UTILIZING MACHINE LEARNING AND OPTICAL SIMULATIONS FOR SPECTROSCOPIC PROBING OF BIOLOGICAL SYSTEMS

Vikas Yadav

Department of Chemistry

Indian Institute of Technology Delhi

May 2025

UTILIZING MACHINE LEARNING AND OPTICAL SIMULATIONS FOR SPECTROSCOPIC PROBING OF BIOLOGICAL SYSTEMS

By

Vikas Yadav

Department of Chemistry

Submitted

In fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

Indian Institute of Technology Delhi

May 2025

CERTIFICATE

This is to certify that the thesis entitled "**Utilizing machine learning and optical simulations for spectroscopic probing of biological systems**" being submitted by **Mr. Vikas Yadav** to the **Indian Institute of Technology Delhi** for the award of "**Doctor of Philosophy**" is a record of bonafide research work carried out by him. He has worked under my guidance and supervision and has fulfilled the requirements for the submission of the thesis. To the best of my knowledge, the results contained in the thesis have not been submitted, in part or full, to any other university or institute for the award of any degree or diploma.

Date:

Prof. Soumik Siddhanta
Associate Professor
Department of Chemistry
Indian Institute of Technology Delhi
New Delhi, 110016

ACKNOWLEDGMENTS

First and foremost, I would like to express my profound and sincere gratitude to my supervisor, Prof. Soumik Siddhanta, for his unwavering support, insightful guidance, and immense patience throughout my research. He has been a mentor in the truest sense, consistently pushing me to think critically and approach problems from multiple angles, while also providing the freedom and encouragement to explore my ideas independently. His feedback and mentorship have been crucial in helping me develop the skills and confidence necessary to navigate the complexities of my research.

I would also like to extend my heartfelt thanks to my SRC members, Prof. Siddharth Pandey, Prof. Sameer Sapra, and Prof. Jatindra K. Sahu for giving valuable suggestions and improving my research. I would express my gratitude to the heads of the Department of Chemistry, Prof. Anil J. Elias, Prof. Narayanan D Kurur, Prof Siddharth Pandey, and Prof Selvarajan Nagendran and Central Research Facility (CRF) at IIT Delhi for providing an enriching academic environment and access to the research facilities. My gratitude extends to my lab members cum friends, Ruchi, Arti, Aastha, Himanshu, and Soumya, whose collaboration and camaraderie have been essential to my research. The many discussions, debates, and brainstorming sessions in lab meetings, we had have significantly contributed to my understanding of key concepts and methodologies. A special mention goes to Dr. Abhay Kumar Tiwari, who assisted me with the experimental aspects of Raman spectroscopy and provided technical support with the ML integration, which were both integral to the completion of this work. The friendships I have forged within the lab have made this journey not only intellectually stimulating but also personally rewarding. I would particularly like to acknowledge their constant encouragement and insightful suggestions that helped me refine my experimental techniques and data analysis approaches.

I would also like to acknowledge the collaborative efforts and support from other research groups, Prof. Jatinder Kumar Sahu, Prof. Sajesh P. Thomas and Prof. Harish Barshilia, whose expertise in different areas enriched my understanding and provided valuable perspectives. The interdisciplinary collaborations with Prof. Tapas Goswami and Prof. Sushil Kumar from UPES, have been part of have broadened my research horizons and introduced me to new methods and techniques that were pivotal to the success of some projects.

Additionally, I would like to thank the UGC-CSIR fellowship programs that supported my research throughout my PhD. Without their financial support, this research would not have

been possible. My heartfelt appreciation also goes to my friends Ruchi, Arti, Aastha, Sahil, Arvind, Nikhil, Jyoti, Ashish, Anuprash, Ashu, Sandeep, Sudhanshu, Pooja, Heena, Udisha, Nidhi, Anjali and more to go. To my parents and lovely wife “Monika”, who have always been my pillars of strength, thank you for your unconditional love, support, and encouragement. I am deeply grateful to for their understanding, sacrifices, and encouragement, which have allowed me to dedicate myself wholeheartedly to my research.

Finally, I would like to extend my gratitude to all those who, in one way or another, contributed to my research and throughout my time on campus. The journey toward completing this PhD has been filled with highs and lows, moments of doubt, and moments of discovery, and I am deeply indebted to all those who have supported me throughout. This thesis is as much a reflection of my efforts as it is a testament to the collective support of all the individuals mentioned above.

Thank you all for being an integral part of this journey!!!

Vikas Yadav

ABSTRACT

This thesis presents a comprehensive approach to advancing Raman spectroscopy applications through a combination of machine learning (ML) techniques, portable spectrometer innovations, and substrate optimization for enhanced light-matter interactions. Focusing on three primary objectives—harnessing light scattering for biological analysis, developing cost-effective Raman instrumentation, and optimizing absorption for multipurpose applications—this work addresses the key challenges and limitations of Raman spectroscopy in complex real-world applications.

The primary focus is on enhancing Raman scattering for precise biological analysis, targeting the complexity and subtle variations within biological samples. Here, ML-driven Raman spectroscopy is used to detect DNA mutations and analyze cellular components with high sensitivity. Using Raman spectroscopy, we successfully identified changes in the molecular vibrations of cytosine induced by hydroxylamine, a mutagen that converts cytosine to uracil, leading to genetic mutations linked to various diseases. By incorporating ML algorithms, such as artificial neural networks (ANNs), we enhanced the detection process, achieving rapid and accurate classification of mutation states in DNA samples. This method was further extended to the classification of complex cancer cell lines, where the inherent spectral complexity required advanced chemometric techniques, such as Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), to deconvolute overlapping signals from biomolecular components. Coupled with ML-based classification, these techniques revealed distinct biochemical profiles across different cancer types, offering a robust framework for molecular diagnosis and cancer cell differentiation.

The secondary focus of this thesis is to develop low-cost, portable Raman spectrometers that maintain high spectral resolution despite their compact form, addressing the limitations typically associated with affordability and portability. Low-cost portable spectrometers often suffer from low spectral resolution and signal-to-noise ratio (SNR), leading to the loss of critical Raman features. To overcome these issues, Generative Adversarial Networks (GANs) were employed to reconstruct high-resolution Raman spectra from low-resolution data, allowing for more accurate molecular feature recognition. By mapping low-quality spectral inputs to their high-resolution counterparts, GANs enhanced the functionality of portable spectrometers, bridging the gap between affordability and the need for precision. Additionally, a grating-free spectrometer design was proposed, utilizing diffraction patterns on a CCD

detector instead of fragile optical gratings. Through GAN-based spectral reconstruction, this innovative design offers a robust, cost-effective alternative that maintains high spectral fidelity, making high-resolution Raman spectroscopy accessible for applications in clinical diagnostics, environmental monitoring, and on-site molecular identification.

Finally, to delves into optimizing absorption and light-trapping properties in substrates, creating multi-functional platforms that harness light efficiently for various applications. By designing smart optical substrates that leverage plasmonic and light-trapping properties, we achieved significant enhancement in Surface-Enhanced Raman Scattering (SERS) and Plasmonic Circular Dichroism (PCD), as well as optimized light absorption in the UV-Visible region. Initial work focused on enhancing PCD signals using nanostructured gold assemblies to increase sensitivity through SPASER technology, allowing for the amplification of weak optical signals essential for detecting subtle molecular changes. Subsequently, the study explored silicon-based nanostructures, such as silicon nanopillars, which were optimized to achieve high optical absorption, particularly in the visible spectrum. This substrate design is critical for light-trapping applications in solar energy harvesting, where maximizing absorption across multiple wavelengths improves conversion efficiency. Through simulations and experimental validation, various geometries were tested to tailor the optical properties of these substrates, achieving both broad-spectrum absorption and increased scattering efficiency, essential for applications ranging from spectroscopy to energy harvesting.

