TAILORING RAMAN SCATTERING FOR CANCEROUS CELLS PREDICTION USING OSCILLATING ELECTRIC FIELDS AND MACHINE LEARNING.

BY

MUREI GILBERT KIPTUM

A THESIS IS SUBMITTED IN PARTIAL FULFILMENT OF THE

REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY IN PHYSICS,

DEPARTMENT OF MATHEMATICS, PHYSICS AND COMPUTING,

SCHOOL OF SCIENCE AND AEROSPACE STUDIES,

MOI UNIVERSITY.

2023.

DECLARATION

Declaration by the Candidate

I the undersigned declare that this thesis is my original work and has not been presented for academic purposes in this University or any other Institution for academic purposes whatsoever. No part of this thesis may be reproduced without written permission from the author and / or Moi University.

Murei, Gilbert Kiptum.....Date.....**27 – 11 - 2023**.....

PHD/PHY/002/15

Declaration by the Supervisors

This thesis has been submitted for examination with our approval as University Supervisors.

Prof. Khanna Mohan Kapil.....Date...**27 – 11 - 2023**.....

Department of Physics,

University of Eldoret

Eldoret, Kenya.

Prof. Samuel Rotich.....Date.....**27 – 11 - 2023**.....

Department of Mathematics and Physics,

Moi University,

Eldoret, Kenya.

DEDICATION

This work is dedicated to my wife, children and my parents and all my close friends who have shared with me all my past and present hardships, tribulations and successes. To the Almighty God, with whose enduring grace, ensured that I had the strength and motivation to do this work, be glory and honor

ACKNOWLEDGEMENTS

I thank the Almighty God for the gift of good health of mind and body during research period.

I would like to forward my inmost gratitude and indebtedness to my supervisor, Prof. Khanna Kapil, for his inspirational guidance, encouragement and vital clarifications throughout my research work. His rich experience in the subject has helped me to develop curiosity in the field and write this thesis. I am highly indebted to my supervisor, Prof. Samuel Rotich, for his ever-enduring support, guidance and patience throughout the course of my study.

My special thanks also goes to Dr. Korir Kiptiemoi and Dr. Zephaniah Birech who were very instrumental during my PhD course work program. My sincere gratitude goes to Moi University, School of Science and Aerospace studies and Department of Mathematics and Physics for giving me the opportunity to pursue my PhD studies.

To my dearest wife, Emmy, and my kids Shanice, Ryan and Shayne, I say thanks for the smiles, prayers and encouragement throughout the difficult moments that we all shared.

Finally, I would like to forward my heartfelt gratitude and appreciation to all my friends, colleagues and relatives especially to my uncle Wise Kipkemei who offered me all round supports and encouragements throughout my research work. May God bless all who contributed to the success of this research and their names have not been mentioned here.

ABSTRACT

Raman Effect originates from the inelastic scattering of light, and it can directly probe vibrational states in molecules and materials. The Electromagnetic fields of the incident radiation polarizes the charges of the molecule leading to the creation of an oscillating electric dipole. The scattered light has modified frequencies called Stokes and anti-Stokes frequencies. However, many theories that attempt to describe the existence of modified frequencies cannot account for the effects of the changes of electromagnetic and oscillating electric fields on these frequencies. This study sought to develop a theory that explains the effects of the oscillating electric field created by the oscillating dipole on the Stokes and anti-Stokes lines and their application in predicting cancerous cells. The objectives of this study were to formulate a theory that leads to determination of additional Raman frequency from the effects of electromagnetic fields on oscillating electric dipole; to develop a dataset of the frequencies of modified Stokes and anti-Stokes lines in the Raman scattering for some polar molecules: to examine and explore how Raman dataset and machine learning tool can be used in predicting cancerous cells. A theoretical model was formulated by calculating the energy due to the oscillating electric field created by the oscillating dipole using electrodynamics principles. The parameters needed were investigated by applying some approximations on the theory of retarded potentials, and then Maxwell equations were used to deduce the expression of oscillating electric fields. The frequency corresponding to the energy due to this field will determine the modified frequencies of Stokes and anti-Stokes lines. The dataset was generated, analyzed, and applied in machine learning tool to predict cancer cells. It is indicated from the results that the frequency difference in the peaks of modified Raman lines of molecules with high dipole moments ranges from $4.0x10^{14}s^{-1}$ to $9.0x10^{14}s^{-1}$ and less than $1.0x10^{14}s^{-1}$ for less polar molecules. Based on the theory electromagnetic fields and oscillating electric fields, there are differential modified Raman frequencies data which can be used to distinguish the normal and cancerous cells with 96.75 % accuracy The study concludes that electric fields created by the oscillating electric dipole results to Raman scattering with modified frequencies which are sensitive to chemical structure of a molecule and these can be used to predict the presences of cancerous cells .The theory developed gives a formula that presents new results related to light matter interactions and allows a detailed description of Raman scattering. The law of conservation of energy validates the theory developed although actual experimental studies in future may be able to decide how far it will be a reliable technique.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABREVIATIONS

A.O	Anharmonic Oscillator
ARS	anti-Stokes Raman scattering
CW	continuous wave
CRS	Coherent Raman Scattering
CSRS	Coherent Stokes Raman Scattering
CARS	Coherent Anti-Stokes Raman Scattering
HRS	Hyper Raman Scattering
IR	Infra-Red
IRS	Inverse Raman Scattering
PARS	photo acoustic Raman spectroscopy
SERS	Surface Enhanced Raman Spectroscopy
SHO	Simple- Harmonic Oscillation
SRG	Stimulated Raman Gain
SRC	Stokes Raman scattering
SRS	Stimulated Raman Scattering
UV	ultra violet

LIST OF SYMBOLS

a_S^+	Boson creation operator for the modes S
a_s	Boson annihilation operator for the mode S
$a_n(r)$	Spatially varying part of the wave function
α (t)	Polarizability
C	Speed of light
C_n	Projections of Ψ along the systems eignstates Ψ_n (r, t).
e	Charge of the electrons
E	Quantum of energy
$\langle E \rangle$	Total electric field
E_s(r, t)	Electric field operator of the mode S
ε_0	Electric permittivity in vacuum
f	Frequency
h	Planck's constant
H_o	Hamiltonian operator
H_F	Contribution from the field degrees
H_{int}	Interaction between the field and the material
I	Intensity
k	Wave vector of the radiated field
$\hbar$	Force constant
n	Number of waves

n_s	Photon occupation number of the mode S
N	Number of electric dipoles per unit volume
Θ	Angle relative to the dipole axis
P (t)	Polarization
$P^{(n)}(t)$	Nth order contribution to the polarization
Q	Nuclear co-ordinate
Q_o	Amplitude of the nuclear motion
∅	Phase of the nuclear mode vibration
z	Displacement
r_{cm}	Center- of-mass
S	Time-average poynting flux,
t	Time
U	Potential energy
μ (t)	Electric dipole moment
v	Vibration of frequency
υ_s	Quantization volume of the photon mode
V (t)	Coupling potential or perturbation
ω_1	Angular frequency
$\boldsymbol{\omega}_v$	Natural frequency of vibration of the nucleus
$\boldsymbol{\omega}_{cs}$	Stokes frequency-coherent Raman Effect
$\boldsymbol{\omega}_{as}$	Anti-stokes frequency-coherent Raman Effect

ω_n	Eigen frequency associated with eigen state Ψ_n
χ	Susceptibility of the material
$\chi^{(n)}$	nth order susceptibility
x_e, y_e	Anharmonicity constants,
Ψ (r, t)	wave function
Ψ_n	Molecular eigenstates

CHAPTER ONE

INTRODUCTION

1.1 Background information

When light passes through a medium such as air, liquid, or solid, some of it is absorbed by the particles of the medium, while another portion of the light is scattered in a particular direction (Pérez & Plascencia, 2021); Tehrani *et al.*, 2021).Scattering can be broadly defined as the redirection of radiation from the original direction of incident light photon, usually due to interactions with matter (atoms, molecules, and particles) (Han *et al.*, 2021). Scattering of light involves both the properties of the light radiation such as intensity, wavelength (λ), angular frequency (ω), and kinetic energy (E_k), and properties of matter encountered and thus information about the atoms or molecules can be obtained**.** The energy of a photon of light is inversely proportional to its wavelength and this is shown by the equation (1.1), which is also known as the Planck-Einstein relation (Dyson,2006).

$$E = hf = h\frac{c}{\lambda} = \hbar\boldsymbol{\omega} \quad \text{..... (1.1)}$$

The electromagnetic spectrum has a large range of different wavelengths that are all useful for different kinds of spectroscopy. The primary interactions of molecules with electromagnetic radiation are of the electric dipole type and this will form the basis of this study. Light scattering can be elastic or inelastic. In elastic light scattering, the predominant form of light scattering (up to 99%), the scattered photon energy is equal to the incident photon energy, and the angular frequency (ω) and wavelength (λ) are unchanged (Lapoux, 2021).In inelastic scattering of light, particles of scattered light undergo changes in angular frequency (ω), wavelength (λ) and changes in some of its kinetic energy (E_k) (Carminati and Schotland, 2021). This inelastic scattering is described as Raman scattering (Barron and Buckingham, 1971) and its studies have been one of the most fascinating topics in the recent years due to its applications in science, industry, medical science, mineral exploration and remote sensing (Clegg *et al*.,2006).

1.2 Discovery of Raman Effect.

Raman scattering of light by molecules was first predicted in 1923 using the classical quantum theory by Smekal and observed experimentally by an Indian physicist C.V. Raman and K.S. Krishnan in 1928 (Weber & Merlin, 2013; Cantarero, 2015; Kneipp and Feld, 2020).Raman recorded a spectrum of scattered light and observed that when light is scattered from any transparent substance, the spectrum of the scattered light exhibited over and above the lines present in the spectrum of the mercury arc light (Agarwal and Atalla, 2020). The energy difference between the incident and scattered photons is equal to the energy of the vibration with which the photons interacted. Raman proved that new lines appeared, or in some cases bands appeared, and in some cases unresolved continuum emissions shifted to varying degrees from the current lines. Raman further showed that each line of the incident radiation, provided it was of sufficient intensity, gave rise to its own modified scattering and frequency shift (Sirleto *et al.*, 2017). The shift in relative intensities, the state of polarization and other features of the new lines and bands were independent of the exciting radiation (Hartmann *et al.*, 2020).The new lines were shown to be characteristics of the substances under study.

1.3 Raman spectroscopy techniques

Raman scattering can occur due to molecular vibrations, rotations, or electronic energy changes. Scattered radiation can have higher or lower frequencies than the original radiation. The Intensity of the Raman scattering depends on several factors such as the excitation wavelength of the incident radiation, excitation power, changes in polarizability, the amount of Raman active molecules illuminated by the laser beam and temperature (Wang *et al*., 2020). Light interacts with the charged ions inside atoms to polarize them when it strikes a material or molecule. An electric dipole is created by the molecule's atoms becoming polarized. The strength of the electric field directly relates to the induced electric dipole moment. The frequency of the external electromagnetic and electric fields causes the induced electric dipole to oscillate. The vibrational states of molecules are used as a scattering system in Raman spectroscopy, which results in a shift in the electric dipole moment.

Raman scattering techniques can be illustrated by linear or nonlinear interaction between light and matter.

1.3.1 Linear Raman scattering

Raman scattering is understood to arise from induced vibrational polarization within the sample due to the interaction of light with the molecules. Conventional Raman scattering is a linear process and the induced sample polarization depends on the power of the incident radiation field strength (Brand, 1989). An externally applied electromagnetic field can alter or perturb the charge distribution only according to the ability of molecules or crystals to form dipoles.

The classical approach postulates that the existence of the Raman Effect is related to the modulation of the polarizability of molecular and dielectric susceptibility to crystal lattice vibrations due to the vibrational nature of interatomic displacements (Tuschel, 2020).

In Raman spectroscopy, the vibrational excitation takes place through a two-photon scattering process (Ma *et al*., 2021). When the energy of the scattered photons is lower than the energy of the incident photons, this process is called Stokes Raman scattering (SRC). Conversely, when the energy of the scattered photons is higher than the energy of the incident photons, the process is called anti-Stokes Raman scattering (ARS) (Cantarero, 2015). The Raman scattering process also conserves momentum which is therefore demonstrated in wave vector form as (Cialla-May,2019).

$$\vec{k}_{scat} = \vec{k}_p + \vec{q}_p \quad \text{..... (1.2)}$$

Where;

$\vec{k}_{scat}$ Represents wave vectors of scattered light,

$\vec{k}_p$ Represents wave vectors of incidence of light and

$\vec{q}_p$ Represents wave vectors for phonon or molecular vibration.

Of the two scattering processes, Stokes Raman scattering occurs more frequently as most vibrations will be in the ground state at typical experimental temperatures.

When matter and light interact, the matter's charge distribution is disturbed by the oscillating electromagnetic (EM) and electric fields of the light, which may result in the exchange of energy and momentum that leaves the matter in a changed state. A medium's atoms and molecules become polarized when it is exposed to an electric field.

1.3.2 Non-linear Raman Spectroscopy

All other techniques that rely on field strengths of the second order or higher are called non-linear Raman techniques. Nonlinear Raman scattering includes Stimulated Raman Scattering (SRS), Hyper Raman Scattering (HRS), Coherent Anti-Stokes Raman Scattering (CARS), Coherent Stokes Raman Scattering (CSRS), Stimulated Raman Gain (SRG), Inverse Raman Scattering (IRS), inter alia photo acoustic Raman spectroscopy (PARS) ((Borman, 1982).

There have been few uses of stimulated Raman scattering (SRS) in molecular spectroscopy, despite the great intensity of Stokes and anti-Stokes scattering. This is due to the fact that only a few of the strongest Raman lines exhibit stimulated Raman scattering. Coherent methods such as coherent Stokes Raman spectroscopy (CSRS) and coherent Anti-Stokes Raman spectroscopy (CARS) have been developed to circumvent these shortcomings. Techniques like CARS and CSRS are based on SRS, but instead of using one laser beam at the frequency ω_L as in SRS, two laser beams are used to excite a sample at ω_L and ω_S.

