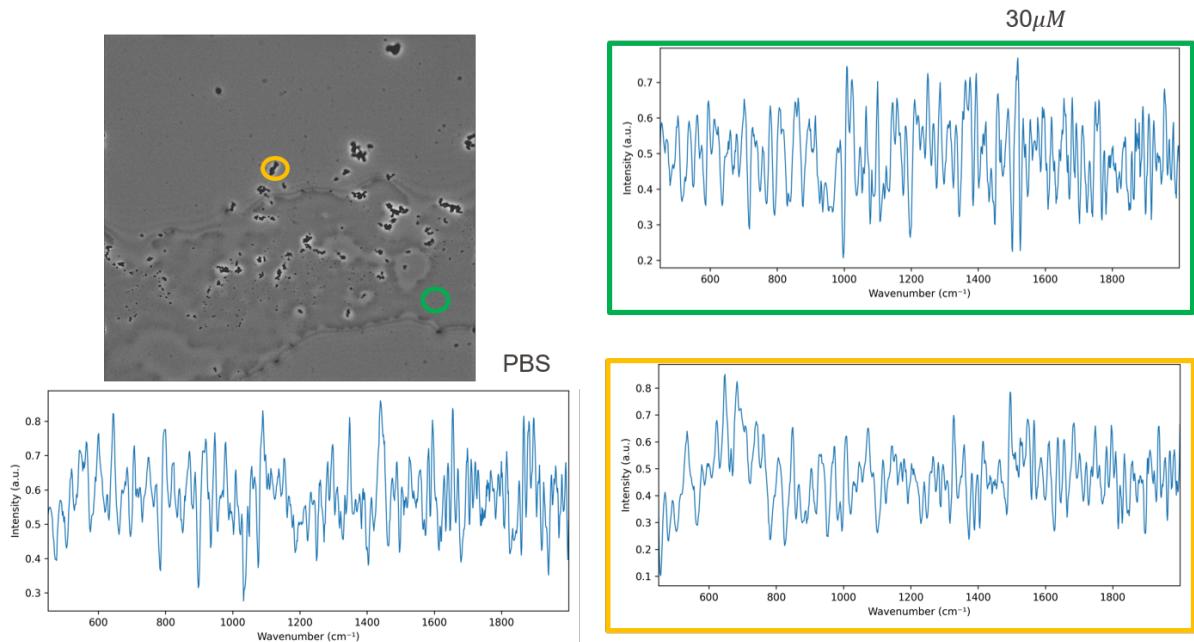


(a)  $30\mu\text{M}$  creatinine on bare glass without SERS enhancement.



(b) PBS sample ( $0\mu\text{M}$  creatinine) on bare glass as control.

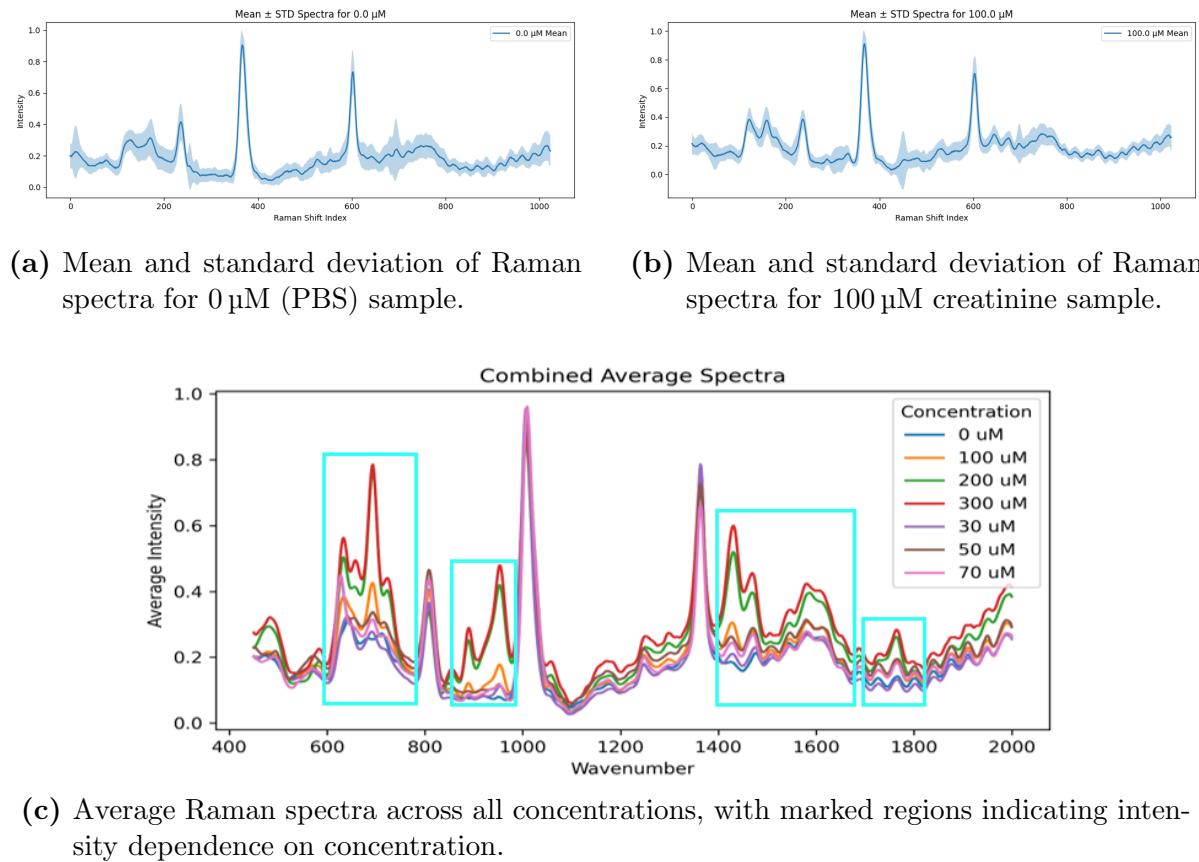
**Figure 3.2:** Raman measurements of bare solution on glass substrate without SERS:  
(a)  $30\mu\text{M}$  creatinine and (b) PBS.