Overall, this thesis provides a robust framework for advancing Raman spectroscopy, integrating ML, innovative spectrometer designs, and optimized substrates to address current limitations in sensitivity, affordability, and versatility. By enhancing scattering and absorption processes through integrated design and advanced computational methods, this work opens new possibilities in molecular diagnostics, pharmaceutical analysis, energy harvesting, and biomedical applications, contributing to the development of versatile, high-performance optical platforms.

सारांश

यह शोध मशीन लर्निंग (एमएल) तकनीकों, पोर्टेबल स्पेक्ट्रोमीटर नवाचारों और बेहतर प्रकाश-पदार्थ इंटरैक्शन के लिए सब्सट्रेट अनुकूलन के संयोजन के माध्यम से रमन स्पेक्ट्रोस्कोपी अनुप्रयोगों को आगे बढ़ाने के लिए एक व्यापक दृष्टिकोण प्रस्तुत करता है। तीन प्राथमिक उद्देश्यों पर ध्यान केंद्रित करते हुए - जैविक विश्लेषण के लिए प्रकाश बिखराव का उपयोग करना, लागत प्रभावी रमन उपकरण विकसित करना और बहुउद्देशीय अनुप्रयोगों के लिए अवशोषण को अनुकूलित करना - यह कार्य जटिल वास्तविक दुनिया के अनुप्रयोगों में रमन स्पेक्ट्रोस्कोपी की प्रमुख चुनौतियों और सीमाओं को संबोधित करता है।

प्राथमिक ध्यान सटीक जैविक विश्लेषण के लिए रमन बिखराव को बढ़ाने पर है, जैविक नमूनों के भीतर जटिलता और सूक्ष्म विविधताओं को लक्षित करना। यहाँ, डीएनए उत्परिवर्तन का पता लगाने और उच्च संवेदनशीलता के साथ सेलुलर घटकों का विश्लेषण करने के लिए एमएल-संचालित रमन स्पेक्ट्रोस्कोपी का उपयोग किया जाता है। रमन स्पेक्ट्रोस्कोपी का उपयोग करके, हमने हाइड्रॉक्सिलमाइन द्वारा प्रेरित साइटोसिन के आणविक कंपन में परिवर्तनों की सफलतापूर्वक पहचान की, एक उत्परिवर्तजन जो साइटोसिन को यूरैसिल में परिवर्तित करता है, जिससे विभिन्न रोगों से जुड़े आनुवंशिक उत्परिवर्तन होते हैं। कृत्रिम तंत्रिका नेटवर्क (एएनएन) जैसे एमएल एल्गोरिदम को शामिल करके, हमने डीएनए नमूनों में उत्परिवर्तन स्थितियों के तेज़ और सटीक वर्गीकरण को प्राप्त करते हुए पता लगाने की प्रक्रिया को बढ़ाया। इस पद्धति को जटिल कैंसर कोशिका रेखाओं के वर्गीकरण के लिए आगे बढ़ाया गया, जहां अंतर्निहित वर्णक्रमीय जटिलता के लिए उन्नत कीमोमेट्रिक तकनीकों की आवश्यकता थी, जैसे कि मल्टीवेरिएट कर्व रेज़ोल्यूशन-अल्टरनेटिंग लीस्ट स्क्वेयर्स (एमसीआर-एएलएस), जैव-आणविक घटकों से ओवरलैपिंग संकेतों को विघटित करने के लिए। एमएल-आधारित वर्गीकरण के साथ, इन तकनीकों ने विभिन्न कैंसर प्रकारों में अलग-अलग जैव रासायनिक प्रोफाइल का खुलासा किया, जो आणविक निदान और कैंसर कोशिका भेदभाव के लिए एक मजबूत ढांचा प्रदान करता है।

इस थीसिस का द्वितीयक फोकस कम लागत वाले, पोर्टेबल रमन स्पेक्ट्रोमीटर विकसित करना है जो अपने कॉम्पैक्ट रूप के बावजूद उच्च स्पेक्ट्रल रिज़ॉल्यूशन बनाए रखते हैं, जो आमतौर पर सामर्थ्य और पोर्टेबिलिटी से जुड़ी सीमाओं को संबोधित करते हैं। कम लागत वाले पोर्टेबल स्पेक्ट्रोमीटर अक्सर कम स्पेक्ट्रल रिज़ॉल्यूशन और सिग्नल-टू-शोर अनुपात (एसएनआर) से ग्रस्त होते कम गुणवत्ता वाले स्पेक्ट्रल इनपुट को उनके उच्च-रिज़ॉल्यूशन समकक्षों के साथ मैप करके, GAN ने पोर्टेबल स्पेक्ट्रोमीटर की कार्यक्षमता को बढ़ाया, जिससे

वहनीयता और सटीकता की आवश्यकता के बीच की खाई को पाटा जा सका। इसके अतिरिक्त, एक झंझरी-मुक्त स्पेक्ट्रोमीटर डिज़ाइन प्रस्तावित किया गया था, जो नाजुक ऑप्टिकल झंझरी के बजाय CCD डिटेक्टर पर विवर्तन पैटर्न का उपयोग करता है। GAN-आधारित स्पेक्ट्रल पुनर्निर्माण के माध्यम से, यह अभिनव डिज़ाइन एक मजबूत, लागत प्रभावी विकल्प प्रदान करता है जो उच्च स्पेक्ट्रल निष्ठा बनाए रखता है, जिससे नैदानिक निदान, पर्यावरण निगरानी और ऑन-साइट आणविक पहचान में अनुप्रयोगों के लिए उच्च-रिज़ॉल्यूशन रमन स्पेक्ट्रोस्कोपी सुलभ हो जाती है।

अंत में, सब्सट्रेट में अवशोषण और प्रकाश-फँसाने वाले गुणों को अनुकूलित करने के लिए, बहु-कार्यात्मक प्लेटफ़ॉर्म बनाना जो विभिन्न अनुप्रयोगों के लिए प्रकाश का कुशलतापूर्वक उपयोग करते हैं। प्लास्मोनिक और लाइट-ट्रैपिंग गुणों का लाभ उठाने वाले स्मार्ट ऑप्टिकल सब्सट्रेट को डिज़ाइन करके, हमने सरफेस-एन्हांस्ड रमन स्कैटरिंग (SERS) और प्लास्मोनिक सर्कुलर डाइक्रोइज़्म (PCD) में महत्वपूर्ण वृद्धि हासिल की, साथ ही UV-विज़िबल क्षेत्र में अनुकूलित प्रकाश अवशोषण भी हासिल किया। प्रारंभिक कार्य नैनोस्ट्रक्चर्ड गोल्ड असेंबलियों का उपयोग करके पीसीडी संकेतों को बढ़ाने पर केंद्रित था ताकि एसपीएएसईआर तकनीक के माध्यम से संवेदनशीलता बढ़ाई जा सके, जिससे सूक्ष्म आणविक परिवर्तनों का पता लगाने के लिए आवश्यक कमजोर ऑप्टिकल संकेतों के प्रवर्धन की अनुमति मिल सके। इसके बाद, अध्ययन ने सिलिकॉन-आधारित नैनोस्ट्रक्चर, जैसे सिलिकॉन नैनोपिलर की खोज की, जिन्हें विशेष रूप से दृश्यमान स्पेक्ट्रम में उच्च ऑप्टिकल अवशोषण प्राप्त करने के लिए अनुकूलित किया गया था। यह सब्सट्रेट डिज़ाइन सौर ऊर्जा संचयन में प्रकाश-फँसाने वाले अनुप्रयोगों के लिए महत्वपूर्ण है, जहाँ कई तरंग दैर्ध्य में अवशोषण को अधिकतम करने से रूपांतरण दक्षता में सुधार होता है। सिमुलेशन और प्रयोगात्मक सत्यापन के माध्यम से, इन सब्सट्रेट्स के ऑप्टिकल गुणों को तैयार करने के लिए विभिन्न ज्यामिति का परीक्षण किया गया, एकीकृत डिजाइन और उन्नत कम्प्यूटेशनल विधियों के माध्यम से प्रकीर्णन और अवशोषण प्रक्रियाओं को बढ़ाकर, यह कार्य आणविक निदान, फार्मास्युटिकल विश्लेषण, ऊर्जा संचयन और जैव चिकित्सा अनुप्रयोगों में नई संभावनाओं को खोलता है, तथा बहुमुखी, उच्च-प्रदर्शन ऑप्टिकल प्लेटफार्मों के विकास में योगदान देता है।