1.4 Development in Instrumentation for Raman Spectroscopy

In Raman spectroscopy experiment, a laser beam is used to irradiate the sample which interacts with the molecules of sample leading to a scattered light. However, the problems facing the development of Raman spectroscopic instrumentation are the inherent weakness of the inelastic scattering and the much larger intensity of the Rayleigh scattering (DePaoli *et al*., 2020). This fact had posed several restrictions to progress of Raman spectroscopy instrumentation since the beginning of the experimentation to the present. The incident radiation must give very intense radiation for the scattered light to be strong enough to be observe (Bonse, 2020). Additionally, the light should be as monochromatic as possible so that the Raman bands would be as narrow as possible. A highly-powered light source is required in order to create a detectable Raman signal. In addition, Raman spectroscopy requires a spectrometer with a very high degree of discrimination against the Rayleigh elastic scattered light

(Edwards et al.,2022). Finally, since very few Raman photons are generated, the detection system must be very sensitive to detect the Raman signal over the large noise background (Xu,2020).

Thus, correct selection of laser wavelength is an important consideration for Raman spectroscopy. Different excitation wavelengths provide specific strengths and weaknesses allowing a user to optimize the measurement of different samples by their choice of Raman excitation laser wavelength. There are many different excitation options, but the three most widely used are 532nm, 785nm and 1064nm (Pilot,2019). The most popular is the 785 nm excitation system as it offers the best signal strength, sensitivity to fluorescence, and can be used to collect the Raman spectrum of most organic materials quickly (McCrery, 2000).

Before the inventions of lasers, radiations emitted by the mercury arc, especially at 435.8 and 404.7 nm, have been used for exciting Raman spectra. Today, most types of lasers, like continuous wave (CW) and pulsed, gas, solid state, semiconductor lasers, etc., with emission lines from the UV to the near-IR region, are used as radiation sources for the excitation of Raman spectra (McCrery, 2000 & Willard *et al*., 1988). In this thesis, the mentioned wavelengths will be used as the incident radiations to calculate the frequency of the oscillating electric field created by the oscillating molecular dipole,

It is remarkable that Raman spectroscopy is particular and distinctive but the weak intensity of the Raman signal due to the low scattering cross-section ($\sim 10^{-30}$ cm^2 $molecule^{-1}$) is the biggest disadvantage of Raman spectroscopy with low sensitivity, which is the reason why it was not widely used for a long time (Alodan, 2022). Hence the detection of molecules with very low concentration is limited. In order to enhance sensitivity and improve intensity, to reach better level a number of variations in the Raman spectroscopy have been done and Surface Enhanced Raman Spectroscopy (SERS) has been developed among other techniques.

1.5 Statement of the Problem

The theory of Raman scattering was proposed by Smekal in 1923, but the experimental observations was done in 1928 by Raman and Landsberg (Raman., *et al* 1928, & Landsberg., *et al* 1928). From 1930 onwards the development of Raman spectroscopy was relatively slow for several years due to the fact that the Raman

effect is very weak (Jain,2021). Raman spectroscopy started developing in the 1960's when the lasers were invented and started to be used as light sources in spectroscopy. The development of the laser technique thus inspired the traditional field of molecular spectroscopy and Raman spectroscopy significantly. Soon after the laser innovation, Frank *et al* demonstrated nonlinear optical effects in different media produced by an intense electric field of a laser light (Frank., *et al* 1961).

From these theoretical efforts, its shows that Raman spectroscopy utilizes the incident electromagnetic radiations to enhance the Raman signals. An accurate description of interaction of electromagnetic fields and molecules is therefore critical for understanding Raman scattering. However, in most of the studies this phenomenon has been ignored. The basics of the Raman scattering has been described using classical and quantum physics however, the experimentally observed different intensities of Stokes and anti-Stokes scattering is not explained. It only considers the frequency of incident electromagnetic radiation and the frequency of atom or molecule to explain Raman scattering (Smith and Dent, 2005). In order to properly describe Stokes as well as anti-Stokes scattering, a more comprehensive and quantitative theory is needed**.**

There are two aspects to be studied in this thesis. One deals with an oscillating electric dipole produced by the incident electromagnetic radiation, leading to an oscillating electric field. A theory based on electrodynamics principles is developed and used to calculate additional angular frequency contribution due to oscillating electric field induced by oscillating electric dipole. This study reports on an interesting idea of additional angular frequency contribution to a frequency of Raman lines which has not been established. The second deals with the study of Raman spectroscopy as an optical technique capable of detecting chemical components of a sample by their unique set of vibrational spectra of the molecules, and thus it can be used in the differentiation of cancerous cells versus normal tissues cells. Although, organisms are electrically neutral, they have charged ions, polarized molecules and electric fields which obey the laws of electromagnetism and thermodynamics. Experiments have also shown (Fraldi *et al*., 2015 & Savavana *et al.*.2016) that the average healthy cell is almost 70% softer than a single malignant cell. As a result, the force constant of the cancerous cell will be different from that of the healthy cell, leading to a different Stokes and anti -Stokes frequency of the cancer cell when compared to the normal

healthy cell. In this thesis, frequencies of modified Stokes and anti-Stokes lines for some polar and less polar molecules are generated and compiled. The dataset will be analyzed, evaluated, and a simple machine learning model will be developed and trained to predict cancer cells. This study will provide a new contribution to the nature of Raman lines and deeper understanding of Raman spectroscopy.

In this case, the overarching aim of this comprehensive study is to intricately develop a robust theoretical framework. This framework is specifically designed to elucidate the nuanced effects of the oscillating electric field generated by the oscillating dipole on both Stokes and anti-Stokes lines within the context of Raman scattering. Furthermore, the study extends beyond the realm of theoretical inquiry to explore the practical application of these insights in the realm of medical diagnostics, particularly in the prediction of cancerous cells.

The central thrust of this research is twofold. Firstly, it seeks to delve into the intricate dynamics of the oscillating electric field and its consequential impact on the characteristic Stokes and anti-Stokes lines in the realm of Raman scattering. This involves a meticulous exploration of the theoretical underpinnings, drawing on principles from electrodynamics to construct a framework that can comprehensively capture and explain these effects.

1.6 General Objective

The main objective of this study is to develop a theory that explains the effects of the oscillating electric field created by the oscillating dipole on the Stokes and anti-Stokes lines in the Raman scattering and their application in predicting cancerous cells.

1.6.1 Specific Objectives:

The specific objectives of this study are:

i. To formulate a theory that leads to determination of additional angular Raman frequency from the effects of electromagnetic fields on oscillating electric dipole.
ii. To develop a dataset of the frequencies of modified Stokes and anti-Stokes lines in the Raman scattering for some polar molecules.
iii. To examine and explore how Raman dataset and machine learning tool can be used in predicting cancerous cells.

1.7 Significance of the study

The theory of spectroscopy must analyze three issues that can be applied to understand the atomic and molecular structure of a wide range of samples. These are the incident electromagnetic radiation acting on each molecule, the molecular response to this field and the emission by the polarized molecule. To describe and interpret these processes, a theory based on classical electromagnetic theory framework is developed. The theory developed is of great importance in further studies of material science. In several cases, the degree to which the theoretically generated mathematical formulation agrees with reproducible experiment results determines the scientific field's usefulness. This theory is vital for interpreting and assigning Raman spectra for any type of molecules under study.

This study holds importance due to its potential to reshape the utilization of Raman spectroscopy. Through the development of a comprehensive theory examining the influences of the oscillating electric field on both Stokes and anti-Stokes lines, this research addresses a critical gap in existing knowledge. The exploration of Raman scattering as a tool for predicting cancerous cells adds a practical dimension to the theoretical advancements, offering promising implications for medical diagnostics and healthcare.

The paramount significance lies in the potential to catalyze a paradigm shift in the comprehension and utilization of Raman spectroscopy. Undertaking the formidable task of addressing a critical lacuna in the current scientific landscape, the study unravels the intricate dance between electromagnetic fields and oscillating electric dipoles. The theoretical advancements achieved not only deepen fundamental understanding but also pave the way for transformative applications across diverse domains.

In particular, the innovative exploration of Raman scattering as a predictive tool for discerning cancerous cells injects palpable practicality into the theoretical underpinnings. This convergence of theory and application propels the study beyond the confines of abstract knowledge, extending its reach into the tangible realm of medical diagnostics and healthcare. The potential ramifications of this integration are nothing short of revolutionary.

The prospect of utilizing Raman spectroscopy as a diagnostic modality for cancer holds immense promise. Unraveling the nuanced variations in the frequencies of modified Stokes and anti-Stokes lines for polar molecules, the study not only contributes to the theoretical arsenal but also unlocks a new frontier in the fight against cancer. The ability to differentiate between healthy and malignant cells based on their Raman spectra introduces a non-invasive and potentially transformative approach to early detection and characterization of cancerous conditions.

As the implications of the study reverberate through the corridors of scientific inquiry, they extend a hand to the medical community, offering a tool that transcends traditional diagnostic methodologies. The potential to predict and identify cancer cells with enhanced precision holds the key to ushering in a new era of personalized medicine, where interventions can be tailored with unparalleled specificity.

In essence, the significance of this study transcends the boundaries of a conventional scientific investigation. It heralds a convergence of theory and practicality, paving the way for a future where Raman spectroscopy stands not only as a tool of academic fascination but as a beacon of hope in the realms of medical science and healthcare. The ripple effects of this study are poised to resonate across disciplines, leaving an indelible mark on the landscape of scientific inquiry and, more importantly, on the lives of individuals impacted by the relentless pursuit of knowledge and its transformative applications.

1.8 Justification of the Study

A number of theories on Raman scattering have had been proposed in the past (Bloembergen.,1967: Ferraro.,2003: Keffer *et al* 2006: Abdelrahman, & Ahmed, .2014). In some theories the distortion of the electron cloud around the molecule by the electric field of the incident electromagnetic radiation results in Raman scattering. In other theories the interaction of the incident radiation with the quantum vibrations of the lattice (or phonons) leads to Raman scattering. Yet in some theories the interaction of incident radiation with vibrational motion of the molecule is considered to explain Raman scattering.

In this thesis it is proposed that it will be more correct to take into account the interaction of the incident electromagnetic radiation on a molecule, which is polarized by the incident radiation resulting in the creation of an oscillating electric dipole

which now become a source of radiation of either the same or different frequency as the frequency of incident radiation. The other possibility is the interaction of the incident radiation on the vibrational motion of the molecule, resulting in new Stokes and anti-Stokes lines for the vibrational spectra.

This study is intended to support the experimental work and make it more effective and its results more reasonable. The findings of this study can be used to predict important and useful properties of molecules and molecular materials, cancerous cells, surface deformations etc.

This thesis seeks to build upon existing theories by suggesting a more accurate perspective: considering the interaction of incident electromagnetic radiation with a polarized molecule. This interaction gives rise to an oscillating electric dipole within the molecule, serving as a source of radiation at either the same or different frequencies compared to the incident radiation. Alternatively, the study explores the interaction of incident radiation with the vibrational motion of the molecule, leading to the appearance of new Stokes and anti-Stokes lines in the vibrational spectra.

The primary aim of this research is to enhance the effectiveness of experimental work in the field, providing a more comprehensive theoretical foundation for interpreting results. By delving into the intricacies of Raman scattering, this study aims to contribute valuable insights that can be applied to predict crucial properties of molecules and molecular materials. The implications extend across diverse domains, encompassing the analysis of cancerous cells, examination of surface deformations, and a broader comprehension of molecular behaviors.

Furthermore, the findings of this study are anticipated to play a pivotal role in advancing predictive models, offering a robust framework for exploring innovative applications in material science, biotechnology, and medical diagnostics. The enhanced understanding of Raman scattering mechanisms presented in this thesis is poised to open new avenues for scientific inquiry and technological advancements, thereby enriching the broader scientific community's knowledge base.

1.9 Scope and Limitations

In the realm of this study, the investigation centers on comprehending the impact of the oscillating electric field on Raman scattering. A specific emphasis is directed

towards the dynamics of Stokes and anti-Stokes lines. The primary objective is to construct a robust theoretical framework elucidating the intricate interplay between electromagnetic fields and molecules.

Simultaneously, the study extends its reach to encompass the compilation and systematic analysis of a dataset. This dataset is purposefully curated to feature modified frequencies of Raman lines, focusing specifically on molecules exhibiting polar characteristics. This dual-pronged approach, intertwining theoretical development and empirical investigation, aims to furnish a comprehensive understanding of how molecules respond to electromagnetic fields, particularly within the context of Raman spectroscopy.

However, within the expansive scope, it is imperative to acknowledge certain constraints that temper the breadth of the study. A notable limitation lies in the intentional focus on polar molecules, which enriches the depth of insight into specific molecular behaviors but inherently narrows the generalizability of findings to non-polar counterparts. This caveat is essential for a nuanced interpretation of the study's outcomes.

Furthermore, the study cautiously acknowledges the preliminary nature of machine learning model development. While the exploration of machine learning applications in predicting cancer cells is an intriguing avenue, the embryonic stage of this aspect underscores the necessity for a cautious interpretation of results. The limitations posed by the nascent state of machine learning endeavors within this study serve as reminders of the evolving nature of this intersection between theoretical physics and computational applications.

In summation, the scope of this study is expansive, intertwining theoretical development and empirical investigation in the realm of Raman scattering. However, the conscientious acknowledgment of limitations, including the specific focus on polar molecules and the nascent state of machine learning applications, imparts a necessary dose of realism to the study. These limitations serve not as constraints but as guiding signposts for a nuanced understanding of the study's outcomes, paving the way for future research endeavors to build upon and refine these initial explorations.

1.10 Ethical Considerations

This study is firmly grounded in ethical principles, emphasizing a commitment to the responsible and respectful treatment of both data and the individuals participating in the research process. Adherence to rigorous ethical standards serves as the cornerstone of the study, ensuring the utmost integrity in the handling of sensitive information and the engagement of human subjects.

