UNIVERSITY OF CALIFORNIA,

IRVINE

Development and Analysis of Plasmonic Nanomaterials for Biosensors

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Engineering

by

Vinusha Reddy Kura

Thesis Committee:

Associate Professor Regina Ragan, Chair

Assistant Professor Allon Hochbaum

Professor James Earthman

2018

TABLE OF CONTENTS

LIST OF FIGURES ..... **iv**

LIST OF TABLES ..... **vi**

ACKNOWLEDGMENT ..... **vii**

ABSTRACT OF THE THESIS ..... **viii**

CHAPTER 1. INTRODUCTION **1**

1.1. RAMAN SCATTERING **3**

1.2. SURFACE ENHANCED RAMAN SPECTROSCOPY (SERS) **4**

1.3. METABOLOMICS **7**

1.4. SPECTRAL ANALYSIS OF SERS DATA **9**

1.5. THESIS ORGANIZATION **11**

CHAPTER 2. SERS SUBSTRATES PREPARATION **12**

2.1. SYNTHESIS OF GOLD NANOPARTICLES **14**

2.1.1. Experiment 16

2.1.2. Results 17

2.2. FUNCTIONALIZATION OF GOLD NANOPARTICLES **19**

2.2.1. Experiment 20

2.3. FABRICATION OF NON-LITHOGRAPHIC SERS SUBSTRATES **20**

2.3.1. Experiment 24

2.3.2. Results 27

CHAPTER 3. SENSOR CHARACTERIZATION **29**

3.1. DETECTION OF METABOLITES **29**

3.1.1. Experiment 32

3.1.2. Results 33

3.2. CONCENTRATION STUDIES **34**

3.2.1. Experiment 35
Results 36
CHAPTER 4. SPECTRAL ANALYSIS USING MACHINE LEARNING 38
4.1. INTRODUCTION TO MACHINE LEARNING 38
4.2. APPLICATIONS OF MACHINE LEARNING IN BIOLOGY 39
4.3. DEEP LEARNING 40
4.4. ARTIFICIAL NEURAL NETWORKS (ANN) IN BIOLOGY 43
4.5. BUILDING AN ANN PREDICTION MODEL 44
4.5.1. Training with Feed-forward Neural Networks (FF-NN) 44
4.5.2. Training with Convolutional Neural Networks (CNN) 46
4.6. SERS DATA ACQUISITION AND BUILDING A PREDICTION MODEL 47
4.6.1. Using Feed-forward Neural Network (FF-NN) 48
4.6.2. Using Convolutional Neural Network (CNN) 51
4.7. MULTIPLEX ANALYSIS USING ANN 54
4.7.1. Experiment 56
4.7.2. Results 56
4.8. CONSIDERATIONS FOR THE BIANALYTE APPROACH AND FUTURE WORK 58
CHAPTER 5. CONCLUSION 60
REFERENCES 61

LIST OF FIGURES

Figure 1.1. A schematic representation of the energy-levels showing the states involved in Raman spectra.....**3**

Figure 1.2. Electric field variation in Localized surface plasmon of spherical noble metal nanoparticles……**5**

Figure 2.1. (a) UV-Vis Absorption spectrum for 40 nm Gold nanoparticles. (b) Resultant size distributions of the colloidal gold nanoparticle solution derived using DLS…….**17**

Figure 2.2. Schematic representation of lipoic acid functionalized Au NPs [65]……**19**

Figure 2.3. A schematic diagram explaining EHD attractive forces in a carboxylic acid-functionalized Au nanospheres colloidal solution near a Si substrate, self-organized with PS-*b*-PMMA diblock copolymer. Inset showing the EHD flow generated from inhomogeneous field at the working electrode surface through polarization of the 'seed' monomers, increasing the probability of nanospheres – nanosphere collisions to immobilize the Au nanoparticles through anhydride bridging [71]…….……**22**

Figure 2.4. Carbodiimide chemistry pathways showing covalent linkages between carboxylic acid functionalized Au nanoparticels and amine-functionalized surface and between nanospheres [71].……**23**

Figure 2.5. (a) SEM image of Au clusters on SERS substrates. (b) SERS spectrum of 10^{-3} M BZT (c) Normalized SERS intensity map of BZT's 1573 cm^{-1} vibrational band across a 100 μm ×100 μm area with RSD of 10.4%. Inset depicting the measured SERS EF mean value to be 1.4×10^9 [77]……**27**

Figure 3.1. SERS spectrum of 1 $\mu g.mL^{-1}$ of (a) o-cresol (b) indole (c) benzoxazole (d) 2-mercaptobenzoxazole (e) 2- aminoacetophenone……**...33**

Figure 3.2. SERS spectral analysis of (a) indole and (b) 2-AAP for concentrations (i) 1 pg.mL $^{-1}$ (ii) 10 pg.mL^{-1} (iii) 100 pg.mL $^{-1}$ (iv) 1 ng.mL^{-1} (v) 10 ng.mL $^{-1}$ (vi) 100 ng.mL $^{-1}$ (vii) 1 μg.mL $^{-1}$ (viii) 10 μg.mL $^{-1}$ (ix) 100 μg.mL $^{-1}$ (x) 1 mg.mL $^{-1}$ **..…….....36**

Figure 4.1. Structure of a neural network consisting of a hidden layer that connects the input layer to the output layer [111]….……**41**

Figure 4.2. Architecture of a perceptron involving information processing. Associated with weighted parameters that can be tuned with activation functions……**42**

Figure 4.3. (a) Accuracy and (b) loss prediction of indole during the training of FF-ANN. (c) Predicted concentrations of Indole (water depicted by a grey bar) using FF-ANN……...**50**

Figure 4.4. (a) Accuracy and (b) loss prediction of 2-AAP during the training of FF-ANN. (c) Predicted concentrations of 2-AAP (water depicted by a grey bar) using FF-ANN……...**51**

Figure 4.5. (a) Accuracy and (b) loss prediction of indole during the training of Convolutional ANN. (c) Predicted concentrations of indole (water depicted by a grey bar) using Convolutional ANN……**53**

Figure 4.6. (a) Accuracy and (b) loss prediction of 2-AAP during the training of Convolutional ANN. (c) Predicted concentrations of 2-AAP (water depicted by a grey bar) using Convolutional ANN……**54**

Figure 4.7. Multiplex detection involving two output nodes where indole is predicted……**57**

Figure 4.8. Multiplex detection involving two output nodes where 2-AA is predicted……**57**

LIST OF TABLES

Table 2.1. Scattered distribution report of the gold nanoparticle by intensity using DLS……...**18**

Table 3.1. Chemical structure of metabolites chosen for SERS analysis…….**31**

ACKNOWLEDGMENT

I am indebted to my committee chair, Dr. Regina Ragan, for providing me the opportunity to join her group and carry out research in her lab. She has always been supportive of my research and career goals, and this work would not have been possible without her encouragement and support.

I would like to thank my committee member, Dr. James Earthman, for his passionate participation and inputs in the validation of this research project.

I am grateful to my fellow labmates; Cuong and Will, who have invested a lot of time and effort in guiding me to achieve my project goals. I am also thankful to my family and friends, whose love and guidance are with me in whatever I pursue.

ABSTRACT OF THE THESIS

Development and Analysis of Plasmonic Nanomaterials for Biosensors

By

Vinusha Reddy Kura

Master of Science in Engineering

University of California, Irvine, 2018

Associate Professor Regina Ragan, Chair

Optical bio-nanosensors are powerful alternatives to conventional analytical techniques because they enable direct, real-time and label-free detection of many biological and chemical substances. Surface enhanced Raman scattering (SERS) spectroscopy is a label-free optical biosensing technique that provides target specific information, enables real-time monitoring of continuous flows and multiplex detection of small volatile molecules. For example, the detection of small molecule metabolites has proved to be crucial in study of microbial activity and early detection of diseases. In the first part of this research, I prepare SERS substrates designed to improve the analytical capabilities of SERS. First, synthesis of uniform Au nanopsheres via a kinetically controlled seeded growth method is performed. Second, chemically assembled Au nanoparticle clusters on self-organized templates are characterized as SERS substrates. The enhancement of SERS signals by controlling nanometer gap spacings between plasmonic nanospheres using the length of the chemical cross linker enables detection of various metabolites down to 1 $pg.mL^{-1}$.

In the second part of this work, I discuss the advantages of using deep learning algorithms for quantification of metabolites. Feed-forward neural networks and 1 D- Convolutional neural

networks are trained on indole and 2-aminoacetophenone SERS spectral data to get robust quantification of concentrations ranging from 1 $pg.mL^{-1}$ to 1 $mg.mL^{-1}$. The models are further assessed for future applications of SERS sensors for multianalyte detection in biological fluids.

CHAPTER 1. INTRODUCTION

A biosensor is an analytical device used for detection of molecules, that combines a biological recognition mechanism with a physical transduction technique [1]. They have applications in drug discovery, disease monitoring, detection of disease causing micro-organisms, or biomarkers that are indicators of a disease in biological fluids. An important criterion for biological or chemical sensing is the sensitive detection of trace compounds in complex biological samples. Several techniques have been explored to provide unambiguous identification and precise quantification of these compounds [2].

Optical biosensors are powerful tools for detection and analysis of compounds in vast areas of applications such as health-care, pharmaceuticals and biomedical research. These sensors can perform fast, real- time remote sensing while enabling multiplexed detection of compounds. Optical biosensing can incorporate the following transduction mechanisms: fluorescence based detection, surface plasmon resonance (SPR) and surface enhanced Raman scattering (SERS). Fluorescence involves labeling target molecules with fluorescent tags or dyes; the fluorescence intensity indicates the concentration of target molecules and the strength of the interaction between the target and biorecognition molecules. The limit of detection for proteins in human serum albumin was determined to be 0.3 $\mu g.mL^{-1}$using fluorescence spectroscopy [3]. Although this technique is highly sensitive and selective; the process of labeling is challenging and it interferes with the function of the biomolecules. Quantification becomes difficult due to fluorescence signal bias as the number of fluorophores on each molecule cannot be precisely controlled [4]. In addition, fluorophore labels are susceptible to photobleaching effect. Moreover, these techniques

involve the overlapping of the broad fluorescence peaks, resulting in difficulty in deconvoluting mixtures of signals limiting multiplexed detection [5]. On the other hand, SERS has a narrower spectral width of Raman peaks minimizing their overlapping, hence enabling multiplex detection compared to fluorescence based detection techniques [6].

Surface plasmon resonance is a popular approach that relies on detecting the variation in refractive index (RI) due to biomolecular binding events at metal/ dielectric interfaces by observing changes in plasmon coupling angle or wavelength. Typically, a label- free SPR biosensor has a LOD of 10 $pg.mL^{-1}$ [7]. However, SPR cannot easily discriminate between the target and non-specifically bound molecules [8]. In contrast, SERS can produce highly sensitive molecularly target specific information relative to SPR.

SERS biosensors are used in detection of various biological samples and diseases including different cancers [9], Parkinson's disease [10] and Alzheimer' s disease [11]. Hence, SERS has emerged as an important label-free optical sensing method for *in vitro* and *in vivo* medical diagnostics as it is capable of achieving single sensitivity in highly diluted solutions. The physical mechanisms enabling transduction of molecular signals is discussed in the following sections. However, common challenges in using SERS include (1) creation of SERS active substrates with low limit of detection and sample to sample reproducibility; (2) difficulty in rapid interpretation of the SERS spectra as Raman bands may have low signal to noise. The aim of the project is to address the above two problems by using a hierarchical chemical self- assembly method for fabricating SERS substrates to achieve necessary enhancements for reproducible detection of biological samples and employing machine learning algorithms to robustly quantify concentrations of biological samples down to 1 $pg.mL^{-1}$ respectively.

1.1. RAMAN SCATTERING

Raman spectroscopy is a powerful spectroscopic technique that provides structural fingerprint of molecules allowing for their identification. It relies on the inelastic scattering of light by vibrational or rotational modes of chemical bonds within the molecule. Raman shifted photons can either be of higher (Anti-stokes) or lower energy (Stokes) from excitation relative to the vibrational energy levels in the ground electronic state (shown in Figure 1.1). The intensity of Raman effect correlates with molecular polarizability during vibrational transitions. Raman spectroscopy requires no special dyes and is mostly a non-destructive and non-contaminating technique [12].

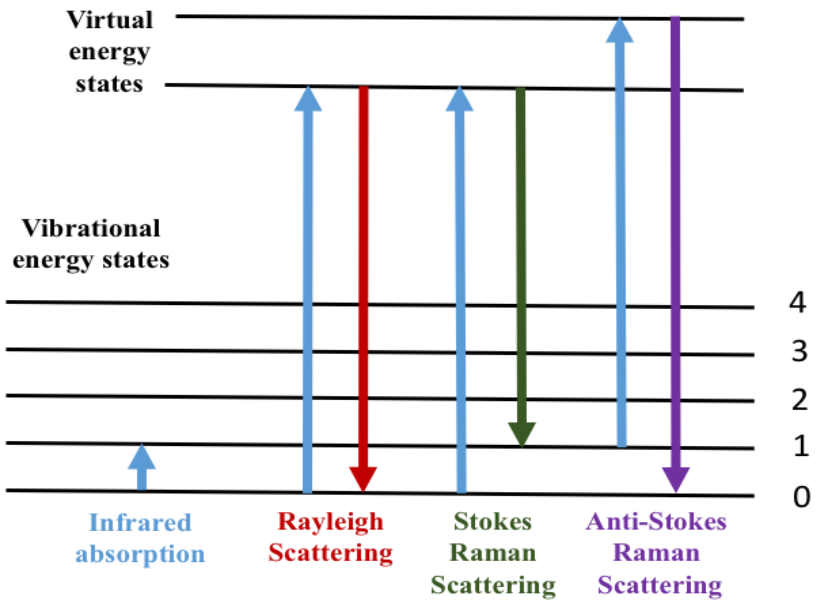


Figure 1.1. Schematic representation of the energy-levels showing the states involved in Raman spectra.

In Raman spectroscopy, the low cross section for inelastic scattering gives rise to low intensity of measured signal limiting sensitive detection of molecules. This has restricted the use of this spectroscopy techniques in molecular biology. A mechanism that enhances the cross section for

Raman scattering can be observed in the presence of rough metallic nanostructures or nanoparticles. This phenomenon is known as Surface Enhanced Raman Spectroscopy (SERS).

1.2. SURFACE ENHANCED RAMAN SPECTROSCOPY (SERS)

SERS was first observed in 1974 by Fleischman et al. as a large Raman signal from pyridine adsorbed onto roughened silver electrode surfaces [13]. The enhancement of Raman scattering intensities was attributed to electromagnetic (EM) mechanisms and chemical mechanisms (CM) by Jeanmaire and Van Duyne respectively [14]. The two enhancement mechanisms, EM and CM, are considered to account for this phenomenon which can be easily understood by the equation (1.1) for an induced dipole moment μ [15].

$$\mu = \alpha E \tag{1.1}$$

where α is the polarizability of the molecule, E is the applied electric field. This signifies that that intensity of Raman scattering is directly proportional to the square of induced dipole moment that can be written as a product of polarizability (CM) and the incident electromagnetic (EM) field intensity as given in the above equation [12].

For molecules chemisorbed on SERS active surface, CM involves establishment of a charge transfer state between the molecule and metal surface, thereby inducing a change in polarizability α of the molecule. The CM phenomena influences the band positions, Raman modes and SERS intensity for a single crystal surface by the following factors [16]: (1) The orientation of the molecule in relation to the metallic surface affects SERS intensity due to surface selection rules. Raman modes perpendicular to the surface are preferably enhanced. (2) Only Raman modes close

to the vicinity of approximately 10 nm from the metal surface are enhanced significantly. This can be attributed to the evanescent wave or field characteristics of the electric field. (3) Sulfur, nitrogen and oxygen show high affinity to silver and gold. Such chemical groups bind easily to the metal because of which their vibrations are dominant in SERS spectrum. In SERS, while the CM factor accounts for two to three orders of magnitude in enhancement, the largest contribution to SERS cross sections comes from EM enhancement factor [17].

Under the influence of local surface plasmon resonance mode (LSPR), the EM in SERS is greatly enhanced. LSPR is a collective oscillation of conduction band electrons in metal nanoparticles excited by the electromagnetic incident light. The interactions between the light wave and surface electrons in the conduction band of nanoparticles give rise to localized plasmon oscillations with resonant frequency (shown in Figure 1.2) that depends on the nanoparticle size, composition, geometry, environment and gap.

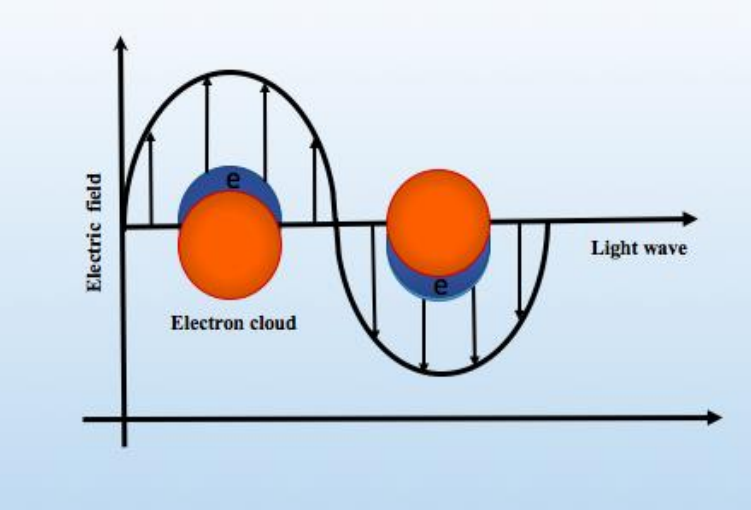


Figure 1.2. Electric field variation in localized surface plasmon of a spherical noble metal nanoparticle.

The large EM effects in SERS occur from excitation of LSPRs when a molecule is adsorbed on the surface of the nanoparticles or is in the vicinity of 'hot spots', as they provide amplification of electric field, EM enhancements, in the vicinity of nanostructured surfaces. In other words, the interaction between the electric field of the conduction band of nanoparticles assembled on the substrate with the coulombic field of adsorbent molecules results in distortion of the molecule giving rise to overlapping of the two electric fields in order to increase the effective cross section of inelastic scatter. Hence, EM in SERS is amplified significantly and the fixed Raman signal at detector scales ~ 4 orders of magnitude of the field enhancement [19].

Nanoparticles that support LSPR play a significant role in SERS as these structures contribute 10^4-10^6 to the enhancement factor [18]. The nanoparticles typically include noble metals such as Au and Ag as their energy levels of d-d transitions exhibit plasmon resonance in visible range of spectrum. Although Ag exhibits the sharpest and strongest vibrational bands among all metals, Au is preferred for biological applications due to its biocompatibility and the thiol-gold functionalization for immobilization of various biomolecules [20, 21].