**Figure 3.3:** Microscope image (top-left) and measurements of  $70\text{nm}$  Au nanostars sample +  $30\mu\text{M}$  Creatinine, illustrating nanoparticles aggregation and SERS hotspots. green - non hotspot measurement; Orange - hotspot measurement

Plotted in Figure 3.4a, is the average over all 136 measurement spots of the PBS ( $0\text{ }\mu\text{M}$  creatinine) sample. This average spectrum shows the typical PBS peaks on the SERS substrate. In Figure 3.4b is a plot of the mean and standard deviation spectrum for a  $100\text{ }\mu\text{M}$  creatinine in PBS dried sample on the SERS substrate. Here, we observe small changes in peak intensities around  $600\text{ cm}^{-1}$ ,  $650\text{ cm}^{-1}$ , and  $1100\text{ cm}^{-1}$ , suggesting concentration-dependent signal variations.

Finally, in Figure 3.4c, is the average spectra for all sample concentrations (labels) on the substrate. In this plot, there is visible intensity-to-concentration dependence in the highlighted (cyan) regions, with signal intensity generally increasing with concentration. While this trend is subtle and difficult to assess purely by eye, the figure demonstrates that creatinine is being quantitatively detected and that some concentration-dependent information is clearly visible in the spectra. It is important to note that this visual observation is only a qualitative indication. The neural network model is designed to analyze the entire spectrum in a black-box manner, potentially capturing subtle patterns that are not obvious to human inspection, that encode meaningful concentration information for regression.



**Figure 3.4:** SERS spectral analysis. Top: (a) mean and variability for PBS ( $0\text{ }\mu\text{M}$ ) and (b) for  $100\text{ }\mu\text{M}$  samples, showing individual concentration features. Bottom: (c) overall averaged spectra across all concentrations to highlight global trends.

For outlier removal, several metrics were used to identify and suppress high-variance, poor-quality spectra. These included signal-to-noise ratio, peak count, reconstruction error, and smoothness metrics. Spectra failing these criteria were excluded from the

dataset to improve overall data quality and ensure reliable inputs for the model across training, validation, and test sets [Figure A.1]. (Note: no more than 8 of the 136 spectra were removed from any sample (label))

Exploratory data analysis was also performed to investigate whether meaningful structure was present in the spectral data. For example, Principal Component Analysis (PCA) was applied as an unsupervised ML method to visualize the data. In Figure A.2a, each point is colored according to its true concentration label. However, the groups corresponding to different concentrations are not well-separated in the two principal component dimensions, indicating limited discriminatory power in this reduced space.

The silhouette score is defined as  $s = \frac{b-a}{\max(a,b)}$ , where  $a$  is the mean intra-cluster distance (how close a sample is to other points in the same cluster), and  $b$  is the mean nearest-cluster distance (how close it is to points in the nearest neighboring cluster). A silhouette score close to 1 indicates well-clustered samples, while a score near 0 suggests overlapping clusters, and negative values imply misclassification.

Additionally, K-Means clustering was tested with  $k = 7$  (matching the number of concentration labels, supervised ML method). When the cluster assignments were overlaid on the PCA scores, the results showed poor alignment with the true labels, further illustrating that this clustering approach was not robust [Figure A.2b].

Feature importance analysis was also attempted by evaluating wavelength-wise importance scores, but no single wavelength exhibited an importance greater than 0.01, suggesting that no single Raman shift region dominates the prediction and that the concentration signal is distributed across the spectrum.

Lastly, a Support Vector Machine (SVM) classifier was trained on the data, yielding Figure A.2d. This approach showed relatively positive results, with an accuracy of approximately 0.6. This observation originally motivated the idea of applying Support Vector Regression (SVR) on the penultimate-layer embeddings of the neural network. By extracting these lower-dimensional, learned representations, SVR could potentially achieve better regression performance on the continuous concentration prediction task.