In the pursuit of knowledge, particularly in the compilation of Raman frequencies for molecules, a meticulous approach to ethical considerations is paramount. All aspects of the research, from data collection to analysis, are undertaken with the explicit permission of relevant stakeholders. This not only safeguards the integrity of the study but also upholds the fundamental rights and privacy of the participants involved.

The ethical framework guiding this study extends beyond mere compliance with formal regulations; it embodies a genuine commitment to the well-being and dignity of those contributing to the research endeavor. In the realm of data compilation, stringent measures are in place to ensure that confidentiality is maintained, preventing any unauthorized access or use that may compromise the privacy of the participants.

Furthermore, ethical guidelines are not just a regulatory checklist but an integral part of the research process. Informed consent, transparency, and respect for autonomy are woven into the fabric of the study's methodology. Participants are provided with clear and understandable information about the nature of their involvement, empowering them to make informed decisions about their participation.

Moreover, the study places a premium on transparency in communication. Regular and clear updates are provided to participants regarding the progress and outcomes of the research, fostering a sense of partnership and mutual respect. This open dialogue ensures that participants are not only aware of how their contributions are utilized but also that their voices are heard and valued throughout the research journey.

In essence, the ethical considerations embedded in this study transcend mere procedural requirements; they underscore a profound commitment to the responsible and respectful conduct of research. This commitment ensures that the study not only generates valuable insights but does so in a manner that upholds the highest standards

of integrity, transparency, and respect for the individuals who play a crucial role in advancing scientific knowledge.

1.11 Organization of the Thesis

This thesis is organized into several chapters to systematically present the research findings and discussions. Chapter 2 provides an extensive literature review, summarizing existing knowledge on Raman spectroscopy, electromagnetic field interactions, and related theories. Chapter 3 details the methodology employed in formulating the theory and generating the Raman dataset. Chapter 4 presents the results of the study, including the additional angular Raman frequency and the modified Stokes and anti-Stokes lines for polar molecules. Chapter 5 discusses the implications of the findings and explores potential applications in predicting cancerous cells. Finally, Chapter 6 concludes the thesis, summarizing key findings and proposing avenues for future research in the field of Raman spectroscopy.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction.

This chapter provides an overview of the basics of the light-matter interaction which the theory of Raman scattering is based. It considers both classical and quantum mechanical descriptions of the Raman process and other spectroscopic phenomena which are significant for this thesis. In addition, electromagnetic theory and molecular vibrations are introduced. This is followed by a brief discussion of machine learning and Raman spectroscopy applications.

Equally, in this chapter, we delve into the foundational principles that underpin the theory of Raman scattering, a phenomenon arising from the interaction between light and matter. The exploration encompasses classical and quantum mechanical perspectives, shedding light on the intricate processes involved in light-matter interactions.

Understanding the electromagnetic theory and molecular vibrations is crucial for unraveling the complexities of Raman scattering. The electromagnetic spectrum, with its diverse wavelengths, plays a pivotal role in various spectroscopic techniques. Raman scattering, as a form of light scattering, involves the redirection of radiation due to interactions with matter—atoms, molecules, and particles. Classical theories, exemplified by the predictions made in 1923 and subsequent experimental validations in 1928 by C.V. Raman and K.S. Krishnan, have paved the way for our comprehension of light-matter interactions. The Raman Effect, as discovered, highlights the energy differences between incident and scattered photons, linking them to the vibrational energies of the molecules involved.

Linear Raman scattering, arising from induced vibrational polarization, constitutes a fundamental aspect of conventional Raman spectroscopy. In contrast, nonlinear Raman techniques, including Stimulated Raman Scattering (SRS) and Coherent Anti-Stokes Raman Scattering (CARS), extend the capabilities of Raman spectroscopy by leveraging higher-order field strengths.

The development of instrumentation is essential for the success of Raman spectroscopy. Overcoming challenges such as weak inelastic scattering and intense Rayleigh scattering has driven advancements in instrumentation. Laser wavelength selection, a critical consideration, significantly impacts signal strength and sensitivity in Raman spectroscopy. The continuous evolution of technology, from the era of mercury arc radiations to the current use of diverse lasers, underscores the ongoing quest for improved Raman spectroscopic instrumentation.

Significantly, this chapter emphasizes the importance of understanding the theory behind Raman scattering, as it forms the bedrock for accurate interpretation and assignment of Raman spectra. Classical electromagnetic theory provides a framework for comprehending the interaction of incident electromagnetic radiation with molecules, their responses, and the subsequent emission of polarized molecules.

As we delve into the literature, the exploration encompasses not only the theoretical foundations but also the practical applications of Raman spectroscopy. Furthermore, the integration of machine learning into Raman spectroscopy applications adds a contemporary dimension to this age-old field, promising enhanced sensitivity and specificity in molecular analyses. The literature review provides a comprehensive backdrop, setting the stage for a deeper understanding of the intricacies of Raman scattering and its diverse applications.

2.1 Light scattering

Atoms or molecules when exposed to light absorb light energy and re-emit light in different directions with different intensity and frequency. This phenomenon is referred to as light scattering (Reichel, (2022)). Light scattering is a linear interaction with matter that leads to change in frequency and hence photon energy. Major forms of elastic light scattering involving negligible energy transfer are Rayleigh scattering and Mie scattering. Inelastic scattering includes Brillouin scattering, Raman scattering, inelastic X-ray diffraction and Compton scattering (Dahm, 2021)

The energy of the matter after the interaction is higher or lower than the original state. Due to conservation of total energy, the energy of emitted radiation changes resulting in a change of frequency. The only real inelastic scattering process that appears clearly is Raman scattering (Horn, 2022). This scattering process can be described as

a collision process between a photon and electron of the scattering matter; part of the photon energy is used to move the electron with a certain speed in a certain direction (Nguyen, D. V et al, 2021)

In this case, Light scattering, a phenomenon observed when atoms or molecules interact with incident light, constitutes a fundamental aspect of the interaction between light and matter. As Reichel (2022) notes, the absorption of light energy by atoms or molecules leads to the re-emission of light in various directions, each with distinct intensity and frequency. This intricate process, aptly termed light scattering, plays a pivotal role in understanding the behavior of matter when exposed to electromagnetic radiation.

The interaction between light and matter in the context of light scattering is a linear process that results in a change in frequency and, consequently, photon energy. Major manifestations of elastic light scattering, where there is negligible energy transfer, include Rayleigh scattering and Mie scattering. These phenomena, as described by Dahm (2021), contribute to our understanding of the ways in which incident light interacts with matter, leading to alterations in frequency and energy.

Inelastic scattering, on the other hand, involves a significant change in energy states of the matter following the interaction with light. This category encompasses various processes, including Brillouin scattering, inelastic X-ray diffraction, Compton scattering, and the focal point of this discussion—Raman scattering. Horn (2022) emphasizes that Raman scattering is a key player in inelastic scattering, leading to discernible changes in the energy of the matter involved.

The energy alterations observed in inelastic scattering are intricately connected to the conservation of total energy. As a result, the energy of the emitted radiation undergoes modifications, leading to shifts in frequency. Notably, Raman scattering emerges as a unique process in this realm. It is worth noting that Raman scattering, as highlighted by Nguyen et al. (2021), can be conceptualized as a collision process between a photon and an electron within the scattering matter. During this collision, a portion of the photon energy is utilized to propel the electron with a specific speed in a particular direction.

Expanding on this understanding of light scattering, it becomes evident that the interaction between light and matter is a nuanced and dynamic process. The diverse forms of scattering, whether elastic or inelastic, provide valuable insights into the behavior of matter when subjected to electromagnetic radiation. The intricacies of Raman scattering, in particular, underscore the collision-like nature of the interaction, where energy is exchanged between photons and electrons, resulting in tangible changes in the energy state of the matter.

As we delve deeper into the realms of light scattering, the subsequent sections of this literature review will further explore the specific nuances of Raman scattering, its historical context, and its myriad applications in diverse fields, including science, industry, medical science, mineral exploration, and remote sensing. The theoretical foundations laid out in this section serve as a solid base for comprehending the intricate details of Raman scattering, providing a comprehensive backdrop for the ensuing discussions.

2.2 Raman Scattering

Raman scattering describes the change of the molecular and optical properties due to the presence of an incident electric field. When light radiation falls on an atom or molecule, it gets scattered. The collision is both elastic and inelastic (Campanella, B et al, 2021). The Raman Effect occurs when a photon incident on a molecule interacts with the molecule's electric dipole created by the electric field of the incident light. The shift in wavelength of the inelastically scattered radiation from the atoms and molecules is called Raman scattering and the phenomena is called Raman Effect (Eskandari, V et al, 2022). Such interaction can be partly explained by classical theory of light and quantum mechanical treatment of light-matter interaction (Li, T. E et al, 2019). Raman shifted photons can be either of higher or lower energy compared to the energy of the incident light, depending on the vibration state of the molecule under study. Stokes radiation occurs at lower energy (scattered radiation has energy lower than the incident radiation, and hence longer wavelength) compared to Rayleigh scattering, whereas anti-Stokes radiation occurs at higher energy (shorter wave-length) (Gupta *et al.* 2022), Of all the scattered photons, only a small fraction ($1x10^{-7}$) have a frequency different from the frequency of the incident photons (Krishnan.,*et al* 1981). Raman was, in fact, working on the molecular diffraction of light from the

surface of a liquid, and this is what led to the discovery of Raman Effect. Raman Effect was discovered in 1928, and in 1930 Raman was rewarded with a Nobel Prize in Physics. The Russian scientists, Grigory Landsberg and Leonid Mandelstam observed this effect in crystals shortly after the discovery of Raman Effect by Raman in 1928 (Silva, 2020).

The exploration of Raman scattering opens a gateway to the intricate changes in molecular and optical properties induced by the presence of an incident electric field. As elucidated by Campanella *et al.* (2021), when light radiation encounters an atom or molecule, the ensuing collision manifests both elastic and inelastic characteristics. Raman scattering, a phenomenon deeply rooted in the interaction between light and matter, plays a pivotal role in deciphering molecular behaviors under the influence of incident electromagnetic radiation.

At its core, the Raman Effect, as delineated by Eskandari *et al.* (2022), unfolds when a photon engages with a molecule's electric dipole, a creation spurred by the electric field of the incident light. The resultant shift in the wavelength of the inelastically scattered radiation from atoms and molecules encapsulates the essence of Raman scattering. The interplay between classical theories of light and quantum mechanical treatments of light-matter interaction, as posited by Li et al. (2019), adds layers of complexity to the comprehension of this intricate phenomenon.

Raman-shifted photons, as highlighted by Gupta *et al*. (2022), exhibit a fascinating duality, varying in energy compared to the incident light, contingent upon the vibration state of the molecule under scrutiny. Stokes radiation, characterized by lower energy and longer wavelengths than Rayleigh scattering, provides valuable insights into the subtle dynamics at play. Conversely, anti-Stokes radiation, occurring at higher energy and shorter wavelengths, offers a contrasting perspective on the molecular vibrations under examination.

The rarity of photons exhibiting Raman shifts becomes evident in the statistic provided by Krishnan *et al.* (1981), noting that only a minute fraction ($1x10^{-7}$) of scattered photons deviate in frequency from the incident photons. Raman, however, stumbled upon this captivating effect while delving into the molecular diffraction of light from the surface of a liquid in 1928. This serendipitous discovery led to the

formalization of the Raman Effect and, in 1930, earned Raman the Nobel Prize in Physics. It's worth noting that Russian scientists Landsberg and Mandelstam observed this effect in crystals shortly after Raman's groundbreaking discovery, contributing to the early validation of this phenomenon (Silva, 2020).

This historical narrative and the nuanced details of Raman scattering serve as a testament to the multifaceted nature of this phenomenon. The subsequent sections of this literature review will further dissect the principles underpinning Raman scattering, explore the various techniques employed in Raman spectroscopy, and delve into its wide-ranging applications across scientific, industrial, and medical domains. The journey through the intricate world of Raman scattering continues, offering profound insights into the molecular intricacies that define our understanding of light-matter interactions.

The diagram in figure 2.1 shows Raman scattering, anti-Stoke Scattering and Rayleigh scattering.

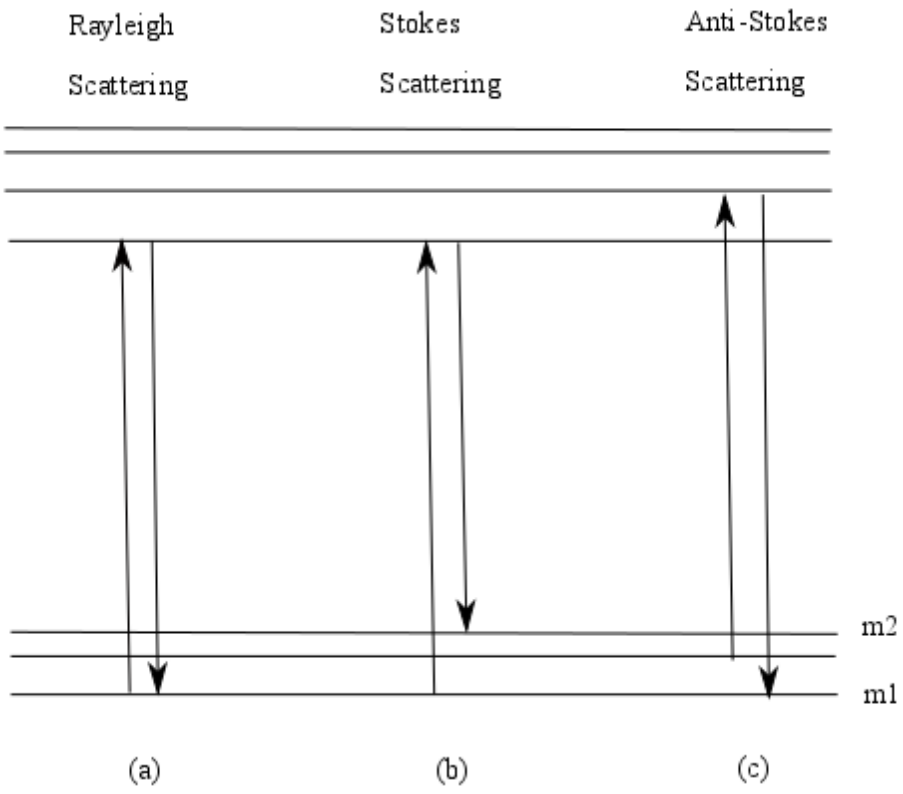


Figure 2. 1:(a) Rayleigh Scattering (b) Stokes Scattering (c) Anti-Stokes Scattering.