In fact, larger EM enhancements were observed for nanoparticle aggregates in relation to individual nanoparticles. This is because when nanoparticles are brought in close proximity with separation smaller than the wavelength of incident light to form aggregates, the transient dipoles couple and give rise to strong EM fields that interfere coherently in the hot spot regions. And with appropriate tuning of size, shape and material of nanoparticles; the aggregation of the nanoparticles red shifts the resonant wavelength of the coupled particles to near infra-red (NIR), enabling cellular and in vivo SERS applications where the tissue requires minimum absorption of incident light [22].

In metal nanoparticle aggregates, the EM enhancements are confined within nanometer length scale regions near the surface of the particles and decay significantly thereafter. [23]. Studies have shown that the highest SERS enhancement was obtained at hotspots between the nanoparticles with gaps less than 10 nm [24]. In fact, experimental studies for single DNA tethered molecules showed single molecule SERS detection limit when Au nanoparticles were engineered with spacing less than 1nm [25].

The SERS enhancement factor (EF) also strongly depends on conditions such as excitation wavelength and the analyte that is being investigated. This is typically experimentally measured according to equation (1.2) below.

$$EF = \frac{I_{SERS}/N_{surface}}{I_{RFM}/N_{bulk}} \tag{1.2}$$

Here, I_{SERS} and I_{RFM} refer to the Raman signal of SERS and free molecule respectively. $N_{surface}$ and N_{bulk} represent the number of molecules that are within the excitation volume of the metal surface and bulk respectively. Generally, very weak vibrational Raman modes, having cross section 10^{-31} - 10^{-21} cm^2 $molecule^{-1}$, can be observed when EF ranges from 10^4 to 10^{12} [26].

1.3. METABOLOMICS

Metabolomics is a rapidly emerging field that deals with the study of chemical processes involving small molecules also known as metabolites that act as intermediates and products of metabolism in biological cells, tissues or organism. The detection of metabolic pathway in a bioprocess and systematic quantification of metabolites can be crucial in the study of microbial activity and early

detection of diseases [27]. For example, determination of concentrations of 2,3 butanediol metabolites from breath gas can enable early detection of exacerbations of cystic fibrosis [28,29]. Currently, the major tools used for determination of metabolite concentrations are mass spectroscopy (MS) coupled with high-performance liquid chromatography (HPLC) or nuclear magnetic resonance (NMR). However, MS and NMR measurements are expensive, time consuming and destructive for biological samples [39]. A time resolved and alternative approach to monitor metabolite activity investigated here is based on label-free surface enhanced Raman scattering (SERS).

In SERS, the enhancement of vibrational scattering signals of the target molecules occurs due to excitation of localized metal surface plasmon resonance (LSPR) in nearby plasmonic particles as discussed in the prior section. The benefits for metabolite detection using SERS include 1) high sensitivity i.e., effective change in output of the sensor per unit change in the parameter being measured, 2) 'fingerprinting' ability to produce distinct spectra from molecules of similar structure, and 3) eliminating the use of expensive reagents and time consuming processing of biological samples unlike other techniques such as immunoassays or polymerase chain reaction (PCR). For dilute or aqueous samples, SERS gives rise to minimal background signal as water has a very small Raman scattering cross-section [5]. Also a single analyte can be identified in one run in SERS as the detection is usually carried out in static configuration, where the analytes are retained on the surface of a thin film.

The biological relevant concentration of metabolites usually ranges from nM to mM levels in biofluids [31, 32]; with modern developments in microfluidics and optical sensing, the high sensitivity and high resolution of SERS enables real-time monitoring of continuous flows and multicomponent analysis in a complex biological system [33]. Hence, the ability to detect

metabolite concentration in clinical samples such as blood and urine using SERS can pave ways towards development of personalized medicine [30].

1.4. SPECTRAL ANALYSIS OF SERS DATA

SERS has the ability to provide a 'molecular fingerprint' by investigating the diverse array of vibrations that a single molecule exhibits to differentiate between various molecules. For spectroscopies like these, extensive vibrational information of the analyte is produced. Conventional quantitative SERS spectral analysis is generally carried out with a single vibrational band that omits the rich vibrational information contained in the spectrum and in return, loses the sensitivity when other molecules are present like in the case of complex biological fluids. [34, 35-36]

In recent times, analysis of Raman spectral data is being subjected to a data driven approach using statistical machine learning methods [37, 38]. Machine learning (ML) is a method of data analysis that allows the system to automatically learn with data and become more accurate in predicting outcomes without being explicitly programmed. There are two ways to handle this learning process: supervised learning and unsupervised learning. In supervised learning, the algorithm is fed with pre-defined training data i.e. the training data includes the desired outputs. It is the most mature and studied approach in ML as learning with supervision is easier compared to learning without supervision. Unsupervised learning does not include any training data. The model is fed with unlabeled dataset and is asked to determine on its own on how to cluster or categorize the data.

The ML methods in SERS spectroscopy have been used for identification and quantification of

different chemical compounds. Yet ML models, mainly use quantitative multivariate analysis, particularly, partial least square (PLS) regression. This has been shown to improve quantitative detection in SERS due to its ability to decompose spectra into loading vectors wherein these vectors maximize the variance among different classes (analyte concentrations), while minimizing the variance within each class [36]. However, the disadvantage of PLS is the requirement of large number of variables to distinguish between two or more classes/groups. Alternative methods that use non-linear regression models are to be employed to achieve good prediction accuracy with small number of variables that are responsible for the separation between classes. Therefore, deep learning models are being explored for quantification of analytes using SERS spectral data [40]. All these models require non-trivial preprocessing such as smoothening and baseline correction; also methods such as principal component analysis (PCA) for good prediction accuracy [39].

Deep learning methods use artificial neural networks. Artificial Neural Networks (ANN) is a set of algorithms that is modeled loosely on the human brain [41]. It consists of millions of simple processing nodes or 'neurons' that are densely interconnected. Neural networks perform layers of non-linear feature transformations by morphing input data with weighted sums and activation functions into spaces more preferable than classification [42]. There have been various methods proposed for quantitative studies using neural networks such as nanoporous silicon oxide immunosensor with ANN analysis to report quantification differentiation of Hep-B and Hep-C virus in blood [43] and simultaneous quantification of the drug nicotine and its major metabolites cotinine and *trans*-3'-hydroxycotinine using ANN [44]. Hence, precise multiplex detection of biomolecules is possible for SERS in combination with ANN.

1.5. THESIS ORGANIZATION

Chapter 2 is devoted to understanding the mechanism and preparation of gold nanoparticles via seeded growth method. It introduces fabrication of non-lithographic SERS substrates using electrohydrodynamic flow. Chapter 3 investigates the potential of SERS substrates for detection and quantification of different metabolites at biological relevant concentrations. In chapter 4, the advantages of deep learning for quantitative detection of metabolites are investigated. In addition to comparison of feed forward neural networks and 1 D- convolutional neural networks, the performance of these models are demonstrated for multiplex detection of molecules. Finally, the key factors to build a multi-analyte sensing platform are discussed.

CHAPTER 2. SERS SUBSTRATES PREPARATION

For successfully detecting analytes, it is crucial that we have suitable SERS active substrates. These substrates must have metal surface features to provide high enhancement as well as be reproducible and robust with a good lifetime.

As discussed in Chapter 1, the EM effect in SERS is dependent on the on the nanostructured surface. Here, we use aggregates of metal nanoparticles that was motivated in the previous section. The optical properties are a function of the single particle properties and the aggregate structure. For isolated metal nanoparticles; the material, the particle size and the shape of nanoparticles affects the LSPR frequency. First consider the polarizability of a single metal nanoparticle which can be explained in the electromagnetic model discussed in the following equation (2.1) [45]. The polarizability of a small metal sphere with dielectric function $\varepsilon(\lambda)$ and radius R surrounded by vacuum is given by [46]:

$$\alpha = \frac{R^3\left(\varepsilon_b\omega^2 - \omega_p^2\right) + i\omega\gamma\varepsilon_b}{\left[(\varepsilon_b + 3)\omega^2 - \omega_p^2\right] + i\omega\gamma(\varepsilon_b + 3)} \qquad (2.1)$$

ε_b : The contribution of interband transitions to the dielectric function

ω_p: Metal plasmon resonance frequency

γ : Electron scattering rate

Here, the electron scattering rate (γ) is inversely proportional to the mean free path of the electron and therefore, also inversely proportional to the dc conductivity of the metal. The EM enhancement

is reduced when γ is large either due to poor conductivity of the metal or extremely small metal nanofeatures that contribute to significant electron scattering at the metal's surface. Metals whose dielectric properties are modified by interband transitions have high ε_b value that further have wide resonance and low SERS enhancement. Transition metals with low conductivity (high γ) and high interband contribution to dielectric function (large ε_b) have low SERS enhancement. For this reason, noble metal nanoparticles metal nanoparticles with optimum conductivity and interband transition values such as gold and silver are considered to be ideal for our SERS substrates. We have chosen Au nanopartilces for our work as it has attracted considerable interests in its use in biotechnology due to its biocompatibility and strong excitation to the IR region of light [54].

In terms of aggregate structure, the most common SERS active substrates can be classified into three types: [47]

(1) Metal nanoparticles in a suspension

(2) Metal nanoparticles immobilized onto a solid substrate

(3) Nanostructures fabricated directly on solid substrates using nanolithography and template-based synthesis

For metal particle synthesis, relative to preparation of nanorods, nanotriangles and nanostars; nanosphere synthesis is an easy, cheap and a safe method for producing metal nanoparticles. Hence, a kinetically controlled colloidal seeded growth synthesis to achieve Au nanospheres has been introduced in the next section. And we have already determined, higher SERS enhancement is observed for nanoparticle aggregates rather than individual particles where average SERS

signals from nanostructures can be attributed to the number of intrinsic 'hotspots' per particle [48]. For these aggregated nanoparticles, fabrication of discrete assemblies at nanoscale dimensions with controlled nanometer gap spacing at high densities over large areas of solid substrates is required. Hence, a cheap and efficient non-lithographic SERS substrate has been introduced in the following sections by hierarchical self-assembly of the Au nanospheres on a diblock copolymer template.

2.1. SYNTHESIS OF GOLD NANOPARTICLES

The preparation of gold nanoparticles can be done via chemical, sonochemical or photochemical synthesis methods. [49] Wet chemical synthesis typically reduces gold ions in a solution, using reducing agents such as sodium citrate, sodium borohydride and hydrazine. Stronger reducing agents like sodium borohydride produce smaller particles whereas weaker reducing agents such as sodium citrate produce larger particles. Hence, the nanoparticle size is controlled by the strength and concentration of the reducing agent [47]. This is because, the reduction in amount of sodium citrate will reduce the amount of citrate ions available for stabilizing the particles which will in fact cause the small particles to aggregate into bigger ones. The chemical reduction method used here for synthesis of gold nanoparticles is commonly known as Turkevich method that uses sodium citrate as its reducing agent [50]. Moreover, Turkevich method is an aqueous synthesis protocol that is eco-friendly.

Surfactants which can be termed as 'Capping agents' play a critical role in influencing nanoparticle growth steps. Capping agents such as sodium citrate, cetrimonium bromide (CTAB) and

polyethylene glycol (PEG) bind to the surface of the nanoparticle to prevent agglomeration using repulsive or steric forces. [47,52]. They prevent the metals from forming a continuous phase instead of discrete nanoparticles by forming an adsorption layer on the particle surface. The surfactants also allow controlled growth of nanostructures on crystal planes by stabilizing specific crystal planes to grow a wide variety of nanoparticle shapes such as nanorods, nanospheres, nanowires and nanostars [53]. In our work, sodium citrate functions as both the reducing agent and the surfactant.

Size controlled growth of the nanoparticles involves two mechanisms namely, LaMer growth and aggregative growth [51]. Size growth in LaMer growth occurs via heterogeneous nucleation process on seed or pre-formed nanoparticles. Whereas, the aggregated growth occurs via agglomeration of primary nanoparticles. Lamer growth has been employed in our work where in the nucleation and growth steps determine the size distribution of the nanoparticles. In the seed stage, the metal atoms combine to form clusters known as crystal nuclei. The crystal nuclei, generally known as seeds grow in size to form nanoparticles.

In the previous chapter, it has been shown that the size of the gold nanoparticles can have a significant effect on EM EF. From previous work, the highest EF was observed for particle sizes whose LSPR frequencies are between in between excitation wavelength and the vibrational mode [55]. For example, an optimum gold nanoparticle size for the maximum enhancement was found to be around 50 nm for 647 nm laser [54]. In this thesis, laser excitation, 785 nm was chosen as it eliminates sample fluorescence from organic molecules and it also has low attenuation with water. Hence, a size of ~ 40 nm gold nanoparticles are excited using a 785 nm laser source in this work.

2.1.1. Experiment

Materials:

$HAuCl_4.3H_2O$ (99%) and trisodium citrate (99%) were purchased from Sigma Aldrich. Milli-Q water was used in all experiments. All glassware was cleaned with aqua regia, rinsed with acetone, isopropyl alcohol, and deionized water prior to nanoparticle synthesis.

S***ynthesis of Au Seeds:***

A solution of 5.5 mL of 60 mM sodium citrate in Milli-Q water (145.5 mL) was heated on a hot plate for 15 min in a 250 mL three-necked round bottom flask under vigorous stirring. A reflux condenser was utilized to prevent the evaporation of the solvent. Once boiling had commenced, 1 ml of HAuCl4 (25 mM) was injected using a syringe. It was observed that the solution changed from yellow to wine red. Here the resulting particles are coated with negatively charged citrate ions, hence are well suspended in H_2O. [56]

Seeded Growth of Au NPs of up to 40 nm in Diameter:

Immediately after the synthesis of Au seeds, the reaction was cooled to 90 °C in the same vessel. Then 1 mL of sodium citrate (60 mM) was injected using a syringe, followed by 1 mL of $HAuCl_4$ (25 mM) after 2 min. After 30 min, 1 mL of sodium citrate and 1 mL of HAuCl4 were added similarly. This step was repeated until 40 nm Au NP were obtained which were confirmed using UV- Vis spectroscopy and dynamic light scattering (DLS). Typically, 5-6 growth steps gave us 40 nm Au NP depending on the concentration of each generation of NPs which was approximately the same as the original seed particles (3×10^{12} NPs/mL). [50] After cooling, the resulting nanoparticle solution was stored in a brown bottle.

Characterization Techniques:

UV-Vis Spectroscopy: UV-Visible spectra were acquired using Agilent Cary-60 absorption spectrometer. 1 mL of Au NPs was diluted to 10 mL and placed in a disposable cuvette. Spectral analysis was performed in the 300 to 800 nm range at room temperature.

Dynamic Light Scattering (DLS): The particle size of the synthesized NPs was obtained using Malvern Zetasizer ZS Nano DLS. The Au NPs sample (1 mL) was diluted to 10 mL and placed in a disposal plastic cuvette. Automatic measurement duration and 3 such measurements were chosen at a temperature of 25 °C to obtain the average particle size and poly dispersity index (PDI) of our samples.

2.1.2. Results

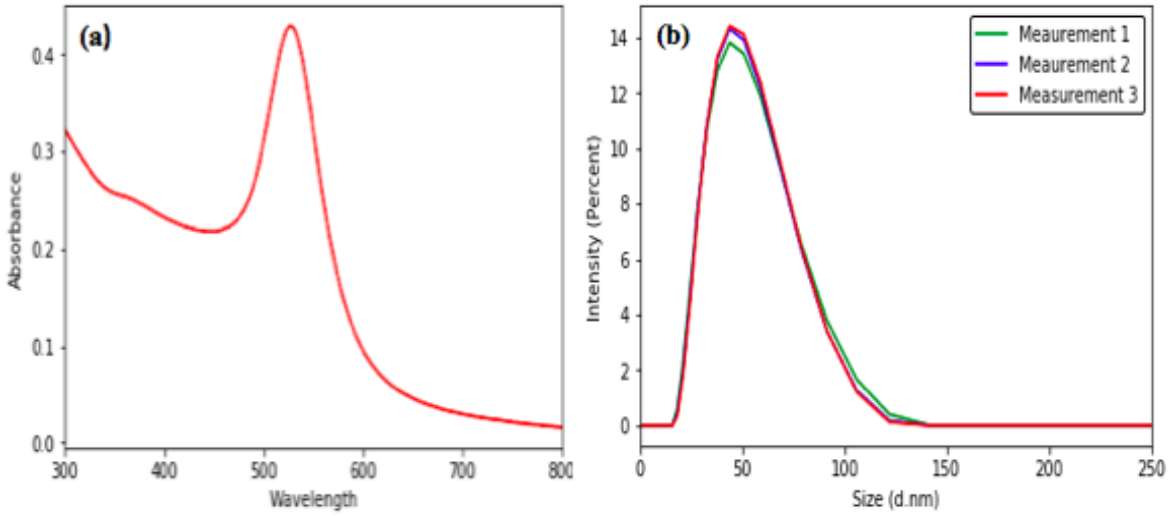


Figure 2.1. (a) UV-Vis Absorption spectrum for 40 nm Gold nanoparticles. (b) Resultant size distributions of the colloidal gold nanoparticle solution derived using DLS

Table 2.1. Scattered distribution report of gold nanoparticles using DLS.

	Z- Ave (nm)	**PDI**
Measurement 1	38.01	0.231
Measurement 2	38.24	0.229
Measurement 3	38.18	0.232

The experimentally measured absorbance by the UV-Vis spectrometer for a particle diameter of 40 ± 7 nm and a particle concentration of 0.16 mg.mL $^{-1}$ is 526.98 nm as given in Figure 2.1.a. This characteristic absorption spectrum is related to the surface plasmon resonance of the 40 nm nanoparticles. Our result is similar to the peak SPR wavelength of 526 nm for 40 nm spherical gold nanoparticles in [57]. Additionally, the monodispersed gold nanoparticles prepared as colloids, were measured using DLS as shown in Figure 2.1 (b). Analysis of the obtained results indicate that the monodisperse nanoparticles have a size of 40±7 nm with a polydispersity index (PDI) of 0.281. The hydrodynamic diameter of the theoretical sphere that diffuses with an equal speed as that of the measured nanoparticle is measured as the size using this technique. This size is not only associated with the metallic core of the nanoparticles but also the substances adsorbed on the surface of the nanoparticles and the thickness of the electrical double layer, moving along with the particle whose size and effect depends on the substances present in the colloid and surface of the nanoparticle [58].

2.2. FUNCTIONALIZATION OF GOLD NANOPARTICLES

Nanoparticles with diameters of 10-100 nm can be synthesized using the Turkevich method wherein they are capped with citrate ions. Based on the affinity of gold for thiol groups, thiol-containing capping agent is then added for efficient functionalization of the particle surface [59]. This is because the covalently bound self-assembled monolayer (SAM) of lipoic acid play several essential roles, like acting as a stabilizing ligand, preserving the center core and making further functionalization steps possible [60, 61]. In addition, the use of lipoic acid results in stable water-dispersible gold nanoparticles and reduces uncontrolled aggregation in solution [62,63]. In Figure 2.2, it is observed that the endocyclic disulphide molecules efficiently displaces the chloride and the citrate ions physisorbed on the surface of gold nanoparticles. This resulting SAM ensures an electrostatic stabilization of the gold nanoparticles at high pH because lipoic acid is negatively charged and prevents aggregation of gold nanoparticles due to repulsion [64].