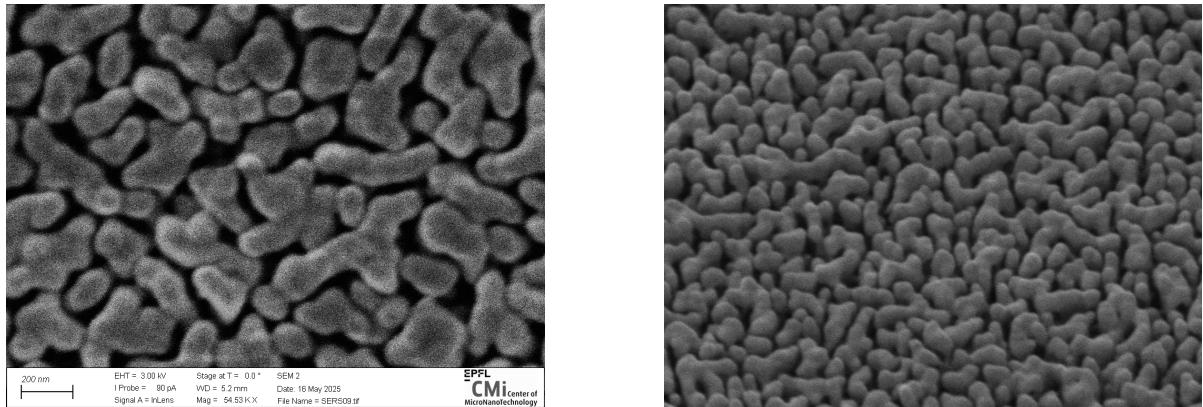
## SERS substrate characterization

The SERS substrate was analyzed using SEM imaging and white-light spectral characterization. In Figure 3.5, we can see a homogeneous, random distribution of gold particle “blobs” across the substrate surface. These structures appear as irregularly shaped gold features with dimensions typically in the 100–200 nm range.

The reflection spectrum of the substrate was also measured. After compensating for the spectral features of a gold mirror (used as a reference), the substrate’s reflection spectrum showed a smooth and regular profile, as expected. This indicates that no significant spectral pattern or artifact is introduced by the substrate itself into the measured Raman signal, ensuring that the substrate does not bias or distort the spectra [Figure A.3a, Figure A.3c].

It is worth noting that the measured blob sizes (approximately 100–200 nm, sometimes approaching 300 nm) are consistent with nanoparticle responses described in Figure A.4b [from Cytodiagnostics]. For example, as shown in this figure, nanoparticles around 150 nm in size exhibit plasmonic peaks near 650–700 nm, which aligns with the 785 nm excitation used in this experiment. Although the substrate particles have irregular shapes and size variation, their spectral properties remain suitable for 785 nm

excitation. Importantly, the reflection measurements confirmed no significant attenuation around these wavelengths, validating that the substrate is well-suited for Raman measurements in this configuration.



(a) SEM image of the Thorlabs SERS substrate at normal incidence, showing the surface morphology and size.

(b) SEM image of the Thorlabs SERS substrate at an angle, highlighting nanoparticle topography.

**Figure 3.5:** SEM characterization of Thorlabs SERS substrates: (a) viewed flat (normal incidence) and (b) viewed at an angle to emphasize surface structure and nanoparticle distribution.

## Deep Neural Network: Direct Predictions in Closed Loop

In [Figure 3.6](#), are the closed-loop prediction results of the neural network when trained and evaluated directly as a regression model (linear last layer activation function). The model's predictions of concentration are poor, with a root mean square error (RMSE) exceeding  $115 \mu\text{M}$  and an  $R^2$  score of approximately  $-0.4$ . These metrics are actually worse than a naïve baseline that simply predicts the average concentration of all samples. This poor performance led to exploring SVM/SVR methods applied to the penultimate-layer embeddings extracted from the neural network [[Figure 2](#)]. The RMSE is defined as  $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ , and the  $R^2$  metric as  $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ . where  $y_i$  are the true concentration values,  $\hat{y}_i$  are the predicted values,  $\bar{y}$  is their mean, and  $n$  is the number of measurements.

## Results from the Neural Network + SVR

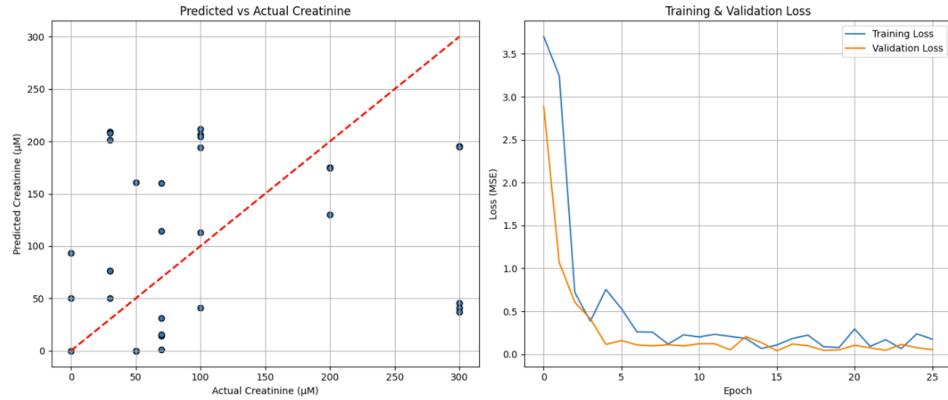
By contrast, the predictions using SVR (Support Vector Regression) on the extracted neural network embeddings showed substantially improved performance. In [Figure 3.7](#) and [Figure 3.8](#), are two plots demonstrating a clear correlation between the true concentrations and the predicted concentrations. The SVR approach achieved regularly an RMSE of less than (or approx.)  $20 \mu\text{M}$  and an  $R^2$  score exceeding  $0.95$ .