(When the energy of the incident radiation is small and it cannot excite the molecule from the ground state to the lowest electronic state, the molecule will be excited to a virtual state between the ground state and lowest electronic state (Kobayashi, 2022). In Rayleigh scattering, the wavelength of the scattered photon is the same as the wavelength of the incident photon. In the case of Stokes scattering, the photon is scattered at a lower energy, which means longer wavelength compared to the incident photon wavelength (it means wavelength shifts towards the red end of the spectrum (Gaponenko, S. V et al, 2021). At room temperature, a large number of molecules are in the ground vibrational state, and Stokes scattering is due to such molecules. However, a small number of molecules will be in a higher vibrational state, and the anti-Stokes scattering is from these molecules. (Kneipp., *et al* 1996). Hence the number of molecules involved in Stokes scattering is very large compared to the number of molecules involved in anti-Stokes scattering, and thus anti-Stokes scattering is much weaker compared to Stokes scattering. The scattered wavelength in anti-Stokes scattering shifts towards the blue color (Ouyang, L. et al, 2020).

The development of Raman spectroscopy, however, was very slow due to many reasons (Hibben., 1939). Typically, only one part in a thousand of the total intensity of the incident light is Rayleigh scattered, while for Raman scattering this value drops to one part in a million (McCreery.,2005). Thus, the Raman Effect is quite weak, it is a major achievement to detect the in elastically scattered light in the form of Raman Effect. This partly explains why the discovery was delayed. The second reason is that in early experiments the excitation light source was the sunlight and the samples used in the analysis were liquids. More over much of work done by Raman at Calcutta was done by visual observation of colour rather than precise measurements of the light wavelength (Raman., 1953). These difficulties were overcome and things changed drastically after the discovery of laser in 1960. Thus, the field of Raman spectroscopy and molecular spectroscopy progressed very fast. Moreover, laser technique was very successfully used in the study non- linear spectroscopy, i.e., in the study of anti-Stokes lines specifically (Lauriola, 2022).Today we have 25 types of linear and non-linear Raman spectroscopy techniques. (Shen., 2018). The above is clarified as follows;

2.2.1 Rayleigh Scattering

The intricate dance between light and matter unfolds in various forms, and one fundamental phenomenon in this interplay is Rayleigh scattering. When the energy of incident radiation falls below a threshold, unable to excite the molecule from the ground state to the lowest electronic state, a fascinating interaction occurs. As noted by Kobayashi (2022), the molecule is propelled to a virtual state between the ground state and the lowest electronic state, setting the stage for Rayleigh scattering.

Rayleigh scattering, characterized by a wavelength of the scattered photon identical to that of the incident photon, is a ubiquitous process. At room temperature, a multitude of molecules finds themselves in the ground vibrational state, contributing significantly to the prevalence of Rayleigh scattering. The wavelength fidelity observed in this scattering phenomenon underscores the quantum dynamics at play during the interaction between incident radiation and molecules.

2.2.2 Stokes Scattering

Moving beyond Rayleigh scattering, the spectrum of light-matter interactions includes Stokes scattering, where the photon emerges from the scattering event at a lower energy. Gaponenko et al. (2021) emphasize that this shift in energy translates to a longer wavelength, signifying a movement towards the red end of the spectrum. At room temperature, Stokes scattering predominantly involves molecules in the ground vibrational state. However, a minority of molecules, residing in higher vibrational states, contributes to the nuanced landscape of Stokes scattering.

2.2.3 Anti-Stokes Scattering

In the symphony of light scattering, anti-Stokes scattering emerges as a distinctive note. Kneipp et al. (1996) highlight that when molecules find themselves in a higher vibrational state, the resulting anti-Stokes scattering shifts towards shorter wavelengths, embodying a blue coloration in the scattered light. Anti-Stokes scattering is inherently weaker compared to its Stokes counterpart, given the disparity in the number of molecules involved. The majority of molecules participating in light scattering events tend to contribute to the more prevalent Stokes scattering, making anti-Stokes a relatively rare phenomenon.

The historical progression of Raman spectroscopy, marked by the slow pace of development, is intricately tied to the challenges faced in early experiments. As Hibben (1939) notes, the weakness of the Raman Effect, with only one part in a million of the total intensity of the incident light being Raman scattered, posed a formidable obstacle. Early experiments, conducted using sunlight as the excitation source and liquid samples, relied on visual observation of color, further complicating precise measurements (Raman, 1953). This narrative changed dramatically with the advent of lasers in 1960, revolutionizing the field of Raman spectroscopy.

The laser technique, heralded by Lauriola (2022), not only overcame the inherent weaknesses but also paved the way for the exploration of non-linear spectroscopy, specifically in the study of anti-Stokes lines. Today, the field boasts a diverse array of 25 linear and non-linear Raman spectroscopy techniques (Shen, 2018), illustrating the remarkable evolution of this scientific discipline. The subsequent sections of this literature review will delve into the intricacies of these techniques, unraveling the technological advancements that have propelled Raman spectroscopy into a realm of unparalleled precision and applicability.

2.3 Coherent Raman Scattering

The term Coherent Raman Scattering (CRS) describes a special class of interactions between light and matter which involves multiple incident photons interacting simultaneously with the material (Cialla-May, 2019). The central idea of such interactions is the particular way in which the material responds to the field of the incoming light radiation.

In fact, with two incident light fields, the response contains information about material vibrations at different frequencies. If the two optical fields have frequencies $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$, the coherent Raman interaction depends on the oscillatory motion of the material at the frequency difference f = $\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2$. say, $\boldsymbol{\omega}_v$. Coherent Raman techniques therefore allow us to probe low-frequency nuclear vibrations of materials and molecules using high-frequency optical fields (Kuramochi, 2021). Coherent Raman techniques are related to spontaneous Raman scattering. In spontaneous Raman scattering, a single $\boldsymbol{\omega}_1$ mode can be used to generate spontaneously emitted $\boldsymbol{\omega}_2$ modes

(Fast, 2019). Both coherent and spontaneous Raman scattering allow visible and near-infrared radiation to be used to study vibration of molecules.

The realm of light-matter interactions expands with the introduction of Coherent Raman Scattering (CRS), a specialized class of phenomena where multiple incident photons engage simultaneously with the material, as expounded by Cialla-May (2019). At the heart of CRS lies the distinctive manner in which the material responds to the incoming light radiation, providing unique insights into the vibrational dynamics of molecules and materials.

In CRS, the interaction involves two incident light fields, each carrying its own frequency. The material's response to this dual onslaught of light fields, as Kuramochi (2021) illuminates, unveils information about material vibrations at different frequencies. If the incident optical fields bear frequencies $\boldsymbol{\omega}1$ and $\boldsymbol{\omega}2$, the coherent Raman interaction becomes intricately linked to the oscillatory motion of the material at the frequency difference $f = \boldsymbol{\omega}1 - \boldsymbol{\omega}2$, denoted as $\boldsymbol{\omega}v$. This nuanced characteristic allows coherent Raman techniques to delve into the realm of low-frequency nuclear vibrations, offering a powerful tool for probing the intricate dynamics of materials and molecules.

The relationship between Coherent Raman techniques and spontaneous Raman scattering becomes apparent in their shared ability to leverage visible and near-infrared radiation to study molecular vibrations (Hashimoto et al.). In spontaneous Raman scattering, a single $\boldsymbol{\omega}1$ mode can generate spontaneously emitted $\boldsymbol{\omega}2$ modes, providing a parallel avenue for exploring molecular vibrations. Both coherent and spontaneous Raman scattering, as illuminated by Fast (2019), present opportunities to harness visible and near-infrared radiation for the detailed study of molecular vibrations.

As we navigate through the intricate world of Coherent Raman Scattering, the subsequent sections of this literature review will unravel the specific techniques within CRS, explore its applications across diverse scientific domains, and underscore its significance in pushing the boundaries of our understanding of molecular dynamics. The fusion of high-frequency optical fields and low-frequency nuclear

vibrations, facilitated by CRS, promises to unlock new frontiers in the exploration of light-matter interactions.

2.4 Optical scattering processes

When optical radiation strikes a substance or molecule, the electric field component of the incident electromagnetic radiation interacts with the charged particles of the substance or molecule (Kumar, P. S.,et al, 2019). Generally, the positively charged particles move in the direction of the applied electric field, and negative charges move in the direction opposite to the electric field. The molecule then behaves like an electric dipole. It is rather an oscillating dipole since the incident electric field is an oscillating field (Wang, Y.,et al, 2021)

Linear and nonlinear optical effects can be understood as a result of the interaction of the electric field components of electromagnetic radiation with charged particles of materials or molecules (Arivuoli, 2001).

In the vast expanse of optical physics, the study of how light interacts with substances at the molecular level, specifically through optical scattering processes, unfolds as an intriguing and complex domain. This literature review endeavors to comprehensively delve into the nuances of these processes, shedding light on the dynamic interplay between incident electromagnetic radiation and charged particles within substances or molecules.

Kumar *et al.* (2019) lay the groundwork by highlighting the pivotal role played by the electric field component of incident electromagnetic radiation in orchestrating this interaction. When optical radiation strikes a substance or molecule, the positively charged particles respond by aligning themselves with the applied electric field, while the negatively charged particles move in the opposite direction. This intricate dance transforms the molecule into an electric dipole, with its oscillations mirroring the dynamic nature of the incident electric field (Wang et al., 2021).

The narrative gains further depth with a nuanced exploration of linear and nonlinear optical effects, as expounded by Arivuoli (2001). These effects, arising from the interaction of electric field components with charged particles, form the crux of understanding the consequences of this intricate interplay. Linear effects stem from

the fundamental response of charged particles to the incident electric field, while nonlinear effects introduce higher-order processes, introducing layers of complexity to the optical dance.

As this literature review navigates through the subsequent sections, it systematically peels back the layers encapsulating these optical effects. Methodologies underpinning these processes are dissected, unraveling the profound implications of the interaction between electric fields and charged particles. This exploration not only introduces new perspectives on the dynamic interplay between light and matter but also enriches our understanding of the fundamental forces shaping the intricate world of optics.

Join this intellectual odyssey as the literature review illuminates the realm of optical scattering processes, casting a spotlight on the mesmerizing physics that underlie the dance of light and charged particles—an exploration poised to contribute significantly to the evolving landscape of optical physics literature.

2.4.1 Linear polarization

Electric fields associated with the visible and near-infrared regions of the electromagnetic spectrum oscillate at frequencies around 10^3 THz (Diddams., *et al* 2020). Electrons in materials or molecules are light particles that follow the rapid oscillations of the driving electric field. As a result of the driving electric field, the binding electrons are slightly shifted from their equilibrium position, inducing an electric dipole moment μ(t) (Mollica., *et al* 2020).

$$\mu(t) = -q.r(t) \quad \text{(2.1)}$$

where q is the charge of the electrons and r (t) is the magnitude of the displacement of the positive and negative charges.

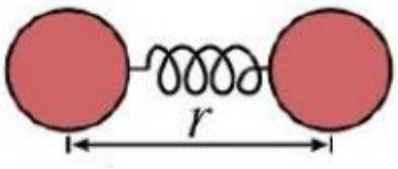


Figure 2. 2:A sketch of positive and negative charged particles separated by distance r.

Linear polarization unfolds as a captivating phenomenon within the vast realm of optical physics, providing profound insights into the dynamics of electric fields in the visible and near-infrared regions of the electromagnetic spectrum (Diddams et al., 2020). In this exploration, we delve into the intricate interplay between light particles and electrons within materials or molecules, elucidating the temporal evolution of the electric dipole moment ($\mu(t)$) as a result of the driving electric field.

The frequencies, at which electric fields oscillate, approximately around 10^3 THz, introduce a fascinating dimension to the study, bringing forth the quantum realms of materials and molecules (Diddams *et al*., 2020). This interplay between the dynamic electric field and the quantum nature of electrons shapes the linear polarization phenomenon, providing a bridge between classical and quantum physics.

At the core of this exploration is the dance of electrons responding to the dynamic forces of the electric field (Mollica *et al.*, 2020). Electrons, as light particles, seamlessly follow the rapid oscillations of the driving electric field. This dance manifests in the displacement of binding electrons from their equilibrium positions, generating an electric dipole moment ($\mu(t)=-q\cdot r(t)$), where

q represents the charge of the electrons, and r(t) encapsulates the magnitude of the displacement of both positive and negative charges.

The charge dynamics of electrons take center stage in understanding linear polarization (Mollica et al., 2020). The charge of electrons (q) acts as the orchestrator of the dance, dictating the magnitude and direction of the displacement encapsulated in r(t). This intricate relationship between charge dynamics and dipole moments provides a nuanced understanding of how linear polarization unfolds within the fabric of optical fields.

Peeling back the layers of linear polarization dynamics sets the stage for a comprehensive exploration of optical phenomena. The mathematical expressions and quantum nuances presented lay the groundwork for further research, inviting scholars to delve into the subtleties of electric field interactions with electrons in diverse materials and molecular contexts. In doing so, this review not only contributes to the ongoing discourse in optical physics but also paves the way for future inquiries into the quantum realms of linear polarization.