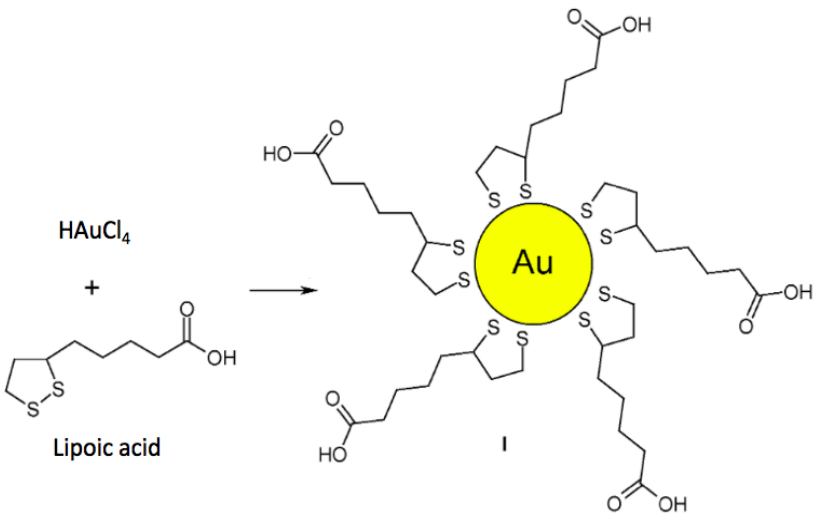


Figure 2.2. Schematic representation of lipoic acid functionalized Au NPs [65]

2.2.1. Experiment

The basicity of citrate-stabilized gold solution was centrifuged at 1700 RCF for 30 min and the supernatant was removed and replaced with pH 11 adjusted Deionized water (0.1M of NaOH) as high basicity reinforces static stabilization [66]. 0.1 mM of lipoic acid (0.1 M dissolved in ethanol) was added to 40 nm citrate stabilized gold nanoparticle solution. This lipoic acid functionalized solution was stirred at 500 rpm for at least 24 hours at room temperature. The nanoparticles were then centrifuged at 1700 RCF for 30 min again, after which the supernatant was rinsed with deionized water to ensure sufficient removal of unbound ligand. The nanoparticles were then stored in a brown bottle until use.

2.3. FABRICATION OF NON-LITHOGRAPHIC SERS SUBSTRATES

The development of novel and efficient substrates is a primary focus of research aimed at improving the analytical capabilities of SERS. Conventional methods to produce SERS films make use of top-down lithography techniques such as electron beam lithography, focused ion beam milling and atomic layer deposition to deposit Au nanoparticle aggregate like clusters which exhibit 1-5 $\times$ 10^8 signal enhancement [67-69]. These top down methods prove to be disadvantageous due to complex, time consuming and expensive processes involved [67]. Previously, self-assembly techniques have been used to grow functionalized gold nanospheres on top of templated block-copolymer substrates achieving an EF $\approx$ 10^8-10^9 [70,71]. The diblock copolymers are a low-cost, self-organized chemical template that can be used to form nanoparticles clusters over large areas without the use of lithography and vacuum phase metal depositions [72]. This method can form a uniform layer to achieve substantial signal enhancements over the surface

with the capability to reuse the substrates for small molecules. Our aim is to fabricate SERS active substrates wherein monodisperse, colloidal gold nanoparticle clusters with sub-nanometer interparticle spacing are attached on functionalized diblock copolymer thin films.

Short range driving forces such as electrostatic interactions, van der Waals forces and chemical cross-linking can control the interparticle gap spacing and have the potential to produce controlled architectures from colloidal architectures [70]. Yet, these cross linking methods have difficulty in achieving high yield of Au nanosphere clusters (oligomers) that are dense and uniform. Hence, we follow Thrift et al. method [shown in Figure 2.3] and use induced charged electroosmosis, also known as electrohydrodynamic (EHD) flow which is a longer range driving force that facilitates chemical cross-linking [73,74]. EHD is an electrokinetic method that promotes lateral flow field and close-packing of particles in colloid on the plane of the working electrode under a dc/ac potential in presence of a counter electrode [67]. Here, the EHD flow brings the 40 nm Au nanoparticles together and the anhydride cross-linking yields 0.9 nm gap interparticle spacing.

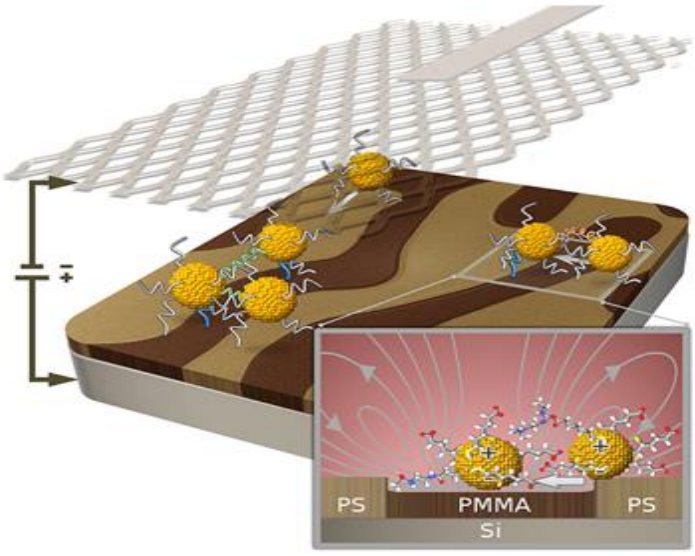


Figure 2.3. Schematic depicts the EHD attractive forces in a carboxylic acid-functionalized Au nanospheres colloidal solution near a Si substrate self-organized with PS-*b*-PMMA diblock copolymer (working electrode). Inset shows the EHD flow generated from inhomogeneous field at the working electrode surface through polarization of the 'seed' monomers, then increasing the probability of nanospheres - nanosphere collisions to immobilize the Au nanoparticles through anhydride bridging. [71]

The two dimensional growth of Au nanosphere oligomers over 1 cm^2 area can be achieved via two step seeded growth method where in EHD flow driven oligomer growth of Au nanosphere seeds is followed by chemical cross-linking for uniform gap spacing [70,71].

1. In the first step, carboxylic acid end groups on the Au nanospheres are electrophoretically driven towards a doped Si working electrode coated with a polystyrene (PS)-*b*-poly-(methyl methacrylate) (PMMA) diblock copolymer template. EDC cross-linking chemistry referred to as carbodiimide cross linking chemistry couples the carboxylic acid moieties with amine functionalized PMMA regions, immobilizing nanoparticles via peptide bond formation. These particles serve as monomer seeds for cluster growth.

2. The second step involves the generation of EHD flow to promote oligomer growth using Au nanosphere monomer seeds on the working electrode. The applied field polarizes the monomer seeds that enables the EHD flow to entrain nearby nanospheres towards the source of perturbation (Au seed) in the plane of the electrode, leading to the growth of oligomers. Here, the transient clusters are immobilized due to the close proximity of carboxylic acid groups in the nanospheres through a different carbodiimide cross linking chemistry. The carboxylic acid terminated Au nanospheres form anhydride bridges with one another, ultimately freezing the clusters in place while maintaining a uniform gap spacing of 0.9 nm between the particles.

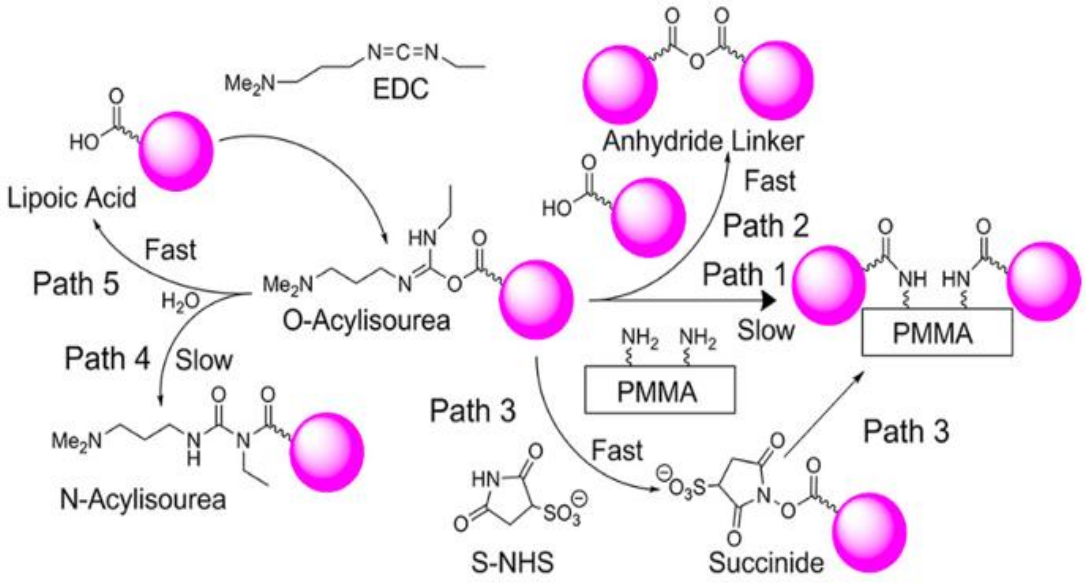


Figure 2.4. Carbodiimide chemistry pathways yield covalent linkages of carboxylic acid functionalized Au nanoparticels to the amine-functionalized surface via paths 1,3 and covalent linkages between nanospheres via path 2. [71]

The mechanism of carbodiimide crosslinking reaction involved in the deposition process is explained in Figure 2.4. Path 1 illustrated and labeled in the figure shows that EDC is used to cross-link carboxylic acid ends of nanospheres and primary amines on PMMA regions. This

mechanism can form oligomers that are aligned with PMMA domains by chemically binding the nanospheres to existing monomers. But carbodiimide cross linking cannot be the driving force for self -organization as electrostatic repulsion between monomers would favor monomers formation on the PMMA domain. Hence, during the deposition process, EDC and s-NHS are used where in EDC reacts with carboxylic acid end groups of nanospheres to form O- acylisourea esters and then to form peptide bonds; S-NHS is added to prevent the formation of N-acylisourea (Path 4) that leads to unreactive nanospheres by consuming reactive o-acylisourea on the surface of nanospheres. But as the concentration of EDC/s-NHS is relatively small, there is also bridging of carboxylate groups on the nanospheres to form anhydride groups (Path 2). In fact, the measured gap spacing between clusters is 0.9 nm in TEM images which is consistent with the length of the anhydride bridge [75]. The formation of anhydride bridges when the carboxylic ends of nanopsheres are brought together lock the transient structures that occur due to EHD flow. The carbodimidde crosslinking of monomers bound to PMMA domains serve as perturbations to the local field and act as nucleation sites to attract close by particles via EHD flow and link them with anhydride bridges to form larger oligomers. EHD flow has the ability to drive oligomerization by forming anhydride bridges between Au nanospheres in such way that uniform gap spacing between the oligomers can be obtained, hence enhancing the optical properties of the substrates.

2.3.1. Experiment

Materials:

Random copolymer poly(styrene-*co*-methyl-methacrylate)- α-hydroxyl-ω-tempo moiety (PS-*r*-PMMA) (M_n = 7400, 59.6% PS) and diblock copolymer poly(styrene-*b*-methyl methacrylate) (PS-*b*-PMMA) (M_n = 170-b-144 kg mol^{-1}) were purchased from Polymer Source, Inc. (Dorval,

Canada), Si(001) wafers with a resistivity of 0.004 ohm-cm were purchased from Virgina Semiconductor (Frederickburg, VA, USa), hydrofluoric acid (HF) was purchased from Fisher Scientific (Pittsburgh, PA, USA). 2-(N-morpholino)ethanesulphonic acid (MES) 0.1 M buffer, 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide hydrochloride (EDC), and N-hydroxy sulfosuccinimide (s-NHS) were purchased from Pierce (Rockford, IL, USA). Dimethyl sulfoxide (DMSO), ethylenediamine, benzenethiol, toluene, ethanol, isopropyl alcohol (IPA), and 52-mesh Pt gauze foil were all purchased from Sigma-Aldrich (St. Louis, MO, USA). Nanopure deionized water (DI) (18.2 MΩ cm^{-1}) was obtained from a Milli-Q Millipore system.

Nanoantenna Oligomer Substrate Preparation:

Lamella PS-*b*-PMMA block copolymer solution in toluene (1 wt%) is spin coated onto a HF-cleaned, heavily doped Si wafer at 3000 rpm for 45 s and annealed at 198 ° C for 72h in vacuum conditions followed by rinsing with toluene. Then PS-*r*-PMMA random copolymer solution in toluene (1 wt%) is spin coated onto the Si substrate at 3000 rpm for 45 s and annealed at 198 ° C for 72h in vacuum conditions to produce desired diblock polymer coated substrates [64]. The polymer coated substrates are completely immersed in DMSO for 5 min and then transferred to a vial containing ethylenediamine/DMSO solution (5% v/v) and kept immersed in it for 5 min. In this step, the PMMA regions on Si substrate are selectively functionalized with amine end groups. The substrate coated with functionalized copolymer is then washed with IPA and deionised water for 1 min each and dried with nitrogen.

Next, 3 mL of lipoic acid functionalized 50 nm gold nanosphere solution (0.1 $mg.mL^{-1}$) is added to a 10 mL beaker. s-NHS (20 mM) in MES (0.1 M) buffer and EDC (8mM) in MES (0.1 M)

buffer is freshly prepared and 35 μL of each is added to the beaker and swirled. The beaker is then placed on a hot plate and heated to a temperature of 80° C.

The 1cm × 1 cm functionalized copolymer coated Si substrate is placed vertically into the solution with the help of alligator clips to act as a anode. Similarly, 1cm × 1 cm Pt mesh is placed vertically into the solution to act as a cathode provided it is placed 5 mm parallel to the anode. Note that the alligator clips should not touch the Au solution at any cost. A dc regulated power supply is used to apply a voltage of 1.2 V for 10 min. After the first deposition, the substrate and Pt mesh are rinsed with IPA and deionized water for 1 min each and dried with nitrogen. The beaker is washed with acetone, IPA and deionized water thoroughly. The process is then repeated with the same substrate and fresh nanosphere solution for the second deposition but the concentration of EDC and s-NHS is changed to 25 μL here. Two growth steps are necessary to obtain oligomers on resonance at the 785 nm illumination wavelength used for SERS measurements. The second growth step is performed with reduced EDC and S-NHS concentrations to promote anhydride formation.

Characterization:

Scanning Electron Microscopy:

After nanoantenna oligomers are assembled onto polymer coated Si substrates, images are collected using Magellan 400 SEM.

Surface Enhanced Raman Spectroscopy:

Raman spectral measurements are acquired using a confocal Renishaw InVia micro Raman with laser excitation 785 nm. To estimate optical uniformity of substrates, the fabricates substrates were

immersed in 10^{-3} M BZT solution in IPA for 12 hrs and then rinsed with IPA. The SERS spectral data was collected at 76 μW with exposure time 0.1 s for mapping and individual spectral acquisitions. All measurements use a 60X water immersion objective with 1.2 NA, immersed in DI water.

2.3.2. Results

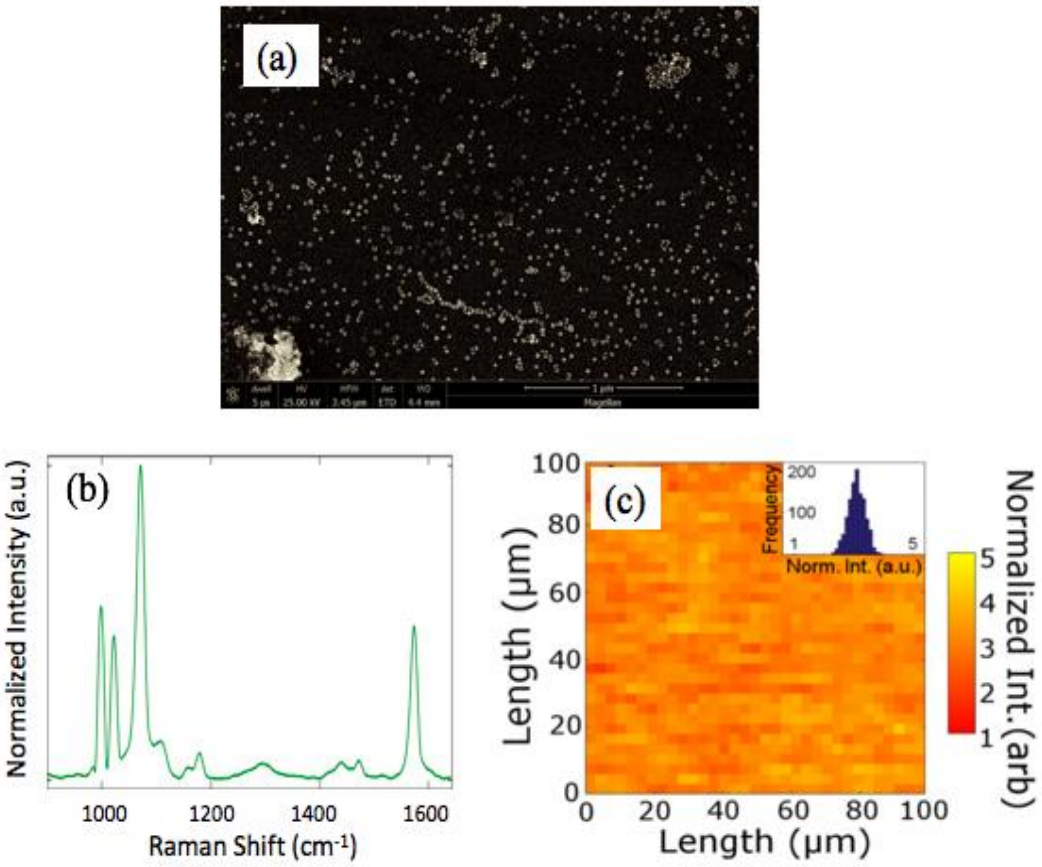


Figure 2.5. (a) SEM image of Au clusters on SERS substrates. (b) SERS spectrum of 10^{-3} M BZT (c) Normalized SERS intensity map of BZT's 1573 cm^{-1} vibrational band across a 100 μm ×100 μm area with RSD of 10.4%. Inset depicts the measured SERS EF mean value of 1.4×10^9 [77].

The SEM image shown in fig.2.4(a). confirms Au cluster formation on the self-organized SERS substrates after the two-step fabrication process. The most typical vibrational modes for the identification of BZT are 1000 cm^{-1}, 1025 cm^{-1}, 1094 cm^{-1} and 1586 cm^{-1} which can be observed in Figure 2.5(b) [77]. The nanoantenna samples were immersed in BZT wherein the characteristic thiol -Au interactions helped verify the uniformity of 40 nm EHD assembly on polymer based substrates. A 100 μm ×100 μm SERS intensity map at BZT vibrational band, 1573 cm^{-1} was studied to get a relative standard deviation (RSD) of 10.4 % across the mapped area [76]. The inset picture (Figure 2.4(c)) shows an average SERS EF value of 1.4×10^9. The low signal deviation is a result of the seeded growth oligomerization and optimized temperature, polymer template and EHD flow. This uniformity enables statistical analysis for quantitative molecular detection through acquisition of large datasets.