## Label Normalization Strategy

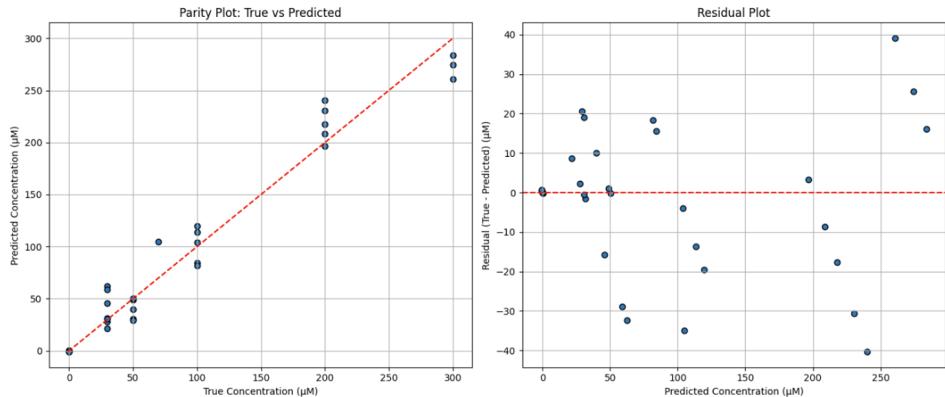
For these experiments,  $\log_{10}$  normalization of the labels was applied. This choice was motivated by the nonlinear spacing of concentrations, particularly the large jumps from  $100$  to  $200$  and  $300 \mu\text{M}$ , compared to the more evenly spaced range from  $0$  to  $100 \mu\text{M}$ .

A separate experiment focused on the 0–100  $\mu\text{M}$  range used maximum-value normalization to preserve the natural linear spacing of these concentrations. In this scenario, the full pipeline—including training, validation, testing, and SVR hyperparameter grid search—was repeated. The results showed regularly an RMSE of less than 10  $\mu\text{M}$  and an  $R^2$  score greater than 0.88 [Figure 3.9].

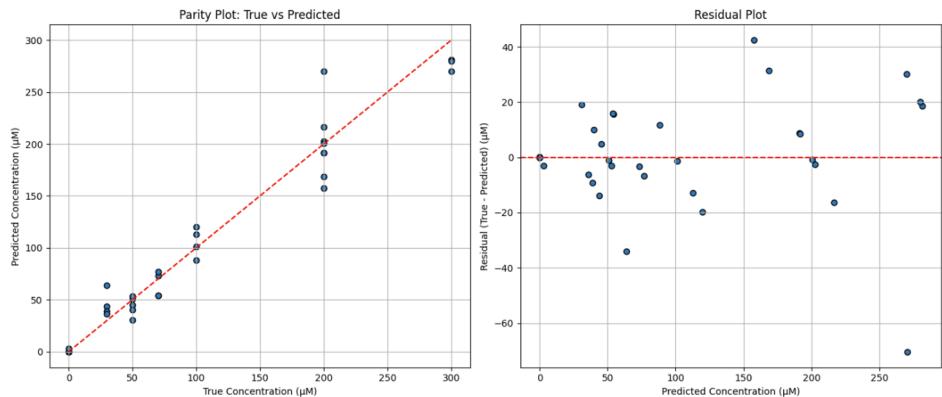
Given the small number of labels and limited number of samples, these results are very promising. With more samples, denser label spacing, and more concentration points, even better predictive performance could likely be achieved.



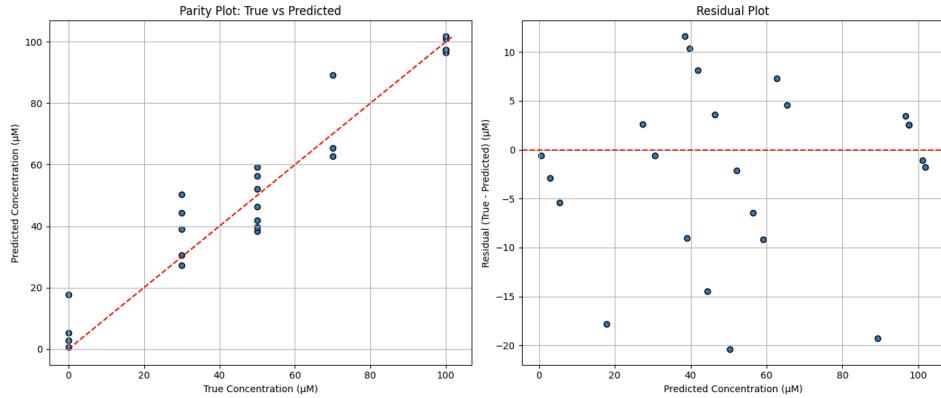
**Figure 3.6:** Closed-loop neural network regression results showing poor performance.



**Figure 3.7:** Regression results using SVR on neural network embeddings.



**Figure 3.8:** Regression results using SVR on neural network embeddings 2 (other run).



**Figure 3.9:** SVR on top of NN regression results on reduced concentration range showing improved accuracy.

## Critical assessment

To validate these results, it is important to recognize several limitations of this experiment.

First, the number of concentration points (labels) is limited and the total number of samples is small. This low data volume reduces the statistical power of model training and evaluation.

Second, the closed-loop deep neural network (i.e., the model trained end-to-end to predict concentration directly) produced poor results. This is typically not a good sign, as it suggests that the underlying spectral features alone may be insufficiently predictive without further feature extraction or regularization.