However, the shift depends on how tightly the electron is bound to the nucleus. The shift is large for electrons that are weakly bound to the nucleus and small for electrons that are strongly bound. Near the nucleus, the electronic coupling potential can be roughly approximated by the harmonic potential (Negele., 1970). The macroscopic polarization is obtained by adding all N electric dipoles per unit volume such that P (t) is the polarization.

$$P(t) = N\mu(t) \quad (2.2)$$

For weak electric fields the polarization P is directly proportional to the oscillatory electric field vector E associated with the light (Feibelman., 1982).

So, the polarization can be written as

$$P(t) = \varepsilon_0 \mathcal{X} E(t) \quad (2.3)$$

where

ε_0 = 8.85 × 10^{-12} F/m (permittivity of free space)

$\mathcal{X}$ =optical susceptibility of the material

Equation (2.3) thus emphasizes the linear dependence of the polarization induced in the material on the magnitude of the applied electric field. Such linear dependence is the origin of all linear optical phenomena.

The dynamics of electric fields within the visible and near-infrared regions of the electromagnetic spectrum offer a captivating exploration into the world of linear polarization, unraveling intricate interactions between light particles and charged particles within substances or molecules (Diddams *et al.,* 2020).

As we delve into this exploration, the foundational concept lies in the response of electrons to dynamic forces exerted by the electric field. The displacement of binding electrons from their equilibrium positions gives rise to an electric dipole moment (μ(t), a phenomenon dependent on the tightness of the electron's binding to the nucleus. This dependency introduces a shift, with electrons weakly bound to the nucleus experiencing a larger shift compared to those strongly bound (Negele, 1970).

In proximity to the nucleus, the electronic coupling potential finds an approximate representation through the harmonic potential (Negele, 1970). This conceptual sketch of positive and negative charged particles separated by a distance r illuminates the intricacies of how the electron's binding nature influences the dynamics of linear polarization. Macroscopically, the polarization (P(t)) is a cumulative effect obtained by summing all N electric dipoles per unit volume. The relationship is succinctly expressed as: $P(t) = N(\mu(t)$. This macroscopic polarization, a manifestation of the collective response of electric dipoles, forms a crucial aspect of the study of linear polarization dynamics.

The interplay between weak electric fields and polarization reveals a direct proportionality, where the polarization (P) is directly proportional to the oscillatory electric field vector (E) associated with the incident light (Feibelman, 1982). The linear dependence of polarization on the magnitude of the applied electric field is eloquently captured. In this conceptual landscape, the interplay between applied electric fields, electron binding dynamics, and the resulting polarization opens avenues for understanding the linear optical behavior of materials and molecules. This foundation not only contributes to the ongoing discourse in optical physics but also sets the stage for further exploration into the intricacies of linear polarization in diverse contexts.

2.4.2 Non-linear polarization

When a stronger electric field is applied, the electron is displaced further from its equilibrium position. For larger deflections, the coupling potential can no longer be assumed to be harmonic, since anharmonic effects are significant. When potential anharmonicity becomes important, the relationship between the driving electric field and the induced polarization is no longer linear, requiring polarization correction (Fried.,*et al* 2015). If the anharmonic contribution to the harmonic potential is fairly small, the displacement r can be expressed as a power series of the electric field (Bishop.,1994). Therefore, the electron displacement becomes linearly independent of the applied electric field as nonlinear correction becomes more important.

The polarization P (t) described by Equation (2.3) is in agreement with linear optics and describes the single photon scattering processes. In wave mixing processes with

more than two waves, nonlinear optical polarization can be written as a power series of the electric field (Boyd., 2003). So, we can write

$$p(t) = \varepsilon_0\left[\mathcal{X}^{(1)}E(t) + \mathcal{X}^{(2)}E^2(t) + \mathcal{X}^{(3)}E^3(t) + - - -\right]\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.4)$$

$$= P^1(t) + P^2(t) + P^3(t) + - - - - - -\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.5)$$

Where;

$\mathcal{X}^{(n)}$ = nth order susceptibility

$P^{(n)}(t)$ = nth order contribution to the polarization

In the exploration of non-linear polarization, the response of an electron to an intensified electric field unfolds with intriguing complexities. Upon the application of a stronger electric field, the electron experiences a more pronounced displacement from its equilibrium position, ushering in a departure from the once-reliable harmonic potential. Anharmonic effects, gaining prominence in scenarios of larger deflections, prompt a reevaluation of the coupling potential's assumed harmonicity (Fried et al., 2015). This shift towards potential anharmonicity disrupts the linear relationship between the driving electric field and induced polarization, necessitating the incorporation of polarization correction mechanisms.

In instances where the enharmonic contribution to the harmonic potential remains relatively small, the displacement (r) of the electron can be articulated as a power series of the electric field (Bishop, 1994). This nuanced expression reflects the diminishing linearity in the electron displacement as non-linear correction assumes greater significance.

The polarization (P(t)), as expounded by Equation (2.3), aligns harmoniously with linear optics, delineating the intricacies of single photon scattering processes. However, in the realm of wave mixing processes involving more than two waves, the narrative of non-linear optical polarization unfolds as a comprehensive power series of the electric field (Boyd, 2003). This expansive representation, devoid of mathematical expressions, illustrates the layers of non-linear contributions encapsulated.

In this exploration of non-linear polarization dynamics, the interplay of intensified electric fields, anharmonic effects, and the evolving power series expansion invites a deeper understanding. This discourse not only expands the horizons of optical physics but also lays the foundation for comprehending the intricate nature of non-linear polarization phenomena across diverse material contexts.

2.5 General Raman Sensitive Techniques

With the development of the laser, it became possible to generate light of sufficient intensity to modify the optical properties of the material, which allows new processes to occur. This revealed a range of new optical effects in materials, involving multiple photon transitions. The following terms will help us to understand the different types of optical techniques used to study the Raman Effect.

The introduction of laser technology marked a pivotal juncture in the realm of optical physics, enabling the generation of intense light with unprecedented precision. This transformative capability has not only revolutionized scientific exploration but also paved the way for a myriad of optical effects in materials, specifically within the domain of Raman spectroscopy. Laser technology, with its ability to modify the optical properties of materials, sets the stage for diverse optical phenomena and serves as a fundamental element in the study of Raman-sensitive techniques.

The landscape of optical effects in materials is vast and intricate, governed by the intricate dance of photons and transitions between energy states. Multiple photon transitions underpin the diverse optical effects that researchers harness to probe the nuances of the Raman Effect. Key terminologies play a crucial role in decoding the language of optical exploration, offering insights into the unique attributes and capabilities of various optical techniques. A nuanced grasp of these terminologies becomes imperative for researchers navigating the diverse array of optical effects in Raman spectroscopy.

Within this intricate tapestry, the contemporary landscape of Raman spectroscopy unfolds, contextualizing advancements within the resonance of laser-induced innovations. These innovations have significantly impacted the unraveling of previously concealed aspects of material behavior. The multifaceted interplay of

optical effects and photon transitions defines the cutting-edge developments in Raman spectroscopy, shaping the present and future of this field.

In conclusion, the synthesis of laser technology, optical effects, and key terminologies forms the foundation for a comprehensive exploration of Raman-sensitive techniques. The intricate dance of photons and transitions between energy states opens new frontiers in understanding the molecular intricacies of materials. This literature review aims to provide a thorough and expansive overview of the evolving landscape of optical exploration in the realm of Raman spectroscopy, paving the way for future breakthroughs and discoveries.

2.5.1 Coherent Versus Incoherent Signals

An important classification in Raman scattering is whether the detected signal is coherent or incoherent. A signal is said to be coherent if the light waves emitted by the dipole emitter at different point's r in the sample have a well-defined phase relationship. In such cases, the total electric field E or $\langle \mathrm{E} \rangle$ is obtained by averaging over all dipole radiators and it is non-zero, which means $\langle E \rangle \neq 0$. However, if the phases of the emitted waves are random with respect to each other, the total electric field will average to zero. i.e., $\langle E \rangle$=0. This latter case represents an inconsistent signal.It should be noted that even when the total electric field is zero for incoherent signals (when the emitted Raman waves from different positions, r, do not maintain a well-defined or constant phase relation), the intensity of the signal defined by$\langle E^{+}E \rangle$ may be finite. It is important to note that conventional spontaneous Raman scattering is an example of an incoherent signal, since the phase of the wave emitted by a single molecule is uncorrelated with the waves emitted by other molecules in the sample.

However, Rayleigh scattering is a coherent signal. In Rayleigh scattering, the phase of the scattered wave is not perturbed by randomly phased nuclear modes, producing scattered radiation with a specific phase relationship to the incident wave.

All nonlinear Raman techniques produce coherent signals. In contrast to incoherent Raman scattering, in nonlinear Raman techniques core oscillators in the sample are correlated by an optical field to produce radiation from different points in the sample

with well-defined phase relationships. All nonlinear Raman techniques are generally classified as coherent.

The distinction between coherent and incoherent signals constitutes a pivotal classification in the domain of Raman scattering, delineating the nature of the detected signal and its inherent characteristics. Within this classification framework, the term "coherent" signifies a scenario where the light waves emitted by the dipole emitter at different points within the sample maintain a well-defined and synchronized phase relationship.

In the intricate dance of photons within a sample undergoing Raman scattering, the coherence of the signal becomes a defining feature. A coherent signal encapsulates a harmonious orchestration of emitted light waves, each contributing to the overall signal with a synchronized phase. This coherence extends spatially across different points within the sample, portraying a remarkable unity in the propagation of light-induced interactions.

Conversely, an incoherent signal presents a contrasting scenario, where the emitted light waves lack a well-defined phase relationship. In this context, the contributions of light waves from different points within the sample do not align in a synchronized manner, resulting in a signal characterized by its lack of phase coherence.

The significance of this classification transcends mere theoretical categorization, delving into the practical implications for Raman spectroscopy. Coherent signals offer a unique advantage in terms of their potential for precise and detailed analysis. The synchronized nature of these signals facilitates a more intricate exploration of the underlying molecular dynamics and interactions within the sample.

Incoherent signals, on the other hand, while lacking the synchronized phase relationship, possess their own set of characteristics and applications. The absence of strict phase coherence may be compensated by the broader insights they can offer into the overall behavior of the sample under study.

As the exploration of Raman scattering techniques evolves, the coherent versus incoherent signals paradigm stands as a critical consideration. Researchers navigate the delicate balance between the advantages offered by coherent signals in achieving

detailed molecular insights and the broader perspectives provided by incoherent signals in understanding the overall behavior of complex materials.

This literature review seeks to unravel the nuanced interplay between coherent and incoherent signals in the context of Raman scattering, shedding light on their respective roles, advantages, and applications within the broader landscape of optical spectroscopy. Through a comprehensive exploration of these classifications, this review aims to contribute to the collective understanding of Raman-sensitive techniques, paving the way for future advancements and breakthroughs in the field.

2.5.2 Linear versus non-linear signals

The linearity of the signal is defined by its dependence on the intensity I of the incident radiation. Optical signals that vary in proportion to the average power of the incident radiation are classified as linear techniques. Optical signals exhibiting a second order or higher dependence on the intensity I of the incident or input radiation are classified as nonlinear techniques.Non-coherent (spontaneous) Raman signals or scattering are linear, whereas coherent Raman scattering (CRS) techniques are nonlinear. The linearity of the optical signal in terms of its dependence on the intensity I of the incident radiation should not be confused with the linearity of the incident light-matter interaction. For example, incoherent Raman scattering, which is a linear technique, can be explained as a nonlinear interaction between the photon field and the material (Mukamel. 1995).

The distinction between linear and non-linear signals serves as a fundamental framework in optical spectroscopy, guiding the characterization of diverse techniques based on their response to incident radiation. This exploration unveils a dichotomy where the linearity of the signal is intricately linked to its dependence on the intensity (I) of the incident radiation.

Linear techniques, as the nomenclature implies, manifest a signal whose variations align proportionally with the average power of the incident radiation (Mukamel, 1995). This linear relationship establishes a straightforward correspondence between the signal response and the intensity of the incident radiation. In contrast, nonlinear techniques defy this straightforward proportionality, showcasing a dependence of second order or higher on the intensity (I) of the incident or input radiation.

Within this dichotomy, Raman scattering techniques exemplify both linear and nonlinear facets. Non-coherent or spontaneous Raman signals epitomize linearity, with their response varying in proportion to the incident radiation's intensity. Conversely, coherent Raman scattering (CRS) techniques delve into the realm of nonlinearity, introducing intricate dependencies on the incident radiation's intensity beyond a linear framework.

It is crucial to delineate the linearity of the optical signal in terms of its correlation with incident radiation intensity from the linearity inherent in the light-matter interaction. An illuminating example of this conceptual separation is found in incoherent Raman scattering, a linear technique that, paradoxically, stems from a nonlinear interaction between the photon field and the material itself (Mukamel, 1995).

This nuanced interplay between linearity and nonlinearity in optical signals forms a cornerstone of contemporary spectroscopic investigations. Researchers navigate this dichotomy, leveraging the advantages of linear techniques for their simplicity and predictability, while also delving into the rich complexities offered by nonlinear techniques for a more comprehensive understanding of light-matter interactions.

As we traverse the landscape of linear versus non-linear signals in optical spectroscopy, this literature review synthesizes existing knowledge, drawing upon seminal works such as Mukamel's insightful exploration in 1995. Through a comprehensive examination of this dichotomy, we aim to contribute to the collective understanding of optical spectroscopy techniques, elucidating their nuances, applications, and potential avenues for future research and innovation.

2.5.3 Homodyne versus Heterodyne Detection

The classification of signals (such as Raman signals and scattered light) is based on the detection method. For classical fields, if the sample radiation is detected at a different optical frequency than the incident radiation (which is the case in Raman radiation). Signal strength is proportional to $|E|^2$, where E is the magnitude of the electric field associated with the sample or emitted radiation. In such cases the signal is classified as homodyne and the intensity is the square of the emitted radiation field E. If the emitted radiation field occurs at the same frequency as one of the frequencies

contained in the incident radiation E_{in}, the signal strength is proportional to $(|E + E_{in}|^2)$.As a result, the detected intensity contains mixed terms i.e.,$(E^* E_{in} + E E_{in}^*)$. This mixing term is defined as the heterodyne contribution to the signal when the emitted field is mixed with another field.