CHAPTER 3. SENSOR CHARACTERIZATION

3.1. DETECTION OF METABOLITES

Metabolomics is the study of small molecules, commonly known as metabolites, within cells, tissues, organisms or biofluids. The metabolic analytes are composed of chemically diverse compounds like amino acids, lipids, sugars and nucleic acid which are typically the end products of gene expression or protein activity [78]. A change in metabolic profiles occurs in a number of key disease related or other cellular pathways, hence their identification and spatial distribution within a biological system is beneficial to gain information related to the current disease or the therapeutic status [79]. In metabolomics, analytical methods such as mass spectroscopy and nuclear magnetic resonance have been employed to investigate the phenotypic properties of biological systems that allows the understanding of the organization of gene functions, the effect of nutrient environment, and the role of cell interaction in multicellular organism. [80, 81]. Recent developments include detection and quantification of breast cancer phenotypic markers expressed on cell surfaces using surface enhanced Raman spectroscopy (SERS) [82]. SERS has proven to be a rapid and extremely sensitive technique for accurate identification and quantification of small biomolecules in complex biofluids. In fact, this fingerprinting technology is most sensitive for short aromatic molecules on metal substrates [83].

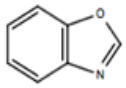
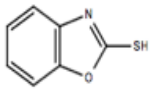
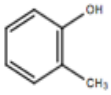
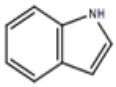
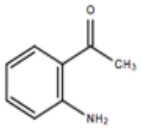
SERS has potential for diverse applications that require the ability to detect specific types of metabolites. Detection of heterocyclic compounds like benzoxazole that exhibit antimicrobial, antiviral and chemotherapeutic activities has proved to be important because nosocomial infections

associated with catheters and other medical implants, caused by bacteria such as *Staphylococcus aureus* and *Staphylococcus epidermis* can be prevented [88]. Also, 2- Mercaptobenzoxazole (2-MBO), a benzoxazole derivative is an important bacteriostatic and fungicidal regulator in plant growth that can act as a corrosion inhibitor due to its coordination with transition metal ions [89].

Production of several *Pseudomonas aeruginosa (PA)* bacteria virulence factors and biofilm formation is coordinated by cell density monitoring mechanism called as quorum sensing [85]. 2-Aminoacetophenone, a secondary metabolite and quorum sensing volatile molecule of *PA* can be detected in the breath of cystic fibrosis (CF) patients using SERS [84]. Similarly, indole that is typically produced by *Escherichia coli* has also been identified as an 'quorum sensing' signal. Indole is a metabolite of amino acid tryptophan and is an important phenotypic characteristic that has been used to differentiate, identify, and diagnose enteric bacterial infections [86, 87].

In addition, the detection and removal of organic componds such as o-cresol in aqueous solutions is crucial as they are detrimental to various species. Bio-sorption, an unconventional technique uses biofilms such as *Arthrobacter viscosus* to remove such organic compounds from aqueous solutions [90].. Keeping these applications in mind, in order to check the capacity of the prepared gold nanoparticles array substrates, the SERS spectra of various biomolecules such as benzoxazole, 2- merceptobenzoxazole, o-cresol, indole and 2- aminoacetophenone were analyzed.

Table.3.1. Chemical structure of metabolites chosen for SERS analysis

Molecule	Structure
Benzoxazole	
2- Mercaptobenzoxazole (2- MBO)	
o-Cresol	
Indole	
2`- Aminoacetophenone (2- AAP)	

3.1.1. Experiment

Substrate Preparation:

The SERS substrates are soaked overnight in 5 different vials containing 1 μg.mL $^{-1}$ solution of benzoxazole, 2- aminoacetophenone, indole, 2- mercaptobenzoxazole and o-cresol. The substrates are cleaned with IPA and DI water for 1 minute and dried with nitrogen before

every measurement. One substrate was used per set of 1 μg.mL $^{-1}$ metabolite measurement.

Spectroscopic Measurements and Procedure:

The SERS enhancement studies is then conducted using Renishaw InVia Raman microscope. Laser wavelength of 785 nm gives maximum electric field enhancement for nanoparticles with 1 nm gap spacing [76]. A 60X water immersion objective with 1.2 NA is chosen for illumination and collection of data. Approximately 1.5 μl of water droplet is transferred onto the SERS substrate and the spectral measurements are acquired for laser power= 7.3 μW and exposure time= 0.5 s.

Spectral Data Processing:

Data processing was performed using MATLAB R2016b (The MathWorks Inc, Natrick, MA) wherein each spectral acquisition is subjected to baseline correction, Savitzky-Golay smoothing, and normalization to average intensity of Si second-order vibrational band (920 – 1045 cm^{-1}), from the substrate. Hence, even if slight intensity variations arise due to shifts in optical collection in the experimental setup, this normalization enables comparison of different samples.

3.1.2. Results

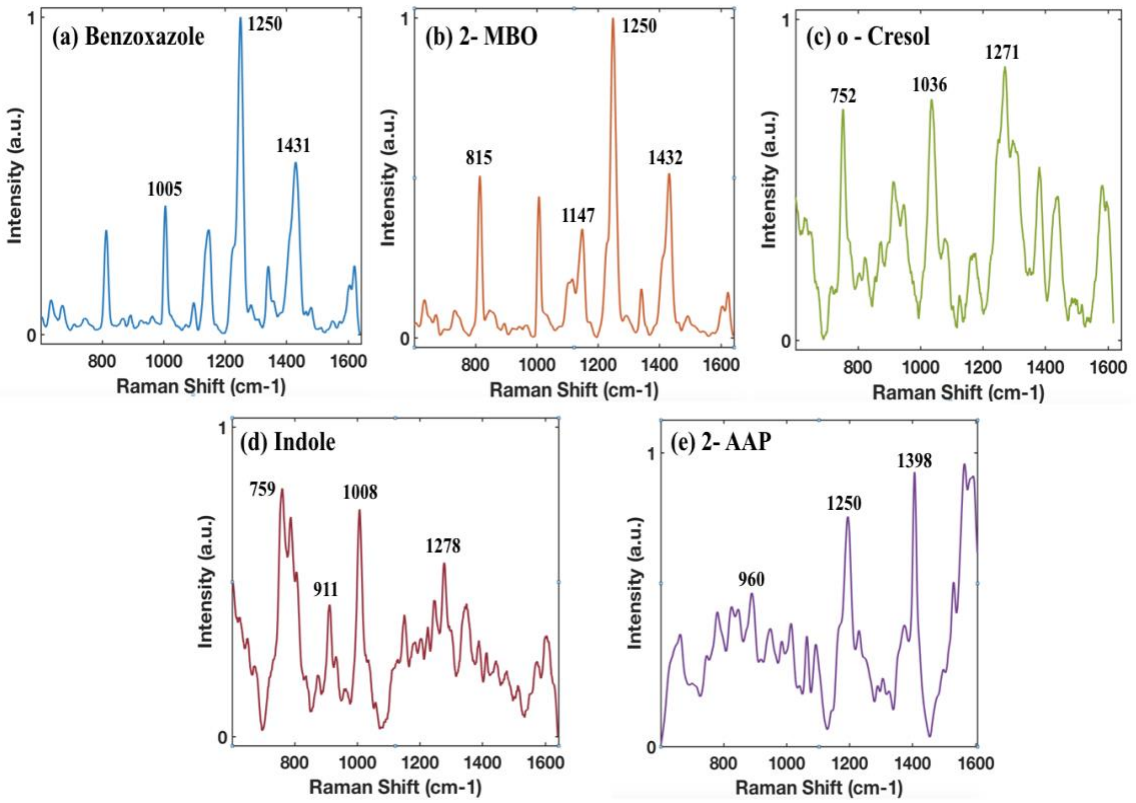


Figure 3.1. SERS spectrum of 1 μg.mL $^{-1}$of **(a)** Benzoxazole **(b)** 2- MBO **(c)** o- Cresol **(d)** Indole **(e)** 2- AAP

In benzoxazole [Figure 3.1(a)] and 2- merceptobenzoxazole [Figure 3.1(b)], Raman bands at 1250 cm^{-1} and 1432 cm^{-1} are observed similar to their SERS spectra on silver hydrosols which represent C-H bending, CNC mode and C=C stretching respectively [80, 92,93]. For indole in in Figure 3.1 (d), specific Raman vibrational modes of an indole ring are examined where in the two strong Raman bands at 759 cm^{-1} and 1008 cm^{-1} are assigned to the indole ring- in phase and out-of-phase breathing modes, respectively [91,94]. The SERS spectra for o-cresol and 2-AAP have been characterized for the first time which show few distinct and clear Raman bands [3.1(c) and 3.1(e)]

Therefore, we were successfully able to achieve the identification of the above metabolites using the fabricated substrate surfaces.

3.2. CONCENTRATION STUDIES

The quantification of metabolites such as proteins, amino acids and sugars can be crucial in the study of microbial activity and early detection of diseases. This is because well-defined concentrations of metabolites allow specific metabolic pathways in a bioprocess thus defining the function of the cell [95]. As discussed in Chapter 1, the major tools used for determination of metabolite concentrations are MS and NMR but SERS is emerging as a time resolved, non-destructive and alternative approach for quantification of metabolites. For example, SERS has applications in assessing drug concentrations in biological matrix where in the accurate quantitative measurement of drugs and their metabolites can be important to establish long-term abuse of illicit materials and drug dosing for legal therapeutics [96].

The limit of detection (LOD) and quantification range of metabolite concentration changes for different molecules in different environments, but a significant level is in nM range. Here, SERS is capable of providing the required enhancements for detections that enable us to observe molecules at biologically relevant concentration. For example, Pyocyanin, a secondary metabolite of PA which is considered a biomarker for life-threating Pseudomonas infections can be detected in aqueous solutions by employing SERS to achieve a limit of detection of 0.5 μM [98]. Further work reported on SERS quantification of pyocyanin display clear Raman bands in SERS spectra from 1 $\mu g \cdot mL^{-1}$ to 100 $\mu g \cdot mL^{-1}$ in Ag colloids, Ag and Au nanorods and Au nanospheres [76,

97]. As the interest of this work lied in determining the LOD for different molecules in biological media; we measured two metabolites namely, indole and 2-AAP for concentrations as low as 1 $pg.mL^{-1}$.

3.2.1. Experiment

Spectroscopic Measurements and Data Processing:

SERS measurements were conducted using Renishaw InVia Raman Microscope. A 785 nm laser was used to excite near the plasmon resonance of Au nanoparticle assemblies. In droplet measurements, as mentioned above, a 60× water immersion objective with 1.2 NA is used for illumination and collection. A drop of solution of interest is transferred onto SERS substrates, and the measurements are acquired with laser power and acquisition time of 7.3 μW and 0.5 s, respectively. SERS substrates are cleaned with IPA and DI water for 1 min each and dried under nitrogen between measurements. One substrate was used per set of measurements with the solution of interest. The calibration mapped SERS data sets for water and 10 concentrations of Indole and 2-AAP are collected. The spectral processing and analysis was performed using MATLAB as mentioned above with each spectrum undergoing baseline correction, smoothening and normalization.

3.2.2. Results

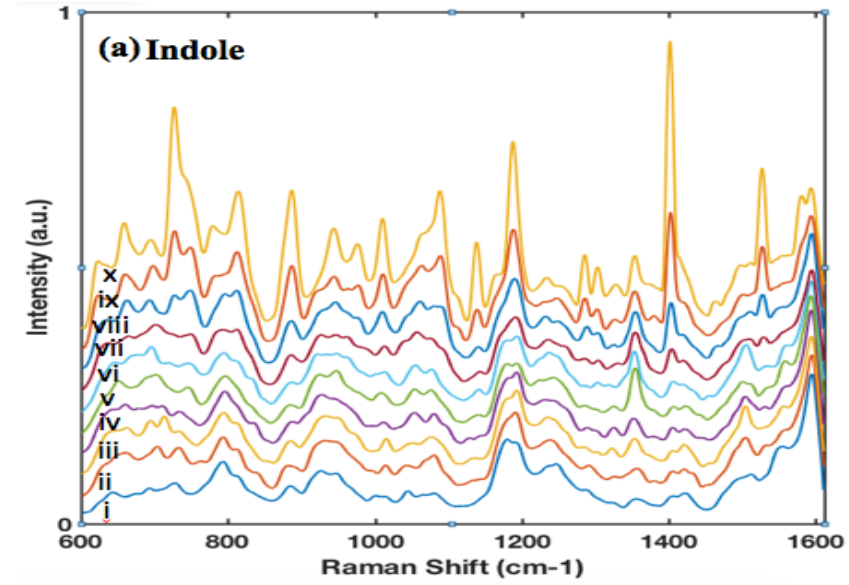


Fig.3.2. SERS spectral analysis of (a) Indole and (b) 2-AAP for concentrations (i) 1 pg.mL $^{-1}$ (ii) 10 pg.mL $^{-1}$ (iii) 100 pg.mL $^{-1}$ (iv) 1 ng.mL $^{-1}$ (v) 10 ng.mL $^{-1}$ (vi) 100 ng.mL $^{-1}$ (vii) 1 µg.mL $^{-1}$ (viii) 10 µg.mL $^{-1}$ (ix) 100 µg.mL $^{-1}$ (x) 1 mg.mL $^{-1}$

From the spectral analysis in Figure 3.2, we can observe that the SERS substrates exhibit detection of indole and 2-AAP at a concentration starting from 1 pg.mL $^{-1}$. And above this concentration, one can observe the molecule signals above the background. For effective analysis on detection of metabolite concentration, our data should show a consistent behavior with the Langmuir molecular adsorption kinetics. This model establishes a linear dependence of intensity on concentration at low concentrations and asymptotic behavior at high concentrations.

Multivariate analysis such as partial least square (PLS) regression was able to achieve a linear response of the measured concentrations of pyocyanin and Coumarin, consistent with Langmuir adsorption kinetics with a LOD of 100 pg.mL $^{-1}$ and 10 pg.mL $^{-1}$ respectively [76, 101]..

CHAPTER 4. SPECTRAL ANALYSIS USING MACHINE LEARNING

4.1. INTRODUCTION TO MACHINE LEARNING

Robust quantification algorithms based on machine learning (ML) can further lower the limit of detection for the metabolites measured on these SERS substrates. ML algorithms facilitate pattern recognition, classification, and prediction [102]. The main objective is to use automated algorithm construction methods to minimize human input or the possibility of human biases for classifying objects or predicting events [103]. Machine learning can be broadly classified into two categories:

Supervised Learning:

In this case, we know the correct answers, and the algorithm iteratively analyzes the training data and produces an inferred function that can be corrected by the user. The learning stops once acceptable results are predicted which are then used for mapping new examples. In other words, the data comes with class labels, and the machine learn how to associate the labeled data (objects) with classes [103]. They are further grouped into two categories: (i) Classification where the output variable is a category (ii) Regression where the output variable is a real value.

Unsupervised Learning:

There is no predefined answer, and there is no teacher. The goal is to learn and discover groups of similar examples within the data. Here, the data is unlabeled, and the machine learns how to define the labels and associate the objects with them [103]. Unsupervised learning is grouped into two

categories: (i) Clustering where inherent groupings in the data are discovered (ii) Association where we want to discover the rules that describe large portions of data

4.2. APPLICATIONS OF MACHINE LEARNING IN BIOLOGY

There has been impressive developments and improvements in existing molecular technologies and facilities over the last few years that has led to an exponential increase in collection of biological information. Hence, there has been a strong motivation to use ML methods for knowledge discovery and data mining to generate models of biological implications [104]. In fact, the human genome project, founded in 1990, involved the complete mapping and understanding of all the genes of human beings. It required massive data which drove the need for computational and statistical tools to organize, maintain, analyze and predict biological results [104].

Machine learning methods such as Markov models, neural networks and clustering methods have been used to predict 1D, 2D, 3D and 4D protein structures from its sequence [105]. Classification has been widely used in image-based screening to classify cell morphologies that are traced by fluorescent markers [102]. Support vector machine (SVM) classification has been used for structure-activity relationship (SAR) analysis of molecular compounds, a technique used by pharmaceutical companies in the drug discovery process which marks the importance of this analysis for commercial purposes [106]. Machine learning is also used in clinical trials, for example, the diagnosis of Alzheimer's disease involves a complex brain morphometric pattern analysis to identify disease-related imaging biomarkers from structural magnetic resonance imaging (MRI) [107]. Machine learning has also extended to metabolomics wherein prediction of molecular characteristics and identification of metabolites from tandem mass spectra is done using

SVM. The molecular properties of the metabolites are predicted to high accuracy; hence they are useful in de novo metabolite identification [108].

4.3. DEEP LEARNING

Deep learning is considered a special case of machine learning algorithm: artificial neural networks (ANN). The central theme of ANN has been inspired by the way a human brain works; it involves many nodes or neurons that are interconnected in layers to form a network. Algorithms associated with deep learning have the potential for big data analysis in order to detect patterns which humans may not perceive. This branch can have varied biological applications like making genomic connections, classify cellular images and in advance drug discovery and development [109].

A neural network consists of an input layer and an output layer with an intermediate layer known as the hidden layer (as shown in Figure 4.1). This hidden layer extracts information by exploiting the structure in the data. The name 'Deep Learning' implies that the network consists of more than one hidden layer [110].

In this network, each layer of nodes train on a distinct set of features based on the output of the previous layer. The deeper the placement of the node in the neural net, the more complex features it can recognize as the nodes aggregate and recombine from previous layers. This is known as feature hierarchy. Most importantly, neural networks are capable of discovering latent structures within unlabeled and unstructured data such as raw media (pictures, texts, video and audio recordings).