कुल मिलाकर, यह थीसिस रमन स्पेक्ट्रोस्कोपी को आगे बढ़ाने, एमएल, अभिनव स्पेक्ट्रोमीटर डिजाइनों और अनुकूलित सब्सट्रेट्स को एकीकृत करने के लिए एक मजबूत ढांचा प्रदान करती है ताकि संवेदनशीलता, सामर्थ्य और बहुमुखी प्रतिभा में वर्तमान सीमाओं को संबोधित किया जा सके। एकीकृत डिजाइन और उन्नत कम्प्यूटेशनल

विधियों के माध्यम से बिखराव और अवशोषण प्रक्रियाओं को बढ़ाकर, यह कार्य आणविक निदान, दवा विश्लेषण, ऊर्जा संचयन और जैव चिकित्सा अनुप्रयोगों में नई संभावनाओं को खोलता है, जो बहुमुखी, उच्च-प्रदर्शन ऑप्टिकल प्लेटफ़ॉर्म के विकास में योगदान देता है।

TABLE OF CONTENTS

CERTIFICATE I

ACKNOWLEDGEMENTS II

ABSTRACT (English) IV

ABSTRACT (Hindi) VI

LIST OF FIGURES XV

LIST OF TABLES XXX

ABBREVATIONS XXXI

Chapter 1: Introduction 1

1.1 Types of Light-Matter Interactions 1

1.2 Scattering: A Detailed Exploration 4

1.2.1 Rayleigh Scattering 4

1.2.2 Raman Scattering 4

1.3 Raman Spectroscopy 6

1.3.1 Raman Scattering Cross-Section 6

1.3.2 Photon-Molecule Interaction and Energy Shift 8

1.3.3 Polarizability Tensor and Induced Dipole Moment in Terms of Classical Mechanics 8

1.3.4 Selection Rules for Raman Scattering 10

1.3.5 Significance of Raman Spectroscopy Across Different Domains 11

1.3.6 The challenges of weak Raman scattering 12

1.4 Methods to enhance weak Raman scattering 13

1.4.1 Experimental-based methods 13

1.4.2 Chemometrics and ML approaches 14

1.5 Chemometrics and ML for better insights and interpretations 16

1.6 Principal Component Analysis (PCA) 17

1.6.1 The Mathematical Basis of PCA 17

1.6.2 Applications of PCA in Raman Spectroscopy 20

1.6.3 Advantages and Limitations of PCA 21

1.7 Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) 21

1.7.1 The Mathematical Basis of MCR-ALS 21

1.7.2 Constraints in MCR-ALS 24

1.7.3 Applications of MCR-ALS in Raman Spectroscopy 27

1.7.4 Advantages and Limitations of MCR-ALS 28

1.8 Other Chemometric Tools for Raman Spectral Analysis 28

1.9 Limitations of Chemometric Tools 28

1.9.1 Assumptions About Data Structure 29

1.9.2 Sensitivity to Noise and Outliers 29

1.9.3 Lack of Interpretability 29

1.10 The Emergence of ML and DL in Raman Spectroscopy 30

1.10.1 Introduction to ML 31

1.11 Artificial Neural Networks (ANNs) 32

1.11.1 Structure of ANNs 32

1.11.2 ANNs in Raman Spectral Analysis 33

1.12 Convolutional Neural Networks (CNNs) 35

1.12.1 Detailed Network Architecture of CNNs 36

1.12.2 Activation Functions in CNNs 39

1.12.3 Applications of CNNs in Chemistry 42

1.13 Generative Adversarial Networks (GANs) 43

1.13.1 Introduction to GAN 44

1.13.2 Detailed GAN Architecture 45

1.13.3 Training Process of GANs 47

1.13.4 Advanced Variants of GANs 47

1.13.5 Applications of GANs in Chemistry 48

1.14 Major Objectives 50

Chapter 2: Machine learning-assisted Raman scattering for detecting DNA mutations 54

2.1 Introduction 55

2.2 Materials and Methods 57

2.2.1 Materials 57

2.2.2 Chemical mutation of the DNA and primer samples 57

2.2.3 Raman spectroscopy of the DNA samples 58

2.2.4 Raman Data Analysis 58

2.3 Results and Discussion 61

2.3.1 Raman spectroscopy of the DNA primers and genomic DNA 61

2.3.2 Induction of mutation in DNA primers and genomic DNA 63

2.3.3 Quantitative analysis of mutagenic reaction on single-stranded DNA primer 66

2.3.4 Quantitative analysis of mutagenic reaction on double-stranded DNA 69

2.3.5 Predicting the number of bases and mutated bases from Raman spectra 72

2.4 Conclusion 74

Chapter 3: Machine learning-assisted Spatial omics and virtual staining for cellular component analysis using Raman scattering 76

3.1 Introduction 77

3.2 Materials and Methods 80

3.2.1 Reagents 80

3.2.2 Cell culture and fixation 80

3.2.3 Raman spectral Acquisitions from cells 80

3.2.4 Pre-processing of Raman spectroscopic data and MCR-based analysis 81

3.2.5 ANN model for the classification of cell lines 82

3.3 Results and Discussion 83

3.3.1 Raman spectroscopic imaging of cancer cells 83

3.3.2 PCA of the hyperspectral Raman imaging 86

3.3.3. MCR analysis of hyperspectral Raman Imaging for spatial profiling 87

3.3.4 Quantitative analysis of the pure cell components from MCR analysis 91

3.3.5 ANN Classification Model for differentiating between cell lines 94

3.4 Conclusion 97

Chapter 4: GAN-driven high-resolution Raman spectral generation for accurate molecular feature recognition 98

4.1 Introduction 99

4.2 Materials and Methods 102

4.2.1 Materials 102

4.2.2 Raman Spectral Acquisition 102

4.2.3 Model building for GAN 103

4.2.4 Classification of GAN-generated spectra using ANN 105

4.2.5 Identification of the generated spectra of organic compounds via spectral barcoding 107

4.3 Results and Discussion 109

4.3.1 Collection and distinction of high- and low-resolution Raman spectra 109

4.3.2 Leveraging GAN for generating high-resolution Raman spectra 114

4.3.3 Identification of pharmaceutical compounds through spectral barcoding 123

4.4 Conclusion 128

Chapter 5: GAN-based reconstruction of Raman spectra from grating-free diffraction patterns 129