Third, all samples were prepared and spotted only once on a single substrate. Although 136 measurement spots were acquired for each concentration, from two different areas on the substrate, this single-batch preparation limits the statistical generalization of the results. Variations across multiple preparation runs or substrates were not evaluated here, which would be essential to evaluate real-world reproducibility.

However, I believe that the results demonstrate strong statistical significance within the scope of this pilot experiment, given that the SVR-enhanced predictions achieved relatively good accuracy and clear concentration dependence, even with these limitations.

## 4 Future & Conclusion

### Further

Based on this experiment, several directions can be pursued to improve reproducibility and extend research.

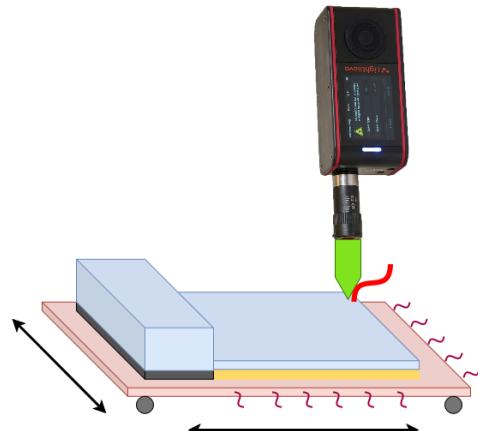
First, the neural network model can be further optimized. This might include adding more relevant inputs, such as specific wavelength peaks, averages over defined spectral regions, or other engineered features that capture key chemical signatures.

A critical next step would be to repeat the experiments multiple times with more samples and more concentration points, increasing the statistical strength of the results and ensuring robust model training.

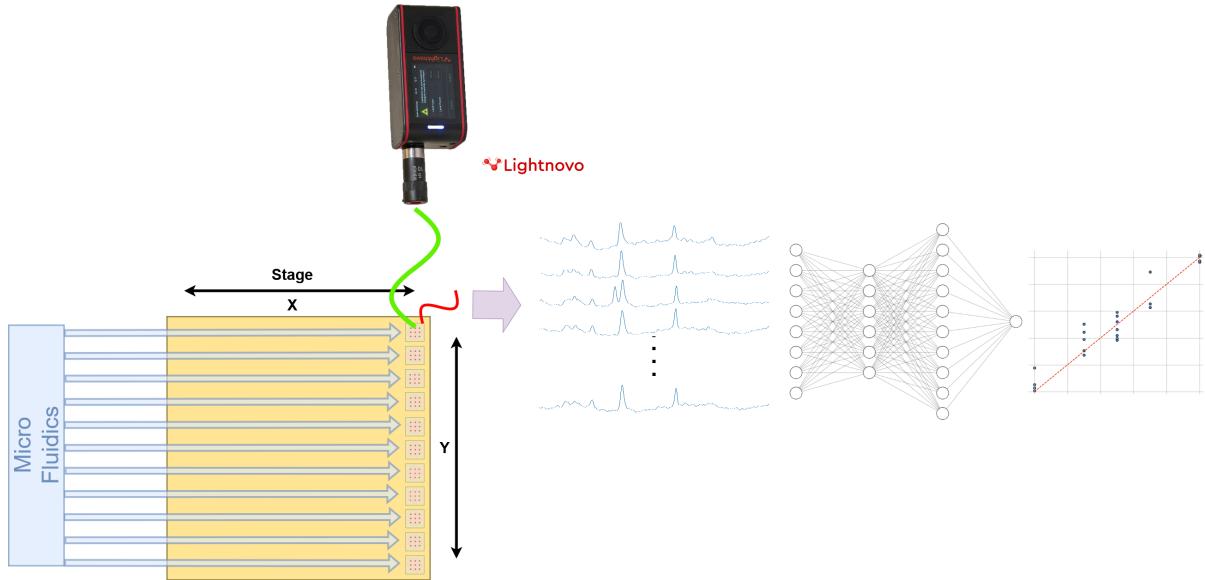
Another important avenue is to test the method in a mixture of serum and PBS, rather than PBS alone, to assess the performance in biologically relevant matrices. This would better simulate real clinical conditions. Additionally using a non-confocal, portable Raman setup (such as the Lightnovo [<https://lightnovo.com/>] portable Raman system) would align better with competition requirements and the goal of building an on-patient monitoring device.

Ultimately, the model should be re-optimized for these real-world specifications. A competition-ready device would also need to incorporate microfluidics to deliver the sample precisely onto the substrate in defined wells or spots. Automation of the entire process will be essential for building a usable and reliable device.

In [Figure 4.1](#), we show a simplified conceptual sketch of what such a device could look like. Although very preliminary, it illustrates the idea of integrating a microfluidic channeled structure with heating 2D piezo stage and properly sized wells on the SERS substrate, forming the basis for a feasible prototype design [Figure 4.2](#) which could perform multiple measurements for each sample spot, once dried.



**Figure 4.1:** 3D conceptual illustration of the proposed portable Raman device prototype.



**Figure 4.2:** Simplified conceptual diagram of the proposed SERS sensing device with integrated microfluidic stage for sample delivery.

## Conclusion

This project focused on predicting creatinine concentrations in a PBS solution using Thorlabs SERS substrates optimized for 785 nm excitation, combined with neural networks based on Transformer architectures. For the concentration range of 0–300, $\mu\text{M}$ , the model achieved an RMSE of less than 20, $\mu\text{M}$  and a  $R^2$  score greater than 0.95. In a narrower 0–100 $\mu\text{M}$  range, results showed an RMSE of less than 10, $\mu\text{M}$  and an  $R^2$  score greater than 0.88.