The advantage of deep learning compared to traditional ML algorithms is automatic feature extraction without human intervention. In other words, algorithms associated with deep learning have the potential for big data analysis in order to detect patterns which humans may not perceive. This branch can have varied biological applications like making genomic connections, classifying cellular images, and in drug discovery and development [110]

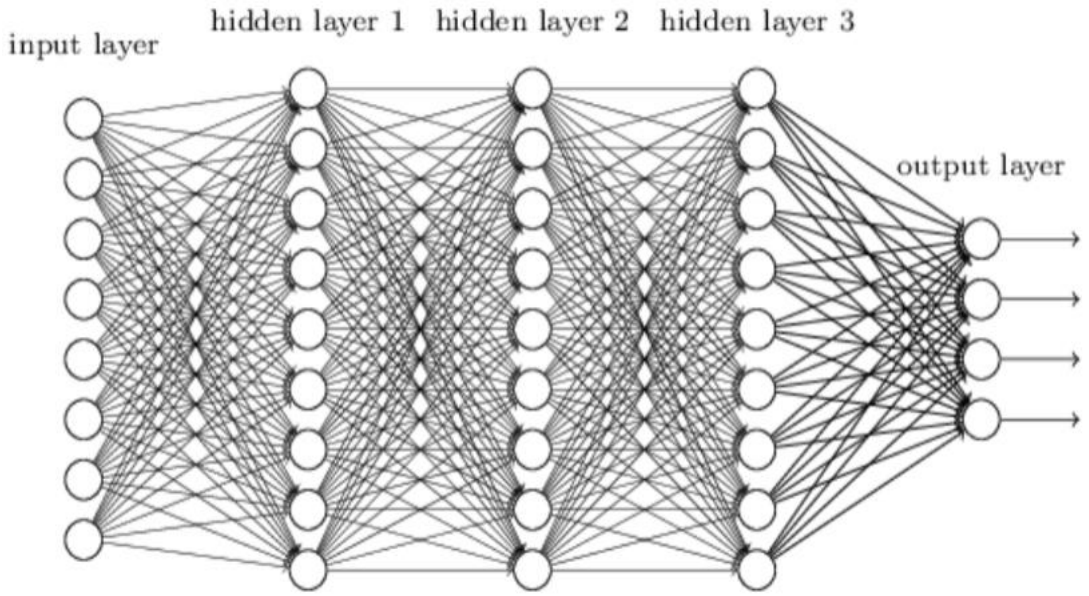


Figure 4.1. Structure of a neural network consisting of a hidden layer that connects the input layer to the output layer [111]

Nodes which are the basic units of computation, receive input from an external source or some other nodes where in each input has a connection weight (w) associated with the relative importance of the inputs present. The node then applies a function ***f*** to the weighted sum of its inputs. The non-linear function ***f***, known as the activation function is used to introduce non-linearity into the output of the nodes as the most of the real-world data is non-linear in nature and we want the nodes or neurons to learn the non-linear representations. Every activation function will take a single number and perform a certain mathematical operation such as threshold, sigmoid,

rectifier and hyperbolic (tanh) functions [112]. These functions are used depending on the output value required as the value can be continuous, binary or categorical.

The main goal of the neural network or perceptron is to reduce the cost function which is a measure of how well a neural network performs with respect to the given training data and the labeled expected output. In the training phase of the network, when input values are provided to the input nodes of the neural network, it provides an output value ($\hat{y}$) based on the weights and biases associated with all the nodes. This output value, $\hat{y}$ is then compared against the known or actual values, y for the provided input values using a cost function (C) where:

$$C = \frac{1}{2}(\hat{y} - y)^2 \tag{3.1}$$

The weights (w) associated with the nodes in the network are updated till a minimum value of cost function (C) is obtained [113].

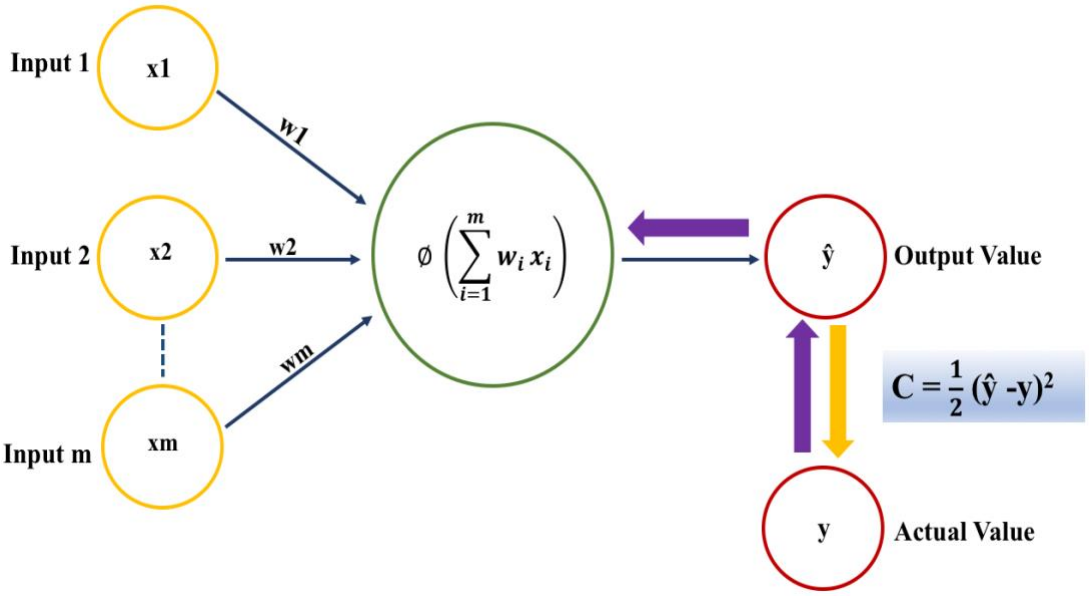


Figure 4.2. Architecture of a perceptron involving information processing. Associated with weighted parameters that can be tuned with activation functions.

4.4. ARTIFICIAL NEURAL NETWORKS (ANN) IN BIOLOGY

ANN can provide interrelationships of large sets of data; hence we can get pattern recognition of complex biological data with millions of inputs. In cancer research alone, ANN has been applied to image processing, outcome prediction, staging and diagnosis. In addition, a microarray experiment which typically consists of 39 human specimens and 8000 genes generate 312,000 data points or gene expression ratios that use ANN for analysis [114]. ANN applications in the field of biology vary from disease detection to diagnosis. Based on gene expression signatures, cancers have been classified to specific diagnostic categories using ANN [115,116]. ANN has been used to predict allergic diseases using single nucleotide polymorphisms data [117]. Drug toxicity risk has also been studied by predicting quinone formation in metabolic oxidation using ANN which enabled accurate prediction of quinone species in non-obvious cases [118].

ANN has further been implemented to increase sensitivity and specificity values of various spectroscopy data such as detection of serum proteins yield using mass spectroscopy for the detection and diagnosis of colorectal cancer [119]. High performance in accuracy, precision, sensitivity and specificity has been observed when principal component analysis (PCA) and ANN classifier models were applied to SERS data of non-structural proteins (NS1), a biomarker for Dengue fever from salivary surface [120]. Cuong *et.al* used multivariate statistical analysis on SERS data to enable robust quantification of pyocyanin starting at concentration 1 ng.mL -1 for early detection of biofilm formation to prevent antimicrobial infections [77]. Such synergistic combinations of spectroscopic measurements and statistical analysis has motivated us to study the SERS spectral analysis of different metabolites using artificial neural networks.

4.5. BUILDING AN ANN PREDICTION MODEL

By using uniform SERS substrates having high signal enhancements, we analyzed the SERS data of indole and 2-AAP using ANN to achieve a resilience quantification of their respective concentrations ranging from 1 pg.mL $^{-1}$ to 1 mg.mL $^{-1}$. Initially, it is important to import the essential packages or libraries that support a variety of deep learning architectures in Python or R. Then data pre-processing is the first step in any data mining process. This step involves feature scaling of imported data using standardization or normalization. Standardization is implemented in this work, which transform attributes with a Gaussian distribution with differing means and standard deviations to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1. Preprocessing of data is essential because, the gathered data might have missing values, impossible data combinations or noisy and unreliable data which makes learning very confusing and inefficient during the training process, affecting the accuracy of the final prediction. [121].

The next step involves the shuffling of all the data points in the data sets and splitting the shuffled data into training set and testing set. Data splitting is an important step in the ANN model, which divides the existing data into training, test and validation sets. The training and test set are labeled where in the training set is used to optimize the model parameters or in other words train the model and the test set is used to cross-validate the training data to avoid over-fitting. The validation set is used after the training to assess the performance of the model [122].

4.5.1. Training with Feed-Forward Neural Networks (FF-NN)

Feed-forward uses supervised learning algorithm where a pattern is presented as an input to learn.

It used gradient descent which is an optimization algorithm to find the values of parameters of a function (f) in order to minimize the cost function by finding the global minimum of the function [123]. As gradient descent is slow for large datasets, an iterative method called stochastic gradient descent is used in FF-NN i.e. this involves updating the parameters or coefficients for each training instant rather than at the end of the batch of instances [124]. The training process is given as [123,125]:

Step 1: The weights are initialized to small numbers close to 0.

Step 2: The first observation in your dataset is inputted where in each feature is one input node.

Step 3: Forward-Propagation: The neurons are activated from left to right and these activations limited by weights are propagated until the predicted result y is obtained.

Step 4: The error is generated after comparing the output value to the predicted value where

Total Error = $\sum \frac{1}{2}$ (target probability – output probability)^2

Step 5: Back-Propagation: The error is back-propagated and the weights are updated by calculating the gradient descent depending on the error obtained. The learning rate determines the extent to which the weights can be updated.

Step 6: Steps 1 to 5 are repeated and the weights are updated after each observation (reinforcement learning) or after a batch of observations (batch learning).

Step 7: When the whole training set is passed through the ANN, it makes an epoch. An optimum number of epochs must be done to get an accurate prediction.

4.5.2. Training with Convolutional Neural Networks (CNN)

When input consists of images having red-blue-green (RGB) encoding, CNN has the ability to arrange its neurons in three dimensions (width, height, depth) to give a 3D output volume of neuron activations. Here, the CNN receives an input of a normal color image as a rectangular box whose width and height are measured by pixels. This is also known as LeNet architecture. It consists of the different layers [126]:

Convolution Layer: This computes the output of neurons by calculating the dot product between their weights and the small input volumes they are connected to obtained by using filters or feature detectors. Many such feature maps are created to obtain the first convolutional layer.

ReLU Layer: An activation function known as the rectifier function [***max (0, x)***] thresholding at 0 is applied during feature mapping to increase non-linearity as the digital images in the real world are highly non-linear.

Max Pooling: This operation is implemented to reduce the spatial distortion and the size of the image. Reducing data volume prevents overfitting which makes this step significant in CNN modelling. The pooling is done by taking an average or sum of pixel values in a chosen matrix. This is also called as downsampling.

Full Connection: A fully connected layer of neurons in CNN has full connections to all activations in the previous layer and the next layer. This layer uses functions such as softmax or cross entropy to classify the input image into various classes using the features based on the training dataset.

The overall training process of CNN is as follows [127]:

Step 1: All the filters and weights are initialized with random values.

Step 2: After acquiring the training data as input, the feed propagation step (Convolution, ReLU and pooling) is executed to find the output probabilities of each class.

Step 3: The total error rate at the error rate is calculated.

Step 4: The error is backpropagated to calculate the gradients of the error with respect to all weights in the neural network and gradient descent is used to update all the weights to minimize the output error.

Step 5: Steps 2-4 are repeated for a specific optimal number of epochs.

4.6. SERS DATA ACQUISITION AND BUILDING A PREDICTION MODEL

The laser wavelength, power and exposure time were carefully selected to avoid any damage to the biological samples including the metabolites and the substrates. An objective lens of 60X was and a laser of 785 nm wavelength at a power of 7.3 μW with exposure time 0.5 s was selected respectively. The mapped SERS data sets including water and 10 concentrations of Indole and 2-AAP concentrations ranging from 1 pg.mL $^{-1}$ to 1 mg.mL $^{-1}$ were acquired. Raman spectra for each concentration was captured for Raman shift between 600 to 1600 cm^{-1} comprising of 1011 Raman shift features. Overall, each dataset consists of 308 spectra with 1011 features.

SERS spectral processing and analysis were performed using numpy, sklearn and keras libraries in Python 3.6.3. In the first step, we apply feature scaling to our imported data of 10 concentrations of Indole/2-AAP by standardizing the values. Next step involves creating labels for the 10 concentrations and log transforming them for easy representation of the y axis in our analysis. From the 308 × 1011 × 10 spectral points collected for all the concentrations, 80 % of them are randomly selected as training set and the remaining withheld as testing set.

An additional step involving principal component analysis (PCA) is implemented on the training set. Every Raman spectrum in the dataset has 1011 Raman shift features. However, most of the features are redundant or insignificant for ANN. Removing these unnecessary features or the noise in this case will reduce the complexity and will only retain the most significant features with maximum variance to differentiate the characteristic peaks for every concentration. Hence, PCA is a dimension reduction tool that is used to reduce the dimensionality of the dataset with large number of correlated variates, while keeping as much as variance possible in the dataset [120]. Note that PCA fit to the training set can be transformed to test set.

We then combined the characteristics of PCA with ANN to predict a set of dependent variables from a large set of independent variables. After compiling the ANN, the dataset for water is imported, pre-processed, PCA fit with training data and predicted before running the validation model.

4.6.1. Using Feed-Forward Neural Network (FF-NN)

Non-augmented data of the ten concentrations of Indole and 2-AAP were separately used for FF-NN. It is important that we select the dense parameters while defining our neural network whose parameters include output dimensions, initializer, activation function (ReLU) and input dimensions. Compiling the NN includes the optimizer to find the best weights to add to the algorithm and loss given by mean squared error. To fit the ANN to the training set, we used 47 epochs and a batch size of 5 and checked the accuracy and loss for each epoch. After predicting Indole/2-AAP concentration, water data was imported and preprocessed with training and the final results are plotted. The predicted concentrations versus actual concentrations were plotted and fitted to determine the R^2 value. The R^2 determines the goodness of the fit by measuring how close the data points are to the fitted regression line.

Results:

The FF-NN makes a prediction of indole and 2-AAP concentrations between 1 pg.mL $^{-1}$to 1 mg.mL $^{-1}$ separately. The accuracy of the prediction is determined by fitting a line of slope 1 to the predicted concentration versus actual concentration plot. The R^2 value for indole was calculated to be 0.93 with loss of 55% and accuracy of 71% for the training set [Figure 4.3 (a) and (b)]. From Figure 4.3 (c)**,** error bars capturing the variations in predictions, indicate that the predictions are not accurate (compared to the actual values). Also, 1 pg.mL $^{-1}$ metabolite concentration coincides with water that questions the ability of this model to learn the data efficiently. When a similar model was used for 2-AAP, the R^2 value obtained was 0.96 [Figure 4.4 (c)] . The loss and accuracy for the training data were both calculated to be 29% and 68% respectively [Figure 4.4 (a) and (b)]. Similar to 2-AAP, significant error bars were present in addition to the predicted value for the lowest concentration of indole overlapping with water. For larger SERS mappings, this issue in the FF-NN model might not allow precise and accurate prediction of metabolite concentrations. Hence, to address this problem, a CNN model was computed for the same spectral data.\

For Indole

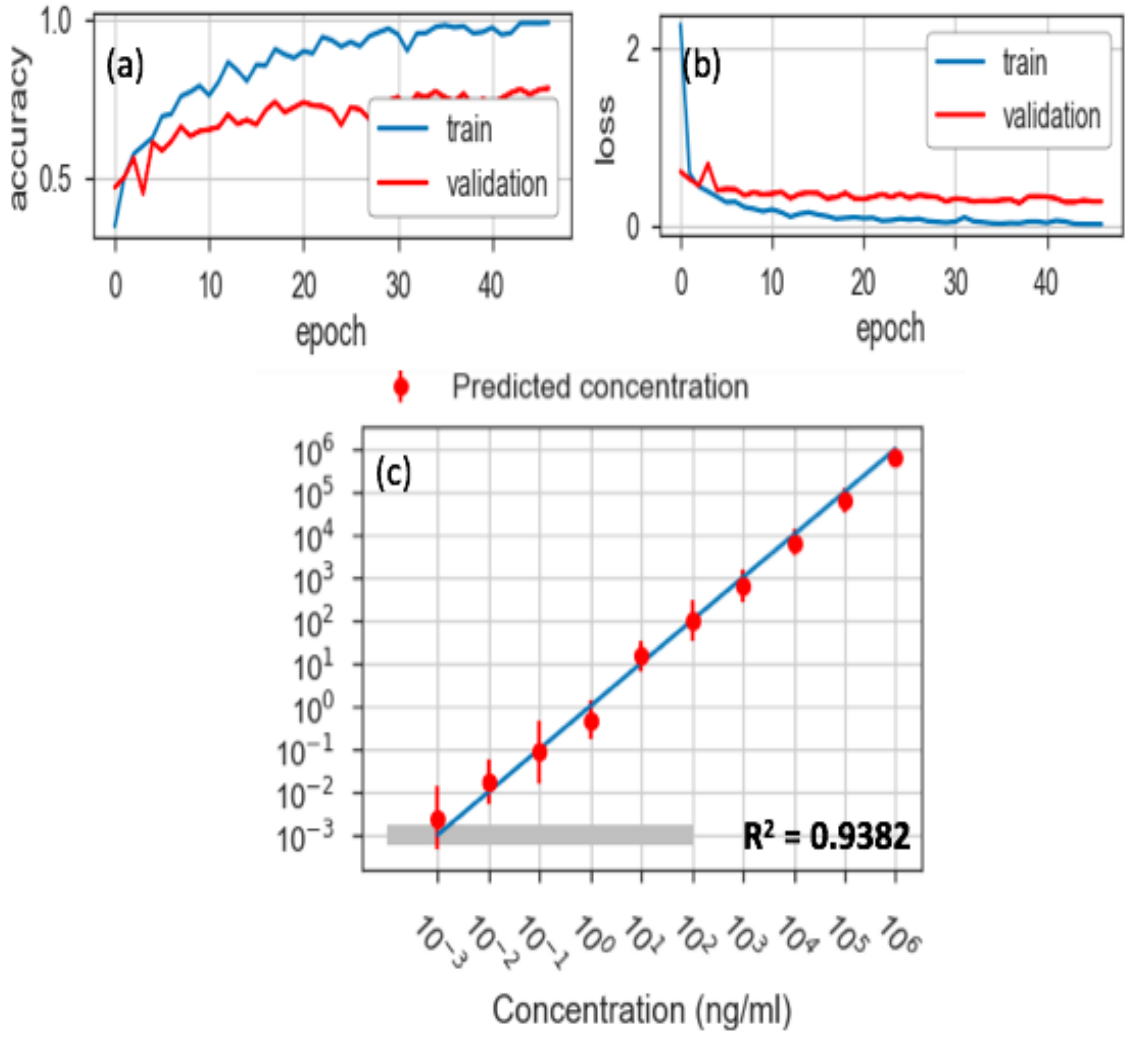


Figure 4.3. (a) Accuracy and (b) loss prediction of indole during the training of FF-ANN. (c) Predicted concentrations of indole (grey bar depicts water) using FF-ANN.