5.1 Introduction 130

5.2 Materials and Methods 132

5.2.1 Materials 132

5.2.2 Raman Spectral Acquisition 132

5.2.3 Data Preprocessing 133

5.2.4 Raman Spectra to Image conversion 133

5.2.5 GAN Model Architecture 135

5.3 Results and Discussion 137

5.3.1 Raman spectral analysis 139

5.3.2 GAN model for the diffraction to fully-resolved Raman spectra 143

5.4 Conclusion 145

Chapter 6: Designing and optimizing plasmonic substrate for light amplification 146

6.1 Introduction 147

6.2 Materials and Methods 148

6.2.1 Simulation parameters 148

6.2.2 Structural parameters 149

6.3 Results and Discussion 149

6.3.1 Extrinsic and intrinsic chiralities of the gold nanorod dimers 149

6.3.2 Chiral amplification in gain-assisted gold nanorod dimers 154

6.3.3 Tuning of chirality enhancement of the gold nanorod dimer 159

6.4 Conclusion 163

Chapter 7: Designing and optimizing smart optical substrate for light trapping applications 164

7.1 Introduction 165

7.2 Materials and Methods 167

7.2.1 Materials 167

7.3 Results and Discussion 169

7.3.1 Optical traits of silicon nanopillars: absorption & reflection 170

7.3.2 Impact of geometrical variations on optical behavior 172

7.3.3 Polarization-dependent optical response of silicon nanopillars 174

7.3.4 Mapping electric field distribution in silicon nanopillars 176

7.3.5 Implications for solar and spectroscopic applications: future perspective 179

7.4 Conclusion 180

Chapter 8: Outlook and Future Work 182

Appendix 184

References 201

Biodata 235

Publications 236

LIST OF FIGURES

Figure 1.1: Light interacting with a sample can undergo processes like scattering, absorption, and reflection, each revealing unique molecular and structural information. By applying these phenomena in spectroscopic techniques, we gain insights into molecular composition, vibrational modes, and chemical bonding within the sample. 2

Figure 1.2: When light scatters from a sample, it produces both Rayleigh and Raman scattering (Stokes and anti-Stokes). Unlike Rayleigh scattering, which is elastic, Raman scattering provides information on the vibrational modes of the sample, making it a valuable tool for determining molecular structure and composition. 5

Figure 1.3: The illustration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) frameworks, highlighting their hierarchical relationship. 30

Figure 1.4: The chronological progression of key milestones in the field of machine learning. 31

Figure 1.5: Demonstration of how kernels operate in CNNs. The figure illustrates the process of applying different kernels to input data (such as an image), highlighting how feature extraction occurs through convolution, resulting in feature maps that capture essential patterns and structures within the data. 36

Figure 1.6: Visualization of how the stride parameter works in CNNs. The figure demonstrates the effect of varying stride values on the movement of the convolutional filter across the input image, showing how larger strides reduce the output size and computational complexity while capturing essential features. 37

Figure 1.7: The figure demonstrates the process of adding extra pixels around the input image to preserve spatial dimensions after convolution operations, ensuring that important features at the edges of the image are not lost and helping to maintain the output size. 38

Figure 1.8: Illustration of how the pooling layer functions in CNNs. The figure shows the process of down-sampling input feature maps through operations like max pooling or average pooling, reducing spatial dimensions while retaining important features, thus improving computational efficiency and model robustness. 38

Figure 1.9: Graphical representation of various activation functions commonly used in ML models (A) ReLU, (B) Leaky ReLU, (C) ELU, (D) Softmax, (E) TanH, and (F) Sigmoid. 40

Figure 1.10: The figure illustrates the interplay between the generator and discriminator, where the generator creates synthetic data, and the discriminator evaluates its authenticity. Through iterative training, both components improve, leading to the generation of high-quality data that closely resembles the real dataset. 45

Figure 2.1: (A-D) Averaged Raman spectra of 32 different primers collected using a 785 nm laser source. The standard deviations in the spectra are shown as shadow region. The primers are divided into four random sets having 8 primers in each set. The shaded region (purple) is the area of interest that contains distinct Raman spectral features of DNA bases, adenine, guanine, thymine, cyto- sine, and uracil. (E-H) The RadViz plot corresponds to the DNA primers from A-H respectively. The color code is similar as represented in A-H. 60

Figure 2.2: Boxplot corresponding to 32 different primers. The primers are divided into four random sets, having 8 primers in each set. The boxplot shows the spectral variations in Raman spectra from each class of primer. The Raman spectra from all the primers were subjected to PCA and the PCA scores were utilized for the spectral variation among the different classes and within the class also. The PCA scores were plotted on the y-axis. 62

Figure 2.3: (A) The schematic of the steps involved in DNA mutation reaction and (B) the mutation reaction mechanism. The HA first attack at the amine group of the cytosine ring to form an intermediate species (I & II), typically hydroxylaminocytosine, through nucleophilic substitution. This intermediate is highly unstable and, upon hydrolysis and subsequent deamination, is transformed into uracil. (C) The Raman spectra acquired from the DNA primer (black) and its mutant (red) were first normalized with respect to the phosphate band at 1089 cm-1 and relative change in the cytosine and uracil base pair were analyzed further. (D) The spectral region from 700-860 cm-1 contains the valuable information about the differential Raman band of guanine (685 cm-1), adenine (730 cm-1), uracil (750 cm-1) and cytosine (780 cm-1). The Raman band of cytosine decreased by nearly 50% after the mutagenic reaction with HA and a subsequent rise in the Raman band of uracil. The peak ratio was first calculated with respect to the normalized phosphate band peak. 64

Figure 2.4: (A) UV-Visible absorbance spectra of DNA primer (F1) without mutation (black) and with mutation (red) showing a blue shift with increase in the absorbance intensity. (B) time-

dependent absorbance spectra of DNA primer show first an increase in the absorbance intensity, then a decrease with time. (C) Absorbance maxima intensity plot for time-dependent absorbance of DNA primer. (D) UV-Visible absorbance spectra of DNA herring sperm without mutation (black) and with mutation (red) showing a blue shift with increase in the absorbance intensity. (E) time-dependent absorbance spectra of DNA herring sperm show increase in the absorbance intensity with time. (F) Absorbance maxima intensity plot for time-dependent absorbance of DNA from herring sperm. 65

Figure 2.5: The reaction mechanism of mutagenic action of HA on cytosine. 66

Figure 2.6: The scatter plot from the PCA component were projected on 2D plane to distinguish between the two class of samples, without mutation (black) and with mutation (red) for (A) DNA primer and (C) for double-stranded DNA. From the scatter plot, the positive and negative peaks are assigned in loadings plot (B) for DNA primer and (D) for double-stranded DNA. The Raman spectral band at 750 cm-1 and 781 cm-1 were analyzed, which shows the cytosine is decreased after post mutation while the uracil is formed. Also, in the comparison between the DNA primer and DNA, the amount of mutation is more prominent in the DNA primer rather than the double-stranded DNA. The similar change is also reflected in the loading plot at the highlighted peaks. 67

Figure 2.7: The RadViz plot from the PCA scores were projected on 2D plane to show the clustering among (A) single-stranded DNA primer, F1 (red) and its mutant (blue). (B) double-stranded genomic DNA herring sperm (red) and its mutant (blue). Each dot on the 2D plane represent one Raman spectra. The spread of points with the class shows the spectral variations and the clustering separates the mutant version from non-mutant sample. 69