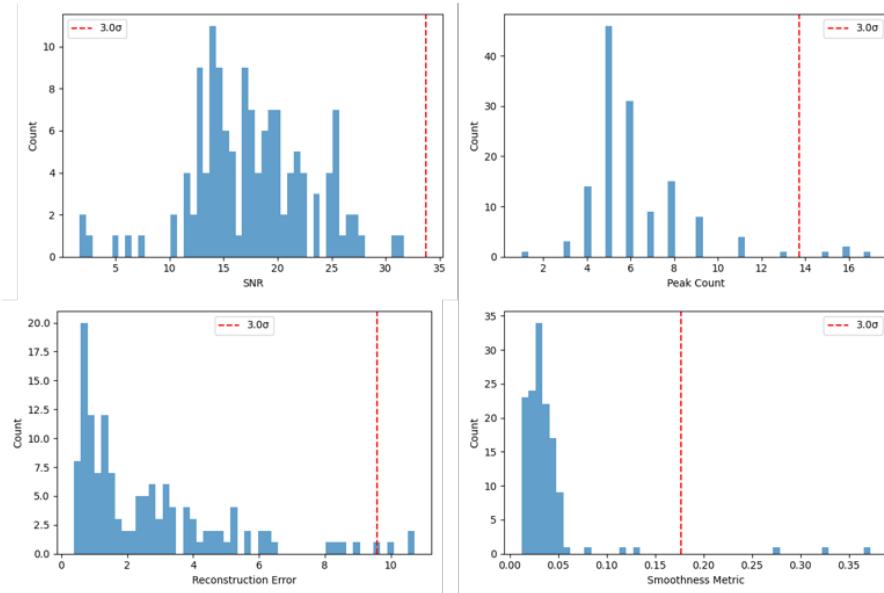
The objective of this project was to assess the feasibility of using SERS and machine learning for label-free quantitative detection of creatinine in PBS. The next step will be to extend the approach to serum samples, acquire more data, and develop an automated measurement and analysis pipeline to enable continuous practical monitoring.

# Bibliography

- [1] W. Zhu, B.-Y. Wen, L.-J. Jie, X.-D. Tian, Z.-L. Yang, P. M. Radjenovic, S.-Y. Luo, Z.-Q. Tian, and J.-F. Li, “Rapid and low-cost quantitative detection of creatinine in human urine with a portable raman spectrometer,” vol. 154, p. 112067.
- [2] L. Zhang, L.-Z. Su, Z. Yu, L.-Y. Su, Y. Zheng, P. Song, Y.-H. Wang, Y.-B. Pan, and B. Zheng, “Quantitative detection of creatinine in human serum by SERS with evaporation-induced optimal hotspots on au nanocubes,” vol. 5, no. 4, pp. 4841–4847. Publisher: American Chemical Society.
- [3] O. C. Koyun, R. K. Keser, S. O. Şahin, D. Bulut, M. Yorulmaz, V. Yücesoy, and B. U. Töreyin, “RamanFormer: A transformer-based quantification approach for raman mixture components,” vol. 9, no. 22, pp. 23241–23251. Publisher: American Chemical Society.
- [4] F. Yang, P. Wen, G. Li, Z. Zhang, C. Ge, and L. Chen, “High-performance surface-enhanced raman spectroscopy chip integrated with a micro-optical system for the rapid detection of creatinine in serum,” vol. 12, no. 8, pp. 4795–4806.
- [5] R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, and L. Fabris, “A review on surface-enhanced raman scattering,” vol. 9, no. 2, p. 57. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- [6] R. Stosch, A. Henrion, D. Schiel, and B. Güttler, “Surface-enhanced raman scattering based approach for quantitative determination of creatinine in human serum,” vol. 77, no. 22, pp. 7386–7392. Publisher: American Chemical Society.
- [7] D. Elmaleh, “Biosense sers project.” [https://github.com/danielmaleh/BioSense\\_SERS](https://github.com/danielmaleh/BioSense_SERS), 2025. GitHub repository.

# A Appendix

## A.0.1 Data ML Stats



**Figure A.1:** Statistical metrics used for outlier removal, including signal-to-noise ratio, peak count, and smoothness.

## Principal Component Analysis (PCA)

PCA is an unsupervised dimensionality reduction technique. It simplifies datasets by transforming correlated variables into fewer, uncorrelated principal components, which retain most of the original data variance. PCA helps visualize high-dimensional data, reduces computation time, and mitigates issues like multicollinearity, making downstream analyses simpler and more interpretable.

## K-Means Clustering

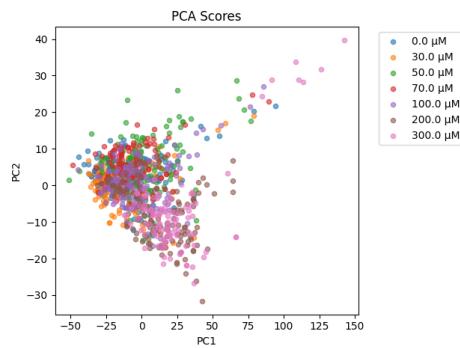
K-Means is an unsupervised machine learning algorithm used to partition data into distinct clusters based on similarity. It groups data points by iteratively assigning them to the nearest centroid (cluster center) and recalculating these centroids until convergence. K-Means helps identify natural groupings or patterns in data without predefined labels.