For 2-AAP:

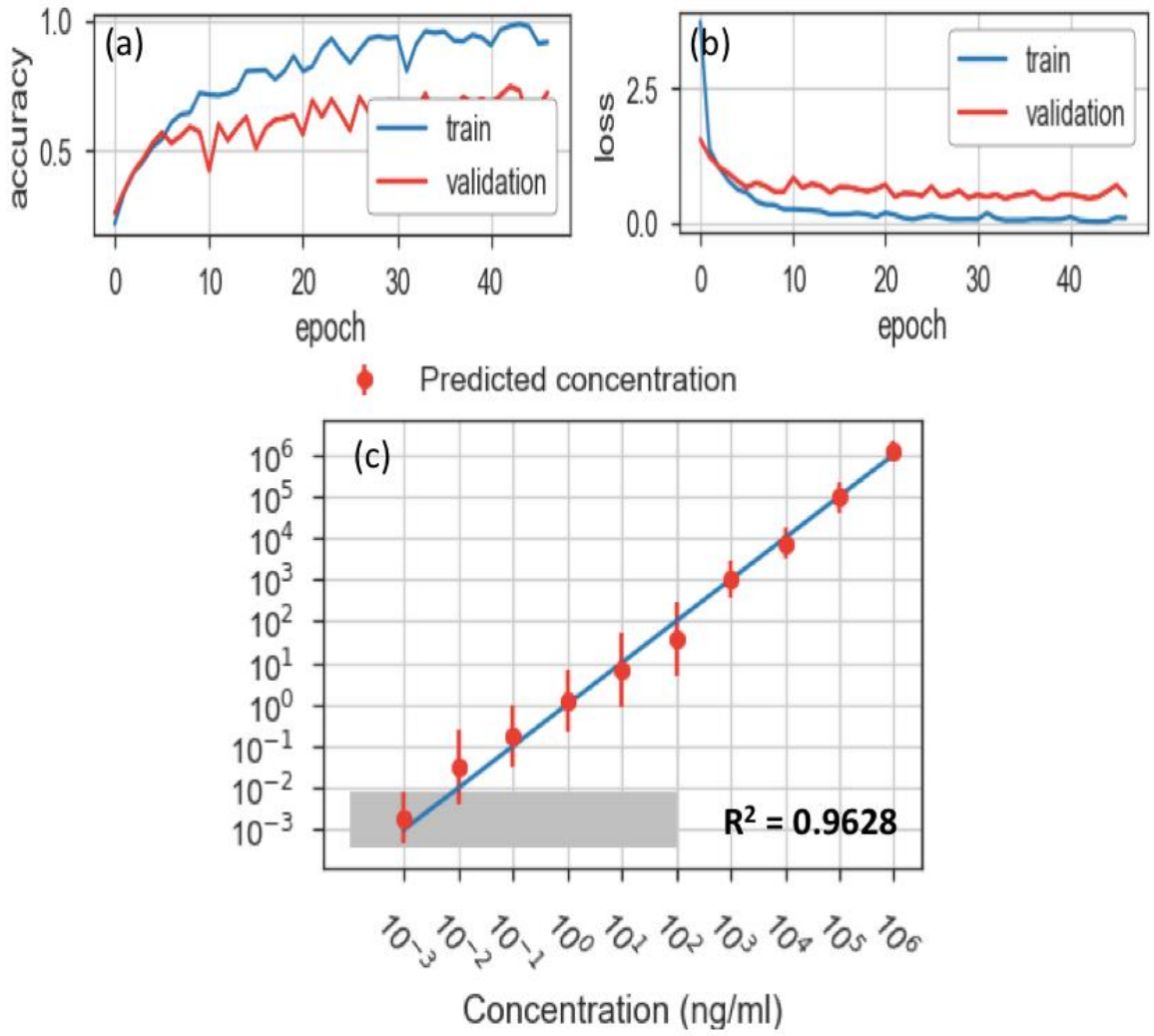


Figure 4.4. (a) Accuracy and (b) loss prediction of 2-AAP during the training of FF-ANN. (c) Predicted concentrations of 2-AAP using FF-ANN**.**

4.6.2. Using Convolutional Neural Network (CNN)

The dataset consisting of 10 concentrations of indole and 2-AAP each was augmented by creating 9x9 images in the CNN model. The 1300 spectral features were shuffled to get 26,998 parameters.

For this algorithm, we have used sequential, convolution 2D, maxpooling, flattening and dense parameters from Keras library. The convolution 2D requires the number of filters, activation function and the input shape to be inputed. The pool size for maxpooling was required to reduce the size of the data. The dense function is also added and then the NN was compiled similar to FF-NN. The training data was fit using 50 epochs and a batch size of 150 to get the desired output. The R^2 value was predicted from the plot of predicted concentration versus actual concentration for comparison with the above FF-NN model.

Results:

The CNN model demonstrates accurate prediction between 1 pg.mL $^{-1}$ to 1 mg.mL $^{-1}$ for both indole and 2-AAP as shown in Figure 4.5 and 4.6 respectively. By fitting the predicted versus actual concentration with a line having a slope of 1 representing perfect predictive capability, we get a R^2 value of 0.97 and 0.99 for indole and 2-AAP respectively. In addition to the R^2 values of the metabolites in the CNN model being higher than that of the FF-NN model, the size of the error bars also decreases significantly. The loss and accuracy values for training data of indole are 0.0238 and 0.996 whereas the loss and accuracy values for training data of 2-AAP are 0.0081 and 1. We can also notice that the predicted concentrations of the metabolites as low as 10 pg.mL $^{-1}$ are well separated from water. Hence, this model exhibits comparatively more sensitivity and specificity to give a robust quantification of different metabolites compared to FF-NN models

For Indole:

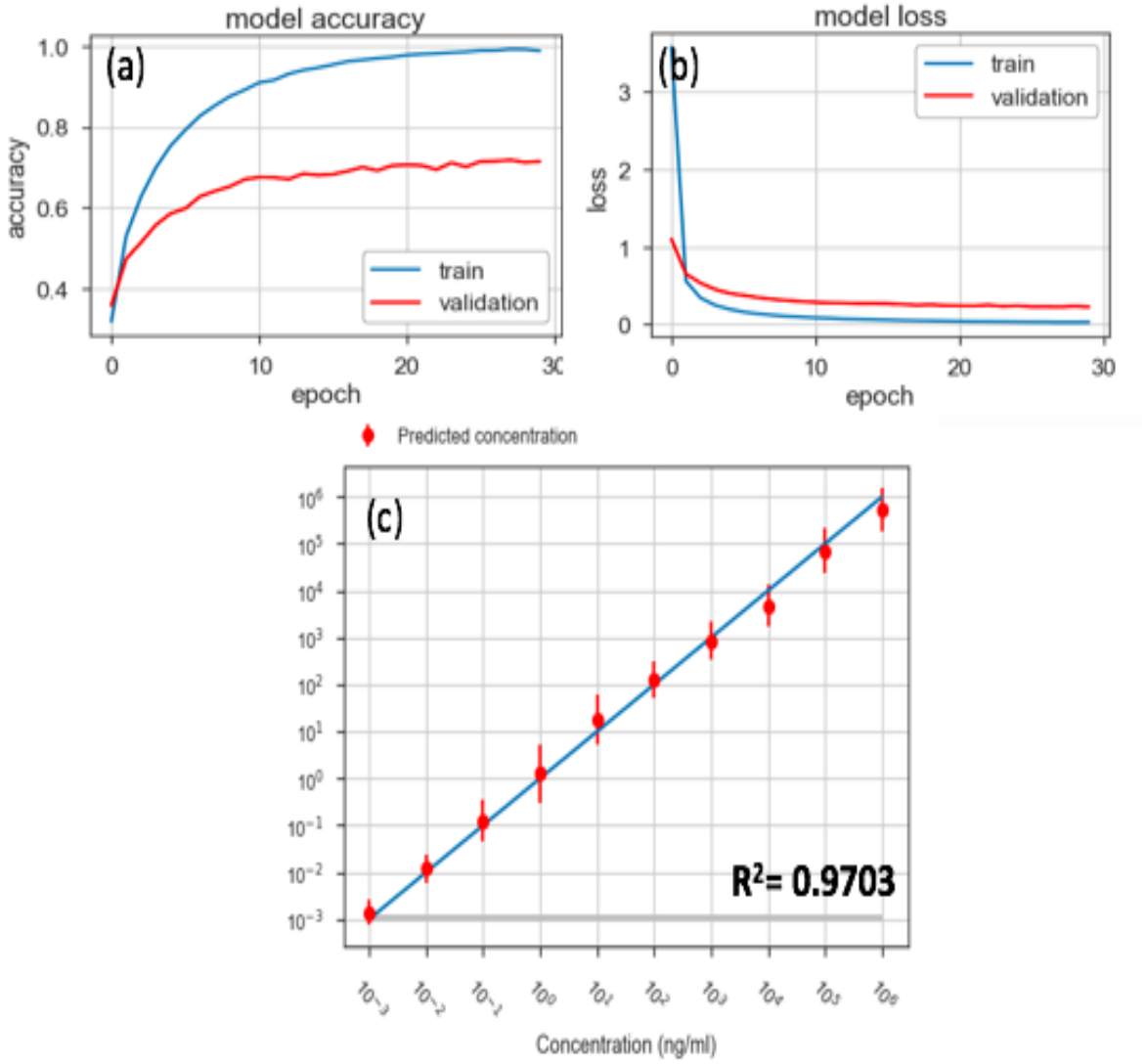


Figure 4.5. (a) Accuracy and (b) loss prediction of indole during the training of Convolutional ANN. (c) Predicted concentrations of indole using Convolutional ANN.

For 2-AAP:

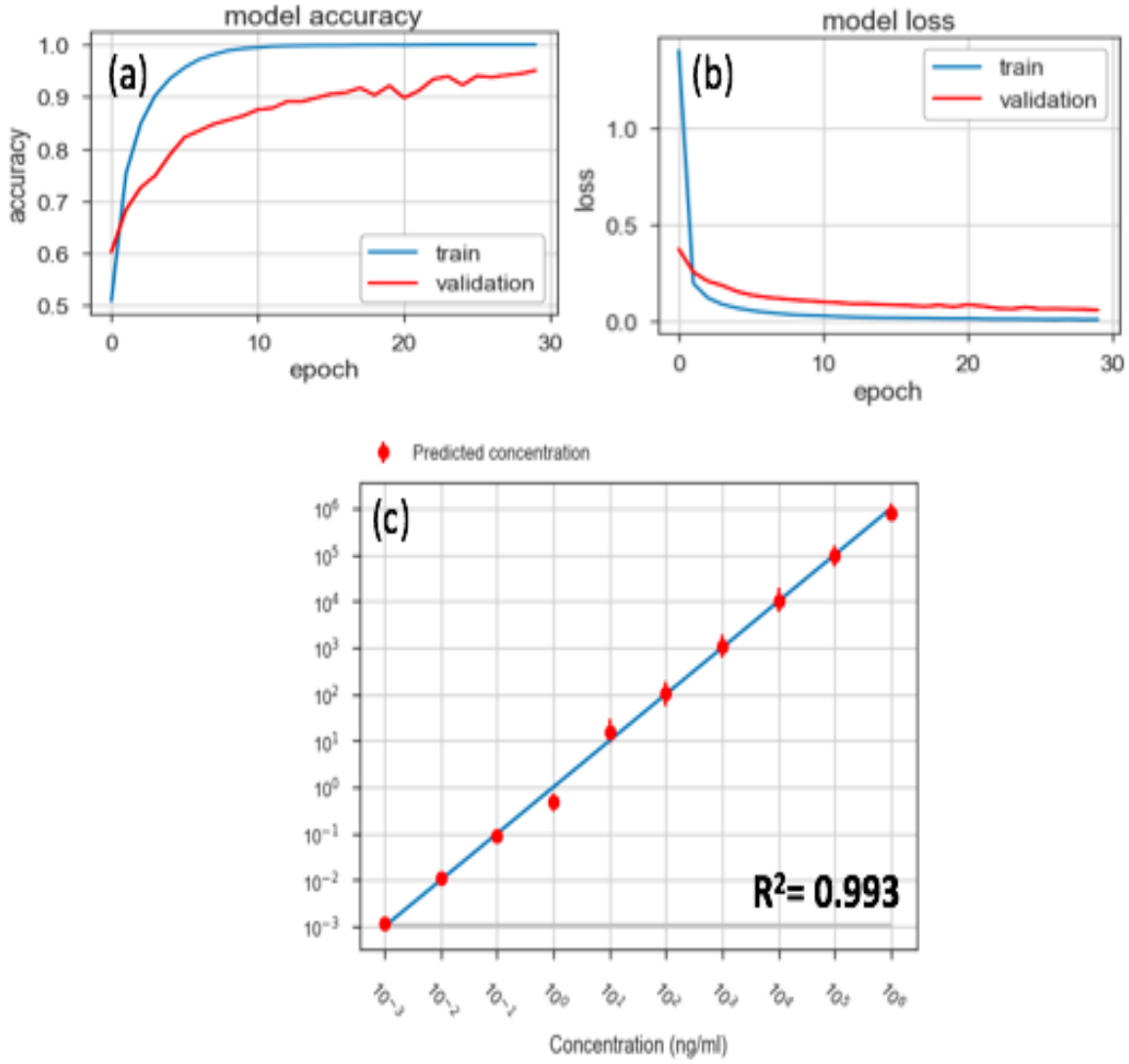


Figure 4.6. (a) Accuracy and (b) loss prediction of 2-AAP during the training of Convolutional ANN. (c) Predicted concentrations of 2-AAP using Convolutional ANN.

4.7. MULTIPLEX ANALYSIS USING ANN

Biological systems are complex i.e. they do not consist of one analyte but many interrelated analytes and it is important that only the most valid, biologically meaningful results from the

complex mixture of compounds is extracted. SERS has attracted attention for its potential to detect multiple analytes simultaneously, also known as multiplexing sensing. Multiplexing eliminates the extra steps required to isolate the analyte of interest from a complex mixture [127]. To identify thiophenol in a mixture of two analytes, a multivariate external control such as an oscillating electric field was used in combination with SERS [128]. External controls were later replaced by statistical models for quantitative multivariate analysis. For example, changes in concentrations of glucose and lactate concentrations were monitored sequentially using SERS and the multivariate analysis was performed using partial least square (PLS) and cross validation method [129]. Yet, absolute determination of analyte levels at very low concentrations proved to be challenging using the above methods. The use of ANN models for multiplex sensing of metabolites is capable of quantifying very low concentration with high sensitivity and precision. Therefore, a SERS sensor coupled with ANN modelling for multiplexing sensing can be useful in metabolic monitoring and diagnosing a wide variety of diseases.

A very popular application of CNN model involves distinguishing between images of two different objects. In the classic ML example of distinguishing between images of cats and dogs, we need to look at their ears, whiskers, tails, furs and so forth. Similarly, for the computer to learn from the training data labeled with discrete values as "dog" or "cat", it needs access to features like ears, whiskers, tails and so forth to distinguish them. These features exist in configuration of group of pixels. The intuition is that the hidden layers in the neural net will extract the features in such a way that lower layers learn simple edges or lines whereas the subsequent higher layers can learn features like shapes. A similar model is implemented to the indole and 2-AAP data and the output is determined.

4.7.1. Experiment

Pure gradients of 2-AAP and indole concentration data were imported separately to train a model with two output modes. The data was augmented by shuffling the two datasets to obtain 52600 images as training data. The training data was fit using 12 epochs and a batch size of 150 to get the two outputs. The output node for 2-AAP was validated by plotting predicted values versus actual values provided indole = 0. The same validation was followed for indole.

4.7.2. Results

Two output nodes were validated separately by plotting the concentrations of the metabolites . When looking at the predicted concentration of indole, the model did not show any traces of 2-AAP in the output node of indole [Figure 4.7] and vice- versa [Figure 4.8]. The model is able to accurately differentiate between the two metabolites as well as quantify them separately in the two output nodes without any crossover effects even after data augmentation. Although these results are promising, more spectra with 2-AAP and indole needs to be collected to further validate this for future work.

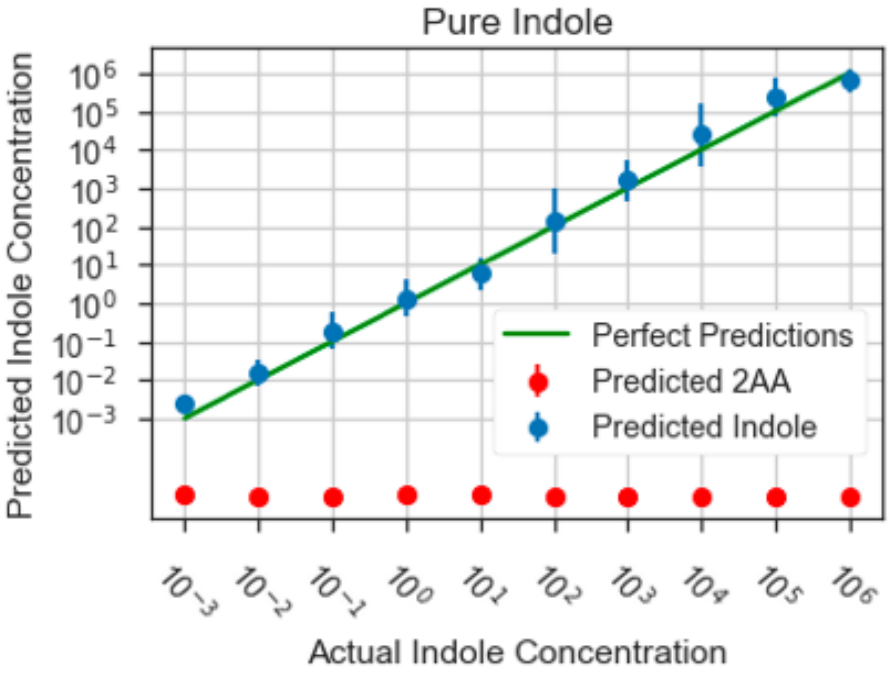


Figure 4.7. Multiplex detection involving two output nodes where indole is predicted.

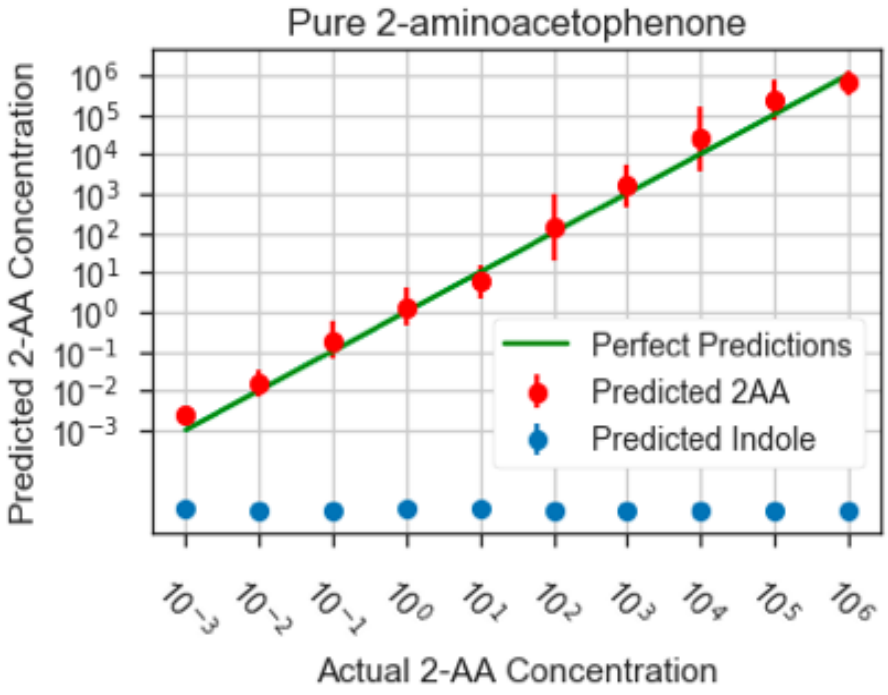


Figure 4.8. Multiplex detection involving two output nodes where 2-AAP is predicted.