Figure 2.8: (A) The Raman spectra acquired from the double-stranded DNA (black) and its mutant (red) were first normalized with respect to the phosphate band at 1089 cm^{-1} and relative change in the cytosine and uracil base pair were analyzed further. (B) The spectral region from 700-860 cm^{-1} contains the valuable information about the differential Raman band of guanine (685 cm^{-1}), adenine (730 cm^{-1}), uracil (750 cm^{-1}) and cytosine (781 cm^{-1}). The Raman band of cytosine decreased by nearly 25% after the mutagenic reaction with HA and a subsequent rise in the Raman band of uracil. The peak ratio was first calculated with respect to the normalized phosphate band peak. 70

Figure 2.9: The peak at 1089 cm^{-1} was normalized before and after mutation and the peak ratio at 1089 cm^{-1} (phosphate) and 781 cm^{-1} (cytosine) and 1089 cm^{-1} and 750 cm^{-1} (uracil) were calculated. (A) The peak ratio for each of the Raman spectra (32 spectra) from single-stranded DNA primer (blue bar) shows an average of 50.102%. The peak ratio for each of the Raman spectra (32 spectra) from double-stranded DNA (green bar) shows an average of 25.310%. (B) The average number of cytosine in single-stranded DNA primer were predicted by the ANN model (black curve) and the actual number of cytosine (red curve) shows a close resemblance between real values and predicted values. 71

Figure 2.10: (A) The schematic shows the regression and classification of sample and its mutant. The PCA reduced components (35 component) were incorporated into the ANN model to find out the number of cytosine bases before and after the mutation. The model shows an overall accuracy of nearly 88% for the regression task. The input data was divided into three parts, and corresponding regression accuracy is shown (B) training, (C) testing, (D) validation, and (E) overall accuracy. (F) The number of cytosine is calculated for DNA primer as well as double-stranded DNA herring sperm shows a decrease of nearly 50 % cytosine in DNA primer and 25% decrease in the DNA. (G) The decrease is also calculated experimentally from the peak ratio and is found to be around 50.102 % for DNA primer and 25.310% for the double stranded-DNA. The prediction from the ANN model effectively matches the prediction from the experimentally calculated values closely. 72

Figure 2.11: The architecture of the ANN model consists of an input layer, three hidden layers with 31 neurons each followed by an output layer of size one. The dimensions of Raman spectra were first reduced to 35 using PCA before incorporating it to ANN model. The output of the model gives the number of cytosine bases present in the sample. 73

Figure 2.12: The error bar plot from the ANN model shows how the output from the model were differ from the actual target values. Lower error shows the more accuracy of the model for the regression. The model was trained on the single-stranded DNA primers to find out the number of cytosine base present in a particular sample and later the trained model was utilized for the calculation of cytosine in the mutated sample. 74

Figure 3.1: (A) The Raman spectra acquired from the three cancer cell lines; AGS, A549, and CAL33 collected with an accumulation time of 15 seconds and three accumulations per spectrum. (B) The RadViz plot generated from PCA scores, highlighting the clustering and distribution of Raman spectra for the three cell lines. Each dot represents an individual Raman

spectrum, and its position reflects the correlation with a particular cell line. The clustering patterns indicate clear separability between the classes, while the spread of points within each cluster corresponds to the spectral variability within the respective sample class. 83

Figure 3.2: The MCR-ALS deconvolute the raw Raman data into pure spectral profiles and the concentration matrix is utilized for the spatial distribution over the cell providing a comprehensive visualization of cellular biochemical composition. The left panel presents spatial maps of the resolved pure components, revealing their precise localization within the A549 cancer cells. Each map highlights the relative abundance of specific components, enabling the identification of distinct biochemical regions, such as the nucleus, endoplasmic reticulum (ER), mitochondria, cell membrane (CM), and cytoplasmic (CP) compartments. The right column displays the deconvoluted Raman spectra associated with these pure components, providing molecular fingerprints for key biochemical entities. The white light (WL) visualization of the cell and corresponding virtual straining (VS) reconstructions derived from the MCR concentration matrix, showcasing the spatial distribution of the resolved components. 85

Figure 3.3: The MCR-ALS deconvolute the raw Raman data into pure spectral profiles and the concentration matrix is utilized for the spatial distribution over the cell providing a comprehensive visualization of cellular biochemical composition. The left panel presents spatial maps of the resolved pure components, revealing their precise localization within the AGS (top) and CAL33 (bottom) cancer cells. Each map highlights the relative abundance of specific components, enabling the identification of distinct biochemical regions, such as the nucleus, endoplasmic reticulum (ER), mitochondria, cell membrane (CM), and cytoplasmic (CP) compartments. The right column displays the deconvoluted Raman spectra associated with these pure components, providing molecular fingerprints for key biochemical entities. The white light (WL) visualization of the cell and corresponding virtual straining (VS) reconstructions derived from the MCR concentration matrix, showcasing the spatial distribution of the resolved components. 89

Figure 3.4: The quantitative distribution of pure cellular components in the AGS, A549, and CAL33 cancer cell lines, derived from the deconvolution of raw Raman spectra into their pure biochemical constituents using MCR-ALS. The top panel (A) highlights the relative abundances of collagen, lipids, and nucleic acids and the below panel (B) highlights the relative

abundances of glycogen, DNA, and uric acid, providing insights into structural and metabolic variations among the cell lines. 93

Figure 3.5: The basic architecture of the ANN model, consists of an input layer followed by one hidden layer with 11 neurons and an output classification layer in the end. The dimensions of the raw Raman spectra is reduced to 45 and incorporated as input to the ANN classification model to cut down the computationally training cost and taken into account only those features which contains the maximum information. The output classification layer have a size of three that corresponds to the total class of the samples. 95

Figure 3.6: (A) The confusion matrix for the ANN classification model, demonstrating an overall accuracy of approximately 99% in differentiating the three cancer cell lines: AGS (1), A549 (2), and CAL33 (3). (B) The ROC curve represents the ANN model's performance in classifying the three cancer cell lines: CAL33, AGS, and A549. The high area under the curve (AUC) values for all classes reflect the model’s robustness and effectiveness in capturing subtle spectral differences, even in noisy Raman datasets. 96

Figure 4.1: The low-resolution (left) and high-resolution (right) 400 Raman spectra each of different samples taken in solid form using a 785 nm laser source. 105

Figure 4.2: The workflow of the GAN mediated transformation of low-resolution to high-resolution Raman spectra. The process involves paired dataset (low- and high-resolution) to train the GAN model and after training, the trained model was tested on the unknown low-resolution dataset to generate the high-resolution counterparts. Inset shows the schematic representation of the high- and low-resolution Raman spectrometer. The fiber optics is for excitation and the surroundings are for collection of the scattered signal from the sample. The labelled components are, long pass filter (LP), band pass filter (BP), mirror (M), and beam splitter (BS). 109

Figure 4.3: Average Raman spectra of different samples accumulated using (A) high resolution and (B) low resolution spectrophotometer. The RadViz plot of the different class of samples using (C) high (D) low-resolution spectrometer. The greater spread represents the more fluctuation in the spectra. (E) Box plot of the different samples using low (left) and high (right) spectrometer. Aspirin (red), ibuprofen (blue), PCM (green), ranitidine (grey), 4-MBA (cyan), and 4-NTP (magenta). 110