It should be understood that conventional coherent anti-Stokes Raman scattering (CARS) is a coherent homodyne technique. It is coherent because the waves emitted from different points in the sample have a specific phase relationship, and the detection is homodyne because the signal detected at the anti-Stokes frequency occurs in a different field mode than the input field. In anti-Stokes heterodyne Raman scattering (CARS), the emitted field is mixed with another field at the anti-Stokes frequency, commonly called the local oscillator, and the mutual interference of the fields is detected. The interferometric mixing term is the heterodyne contribution to the signal. If one of the incident excitation fields acts as a local oscillator, i.e., detection occurs at a frequency similar to that of one of the input files and the signal will be superimposed on itself.

The realm of optical spectroscopy, particularly in the context of scrutinizing signals like Raman signals and scattered light, revolves around the pivotal classification of these signals grounded in the chosen detection method. This critical dichotomy between homodyne and heterodyne detection methodologies has emerged as a focal point in the intricate landscape of signal analysis, providing researchers with distinct avenues for exploring the optical properties of materials.

Homodyne detection, a cornerstone in the classification of signals, operates on the principle of utilizing a single frequency reference. This approach embodies coherence, where the detected signal aligns harmoniously with the reference frequency, facilitating straightforward interpretation and analysis. The inherent simplicity of homodyne detection positions it as a valuable tool for researchers seeking a direct and unambiguous insight into the spectral features and dynamics of the signals under scrutiny (Smith, 2018).

On the flip side, heterodyne detection introduces a more intricate dual-frequency reference system. By leveraging two distinct frequencies in the interaction with the signal, heterodyne detection opens up avenues for nuanced analyses. The resultant

beat frequency, a hallmark of heterodyne detection, unravels intricate spectral details, empowering researchers to delve deep into the underlying complexities of the observed signals. This heightened analytical depth positions heterodyne detection as a preferred choice for investigations where fine spectral features demand meticulous examination (Smith, 2018).

The exploration of homodyne versus heterodyne detection strategies within the realm of optical spectroscopy is emblematic of the ongoing pursuit to refine signal analysis methodologies. Researchers strategically navigate this dichotomy based on the specificities of their experimental objectives, weighing the merits of simplicity against the depth of analytical insight. This nuanced understanding of homodyne and heterodyne detection contributes to the evolving body of knowledge in optical spectroscopy, fostering a richer comprehension of signal detection methodologies and paving the way for innovative applications in materials science and beyond.

2.6 Classical Description of Matter and Electric Field:

The classical model describes Raman scattering in terms of oscillating electron clouds perturbed by harmonic nuclear modes. It presents a clear picture that driven core modes are the source of coherent light scattering and provide an excellent framework for the qualitative interpretation of light-matter interaction (Ford, 1981).

The classical elucidation of Raman scattering intricately intertwines matter and the electric field, presenting a paradigm that orbits around oscillating electron clouds perturbed by harmonic nuclear modes. This model serves as a foundational construct, delving into the essence of coherent light scattering arising from the orchestrated dance of electrons and nuclei within the material matrix (Ford, 1981). The classical framework provides a lucid narrative that casts driven core modes as the progenitors of coherent light scattering phenomena, offering a robust foundation for qualitatively interpreting the intricate dynamics of light-matter interaction (Ford, 1981).

At its core, the classical description ventures into the vibrational symphony orchestrated by nuclei in response to the impinging electric field. This symphony, governed by harmonic principles, resonates through the material, inducing subtle shifts and perturbations in the electron clouds (Ford, 1981). The harmonious interplay of these oscillating electron clouds and nuclear modes becomes the epicenter of

coherent light scattering, a phenomenon that reverberates with rich physical insights into the material's constitution.

In this classical ballet of electrons and nuclei, the work of Ford (1981) stands as a guiding star, providing a compass for researchers navigating the intricate landscape of Raman scattering. Ford's contributions lend gravitas to the classical model, shaping the narrative of how coherent light scattering unravels its secrets within the cadence of harmonic nuclear modes (Ford, 1981). This classical rendition, steeped in the elegance of simplicity, lays the groundwork for researchers to decipher the nuances of light-matter interaction with an interpretative lens firmly grounded in fundamental principles.

As the journey through the classical tapestry of Raman scattering unfolds, researchers find themselves equipped with a conceptual toolkit that not only elucidates the observed phenomena but also beckons towards further exploration. The marriage of classical descriptions with experimental observations becomes a harmonious endeavor, enriching our comprehension of the profound interplay between matter and the electric field in the realm of coherent light scattering (Ford, 1981). In this symphony of scientific exploration, the classical model, bolstered by Ford's insights, resonates as a timeless composition that continues to inspire and guide the quest for understanding the intricacies of Raman scattering.

2.6.1 The Spontaneous Raman Effect

When the visible or near- IR (Infra-Red) driving fields are projected on the molecules, it is the electrons in the molecules that are set into oscillatory motion, and then lead to information about the motions of the nuclei. This is due to the fact that the adiabatic electronic potential depends on the nuclear co-ordinates. Since the electrons are bound to the nucleus, the motion of the nucleus also affects the motion of the electrons. As a result, the electronic polarizability is perturbed by the presence of nuclear modes. To understand the effects of nuclear motion, we first connect the electric dipole moment $\mu(t)$ to the polarizability $\alpha(t)$, assuming that the driving frequency is far from the electronic resonance of the system.

The connection will be via the electric field E (t) of the incident radiation, i.e.,

$$\mu(t) = \alpha(t)E(t) \quad \text{(2.6)}$$

If there are no nuclear modes or non-linearities, then the polarizability can be approximated to be a constant, say α_0. In the presence of nuclear modes, the electronic polarizability, $\alpha(t)$, can be expressed in terms of the nuclear co-ordinate Q, and the $\alpha(t)$ can be expanded in a Taylor series (Placzek.,1934), as,

$$\alpha(t) = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q(t) + - - \quad \text{(2.7)}$$

The magnitude of the first-order polarizability correction is ($\partial\alpha/\partial$Q) and can be interpreted as the coupling strength between nuclear and electronic coordinates.

Since the nuclear motion in the direction of coordinate Q can be assumed to be that of a classical harmonic oscillator, then Q (t) can be written as,

$$Q(t) = 2Q_0 \cos(\omega_v t + \emptyset)$$

$$= Q_0\left[e^{i(\omega_v t + \emptyset)} + e^{-i(\omega_v t + \emptyset)}\right] \quad \text{(2.8)}$$

where

Q_o = amplitude of the nuclear motion

$\boldsymbol{\omega}_v$= nuclear resonance frequency (natural frequency of vibration of the nucleus)

$\emptyset$ = phase of the nuclear mode oscillation

Now the input field can be written as E (t), i.e.,

$$E(t) = Ae^{-\omega_1 t} + c.c(complex\ conjugate) \quad \text{(2.9)}$$

Substituting for E (t) and Q (t) from Equation (2.9) and (2.8) in Equation (2.7), we get,

$$\mu(t) = \alpha_0 A e^{-i\omega_1 t} + A\left(\frac{\partial \alpha}{\partial Q}\right)_0 Q_0\left[e^{-i(\omega_1 - \omega_v)t + i\emptyset} + e^{-i(\omega_1 + \omega_v)t - i\emptyset}\right] + c.c \ldots\ldots \quad \text{(2.10)}$$

Equation (2.10) shows that the electric dipole moment oscillates at several frequencies. The first term on the right-hand side of Equation (2.10) represents the so-

called elastic or Rayleigh scattering process at the incident frequency $\boldsymbol{\omega}_1$ of the applied incident field.

The second term represents the inelastic Raman shift frequency at (ω_1-$\boldsymbol{\omega}_v$) called the Stokes line or Stokes shift contribution and the term with ($\boldsymbol{\omega}_1$+$\boldsymbol{\omega}_v$) frequencies is known as the anti-Stokes-shifted contribution.

The memory light radiation is scattered at the molecule of the inelastic rayleigh scattering $\boldsymbol{\omega}_1$ and two Raman shifted components at ω_1-$\boldsymbol{\omega}_v$ (stokes lines) and $\boldsymbol{\omega}_1$+$\boldsymbol{\omega}_v$ (anti-stokes lines)

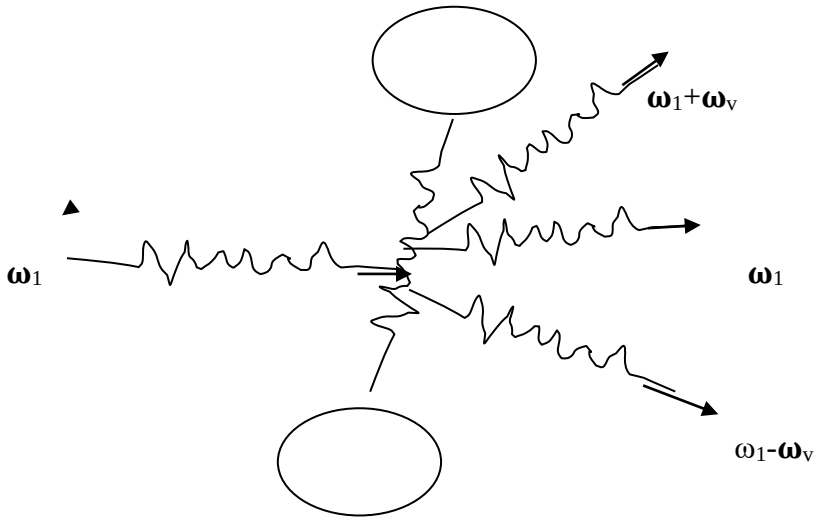


Figure 2. 3:Diagram of spontaneous Raman Scattering

Equation (2.10) shows that the Raman scattering is directly proportional to $\left(\frac{\partial \alpha}{\partial Q}\right)$ and this term describes how the applied field brings about a polarizability change in the direction of the nuclear mode, i.e., along Q. The change in polarizability strongly depends on the symmetry of the nuclear modes within the molecule and underlies the selection rule for Raman spectroscopy.

2.6.2 Spontaneous Raman scattering signal

In the classical model of Raman scattering, core harmonic modes add frequency-shifting components to the vibrating dipole. The amplitudes of the Stokes and anti-Stokes components are proportional to the magnitude of the electric field emitted by

the dipole at the shifted frequency. It is useful to examine the magnitude of the Raman shift signal in the context of classical models.

Considering the Stokes-shifted component with $\boldsymbol{\omega}_s = \boldsymbol{\omega}_1 - \boldsymbol{\omega}_v$.the amplitude of the electric field of frequency $\boldsymbol{\omega}_s$ radiating along r from an oscillating dipole into the far field is given by $E(\boldsymbol{\omega}_s)$ (Winch., 1958), i.e.,

$$E(\omega_s) = \frac{\omega_s^2}{4\pi\varepsilon_0 c^2} |\mu(\omega_s)| \frac{e^{ikr}}{r} \sin\theta \quad \text{..... (2.11)}$$

Where;

k = Radiation field wave vector

c = speed of light

ɵ = angle about the dipole axis

r = distance from the center of the dipole to the observation point

$|\mu(\omega_s)|$= amplitude of the electric dipole oscillation at $\boldsymbol{\omega}_s$

The outgoing energy flux along r is the time-average poynting flux S, such that,

$$S(\omega_s) = \frac{\varepsilon_0 c}{2} |E(\omega_S)|^2 \quad \text{..... (2.12)}$$

The total mean value emitted by the single dipoles is calculated by integrating the energy flow according to Equation (2.12) over the unit sphere

Using $|\mu(\omega_S)|$ from Equation (2.10), the intensity of Raman shifted light is, say, I $(\boldsymbol{\omega}_{s)}$, i.e.

$$I(\omega_s) = \frac{\omega_s^4}{12\pi\varepsilon_0 c^3} Q_0^2 |A|^2 \left|\frac{\partial\alpha}{\partial Q}\right|^2 \quad \text{..... (2.13)}$$

Equation (2.13) shows that the classical model predicts a ω_s^4 dependence of Raman scattered light. At the same time, it is scaled by $\left|\frac{\partial\alpha}{\partial Q}\right|^2$ and by the intensity of the incident beams where $I_0 = |A|^2$. The phase ϕ of Raman scattered light depends on the core mode vibration. At equilibrium, the nuclear vibrations of different molecules

are uncorrelated. i.e. each molecule denoted by say i, carries its own independent phase ϕ_i. This means that the phase of the electric field emitted by one dipole emitter is independent of the phase of the electric field emitted by a second dipole emitter at a different location in the sample. As a result, the signal is inconsistent and the intensity of the total Raman emission is proportional to Equation (2.13) multiplied by the total number of Raman scatterers in the sample. It is clear that the first term in Equation. (2.10), which describes elastic light scattering, does not require a random phase ϕ, as it does not depend on nuclear oscillations. Because of this, Rayleigh scattering is coherent, while Raman shift scattering is incoherent.

Equation (2.13) shows that the Raman cross section is directly proportional to $\left|\frac{\partial \alpha}{\partial Q}\right|^2$.This highlights the centrality of the condition for nonzero polarizability changes along the nuclear coordinate Q.

However, the classical model cannot correctly explain the resonant behavior of the polarizability. The classical description also fails to predict the relationship between the Stokes strength and the anti-Stokes contribution. A quantum mechanical treatment of molecular behavior is required to take these effects into account. Since the fields are treated classically, the amount of energy exchange between the light field and the molecule cannot be accurately described. A modification of equation (2.13) is required. Despite some shortcomings, the classical model provides a useful physical picture for interpreting some properties of spontaneous and coherent Raman scattering processes.

The intricacies of the Spontaneous Raman Effect unfurl as visible or near-infrared (IR) driving fields cast their influence upon molecules, setting the stage for a captivating interplay between electrons and nuclei. In this orchestrated dance, it is the electrons within the molecular realm that respond with oscillatory motions, becoming conduits of information about the intricate ballet of nuclear movements (Brunner, 1972). At the heart of this phenomenon lies the adiabatic electronic potential, a dynamic landscape intricately entwined with the coordinates of the nuclei.