4.8. CONSIDERATIONS FOR THE BIANALYTE APPROACH AND FUTURE WORK

The simultaneous detection of two or more metabolites in a mixture of compounds on a multi-sensing platform can be performed by SERS in combination with ANN provided the following factors are considered:

An important consideration for multianalyte studies using SERS is the selection of analytes provided the two or more molecules should contribute non-negligibly to the spectral counts. The chosen bianalyte partners should have clear spectral differences that lead to unambiguous identification of the individual analytes and their mixtures and the analytes should have comparable order of magnitude. Another essential factor to consider while selecting the analytes is that both of them should be reasonably photostable wherein the SERS events should not skew in favor of the more stable analyte. The bianalyte approach can also be biased depending on the molecular coverage of the analyte which depends on the surface binding affinity of the analytes to the SERS substrate [130]. Hence, the selection process for the bianalyte pair should be strategically assessed and developed.

After careful selection of the bianalyte pair and collection of their SERS spectral data, the data of the two analytes can be systematically mixed in various permutations to train and test the CNN model. The spectral data of a mixture of the two analytes in known concentrations can then be validated using the ANN models. In such a model, there might be various other problems to address as the mixed molecules prediction and quantification might be unable to uncover the complex features related to the spectral data of the individual analytes, often referred to as ANN black boxes. Moreover, the small biological data sets might be susceptible to overfitting especially with

unbiased and noisy data. Hence, we require large training data sets for the model to learn a proper generalization of knowledge contained in the data. With many types of ANNs available, selecting the right and task appropriate architecture is not straight-forward. Finally, the training process of the spectral datasets might be computationally intensive and time consuming for which it requires access to graphics processing units (GPU) making the computation expensive [131].

Appropriate implementation of these factors can lead to promising advancements in SERS based biosensors and computational biology for complex multianalyte systems.

CHAPTER 5 CONCLUSION

A kinetically controlled chemical reduction method involving nucleation and growth steps was carried out to synthesize uniform 40 nm Au nanospheres. SERS substrates were fabricated using EHD flow to drive the growth of Au nanosphere oligomer seeds followed by chemical cross-linking. This non-lithographic fabrication technique yielded uniform nanogap spacings of 0.9 nm that provided uniform SERS signals which in fact proved to be advantageous for detection of metabolites when individual Raman bands were investigated. SERS substrates exhibited the ability to detect indole and 2-AAP in aqueous media in biologically relevant levels of 1 $pg.mL^{-1}$ to 1 $mg.mL^{-1}$. Large area signal uniformity, low excitation power of 7.3 μW and short laser acquisition time of 0.5 s enabled statistical analysis for quantitative molecular detection through acquisition of large datasets.

The procurement of reliable training sets with high and reproducible SERS intensity can give rise to robust spectral analysis of metabolites using deep learning. ANN models such as feed-forward ANN and convolutional ANN were applied and compared to achieve quantification of metabolite concentrations. Convolutional ANN models applied to indole and 2-AAP were able to quantify concentrations as low as 1 $pg.mL^{-1}$ with high sensitivity and precision. The algorithm was further used to obtain individual quantification of metabolites in a mixture of two metabolites attracting attention for its potential for multiplex detection. Thus, the simultaneous detection of various metabolites in a complex biological mixture on a multi-sensing platform using SERS in combination with ANN analysis can be a promising biosensing technique for metabolic monitoring and diagnosis of metabolic diseases.

REFERENCES

1. N. Bhalla, P. Jolly, N. Formisano, and P. Estrela, "Introduction to biosensors," *Essays in Biochemistry*, vol. 60, no. 1, pp. 1-8, 2016.
2. T. Vo-Dinh, "SERS chemical sensors and biosensors: new tools for environmental and biological analysis," *Sensors and Actuators B: Chemical*, vol. 29, no. 1-3, pp. 183–189, 1995.
3. F.-L. Cui, J. Fan, J.-P. Li, and Z.-D. Hu, "Interactions between 1-benzoyl-4-p-chlorophenyl thiosemicarbazide and serum albumin: investigation by fluorescence spectroscopy," *Bioorganic & Medicinal Chemistry*, vol. 12, no. 1, pp. 151–157, 2004.
4. X. Fan, I. M. White, S. I. Shopova, H. Zhu, J. D. Suter, and Y. Sun, "Sensitive optical biosensors for unlabeled targets: A review," *Analytica Chimica Acta*, vol. 620, no. 1-2, pp. 8–26, 2008.
5. S. Abalde-Cela, C. Abell, R. A. Alvarez-Puebla, and L. M. Liz-Marzán, "Real Time Dual-Channel Multiplex SERS Ultradetection," *The Journal of Physical Chemistry Letters*, vol. 5, no. 1, pp. 73–79, Oct. 2013.
6. K. C. Bantz, A. F. Meyer, N. J. Wittenberg, H. Im, Ö. Kurtuluş, S. H. Lee, N. C. Lindquist,S.-H. Oh, and C. L. Haynes, "Recent progress in SERS biosensing," *Physical Chemistry Chemical Physics*, vol. 13, no. 24, p. 11551, 2011.
7. H. Nguyen, J. Park, S. Kang, and M. Kim, "Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications," *Sensors*, vol. 15, no. 5, pp. 10481–10510, 2015.
8. H. Nguyen, J. Park, S. Kang, and M. Kim, "Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications," *Sensors*, vol. 15, no. 5, pp. 10481–10510, 2015.
9. Y. S. Huh, A. J. Chung, and D. Erickson, "Surface enhanced Raman spectroscopy and its application to molecular and cellular analysis," *Microfluidics and Nanofluidics*, vol. 6, no. 3, pp. 285–297, Aug. 2009.
10. A. B. Veloso, J. P. F. Longo, L. A. Muehlmann, B. F. Tollstadius, P. E. N. Souza, R. B. Azevedo, P. C. Morais, and S. W. D. Silva, "SERS Investigation of Cancer Cells Treated with PDT: Quantification of Cell Survival and Follow-up," *Scientific Reports*, vol. 7, no. 1, Mar. 2017.
11. L. Guerrini, R. Arenal, B. Mannini, F. Chiti, R. Pini, P. Matteini, and R. A. Alvarez-Puebla, "SERS Detection of Amyloid Oligomers on Metallorganic-Decorated Plasmonic Beads," *ACS Applied Materials & Interfaces*, vol. 7, no. 18, pp. 9420–9428, 2015.
12. A. Zengin, U. Tamer, and T. Caykara, "A SERS-Based Sandwich Assay for Ultrasensitive and Selective Detection of Alzheimer's Tau Protein," *Biomacromolecules*, vol. 14, no. 9, pp. 3001–3009, Sep. 2013.
13. J. Chan, S. Fore, S. Wachsmann-Hogiu, and T. Huser, "Raman spectroscopy and microscopy of individual cells and cellular components," *Laser & Photonics Review*, vol. 2, no. 5, pp. 325–349, Jan. 2008.
14. M. Fleischmann, P. Hendra, and A. Mcquillan, "Raman spectra of pyridine adsorbed at a silver electrode," *Chemical Physics Letters*, vol. 26, no. 2, pp. 163–166, 1974.
15. L. A. Lane, X. Qian, and S. Nie, "SERS Nanoparticles in Medicine: From Label-Free Detection to Spectroscopic Tagging," *Chemical Reviews*, vol. 115, no. 19, pp. 10489–10529, 2015.
16. D. Cialla, S. Pollok, C. Steinbruker, K. Weber, and J. Popp, "SERS-based detection of biomolecules," *Nanophotonics*, vol. 3, no. 6, Jan. 2014.

17. X. Gu, M. J. Trujillo, J. E. Olson, and J. P. Camden, "SERS Sensors: Recent Developments and a Generalized Classification Scheme Based on the Signal Origin," *Annual Review of Analytical Chemistry*, vol. 11, no. 1, 2018.
18. E. Petryayeva and U. J. Krull, "Localized surface plasmon resonance: Nanostructures, bioassays and biosensing—A review," *Analytica Chimica Acta*, vol. 706, no. 1, pp. 8–24, 2011.
19. A. J. Mcquillan, "The discovery of surface-enhanced Raman scattering," *Notes and Records of the Royal Society*, vol. 63, no. 1, pp. 105–109, 2009.
20. S. Cong, Y. Yuan, Z. Chen, J. Hou, M. Yang, Y. Su, Y. Zhang, L. Li, Q. Li, F. Geng, and Z. Zhao, "Noble metal-comparable SERS enhancement from semiconducting metal oxides by making oxygen vacancies," *Nature Communications*, vol. 6, no. 1, 2015.
21. A. M. Schwartzberg, C. D. Grant, A. Wolcott, C. E. Talley, T. R. Huser, R. Bogomolni, and J. Z. Zhang, "Unique Gold Nanoparticle Aggregates as a Highly Active Surface-Enhanced Raman Scattering Substrate," *The Journal of Physical Chemistry B*, vol. 108, no. 50, pp. 19191–19197, 2004.
22. R. M. Jarvis and R. Goodacre, "Characterisation and identification of bacteria using SERS," *Chemical Society Reviews*, vol. 37, no. 5, p. 931, 2008.
23. S. Abalde-Cela, C. Abell, R. A. Alvarez-Puebla, and L. M. Liz-Marzán, "Real Time Dual-Channel Multiplex SERS Ultradetection," *The Journal of Physical Chemistry Letters*, vol. 5, no. 1, pp. 73–79, Oct. 2013.
24. K.-H. Su, Q.-H. Wei, X. Zhang, J. J. Mock, D. R. Smith, and S. Schultz, "Interparticle Coupling Effects on Plasmon Resonances of Nanogold Particles," *Nano Letters*, vol. 3, no. 8, pp. 1087–1090, 2003.
25. S. M. Adams, S. Campione, J. D. Caldwell, F. J. Bezares, J. C. Culbertson, F. Capolino, and R. Ragan, "Non-lithographic SERS Substrates: Tailoring Surface Chemistry for Au Nanoparticle Cluster Assembly," *Small*, vol. 8, no. 14, pp. 2239–2249, 2012.
26. J.-H. Lee, J.-M. Nam, K.-S. Jeon, D.-K. Lim, H. Kim, S. Kwon, H. Lee, and Y. D. Suh, "Tuning and Maximizing the Single-Molecule Surface-Enhanced Raman Scattering from DNA-Tethered Nanodumbbells," *ACS Nano*, vol. 6, no. 11, pp. 9574–9584, Dec. 2012.
27. A. Jaworska, S. Fornasaro, V. Sergo, and A. Bonifacio, "Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM). A Critical Review," *Biosensors*, vol. 6, no. 3, p. 47, 2016.
28. D. Liu, T. Evans, and F. Zhang, "Applications and advances of metabolite biosensors for metabolic engineering," *Metabolic Engineering*, vol. 31, pp. 35–43, 2015.
29. K. L. Whiteson, S. Meinardi, Y. W. Lim, R. Schmieder, H. Maughan, R. Quinn, D. R. Blake, D. Conrad, and F. Rohwer, "Breath gas metabolites and bacterial metagenomes from cystic fibrosis airways indicate active pH neutral 2,3-butanedione fermentation," *The ISME Journal*, vol. 8, no. 6, pp. 1247–1258, Sep. 2014.
30. V. Shalabaeva, L. Lovato, R. L. Rocca, G. C. Messina, M. Dipalo, E. Miele, M. Perrone, F. Gentile, and F. D. Angelis, "Time resolved and label free monitoring of extracellular metabolites by surface enhanced Raman spectroscopy," *Plos One*, vol. 12, no. 4, 2017.
31. Z. Yu, G. Kastenmüller, Y. He, P. Belcredi, G. Möller, C. Prehn, J. Mendes, S. Wahl, W. Roemisch-Margl, U. Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. Li, H.-E. Wichmann, A. Peters, F. Kronenberg, K. Suhre, J. Adamski, T. Illig, and R. Wang-Sattler, "Differences between Human Plasma and Serum Metabolite Profiles," *PLoS ONE*, vol. 6, no. 7, Aug. 2011.

32. A. A. Aksenov, R. D. Silva, R. Knight, N. P. Lopes, and P. C. Dorrestein, "Global chemical analysis of biology by mass spectrometry," *Nature Reviews Chemistry*, vol. 1, no. 7, p. 0054, May 2017.
33. S. Stryeck, R. Birner-Gruenberger, and T. Madl, "Integrative metabolomics as emerging tool to study autophagy regulation," *Microbial Cell*, vol. 4, no. 8, pp. 240–258, Jul. 2017.
34. P. Mosier-Boss, "Review on SERS of Bacteria," *Biosensors*, vol. 7, no. 4, p. 51, 2017.
35. E. Petryayeva and U. J. Krull, "Localized surface plasmon resonance: Nanostructures, bioassays and biosensing—A review," *Analytica Chimica Acta*, vol. 706, no. 1, pp. 8–24, 2011.
36. Srichan, M. Ekpanyapong, M. Horprathum, P. Eiamchai, N. Nuntawong, D. Phokharatkul, P. Danvirutai, E. Bohez, A. Wisitsoraat, and A. Tuantranont, "Highly-Sensitive Surface-Enhanced Raman Spectroscopy (SERS)-based Chemical Sensor using 3D Graphene Foam Decorated with Silver Nanoparticles as SERS substrate," *Scientific Reports*, vol. 6, no. 1, 2016.
37. S. Grubisha, R. J. Lipert, H.-Y. Park, J. Driskell, and M. D. Porter, "Femtomolar Detection of Prostate-Specific Antigen: An Immunoassay Based on Surface-Enhanced Raman Scattering and Immunogold Labels," *Analytical Chemistry*, vol. 75, no. 21, pp. 5936–5943, 2003.
38. S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, and H. Zeng, "Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis," *Biosensors and Bioelectronics*, vol. 25, no. 11, pp. 2414–2419, 2010.
39. C. V. D. Anibal, L. F. Marsal, M. P. Callao, and I. Ruisánchez, "Surface Enhanced Raman Spectroscopy (SERS) and multivariate analysis as a screening tool for detecting Sudan I dye in culinary spices," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 87, pp. 135–141, 2012.
40. J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, and S. J. Gibson, "Deep convolutional neural networks for Raman spectrum recognition: a unified solution," *The Analyst*, vol. 142, no. 21, pp. 4067–4074, 2017.
41. O. Alharbi, Y. Xu, and R. Goodacre, "Simultaneous multiplexed quantification of nicotine and its metabolites using surface enhanced Raman scattering," *The Analyst*, vol. 139, no. 19, pp. 4820–4827, 2014.
42. J. Wang, X. Liao, P. Zheng, S. Xue, and R. Peng, "Classification of Chinese Herbal Medicine by Laser-Induced Breakdown Spectroscopy with Principal Component Analysis and Artificial Neural Network," *Analytical Letters*, vol. 51, no. 4, pp. 575–586, 2017.
43. C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, "Deep Learning in Label-free Cell Classification," *Scientific Reports*, vol. 6, no. 1, 2016.
44. W. Chakraborty, R. Ray, N. Samanta, and C. Roychaudhuri, "Quantitative differentiation of multiple virus in blood using nanoporous silicon oxide immunosensor and artificial neural network," *Biosensors and Bioelectronics*, vol. 98, pp. 180–188, 2017.
45. O. Alharbi, Y. Xu, and R. Goodacre, "Simultaneous multiplexed quantification of nicotine and its metabolites using surface enhanced Raman scattering," *The Analyst*, vol. 139, no. 19, pp. 4820–4827, 2014.
46. M. Kerker, "Electromagnetic model for surface-enhanced Raman scattering (SERS) on metal colloids," *Accounts of Chemical Research*, vol. 17, no. 8, pp. 271–277, 1984.
47. T. B. Nguyen, T. K. T. Vu, Q. D. Nguyen, T. D. Nguyen, T. A. Nguyen, and T. H. Trinh, "Preparation of metal nanoparticles for surface enhanced Raman scattering by laser ablation

method," *Advances in Natural Sciences: Nanoscience and Nanotechnology*, vol. 3, no. 2, p. 025016, Apr. 2012.
48. Mosier-Boss, "Review of SERS Substrates for Chemical Sensing," *Nanomaterials*, vol. 7, no. 6, p. 142, Aug. 2017.
49. F. Tian, F. Bonnier, A. Casey, A. E. Shanahan, and H. J. Byrne, "Surface enhanced Raman scattering with gold nanoparticles: effect of particle shape," *Anal. Methods*, vol. 6, no. 22, pp. 9116–9123, 2014.
50. J. Polte, T. T. Ahner, F. Delissen, S. Sokolov, F. Emmerling, A. F. Thünemann, and R. Kraehnert, "Mechanism of Gold Nanoparticle Formation in the Classical Citrate Synthesis Method Derived from Coupled In Situ XANES and SAXS Evaluation," *Journal of the American Chemical Society*, vol. 132, no. 4, pp. 1296–1301, Mar. 2010.
51. J. Kimling, M. Maier, B. Okenve, V. Kotaidis, H. Ballot, and A. Plech, "Turkevich Method for Gold Nanoparticle Synthesis Revisited," *The Journal of Physical Chemistry B*, vol. 110, no. 32, pp. 15700–15707, 2006.
52. Q. Yao, X. Yuan, V. Fung, Y. Yu, D. T. Leong, D.-E. Jiang, and J. Xie, "Understanding seed-mediated growth of gold nanoclusters at molecular level," *Nature Communications*, vol. 8, no. 1, 2017.
53. C. Ziegler and A. Eychmüller, "Seeded Growth Synthesis of Uniform Gold Nanoparticles with Diameters of 15−300 nm," *The Journal of Physical Chemistry C*, vol. 115, no. 11, pp. 4502–4506, Mar. 2011.
54. C. M. Phan and H. M. Nguyen, "Role of Capping Agent in Wet Synthesis of Nanoparticles," *The Journal of Physical Chemistry A*, vol. 121, no. 17, pp. 3213–3219, 2017.
55. S. Hong and X. Li, "Optimal Size of Gold Nanoparticles for Surface-Enhanced Raman Spectroscopy under Different Conditions," *Journal of Nanomaterials*, vol. 2013, pp. 1–9, 2013.
56. K. G. Stamplecoskie, J. C. Scaiano, V. S. Tiwari, and H. Anis, "Optimal Size of Silver Nanoparticles for Surface-Enhanced Raman Spectroscopy," *The Journal of Physical Chemistry C*, vol. 115, no. 5, pp. 1403–1409, Dec. 2011.
57. N. G. Bastús, J. Comenge, and V. Puntes, "Kinetically Controlled Seeded Growth Synthesis of Citrate-Stabilized Gold Nanoparticles of up to 200 nm: Size Focusing versus Ostwald Ripening," *Langmuir*, vol. 27, no. 17, pp. 11098–11105, Jun. 2011.
58. A. Zuber, M. Purdey, E. Schartner, C. Forbes, B. V. D. Hoek, D. Giles, A. Abell, T. Monro, and H. Ebendorff-Heidepriem, "Detection of gold nanoparticles with different sizes using absorption and fluorescence based method," *Sensors and Actuators B: Chemical*, vol. 227, pp. 117–127, 2016.
59. E. Tomaszewska, K. Soliwoda, K. Kadziola, B. Tkacz-Szczesna, G. Celichowski, M. Cichomski, W. Szmaja, and J. Grobelny, "Detection Limits of DLS and UV-Vis Spectroscopy in Characterization of Polydisperse Nanoparticles Colloids," *Journal of Nanomaterials*, vol. 2013, pp. 1–10, 2013.
60. L. Vigderman and E. R. Zubarev, "Therapeutic platforms based on gold nanoparticles and their covalent conjugates with drug molecules," *Advanced Drug Delivery Reviews*, vol. 65, no. 5, pp. 663–676, 2013.
61. A. A. Volkert, V. Subramaniam, M. R. Ivanov, A. M. Goodman, and A. J. Haes, "Salt-Mediated Self-Assembly of Thioctic Acid on Gold Nanoparticles," *ACS Nano*, vol. 5, no. 6, pp. 4570–4580, Mar. 2011.