Figure 4.4: The additional low-resolution Raman spectra (10 spectra from each sample class) (A) Aspirin, (C) Ibuprofen, and (E) PCM, where the highlighted part in the low-resolution spectra shows the spectral differences (similarity index of 23.96%, 21.01%, and 23.80%, respectively) within the same drug sample under identical conditions. This low-resolution Raman spectra (A, C, and E) was fed to the trained GAN model to get the high-resolution instance. The highlighted part shows that instead of the spectral differences in low-resolution profile, the model capable to rebuild these lost features and generate the high-resolution counterpart (B) Aspirin, (D) Ibuprofen, and (F) PCM (similarity index of 96.70%, 94.31%, and 96.30%, respectively). (G) The violin plot shows the variation of similarity index in the additional low-resolution and GAN-generated high resolution Raman spectra shows less than 25% similarity between low-resolution profile while the GAN-generated instances shows more than 90% similarity index. (H) The scatter plot from the PCA scores shows the highly random fluctuation (more spread in data points) in the low-resolution profile and minimum in the GAN-generated high-resolution Raman spectra. 111

Figure 4.5: The additional low-resolution Raman spectra (10 spectra from each sample class) (A) Ranitidine, (C) 4-MBA, and (E) 4-NTP, where the highlighted part in the low-resolution spectra shows the spectral differences (similarity index of 15.39%, 22.21%, and 24.05%, respectively) within the same drug sample under identical conditions. This low-resolution Raman spectra (A, C, and E) was fed to the trained GAN model to get the high-resolution instance. The highlighted part shows that instead of the spectral differences in low-resolution profile, the model capable to rebuild these lost features and generate the high-resolution counterpart (B) Ranitidine, (D) 4-MBA, and (F) 4-NTP (similarity index of 96.45%, 93.13%, and 99.57%, respectively). (G) The violin plot shows the variation of similarity index in the additional low-resolution and GAN-generated high resolution Raman spectra shows less than 25% similarity between low-resolution profile while the GAN-generated instances shows more than 90% similarity index. (H) The scatter plot from the PCA scores shows the highly random fluctuation (more spread in data points) in the low-resolution profile and minimum in the GAN-generated high-resolution Raman spectra. 112

Figure 4.6: The schematic of the GAN model consists of two primary components: a generator and a discriminator. The generator is tasked with producing high-resolution spectra from low-resolution inputs, aiming to generate outputs that closely resemble realistic high-resolution spectra. Conversely, the discriminator's role is to differentiate between the output generated by the generator and real high-resolution spectra. The low-resolution Raman spectra is feed to the

generator, after every epoch the generated high-resolution Raman spectra was compared with the real high-resolution counterpart by the discriminator. The discriminator tested the capabilities of generator at every epoch, results in the generation of closely resembled high resolution output by the generator after each epoch. 114

Figure 4.7: Raman spectra from low-resolution (black), high-resolution (red), and GAN generated (blue) of different pharmaceutical and drug molecules, (A) aspirin, (B) ibuprofen, (C) PCM, (D) ranitidine, (E) 4-MBA, and (F) 4-NTP. 400 spectra were collected from each sample. 115

Figure 4.8: The architecture of the ANN classification network having highest accuracy and minimum test loss (model 1) is consists of one input layer, two hidden layers with 17 and 18 neurons respectively, followed by the classification layer in the end. 116

Figure 4.9: Confusion matrices showing the accuracy of the ANN classification network (model 1) for the additional dataset, i.e., GAN-generated high-resolution Raman data with a classification accuracy of 97.5%. The representation for different samples is as follows: (1) aspirin, (2) ibuprofen, (3) PCM, (4) ranitidine, (5) 4-MBA, and (6) 4-NTP. 117

Figure 4.10: (A) Confusion matrices showing the accuracy of the ANN classification network (model 2) for the additional dataset, i.e., GAN-generated high-resolution Raman data with a classification accuracy of 97.4%. (B) ROC of the ANN classification network (model 2) for GAN-generated high-resolution Raman data. The closer the curve is to the one, the more accurate the results will be for classification. The representation for different samples is as follows: (1) aspirin, (2) ibuprofen, (3) PCM, (4) ranitidine, (5) 4-MBA, and (6) 4-NTP. 118

Figure 4.11: (A) Confusion matrices showing the accuracy of the ANN classification network (model 3) for the additional dataset, i.e., GAN-generated high-resolution Raman data with a classification accuracy of 97.3%. (B) ROC of the ANN classification network (model 3) for GAN-generated high-resolution Raman data. The closer the curve is to the one, the more accurate the results will be for classification. The representation for different samples is as follows: (1) aspirin, (2) ibuprofen, (3) PCM, (4) ranitidine, (5) 4-MBA, and (6) 4-NTP. 118

Figure 4.12: The GAN model was trained, validated, and tested using the 2400 low-resolution Raman spectra from 6 different class of samples (random splitting of dataset). The trained model for further tested on the additional low-resolution dataset (never fed to model) to check for the model accuracy to transform the low-resolution instance to a high-resolution Raman

spectrum. The trained GAN model transformed the low-resolution to high resolution in completely unsupervised way. (A) The low-resolution Raman spectra of the ibuprofen drug, where the highlighted part in the low-resolution spectra shows the spectral differences (similarity index of 21% only) within the same sample under identical conditions. (B) This low-resolution Raman spectra (A) was fed to the trained GAN model to get the high-resolution instance. The highlighted part shows that instead of the spectra differences in low-resolution profile, the model capable to rebuild these lost features and generate the high-resolution counterpart (similarity index of 94%). (C) The violin plot shows the variation of similarity index in the additional low-resolution (10 more samples) with similarity index of around 21% and GAN-generated high resolution Raman spectra with similarity index of 94%. (D) The scatter plot from the PCA scores shows the highly random fluctuation (more spread in data points) in the low-resolution profile and minimum in the GAN-generated high-resolution Raman spectra. 119

Figure 4.13: The GAN model was trained using the sample-out method. We have total of 6 different class of samples. We trained the GAN model using multiple sample class and test for the similarity index for the unknown class (sample-out) of the sample (ibuprofen here). As we trained the model with more and more class, the similarity index for the GAN-generated high-resolution Raman spectra of unknown class to real high- resolution instance is increasing in a non-linear way. The model was trained in an unsupervised way, and the GAN-generated Raman spectra began to replicate high-resolution features when the model was trained on more than three distinct sample classes. The GAN-generated high-resolution Raman spectra is compared against the real high-resolution Raman spectra (bottom one). 120

Figure 4.14: (A) SNR calculation of different pharmaceutical and organic molecules using high-resolution spectrometer (square), low-resolution spectrometer (circle), and GAN-generated (triangle) results. (B) ROC curve for the classification of GAN-generated high-resolution Raman spectra for different samples. The closer the curve to one, more accurate will be the classification. 121

Figure 4.15: The generated high resolution Raman spectra from low resolution through GAN model was converted into Raman barcode. These barcodes were employed as inputs to the model, which performed spectral comparison against a spectral library. The model output provided identification of the unknown sample based on spectral matching. The ANN model was evaluated using coefficient of determination (R^2), mean absolute error (MAE), and root

mean square error (RMSE). The spectral similarity was calculated using similarity index (SSIM). 124

Figure 4.16: The process of making a Raman barcode. (A) the Raman spectra of the 4-NTP molecule. (B) using the "findpeaks" function to extract the peak location and FWHM and incorporate that information in the barcode. (C) The Raman barcode of the 4-NTP molecule. 126

Figure 4.17: The RadViz clustering between different class of samples (A) high, (B) low-resolution, and (C) GAN-generated high-resolution Raman spectra showing different clustering efficiency. The clustering efficiency of GAN-generated high-resolution Raman spectra is significantly improved over its low-resolution counterpart. Aspirin (red), ibuprofen (blue), PCM (green), ranitidine (grey), 4-MBA (cyan), and 4-NTP (magenta). 127