## Support Vector Regression (SVR)

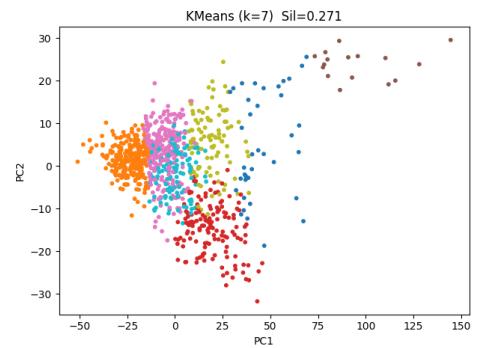
SVR is a supervised machine learning algorithm used for regression tasks. It predicts continuous outcomes by identifying a regression line (or hyperplane in higher dimensions) that best fits the data within a specified margin. Unlike traditional regression methods, SVR aims to minimize errors only beyond a certain threshold, thus focusing primarily on fitting the model within an acceptable tolerance, allowing for robustness against outliers.

### Feature Importance

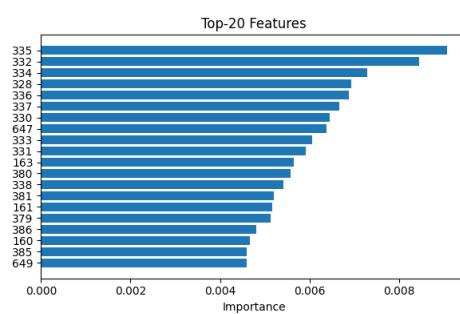
Feature importance refers to methods used to evaluate how influential each feature (variable) is in determining the predictions made by a supervised machine learning model. It helps identify the most relevant variables contributing to model performance, improving interpretability, reducing complexity, and aiding feature selection. Common methods include tree-based feature importance (e.g., Random Forest), permutation importance, and coefficients from linear models.



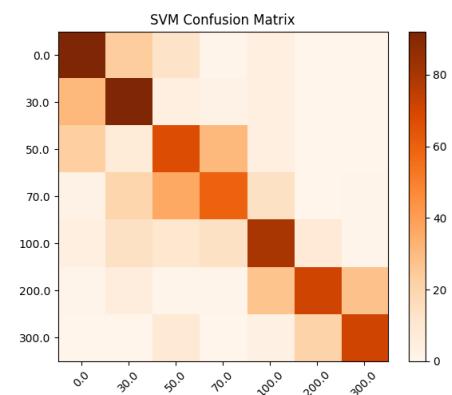
(a) PCA scores colored by true concentration labels.



(b) K-means clustering result projected onto PCA space.



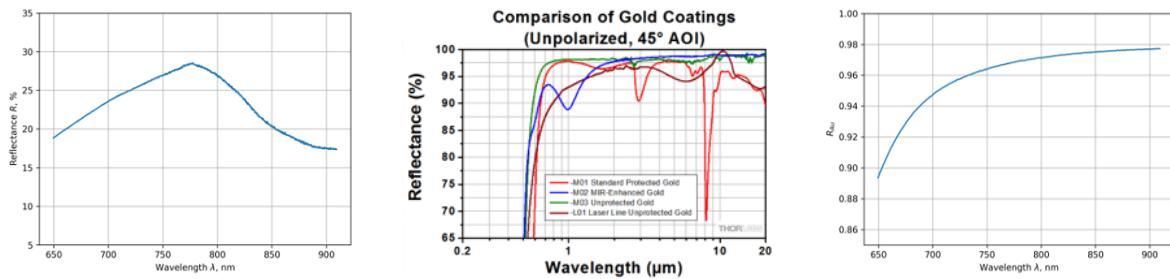
(c) Feature importance analysis across the spectrum.



(d) SVM confusion matrix showing moderate classification accuracy.

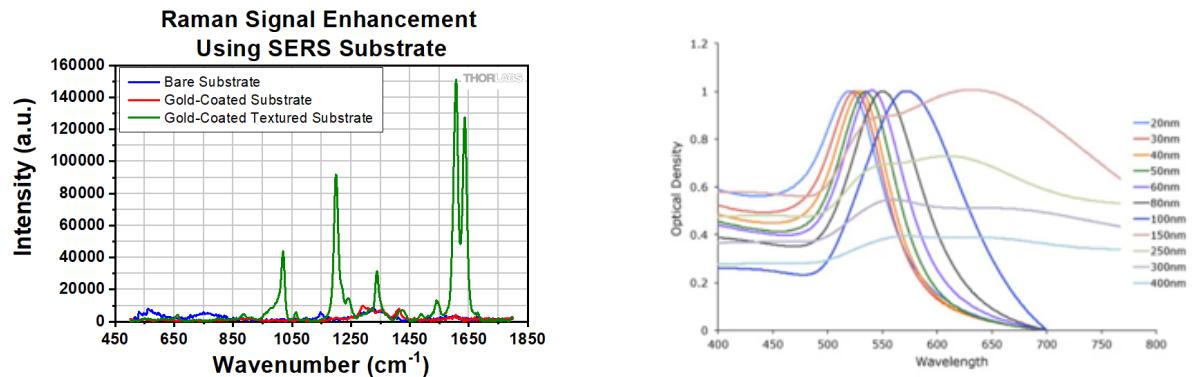
**Figure A.2:** Exploratory data analysis on spectral dataset: (a) PCA visualization with limited group separation; (b) K-means clustering illustrating poor label separation; (c) Feature importance indicating low wavelength-specific predictive power; (d) SVM confusion matrix showing moderate classification performance.