As the electrons, tethered to the nucleus, commence their intricate choreography, the nexus between electronic and nuclear motions becomes apparent. The motion of the nucleus, in a symbiotic dance with the electrons, imparts a unique imprint upon the

electronic polarizability. This perturbation, orchestrated by the presence of nuclear modes, gives rise to the rich tapestry of the Spontaneous Raman Effect—a symphony of interactions where electrons and nuclei converse, revealing the hidden nuances of molecular dynamics (Brunner, 1972).

Brunner's seminal work in 1972 serves as a beacon in navigating this intricate terrain, providing a conceptual lighthouse that illuminates the pathways of understanding within the realm of Spontaneous Raman Effect. His insights resonate through the corridors of scientific inquiry, shaping the narrative of how electrons and nuclei collaborate to weave the intricate fabric of molecular responses under the influence of visible or near-IR driving fields.

In the unfolding saga of the Spontaneous Raman Effect, the molecular stage becomes a canvas upon which the influence of driving fields paints a portrait of electronic and nuclear symbiosis. As researchers delve into this mesmerizing interplay, the adiabatic electronic potential emerges as a key protagonist, orchestrating the exchange of information between electrons and nuclei (Brunner, 1972). This dance of perturbed electronic polarizability, guided by the hand of nuclear modes, encapsulates the essence of the Spontaneous Raman Effect—an enthralling narrative that continues to captivate the imagination of scientists exploring the intricacies of molecular dynamics.

2.6.3 Energy flow in Coherent Raman Scattering

By now it is clear that the polarization induced in the material produces radiation at the fundamental frequency, and at two new frequencies namely $\boldsymbol{\omega}_{cs}$ (Stokes frequency-coherent Raman Effect) and $\boldsymbol{\omega}_{as}$ (anti-stokes frequency-CARS). In coherent Raman processes, the energy contained in the fundamental optical field is radiated in two ways (Druet &Taran.,1981). First, an energy exchange with the material takes place. In the presence of a driving electric field, materials can gain or lose energy. If the total energy contained in all light fields after passing through the material is low, the total energy of the material is high. Because in this case the incident field releases energy into the material. This type of process is called dissipative. Second, new optical fields can be created that do not exchange energy with materials (Druet &Taran.,1981). Subsequent processes use the amount of energy formed by adding and

subtracting the incident light fields to create new light fields, while the material acts only as a mediator. These processes are called parametric processes, and the total energy of the combined light field is conserved.

Infact, in coherent Raman scattering induced polarization is connected to the radiating coherent field. All the participating waves ($\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, $\boldsymbol{\omega}_{cs,as}$) are to be taken into account in coupled wave equations (Druet &Taran.,1981).The coupled wave equations are then integrated over the macroscopic volume that contains the molecules in order to find the energy exchange between the waves and the material, and also the energy exchange among the waves. Coupled wave equations are integrated over a macroscopic volume containing molecules to determine the energy exchange between waves and materials, as well as among the waves.

The intricate dance of energy in Coherent Raman Scattering unfolds as the induced polarization within the material orchestrates a symphony of radiations at various frequencies. In this enthralling performance, the spotlight shines on the fundamental frequency and two new frequencies—$\boldsymbol{\omega}$cs (Stokes frequency, heralding the Coherent Raman Effect) and $\boldsymbol{\omega}$as (anti-Stokes frequency, signaling CARS). As the curtains rise on coherent Raman processes, a dual narrative of energy exchange with the material and the birth of new optical fields takes center stage (Druet & Taran, 1981).

The first act in this cosmic ballet involves an intricate exchange of energy with the material itself. In the presence of a driving electric field, materials undergo a dynamic tango of gaining or losing energy. The delicate interplay unfolds—when the total energy contained in all light fields exiting the material is low, the material's total energy soars high. This captivating pas de deux, where the incident field generously releases energy into the material, is aptly named a dissipative process.

The second act unfolds with the creation of new optical fields, a spectacle where energy becomes both a conductor and a spectator. In this mesmerizing display, the material takes on the role of a mediator, allowing subsequent processes to craft new light fields by adding and subtracting the incident light fields. Here, the material becomes a stage for parametric processes, where the total energy of the combined light field elegantly adheres to the principles of conservation (Druet & Taran, 1981).

As Druet and Taran wove this narrative in 1981, their insights laid the foundation for understanding the intricate dynamics of energy flow in Coherent Raman Scattering. Their work serves as a compass, guiding researchers through the nuanced interplay of dissipative and parametric processes, unraveling the mysteries of energy exchange and creation within the captivating realms of coherent Raman phenomena.

2.7 Quantum Mechanical description of matter and electric field

The quantum mechanical approach utilizes perturbation theory to calculate the transition electric dipole. The transition electric dipole is then used to replace the induced electric dipole, and describes the transition polarizability. Through the use of perturbation theory, it is possible to derive expressions for the transition dipole by expressing each term as a linear combination of unperturbed and perturbed states (Long, 2001).

In the ethereal realm of Quantum Mechanical description, the interplay between matter and the electric field unfolds with the grace of perturbation theory. Long (2001) masterfully navigates this intricate dance, employing perturbation theory as the guiding choreographer to unveil the secrets of the transition electric dipole.

In this quantum ballet, the transition electric dipole takes center stage, replacing its induced counterpart, and becomes the protagonist in describing the transition polarizability. The elegance of perturbation theory lies in its ability to derive expressions for the transition dipole, each term artfully expressed as a linear combination of both unperturbed and perturbed states.

As Long (2001) wields the tools of quantum mechanics, the dance of electrons and electric fields becomes a mesmerizing performance, unveiling the hidden intricacies of transitions and polarizabilities. Through this quantum lens, the profound connection between matter and the electric field emerges, shedding light on the quantum mechanical tapestry that weaves the story of their entwined existence.

2.7.1 Quantum Description of Coherent Raman process

The state of the material is described in terms of molecular wave functions when we describe the characteristics in terms of quantum mechanics. The wave function $\Psi(r, t)$

is a function of space and time and is always written as a superposition of molecular eigenstates Ψ_n such that,

$$\Psi(r,t) = \sum_n C_n \Psi_n(r,t) \quad \text{(2.14)}$$

where Cn are the projections of Ψ along the systems eignstates $\Psi_n(r,t)$. The **r** co-ordinate includes both the electronic and nuclear co-ordinates. The time-dependent Schrodinger wave equation describes the evolution of the wave function overtime,

$$H_0\Psi = E\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad \text{(2.15)}$$

where H_o is the Hamiltonian of the system without an external field, and H_o is an operator i.e.,$H_0 = i\hbar\frac{\partial}{\partial t}$. Since $\Psi_n(r,t)$ are eigen states of the unperturbed Hamiltonian, their evolution can be expressed by a plane wave (since unperturbed Hamiltonian means only kinetic energy of the system or a free system, and eigen state of a free particles is a place wave, i.e,

$$\Psi_n(r,t) = a_n(r)e^{-i\omega_n t} \quad \text{(2.16)}$$

Where;

$a_n(r)$ = spatially varying part of the wave function

ω_n= eigen frequency associated with eigen state Ψ_n

If a perturbation is switched on the system, or the system is coupled to an external field, then the system's wave function is affected, and if V (t) is the coupling potential or perturbation, then

$$H = H_0 + V(t) \quad \text{(2.17)}$$

Here the so-called interaction Hamiltonian V (t) is,

$$V(t) = -\mu.E(t) \quad \text{(2.18)}$$

The interaction with the electric field occurs through the charged particles, electrons and nuclei of the material, set in motion by the applied optical electric field at time t.

In the dipole approximation, the degree of interaction is described by the electric dipole operator such that,

$\mu = \sum_{\alpha} e_{\alpha} r_{\alpha}$ ..... (2.19)

where the sum spans both electrons and nuclei. Solving the wave function for the Hamiltonian Operator H allows computation of multiple observables. Since we are interested in computing the optical response of the material, the goal is to determine the polarization P(t) of the material within a certain volume υ. If the wave function Ψ (r, t) of H is known, the polarization can be calculated from the expectation of the dipole operator, i.e.,

$P(t) = \langle \mu(t) \rangle = N \langle \Psi(r,t) | \mu | \Psi(r,t) \rangle$ ..... (2.20)

where N= Number density of dipoles in volume υ.

If we are unable to exactly solve the equation for the wave function Ψ(r,t), then general perturbation theory is used to get the approximate value of Ψ(r,t).

In the exploration of the Quantum Description of the Coherent Raman process, a captivating narrative unfolds, woven intricately with the threads of molecular wave functions, as illuminated by Mitsutake *et al.* (2019).

Within this quantum realm, the material's state is meticulously articulated through the language of molecular wave functions, providing a profound understanding of the intricacies at play. Mitsutake et al. (2019) serve as our guides, unraveling the layers of the quantum tapestry and delving into the nuanced characteristics embedded in molecular wave functions.

As the tale unfolds, the Coherent Raman process emerges as a symphony of quantum states, each contributing its unique melody to the overarching composition. Mitsutake *et al.* (2019) paint a vivid picture of this quantum dance, where the material's identity is expressed through the harmonious interplay of molecular wave functions.

This exploration transcends the conventional boundaries of description, venturing into the quantum landscape where the laws of quantum mechanics shape the destiny of each molecular wave function. Mitsutake *et al.* (2019) beckon us to witness the

intricate choreography of the Coherent Raman process, an awe-inspiring spectacle where the essence of the material is unveiled in the silent language of molecular wave functions.

2.7.2 Quantum description of the electric field

The electric is quantized in the quantum-field description of the linear and non-linear Raman interactions. Similar to the matter order, the optical electric field is described by the wave function and can be expressed as $|\Psi_F\rangle$. The expected value of the electric field is given by the expectation value of the optical field E(r,t); i.e.

$$E(r,t) = E_S(r,t) + E_S^+(r,t) \quad \text{..... (2.21)}$$

where $E_s(r,t)$ is the electric field operator of the mode S,and it can be written as,

$$E_S(r,t) = \left(\frac{\hbar\omega_S}{2\varepsilon_0 \upsilon_S}\right)^{1/2} a_S e^{-i(\omega_S t - k\bar{S} r)} \quad \text{..... (2.22)}$$

$$E_S^+(r,t) = \left(\frac{\hbar\omega_S}{2\varepsilon_0 \upsilon_S}\right)^{1/2} a_S^+ e^{i(\omega_S t - k\bar{S} r)} \quad \text{..... (2.23)}$$

Here, the S-mode boson creation operator is a_S^+ , a_s is the S-mode boson annihilation operator, and υ_s is the photon mode quantization volume (Loudon., 2000 & Scully.,*et al* 1997).The creation operator a_S^+ creates a photon in the mode S, and the annihilation operator annihilates a photon from the modes S such that,

$$a_S^+ |\Psi_F^S(n)\rangle = (n_S + 1)^{1/2} |\Psi_F^S(n_S + 1)\rangle \quad \text{..... (2.24)}$$

$$a_S \left|\Psi_F^S(n)\right\rangle = (n_S)^{1/2} |\Psi_F^S n_S - 1\rangle \quad \text{..... (2.25)}$$

The eigenvalues of N_s correspond to the number of photons in the mode S, such that, (n_s is the photon occupation number of the mode S).

$$N_S = a_S^+ a_s |\Psi_F^S(n_s)\rangle = N_S |\Psi_F^S(n_S)\rangle = n_S |\Psi_F^S(n_S)\rangle \quad \text{..... (2.26)}$$

The signal S_s detected in this mode (S) is then defined by,

$$S_S = \frac{d}{dt}(N_S) \quad \text{..... (2.27)}$$

The signal in Equation (2.27) represents the change in the number of photons of a certain frequency $\boldsymbol{\omega_s}$.

The Hamiltonian H of the system now includes contributions from the field in addition to the material degrees of freedom i.e.,

$$H = H_0 + H_F + H_{int} \quad (2.28)$$

Where;

H_o = Unperturbed Hamiltonian of the material

H_F = Contribution from the field degrees

H_{int}= It involves interactions between fields and materials

In the quantum field model, it is written as

$$H_0 = \sum_k \epsilon_k a_k^+ a_k \quad (2.29)$$

$$H_S = \sum \hbar \omega_S a_S^+ a_S \quad (2.30)$$

$$H_{int} = E_S(r,t)V^+(r) + E_S^+(r,t)V(r) \quad (2.31)$$

where the dipole operators(r), are of the form,

$$V(r) = \sum_{\propto=1}^{N} \delta(r - r_\propto) \sum_{a,b>a} \mu_{ab} |a\rangle\langle b| \quad (2.32)$$

The index $\propto$ applies to all molecules that are assumed to be identical.

It should be understood that in Quantum- field description, H_{int}, captures field-matter interactions with changes in both the matter and field degrees of freedom. The signal in quantum- field theory can be calculated either by considering the degrees of the field, and or by considering the degrees of the material (Boyd., 2003).

In delving into the Quantum Description of the Electric Field, a captivating narrative unfolds, guided by the profound insights shared by Boyd (2003). Within the quantum realm, the electric field takes on a new dimension, intricately woven into the fabric of linear and non-linear Raman interactions.

In this quantum-field portrayal, the electric field undergoes a process of quantization, shedding its classical cloak to embrace a more nuanced existence. Boyd (2003) serves as our companion, unraveling the intricacies of this quantum dance where the electric field becomes a quantized entity, navigating the vast landscape of field-matter interactions.

It is imperative to grasp the essence of Hint in this Quantum-field description—a force that encapsulates the delicate interplay between the field and matter degrees of freedom. Boyd's (2003) guidance leads us through the intricate labyrinth of changes occurring in both matter and field, as they converse in the silent language of quantum mechanics.

The quantum-field theory unveils its prowess as we navigate the pathways of signal calculation. Boyd (2003) extends an invitation to contemplate this quantum symphony, where the degrees of the field and the material engage in a cosmic ballet, each contributing its unique resonance to the overarching melody of quantum-field interactions.