Figure 5.1: The diffraction pattern formed on the CCD detector of size 255x1024 form various samples. The slit size is set to 10 μm resulting in a small portion of the CCD to exposed. After that cropping down the non-information part and adding an extra dummy layer of size 1x16 so that the final size after cropping would be 256x16. Now, reshaping the cropped image to a square based image of size 64x64. Since for each sample class we have 65 data, combining all these images and finally formed *dataset A*. 134

Figure 5.2: The Raman spectra of various samples using 600 lines/mm grating. The Raman spectra was converted into a 64x64 Raman image. Since for each sample class we have 65 data, combining all these images and finally formed *dataset B*. 135

Figure 5.3: The basic architecture of the GAN model, which consist of a generator and a discriminator. The role of the generator is to take diffraction image without grating (dataset A) as an input and map it to the corresponding Raman image with grating (dataset B). While role of the discriminator is discerned between the Real Raman image and the generated Raman image by generator. During the model training, generator tried to produced Raman Images which are more realistic to the real one, while making it more difficult for the discriminator to discriminate between them. Once the model is full trained, now the generator have the power to fool the discriminator and give fully dispersed Raman spectra from the diffraction pattern. 136

Figure 5.4: (A) Raman spectra of multiple samples acquired using a grating with 1800 lines/mm, showcasing distinct spectral features across different classes. (B) The Raman spectra

were subjected to PCA for dimensions reduction, and the resulting scores were visualized as a RadViz scatter plot. In this plot, each point corresponds to an individual Raman spectrum, enabling the analysis of intra-class similarities and inter-class differences. A wider dispersion of data points indicates greater spectral variability among the samples. (C) The variations in spectral features are further represented through boxplots, providing a quantitative overview of the observed differences. 4-DMAP (black), 4-MBA (red), 4-NTP (blue), PCM (green), ranitidine (magenta), thymine (yellow), and uracil (cyan). 140

Figure 5.5: The convolutional neural network architecture for the classification network consisted of five convolutional blocks contains convolutional layers, batch normalization, relu, and max pooling layer. In the end of the architecture, fully connected layer of size seven is attached to pool out the details from the previous layers and feed the output to softmax layer. The softmax layer show the results in commutative sum of one, which finally gave the class name of the input. 142

Figure 5.6: The CNN classification model was trained using diffraction images from *dataset B*. The dataset was partitioned into 80% for training and 20% for testing. The model achieved an overall classification accuracy of 100%, as demonstrated by the performance metrics on the (A) training dataset and (B) testing dataset. 142

Figure 5.7: The real Raman spectra (black) and GAN generated Raman spectra (red) of various samples (A) 4-DMAP, (B) 4-MBA, (C) PCM, (D) Ranitidine, (E) Thymine, and (F) Uracil. 143

Figure 5.8: The SNR comparison between generated and real spectra showed that the GAN-assisted approach successfully preserved spectral features while managing low SNR. 144

Figure 5.9: Schematic representation of the overall workflow, demonstrating the GAN's ability to convert diffraction images into fully resolved Raman spectra while simultaneously performing classification. 145

Figure 6.1: (A) The simulation setup showing the gold nanorod dimer (dimer-1) and the obliquely incident circularly polarized light (CPL). The gold nanorods are coated with a gain medium of doped silica, and the entire simulation is performed in water as a medium. (B) The computed scattering responses of uncoated gold dimer to LCP and RCP light are shown in black and red, respectively. (C) The computed scattering responses of gold dimer coated with 5 nm silica to LCP and RCP light are shown in black and red, respectively. (D) Plasmonic

hybridization of the gold nanorod dimers. The black and red curves correspond to the uncoated and the 5 nm coated nanorods interacting with LCP light. (E) The energy level diagram corresponding to (D) describing the interaction between the plasmonic modes (ω_0), resulting in a bonding (ω_-) and an anti-bonding mode (ω_+) respectively. The bonding mode redshifts to 1.59 eV and the anti-bonding mode blue shifts to 1.77 eV for the uncoated nanorods (black). Similar shifts have also been observed for 5 nm silica coated nanorods (in Red). (F) The asymmetric and symmetric modes (bonding and anti-bonding modes respectively) shown as charge density distributions. The anti-bonding and bonding modes for uncoated gold nanorods are shown in (i-ii) and for 5 nm coated gold nanorods in (iii-iv). These modes are calculated for LCP light with an oblique incidence of 60^o with y-z axis. 150

Figure 6.2: The circular dichroism spectra of dimer-1 emerging from the difference between the scattering of LCP (black) and RCP (red) light (A) giving rise to a bisignate form of the spectrum (B). 151

Figure 6.3: Calculated pattern of near-field intensity of the gold nanorod dimer-1. (A and B) Shows the near-field intensities when the uncoated and 5 nm silica coated nanostructures are excited by LCP light at their specific bonding wavelengths (774 and 787 nm) respectively. Whereas (C and D) shows the near-field intensities when the uncoated and 5 nm silica coated nanostructures are excited by LCP light at their respective anti-bonding wavelengths (709 and 718 nm). 152

Figure 6.4: Calculated far-field profiles of the gold nanorod dimer-1. (A) Dimer excited by LCP at anti-bonding wavelength of 709 nm, and at bonding wavelength of 774 nm (B). (C) Off resonance excitation at 650 nm of the gold nanorod dimer. 153

Figure 6.5: The circular dichroism spectra emerging from the difference between the scattering of LCP (black) and RCP (red) light (A) at normal incidence and (C) at oblique incidence giving rise to a bisignate form of the spectrum (B) at normal incidence and (D) at oblique incidence. 154

Figure 6.6: Calculated absorbance (red), scattering (black) and extinction (blue) of a gold nanorod dimer system immersed in water and with left circularly polarized light incident at an oblique angle of incidence on dimer-1. Figures (A-C) represents absorbance, scattering and extinction of nanorod systems with k values of 0, 0.05 and 0.104 respectively. (D) The calculated anisotropy factor (g) with different k values. The black and red curves show the

trend of bonding and anti-bonding modes respectively. (E) The calculated anisotropy factors as a function of different wavelengths. (F) The calculated optical cross-section as a function of different values of k-values around the critical point of 0.104. 155

Figure 6.7: (A) Near-field intensity of 5 nm silica coated v-shaped nanorod assembly (dimer-1), and (B) when the gain coefficient, k of the silica layer is at its critical value of 0.104. The excitation was through LCP light at bonding wavelengths. 157

Figure 6.8: The simulation setup showing the twisted gold nanorod assembly showing intrinsic chirality (dimer-2). Figures (B-D) represents absorbance, scattering and extinction of nanorod systems with k values of 0, 0.07 and 0.150 respectively. Calculated absorbance (red), scattering (black) and extinction (blue) are shown for a gold nanorod dimer system immersed in water and with left circularly polarized light incident at normal incidence. (E) The calculated anisotropy factor (g) with different k values. The black and red curves show the trend of bonding and anti-bonding modes respectively. (F) The calculated optical cross-section as a function of different values of k-values around the critical point of 0.150. 157

Figure 6.9: The simulation setup showing the twisted gold nanorod assembly showing intrinsic chirality (dimer-2). Figures (B-D) represents absorbance, scattering and extinction of nanorod systems with k values of 0, 0.07 and 0.147 respectively. Calculated absorbance (red), scattering (black) and extinction (blue) are shown for a gold nanorod dimer system immersed in water and with right circularly polarized light incident at oblique incidence. (E) The calculated anisotropy factor (g) with different k values. The black and red curves show the trend of bonding and anti-bonding modes respectively. (F) The calculated optical cross-section as a function of different values of k-values around the critical point of 0.147. 159