62. Turcu, I. Zarafu, M. Popa, M. Chifiriuc, C. Bleotu, D. Culita, C. Ghica, and P. Ionita, "Lipoic Acid Gold Nanoparticles Functionalized with Organic Compounds as Bioactive Materials," *Nanomaterials*, vol. 7, no. 2, p. 43, 2017.
63. J. M. Abad, S. F. L. Mertens, M. Pita, V. M. Fernández, and D. J. Schiffrin, "Functionalization of Thioctic Acid-Capped Gold Nanoparticles for Specific Immobilization of Histidine-Tagged Proteins," *Journal of the American Chemical Society*, vol. 127, no. 15, pp. 5689–5694, 2005.
64. M. R. Ivanov, H. R. Bednar, and A. J. Haes, "Investigations of the Mechanism of Gold Nanoparticle Stability and Surface Functionalization in Capillary Electrophoresis," *ACS Nano*, vol. 3, no. 2, pp. 386–394, 2009.
65. S. Roux, B. Garcia, J.-L. Bridot, M. Salomé, C. Marquette, L. Lemelle, P. Gillet, L. Blum, P. Perriat, and O. Tillement, "Synthesis, Characterization of Dihydrolipoic Acid Capped Gold Nanoparticles, and Functionalization by the Electroluminescent Luminol," *Langmuir*, vol. 21, no. 6, pp. 2526–2536, 2005.
66. I. Turcu, I. Zarafu, M. Popa, M. Chifiriuc, C. Bleotu, D. Culita, C. Ghica, and P. Ionita, "Lipoic Acid Gold Nanoparticles Functionalized with Organic Compounds as Bioactive Materials," *Nanomaterials*, vol. 7, no. 2, p. 43, 2017.
67. Y. Lin, Y.-T. Tsai, C.-C. Chen, C.-M. Lin, and C.-H. Chen, "Two-Step Functionalization of Neutral and Positively Charged Thiols onto Citrate-Stabilized Au Nanoparticles," *The Journal of Physical Chemistry B*, vol. 108, no. 7, pp. 2134–2139, 2004.
68. S. M. Adams, S. Campione, J. D. Caldwell, F. J. Bezares, J. C. Culbertson, F. Capolino, and R. Ragan, "Non-lithographic SERS Substrates: Tailoring Surface Chemistry for Au Nanoparticle Cluster Assembly," *Small*, vol. 8, no. 14, pp. 2239–2249, 2012.
69. J. Fontana, J. Livenere, F. J. Bezares, J. D. Caldwell, R. Rendell, and B. R. Ratna, "Large surface-enhanced Raman scattering from self-assembled gold nanosphere monolayers," *Applied Physics Letters*, vol. 102, no. 20, p. 201606, 2013.
70. M. Sepaniak, S. Wells, D. Bhandari, N. Lavrik, P. M. Champion, and L. D. Ziegler, "Nanofabrication of Disc on Pillar Substrates for Surface Enhanced Raman Spectroscopy," *AIP Conference Proceedings*, 1267, 273, 2010.
71. W. J. Thrift, C. Q. Nguyen, M. Darvishzadeh-Varcheie, S. Zare, N. Sharac, R. N. Sanderson, T. J. Dupper, A. I. Hochbaum, F. Capolino, M. J. A. Qomi, and R. Ragan, "Driving Chemical Reactions in Plasmonic Nanogaps with Electrohydrodynamic Flow," *ACS Nano*, vol. 11, no. 11, pp. 11317–11329, 2017.
72. W. Thrift, A. Bhattacharjee, A. I. Hochbaum, R. Ragan, C. Nguyen, and K. Whiteson, "Robust SERS spectral analysis for quantitative detection of pycocyanin in biological fluids," *Biosensing and Nanomedicine X*, 2017.
73. J. H. Choi, S. M. Adams, and R. Ragan, "Design of a versatile chemical assembly method for patterning colloidal nanoparticles," *Nanotechnology*, vol. 20, no. 6, p. 065301, 2009.
74. W. D. Ristenpart, P. Jiang, M. A. Slowik, C. Punckt, D. A. Saville, and I. A. Aksay, "Electrohydrodynamic Flow and Colloidal Patterning near Inhomogeneities on Electrodes," *Langmuir*, vol. 24, no. 21, pp. 12172–12180, Apr. 2008.
75. M.-G. Song, K. J. M. Bishop, A. O. Pinchuk, B. Kowalczyk, and B. A. Grzybowski, "Formation of Dense Nanoparticle Monolayers Mediated by Alternating Current Electric Fields and Electrohydrodynamic Flows," *The Journal of Physical Chemistry C*, vol. 114, no. 19, pp. 8800–8805, 2010.
76. W. Thrift, A. Bhattacharjee, M. Darvishzadeh-Varcheie, Y. Lu, A. Hochbaum, F. Capolino, K. Whiteson, and R. Ragan, "Surface enhanced Raman scattering for detection of

Pseudomonas aeruginosaquorum sensing compounds," *Biosensing and Nanomedicine VIII*, 2015.
77. C. Q. Nguyen, W. J. Thrift, A. Bhattacharjee, S. Ranjbar, T. Gallagher, M. Darvishzadeh-Varcheie, R. N. Sanderson, F. Capolino, K. Whiteson, P. Baldi, A. I. Hochbaum, and R. Ragan, "Longitudinal Monitoring of Biofilm Formation via Robust Surface-Enhanced Raman Scattering Quantification of Pseudomonas aeruginosa- Produced Metabolites," *ACS Applied Materials & Interfaces*, vol. 10, no. 15, pp. 12364–12373, 2018.
78. G. Macias, M. Alba, L. F. Marsal, and A. Mihi, "Surface roughness boosts the SERS performance of imprinted plasmonic architectures," *Journal of Materials Chemistry C*, vol. 4, no. 18, pp. 3970–3975, 2016.
79. X. Cai and R. Li, "Concurrent profiling of polar metabolites and lipids in human plasma using HILIC-FTMS," *Scientific Reports*, vol. 6, no. 1, Jul. 2016.
80. G. N. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah, and D. Raftery, "Metabolomics-based methods for early disease diagnostics," *Expert Review of Molecular Diagnostics*, vol. 8, no. 5, pp. 617–633, 2008.
81. J. S. Fletcher, H. L. Kotze, E. G. Armitage, N. P. Lockyer, and J. C. Vickerman, "Evaluating the challenges associated with time-of-fight secondary ion mass spectrometry for metabolomics using pure and mixed metabolites," *Metabolomics*, vol. 9, no. 3, pp. 535–544, Nov. 2012.
82. S. L. Clauson, J. M. Sylvia, T. A. Arcury, P. Summers, and K. M. Spencer, "Detection of Pesticides and Metabolites Using Surface-Enhanced Raman Spectroscopy (SERS): Acephate," *Applied Spectroscopy*, vol. 69, no. 7, pp. 785–793, 2015.
83. S. Lee, H. Chon, J. Lee, J. Ko, B. H. Chung, D. W. Lim, and J. Choo, "Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging," *Biosensors and Bioelectronics*, vol. 51, pp. 238–243, 2014.
84. D. Dressler, Y. Mastai, M. Rosenbluh, and Y. Fleger, "Surface-enhanced Raman spectroscopy as a probe for orientation of pyridine compounds on colloidal surfaces," *Journal of Molecular Structure*, vol. 935, no. 1-3, pp. 92–96, 2009.
85. F. J. Gilchrist, H. Sims, A. Alcock, J. Belcher, A. M. Jones, D. Smith, P. Španěl, A. K. Webb, and W. Lenney, "Quantification of hydrogen cyanide and 2-aminoacetophenone in the headspace of Pseudomonas aeruginosa cultured under biofilm and planktonic conditions," *Analytical Methods*, vol. 4, no. 11, p. 3661, 2012.
86. C. T. Oloughlin, L. C. Miller, A. Siryaporn, K. Drescher, M. F. Semmelhack, and B. L. Bassler, "A quorum-sensing inhibitor blocks Pseudomonas aeruginosa virulence and biofilm formation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 44, pp. 17981–17986, 2013.
87. M. Hu, C. Zhang, Y. Mu, Q. Shen, and Y. Feng, "Indole Affects Biofilm Formation in Bacteria," *Indian Journal of Microbiology*, vol. 50, no. 4, pp. 362–368, 2010.
88. A. Berstad, J. Raa, and J. Valeur, "Indole – the scent of a healthy 'inner soil,'" *Microbial Ecology in Health & Disease*, vol. 26, 2015.
89. C. Darkoh, C. Chappell, C. Gonzales, and P. Okhuysen, "A Rapid and Specific Method for the Detection of Indole in Complex Biological Samples," *Applied and Environmental Microbiology*, vol. 81, no. 23, pp. 8093–8097, 2015.
90. M. Ghaedi, A. Daneshfar, A. Shokrollahi, H. Ghaedi, and F. A. Pili, "Highly Selective and Sensitive Spectrophotometric Determination of Trace Amounts of Silver Ion in Surfactant Media Using 2-Mercaptobenzoxazole," *Annali di Chimica*, vol. 97, no. 9, pp. 971–982, 2007.

91. C. Quintelas, E. Sousa, F. Silva, S. Neto, and T. Tavares, "Competitive biosorption of ortho-cresol, phenol, chlorophenol and chromium(VI) from aqueous solution by a bacterial biofilm supported on granular activated carbon," *Process Biochemistry*, vol. 41, no. 9, pp. 2087–2091, 2006.
92. X. Zou, Y. Wang, W. Liu, and L. Chen, "m-Cresol purple functionalized surface enhanced Raman scattering paper chips for highly sensitive detection of pH in the neutral pH range," *The Analyst*, vol. 142, no. 13, pp. 2333–2337, 2017.
93. Q. Tu, "Surface-enhanced Raman spectroscopy study of indolic molecules adsorbed on gold colloids," *Journal of Biomedical Optics*, vol. 15, no. 2, p. 020512, Jan. 2010.
94. A. Bigotto and B. Pergolese, "Surface-enhanced Raman spectroscopic studies of 2-mercaptobenzoxazole on silver sols," *Journal of Raman Spectroscopy*, vol. 32, no. 11, pp. 953–959, 2001.
95. K. J. Park, C. Wu, A. R. Mercer-Smith, R. A. Dodson, T. L. Moersch, P. Koonath, A. C. Pipino, H.-W. Lu, Y. Yang, V. S. Sapirstein, C. J. Taylor, and A. Niemz, "Raman system for sensitive and selective identification of volatile organic compounds," *Sensors and Actuators B: Chemical*, vol. 220, pp. 491–499, 2015.
96. V. Shalabaeva, L. Lovato, R. L. Rocca, G. C. Messina, M. Dipalo, E. Miele, M. Perrone, F. Gentile, and F. D. Angelis, "Time resolved and label free monitoring of extracellular metabolites by surface enhanced Raman spectroscopy," *Plos One*, vol. 12, no. 4, 2017. 125 C. Darkoh, C. Chappell, C. Gonzales, and P. Okhuysen, "A Rapid and Specific Method for the Detection of Indole in Complex Biological Samples," *Applied and Environmental Microbiology*, vol. 81, no. 23, pp. 8093–8097, 2015.
97. O. Alharbi, Y. Xu, and R. Goodacre, "Simultaneous multiplexed quantification of caffeine and its major metabolites theobromine and paraxanthine using surface-enhanced Raman scattering," *Analytical and Bioanalytical Chemistry*, vol. 407, no. 27, pp. 8253–8261, Jul. 2015.
98. O. Žukovskaja, I. Jahn, K. Weber, D. Cialla-May, and J. Popp, "Detection of Pseudomonas aeruginosa Metabolite Pyocyanin in Water and Saliva by Employing the SERS Technique," *Sensors*, vol. 17, no. 8, p. 1704, 2017.
99. J. Catalán, "The first UV absorption band for indole is not due to two simultaneous orthogonal electronic transitions differing in dipole moment," *Physical Chemistry Chemical Physics*, vol. 17, no. 19, pp. 12515–12520, 2015.
100. J. Huang, P. Liang, J. Xu, Y. Wu, W. Shen, B. Xu, D. Zhang, J. Xia, and S. Zhuang, "Qualitative and quantitative determination of coumarin using surface-enhanced Raman spectroscopy coupled with intelligent multivariate analysis," *RSC Adv.*, vol. 7, no. 77, pp. 49097–49101, 2017.
101. C. Sommer and D. W. Gerlich, "Machine learning in cell biology – teaching computers to recognize phenotypes," *Journal of Cell Science*, vol. 126, no. 24, pp. 5529–5539, 2013.
102. A. L. Tarca, V. J. Carey, X.-W. Chen, R. Romero, and S. Drăghici, "Machine Learning and Its Applications to Biology," *PLoS Computational Biology*, vol. 3, no. 6, 2007.
103. G. G. Dumancas, I. Adrianto, G. Bello, and M. Dozmorov, "Current Developments in Machine Learning Techniques in Biological Data Mining," *Bioinformatics and Biology Insights*, vol. 11, p. 117793221668754, 2017.
104. D. B. Kell, "Metabolomics, modelling and machine learning in systems biology - towards an understanding of the languages of cells. Delivered on 3 July 2005 at the 30th FEBS Congress and 9th IUBMB conference in Budapest," *FEBS Journal*, vol. 273, no. 5, pp. 873–894, 2006.

105. J. Cheng, A. Tegge, and P. Baldi, "Machine Learning Methods for Protein Structure Prediction," *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 41–49, 2008.
106. R. Burbidge, M. Trotter, B. Buxton, and S. Holden, "Drug design by machine learning: support vector machines for pharmaceutical data analysis," *Computers & Chemistry*, vol. 26, no. 1, pp. 5–14, 2001.
107. M. Liu, J. Zhang, E. Adeli, and D. Shen, "Deep Multi-task Multi-channel Learning for Joint Classification and Regression of Brain Status," *Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017*, pp. 3–11, 2017.
108. M. Heinonen, H. Shen, N. Zamboni, and J. Rousu, "Metabolite identification and molecular fingerprint prediction through machine learning," *Bioinformatics*, vol. 28, no. 18, pp. 2333–2341, 2012.
109. S. Webb, "Deep learning for biology," *Nature*, vol. 554, no. 7693, pp. 555–557, 2018.
110. S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717–727, 2000.
111. M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.
112. B. Huang and H. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Transactions on Neural Networks*, vol. 9, no. 1, pp. 224–229, 1998.
113. M. Saerens, "Building cost functions minimizing to some summary statistics," *IEEE Transactions on Neural Networks*, vol. 11, no. 6, pp. 1263–1271, 2000.
114. F. M. Selaru, Y. Xu, J. Yin, T. Zou, T. C. Liu, Y. Mori, J. M. Abraham, F. Sato, S. Wang, C. Twigg, A. Olaru, V. Shustova, A. Leytin, P. Hytiroglou, D. Shibata, N. Harpaz, and S. J. Meltzer, "Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions," *Gastroenterology*, vol. 122, no. 3, pp. 606–613, 2002.
115. J. Khan, J. S. Wei, M. Ringnér, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.
116. N. R. Pal, K. Aguan, A. Sharma, and S.-I. Amari, "Discovering biomarkers from gene expression data for predicting cancer subgroups using neural networks and relational fuzzy clustering," *BMC Bioinformatics*, vol. 8, no. 1, p. 5, 2007.
117. S. Tomida, T. Hanai, N. Koma, Y. Suzuki, T. Kobayashi, and H. Honda, "Artificial Neural Network Predictive Model for Allergic Disease Using Single Nucleotide Polymorphisms Data.," *Journal of Bioscience and Bioengineering*, vol. 93, no. 5, pp. 470–478, 2002.
118. T. B. Hughes and S. J. Swamidass, "Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism," *Chemical Research in Toxicology*, vol. 30, no. 2, pp. 642–656, Feb. 2017.
119. Y.-D. Chen, "Artificial Neural Networks Analysis of Surface-Enhanced Laser Desorption/Ionization Mass Spectra of Serum Protein Pattern Distinguishes Colorectal Cancer from Healthy Population," *Clinical Cancer Research*, vol. 10, no. 24, pp. 8380–8385, 2004.
120. N. H. Othman, K. Y. Lee, A. R. M. Radzol, and W. Mansor, "PCA-SCG-ANN for Detection of Non-structural Protein 1 from SERS Salivary Spectra," *Intelligent Information and Database Systems Lecture Notes in Computer Science*, pp. 424–433, 2017.
121. N. M. Nawi, W. H. Atomi, and M. Rehman, "The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks," *Procedia Technology*, vol. 11, pp. 32–39, 2013.

122. R. R. Picard and K. N. Berk, “Data Splitting,” *The American Statistician*, vol. 44, no. 2, pp. 140–147, 1990.
123. D. Svozil, V. Kvasnicka, and Pospichal Jiří, “Introduction to multi-layer feed-forward neural networks,” *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997.
124. W. Gardner, “Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique,” *Signal Processing*, vol. 6, no. 2, pp. 113–133, 1984.
125. J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
126. P. O. Pinheiro and R. Collobert, “From image-level to pixel-level labeling with Convolutional Networks,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
127. S. Kasera, L. O. Herrmann, J. D. Barrio, J. J. Baumberg, and O. A. Scherman, “Quantitative multiplexing with nano-self-assemblies in SERS,” *Scientific Reports*, vol. 4, no. 1, 2014.
128. S. Walia, A. K. Shah, P. R. Stoddart, M. Bhaskaran, and S. Sriram, “Electric field induced surface-enhanced Raman spectroscopy for multianalyte detection,” *Physical Chemistry Chemical Physics*, vol. 17, no. 11, pp. 7095–7099, 2015.
129. N. C. Shah, O. Lyandres, J. T. Walsh, M. R. Glucksberg, and R. P. V. Duyne, “Lactate and Sequential Lactate−Glucose Sensing Using Surface-Enhanced Raman Spectroscopy,” *Analytical Chemistry*, vol. 79, no. 18, pp. 6927–6932, 2007.
130. A. B. Zrimsek, N. L. Wong, and R. P. V. Duyne, “Single Molecule Surface-Enhanced Raman Spectroscopy: A Critical Analysis of the Bianalyte versus Isotopologue Proof,” *The Journal of Physical Chemistry C*, vol. 120, no. 9, pp. 5133–5142, 2016.
131. P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, “Applications of Deep Learning in Biomedicine,” *Molecular Pharmaceutics*, vol. 13, no. 5, pp. 1445–1454, 2016.