### A.0.2 ThorLabs SERS substrate characterization



- (a) Spectral response normalized to gold mirror reference.  $R = \frac{R_{SERS}(\text{cts.})}{R_{\text{lamp from Au mirror}}} * 100\%$
- (b) Spectral features of gold (red curve). (*Wikipedia*)
- (c) Rectified spectral features after features removal.  $R = \frac{R_{SERS}(\text{cts.})}{R_{\text{lamp from Au mirror}}/R_{Au}} * 100\%$

**Figure A.3:** Spectral characterization of the SERS substrate.

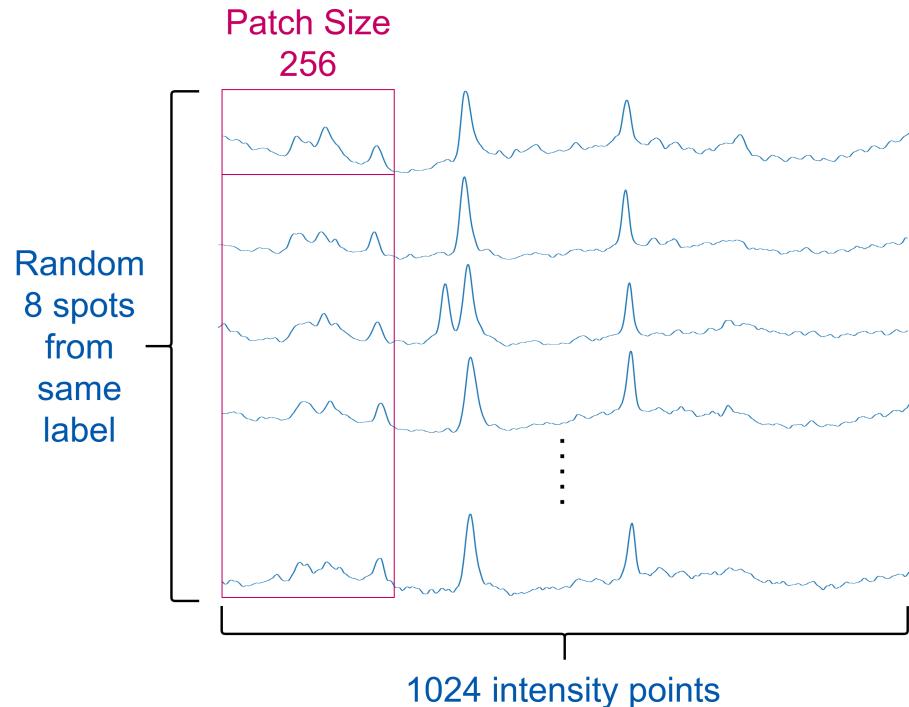


- (a) Data from ThorLabs comparing response with SERS nanostructured gold surface, regular gold mirror and bare substrate. (*ThorLabs.com*)

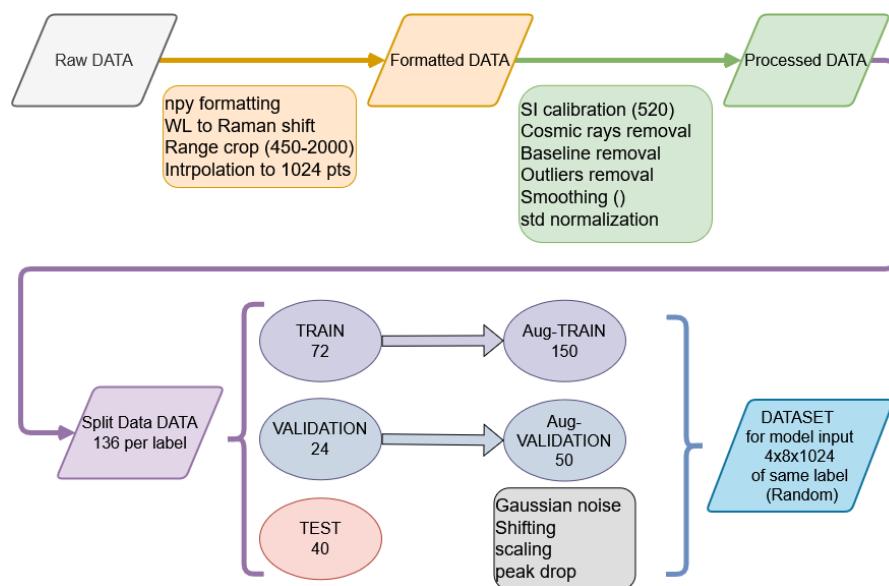
- (b) Supplier data showing relationship between gold nanoparticle size and optical density for characterization reference. (*Cytodiagnostics.com*)

**Figure A.4:** (a) Comparative data for ThorLabs SERS substrate (b) Characterization reference data for gold nanoparticles.

### A.0.3 T-NN model



**Figure A.5:** Illustration of NN input shape and patch size (256 wavenumbers) used for transformer attention layers.



**Figure A.6:** Data processing pipeline from raw Raman spectra to model-ready inputs including preprocessing and augmentation.