Figure 6.10: (A) The PCD spectra at different thicknesses of silica layer on dimer-1 without a gain medium. (B) PCD spectra of the gold nanorod dimer with 15 nm thickness of silica containing different values of gain coefficient, k. (C) The variation of E-field with various thicknesses of silica with and without gain medium. The E-field is calculated from the obliquely incident LCP light on gold nanoparticle dimer at bonding and anti-bonding resonances respectively. (D) The variation of anisotropic factor (g) for both bonding (black) and anti-bonding (red) with the thickness of the gain medium. 160

Figure 6.11: The variation of chiroptical response of the gold nanorod dimer-1 with the angle between the two nanorods (A). (B) Shows the variation of E-field intensity with angle, and (C)

shows the corresponding CD signals. (D) shows the variation of the anisotropic factor (g) with varying value of angle θ for both bonding (black) and anti-bonding modes (red). 161

Figure 6.12: The variation of CD signals with various angles of incidence of LCP and RCP light and also the angle between the gold nanorods with 5 nm gain medium at critical value of gain coefficient k. The arrow represent that the value of gain medium is increasing in that direction from 0 to 0.104. 162

Figure 6.13: Variation of PCD response of dimer-1 with aspect ratio. (A) The blue and red curves are PCD from gold nanorod dimer having same aspect ratios (1:3) but different lengths. The black curve corresponds to the PCD from dimers with 1:4 aspect ratio. The corresponding values of diameter (D) and height (H) and the anisotropy factors are shown in (B). 163

Figure 7.1: Schematic representation of different morphology of SiNPs (A) straight, (B) slanted, and (C) Kinked. 169

Figure 7.2: Total reflectance spectra of SiNPs (A) straight, (B) slanted, and (C) Kinked. We optimize the various parameters in the SiNPs, (A) the reflection profile at various periodicity. In the inclined SiNPs the angle of inclination is varied for both (B) slanted and (C) kinked. 170

Figure 7.3: Total absorption spectra of SiNPs (A) straight, (B) slanted, and (C) Kinked. We optimize the various parameters in the SiNPs, (A) the absorption profile at various diameter and later kept it constant at 120 nm. In the inclined SiNPs the angle of inclination is varied for both (B) slanted and (C) kinked. 172

Figure 7.4: Absorption efficiency of SiNPs, straight, slanted, and kinked in (A) TE and (B) TM polarization. 175

Figure 7.5: Comparison of the absorption efficiency of kinked (A and B) and slanted (C and D) SiNPs in TE (left) and TM (right) polarization shows the higher absorption efficiency of kinked SiNPs over slanted in both polarization. 176

Figure 7.6: The schematic of the light interaction with the SiNPs and how light is entrapped within the SiNPs. 177

Figure 7.7: The electric field distribution with the different morphology of SiNP towards different polarization TE(left) and TM (right) at different wavelength. 178

Figure 7.8: Light entrapment in the different morphology of SiNP at different wavelength. 179

LIST OF TABLES

Table 3.1: The quantitative comparative measurement of different cell component present in the different cell lines. The fractions represent the relative proportions of various pure components present within the different cell types, providing a comparative analysis of their biochemical composition. The biochemical components are visually represented using a color-coded scheme: red indicates components with higher prominence, orange represents components with moderate levels, and green signifies components with relatively equal distribution across the samples. 91

Table 4.1: The best three ANN classification network with their optimized parameters and performance during training and testing with additional GAN-generated high-resolution dataset. 117

Table 4.2: SNR of Raman spectra accumulated using high-resolution (HR), low-resolution (LR), and machine learning (ML) generated for different samples. 122

Table 4.3: Average similarity index (%) of the low-resolution Raman barcode with the high-resolution Raman barcode. 125

Table 4.4: Standard deviation of similarity index of the low-resolution Raman barcode with the high-resolution Raman barcode. 125

Table 4.5: Average similarity index (%) of the GAN-generated high-resolution Raman barcode with the high-resolution Raman barcode. 126

Table 4.6: Standard deviation of similarity index of the GAN-generated high-resolution Raman barcode with the high-resolution Raman barcode. 127

Table 5.1: Average similarity index between the images from dataset A, within and across the class. There are 65 images within each class. 141

Table 5.2: Average of the standard deviation of similarity index between the images from dataset A, within and across the class. There are 65 images within each class. 141

ABBREVIATIONS

AFM Atomic Force Microscopy

AI Artificial Intelligence

ALS Asymmetric Least Square

ANN Artificial Neural Network

API Active Pharmaceutical Ingredient

CCD Charge Coupled Device

CD Circular Dichroism

CARS Coherent Anti-Stokes Raman Scattering

c-GAN Conditional Generative Adversarial Network

CNN Convolutional Neural Network

D Discriminator

DCGAN Deep Convolutional Generative Adversarial Network

DL Deep Learning

DNN Deep Neural Network

DT Decision Tree

ELU Exponential Linear Unit

FDTD Finite Difference Time Domain

FEM Finite Element Method

FISH Fluorescence in situ Hybridization

FNOs Fourier Neural Operators

FWHM Full Width Half Maxima

G Generator

GAN Generative Adversarial Network

GNN Graph Neural Network

GBM Gradient Boosting Machines

HCA Hierarchical Cluster Analysis

HNSCC Head and Neck Squamous Cell Carcinoma

HOMO Highest Occupied Molecular Orbitals

HPC High Performance Computing

HRM High-Resolution Melting

ICA Independent Component Analysis

IHC Immunohistochemistry

IPS Integrated Photonic Systems

IR Infrared

k-NN k-Nearest Neighbours

LCP Left Circular Polarized Light

LMRs Leaky Mode Resonances

LSPR Localized Surface Plasmon Resonance

LSTM Long Short-Terms Memory Network

LUMO Lowest Unoccupied Molecular Orbitals

MCR-ALS Multiple Curve Resolution by Asymmetric Least Square

ML Machine learning

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

NGS Next-Generation Sequencing

NIR Near-Infrared

P2P GAN pixel-to-pixel Generative Adversarial Network

PBC Periodic Boundary Condition

PCs Principal Components

PCA Principal Component Analysis

PCD Plasmonic Circular Dichroism

PET Preservative Efficacy Testing

PINNs Physics-Informed Neural Networks

PINOs Physics-Informed Neural Operators

PLSR Partial Least Square Regression

PML Perfectly Matched Layers

PReLU Parametric Rectified Linear Unit

RadViz Radial Visualization

RCP Right Circular Polarized Light

ReLU Rectified Linear Unit

ROC Receiver Operating Characteristic

RF Random Forest

RNN Recurrent Neural Network

RRS Resonance Raman Scattering

RS Raman Spectroscopy

SEM Scanning Electron Microscopy

SERS Surface-Enhanced Raman Spectroscopy

SGD Stochastic Gradient Descent

SiNP Silica Nanopillar

SNP Single-Nucleotide Polymorphisms

SNR Signal-to-Noise Ratio

SPR Surface Plasmon Resonance

SPASER Surface Plasmon Enhanced Stimulated Emission of Radiations

SSIM Structural Similarity Index

STM Scanning Tunneling Microscopy

SVD Singular Value Decomposition

SVM Support Vector Machine

TE Transverse Electric

TEM Transmission Electron Microscopy

TERS Tip-Enhanced Raman Scattering

TM Transverse Magnetic

UDG Uracil-DNA Glycosylase

UV Ultra-Violet

VAE Variational Autoencoders