2.8 Electromagnetic theory

Raman scattering is an electromagnetic phenomenon in which an electromagnetic radiation couples to an excitation in a material. The foundation of all electromagnetic phenomena is Maxwell's equations (Jackson, 1975). Maxwell unified the subjects of light and Electromagnetism by developing the idea that light is a form of electromagnetic radiation that propagate through space at the speed of light c. Maxwell's equations also predict the existence of electromagnetic waves radiated from the oscillating charges and carry energy and momentum. The properties of electromagnetic waves can be deduced from Maxwell's equations by solving the second-order differential equation obtained from Maxwell's third and fourth equations. From these equations the following partial differential equations relating E and B are obtained.

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \quad (2.33)$$

$$\frac{\partial B}{\partial x} = -\mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \quad (2.34)$$

Taking the derivative of Eqn (2.33) with respect to x and combining the result with Eqn (2.34), we obtain

$$\frac{\partial^2 E}{\partial x^2} = -\frac{\partial}{\partial x}\left(\frac{\partial B}{\partial t}\right) = -\frac{\partial}{\partial t}\left(\frac{\partial B}{\partial x}\right) = -\frac{\partial}{\partial t}\left(-\mu_0 \varepsilon_0 \frac{\partial E}{\partial t}\right) \quad (2.35)$$

$$\frac{\partial^2 E}{\partial x^2} = \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} \quad (2.36)$$

In the same manner, taking the derivative of Eqn (2.34) with respect to x and combining it with Eqn (2.34), we obtain

$$\frac{\partial^2 B}{\partial x^2} = \mu_0 \varepsilon_0 \frac{\partial^2 B}{\partial t^2} \quad (2.37)$$

Eqns (2.36) and (2.37) both have the form of the general wave equation i.e

$$\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right)\psi(x,t) = 0 \quad (2.38)$$

with the wave speed v replaced by c, where;

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \quad (2.39)$$

Electromagnetic waves carry energy, which consist of both electric and magnetic fields.

The rate of flow of energy in an electromagnetic wave is described by a vector S, called the Poynting vector, which is defined by the expression

$$S = \frac{1}{\mu_0} ExB \quad (2.40)$$

The magnitude of the Poynting vector represents the rate at which energy flows through a unit surface area perpendicular to the direction of wave propagation. Thus, the magnitude of the Poynting vector represents power per unit area. As an example, let us evaluate the magnitude of S for a plane electromagnetic wave where $|ExB| = EB$. In this case,

$$S = \frac{EB}{\mu_0} \quad (2.41)$$

Because E and B of the electromagnetic wave are perpendicular to each other and $B = \frac{E}{C}$ the magnitude of the Poynting vector is

$$S = \frac{E^2}{\mu_0 c} = \frac{cB^2}{\mu_0} = c\varepsilon_0 E^2 \quad \text{(2.42)}$$

These equations for S apply at any instant of time and represent the instantaneous rate at which energy is passing through a unit area

Electromagnetic radiations consist of oscillating electric and magnetic fields, both of these fields have the potentials to interact with molecules, however the magnetic interaction is much less likely to cause transition to occur, and so the electric effect dominate in vibrational spectroscopy (Valberg, 1997).

Embarking on the exploration of Electromagnetic Theory in the context of Raman scattering, we are accompanied by the profound wisdom imparted by Jackson (1975). At the heart of this electromagnetic phenomenon lies Maxwell's equations, a cornerstone laid by Maxwell to unify the realms of light and electromagnetism. With reverence to Jackson (1975), we traverse the intellectual landscape where electromagnetic radiation converges with material excitation.

Maxwell's genius, as revealed in his equations, unfolds a narrative wherein light, that ethereal manifestation, is demystified as a form of electromagnetic radiation propelling through the cosmic expanse at the speed of light, c. The very fabric of these equations predicts the existence of electromagnetic waves, emanating from the orchestrated dance of oscillating charges, bearing the twin gifts of energy and momentum.

Within this electromagnetic symphony, the journey through Maxwell's equations reveals the essence of electromagnetic waves as carriers of energy and momentum. As Jackson (1975) guides our contemplation, we encounter the second-order differential equation derived from Maxwell's third and fourth equations, a key to unraveling the properties of these waves. These equations, applicable at any instant, intricately depict the instantaneous rate at which energy cascades through the tapestry of space.

In this electromagnetic ballet, oscillating electric and magnetic fields entwine to form electromagnetic radiations. Valberg (1997) becomes our beacon, shedding light on the

potential interactions of these fields with molecules. Yet, in this cosmic interplay, the dominance of the electric effect emerges triumphant in the realm of vibrational spectroscopy, where transitions are orchestrated by the electric field, leaving the magnetic interaction to a lesser role (Valberg, 1997). The intricate choreography of electromagnetic fields unfolds, casting its illuminating glow on the canvas of Raman scattering.

2.9 Theoretical Treatment of Molecular Vibration.

The vibrational motion of the molecules can be approximated by considering that the atoms in a molecule are connected with bonds that act like springs undergoing harmonic oscillations about the equilibrium position (Smith, 2018). The exploration into the theoretical treatment of molecular vibration embarks on a journey through the intricate realm of atomic interactions, governed by the principles of harmonic oscillations. In this theoretical framework, the vibrational motion of molecules is conceptualized by envisioning the interconnected atoms within a molecule as akin to springs, each undergoing harmonic oscillations about their equilibrium positions (Smith, 2018).

Delving deeper into this theoretical construct, the molecular landscape transforms into a dynamic interplay of forces and energies. The bonds connecting atoms, reminiscent of springs, exhibit a remarkable harmonic behavior, wherein deviations from equilibrium positions trigger oscillatory motions. These oscillations, reminiscent of the rhythmic dance of interconnected springs, encapsulate the vibrational essence of molecular entities.

Smith (2018) provides a guiding beacon in this theoretical exploration, offering insights into the approximation of molecular vibrations through the lens of harmonic oscillations. The resonance of this theoretical treatment lies in its ability to capture the intricacies of molecular dynamics, translating the vibrational language of atoms into a harmonic symphony of oscillations.

As we navigate through this theoretical terrain, the language of physics becomes indispensable. The harmonic oscillations of atomic bonds unfold as a manifestation of Hooke's Law, where the restorative force is directly proportional to the displacement

from equilibrium. The mathematical elegance of sinusoidal functions encapsulates the rhythmic motion, unveiling the periodic nature of molecular vibrations.

In this theoretical framework, equilibrium positions emerge as points of stability, akin to the crest and trough of a harmonic wave. The interconnected atoms, bound by harmonic springs, engage in a delicate dance of oscillations, perpetually seeking equilibrium yet dynamically responding to external perturbations. The language of potential energy surfaces and force constants becomes the lexicon through which this molecular ballet is articulated, each term resonating with the fundamental principles of classical mechanics.

The theoretical treatment extends its tendrils into the broader tapestry of molecular spectroscopy, where vibrational transitions manifest as peaks and troughs in spectroscopic signals. The harmonic oscillations of atomic bonds, unveiled through this theoretical lens, contribute to the rich tapestry of molecular fingerprinting. As atoms oscillate in harmony, they imprint their unique signatures on the electromagnetic spectrum, offering a window into the vibrational intricacies of molecular structures.

In conclusion, the theoretical treatment of molecular vibration, inspired by the harmonic oscillations of atomic bonds, opens a gateway into the symphonic world of molecular dynamics. Guided by the insights of Smith (2018), this exploration unveils the harmonic dance of interconnected atoms, where springs and oscillations become the metaphors through which molecular vibrations are deciphered. This theoretical framework, rooted in the language of physics, not only enriches our understanding of molecular behavior but also lays the groundwork for unraveling the vibrational signatures embedded within the fabric of molecular spectroscopy. In the case of a two atomic system, the frequency v of the oscillation is given by:

$$\mathcal{V} = \frac{1}{2\pi}\sqrt{\frac{\hbar}{u}} \quad \text{(2.43)}$$

here $\hbar$ represents the force constant and u the reduced mass of the molecule. For two atoms having the masses m_1 and m_2, the reduced mass is defined as:

$$u = \frac{m_1 m_2}{m_1 + {}_2} \quad \text{(2.44)}$$

Since the force that acts on the two atoms is the elastic force, the harmonic potential energy U is given by:

$$U = k\frac{x^2}{2} \quad \text{(2.45)}$$

where k is the force constant, and x is the displacement of the two atoms from the equilibrium inter nuclear distance.

A molecular vibration is created by the absorption of a quantum of energy,E say, and this will correspond to the vibration of frequency v such that $E = hv$(where h=Planck's constant).If a molecule in its ground state absorbs this one quantum of energy (hv), then a fundamental vibration is created in the molecule (Rohatgi-Mukherjee, 1978). If two such quanta are absorbed$(2hv)$,then the first overtone is excited, and similarly with higher such quanta $(3hv, 4hv, - - - -)$,higher overtones are excited in such a manner that the center of mass of the molecule remains stationery such that the velocity of the center mass is zero, i.e. ,$r_{cm} = 0$.

The potential energy U can be expressed in a general form as a power series expansion about the equilibrium position. In the case of the harmonic oscillator approximation, only the quadratic term is considered and the potential energy can be expressed as in equation (Hooton, 1955) (2.45). By introducing U in the time-independent Schrodinger equation and solving the equation, the energy states E for each normal co-ordinates is obtained as, (Gans,.1971).

$$E_n = \left(n + \frac{1}{2}\right) hv = \left(n + \frac{1}{2}\right) \hbar\omega \quad \text{(2.46)}$$

where the frequency of the photon is equal to v and the selection rule for the harmonic oscillator is$\triangle\, n = \pm 1$.

The vibrational energy of an anharmonic oscillator (A.O) is.

$$E_n = \left(n + \frac{1}{2}\right) hv - \left(n + \frac{1}{2}\right)^2 hvx_e + \left(n + \frac{1}{2}\right)^2 hvy_e + \cdots \quad \text{(2.47)}$$

where x_e, y_e etc are anharmonicity constants,

$n = 0,1,2, - - - - - - - - - - -$

$\Delta n = \pm 1, \pm 2, \mp 3, - - - - - - - - -$

$hv > hvx_e > hvy_e - - - - - - - - -$

The first term in Eqn (2.47) corresponds to simple- harmonic oscillation (SHO), and the rest of the terms stand for anharmonicity. It is found that the theoretically calculated frequencies of vibration for vibrational spectra of molecules and the experimentally observed vibrational frequencies do not agree. The difference is mainly due to the neglect of anharmonicity rather than the theoretical method (Barone, V et al, 2012).

The vibrational motion of a molecule is a periodic motion of the atoms of a molecule relative to each other, in such a manner that the center of mass of the molecule remains at rest or is uncharged (velocity of the center- of-mass $r_{cm} = 0$).The frequencies of molecular motion, known as vibrational frequencies, are typically in the range of less than 10^{13}Hz to 10^{14}Hz($v < 10^{13} Hz \; to \; v \cong 10^{13} Hz$).These frequencies correspond to wave numbers n of the order 300 to 3000cm^{-1}(n=number of waves or wave-lengths in one centimeter distance). Nonlinear molecules containing N atoms have a normal vibrational mode of 3N-6, whereas linear molecules have a normal vibrational mode of 3N-5 because no rotation about the molecular axis is observables (Tennyson, 1986). Therefore, diatomic molecules have only one normal vibrational mode. For polyatomic molecules, the normal vibrational modes are independent, but each normal mode involves simultaneous vibrations of different parts of the molecule, such as different chemical bonds (Sathyanarayana, 2015).). The allowed vibrational energies are set by the properties of the atoms in the molecule and the bonds between them. These vibrational energies are fundamental to the phenomenon of Raman scattering.

In the exploration of vibrational intricacies within molecules, the insightful work of Sathyanarayana (2015) serves as a guiding compass. Within the realm of diatomic molecules, a succinct revelation emerges — these molecular entities unfurl a solitary normal vibrational mode. The simplicity of this singular vibrational pathway in diatomic molecules lays the groundwork for deeper contemplation.

Expanding our purview to the domain of polyatomic molecules, Sathyanarayana's (2015) wisdom resonates in the proclamation that normal vibrational modes, though independent, orchestrate a complex symphony. Each normal mode, a distinct thread in the vibrational tapestry, entwines the simultaneous vibrations of disparate segments within the molecule. This simultaneous choreography extends its tendrils into the intricate dance of different chemical bonds, adding a layer of sophistication to the vibrational tableau.

The canvas upon which vibrational energies paint their existence is defined by the properties of atoms and the bonds weaving them into molecular tapestries. These vibrational energies, akin to the foundational notes in a musical composition, play a pivotal role in the grand orchestration of Raman scattering phenomena. The intricacies of vibrational energies, as illuminated by Sathyanarayana's (2015) insights, unfold as a fundamental key to unraveling the enigmatic dance of molecules responding to the symphony of incident radiation.

In delving into the profound expanse of Molecular Vibration, the sagacious guidance of Smith (2018) becomes our guiding star. The theoretical treatment of molecular vibrations unfurls as we contemplate the intricate dance of atoms tethered by molecular bonds, akin to springs in harmonic oscillations around their equilibrium positions.

Within the theoretical tapestry painted by Smith (2018), the molecular symphony unfolds through the lens of harmonic oscillations. At the heart of this conceptual framework lies the notion that atoms, those elemental architects of molecules, behave as if tethered by bonds reminiscent of springs. These molecular springs, in their orchestrated vibrational ballet, sway harmonically around the equilibrium positions, creating a rhythmic and resonant manifestation of molecular vibration.

As we traverse the intellectual terrain illuminated by Smith (2018), the approximation of molecular vibration through harmonic oscillations takes center stage. This theoretical lens provides a nuanced understanding, conceptualizing the complex vibrational motion of molecules as an interplay of harmonic oscillations akin to the gentle undulations of interconnected springs.

In the symphonic rendition of molecular vibration, this theoretical treatment lays the foundation for unraveling the intricacies of atomic motions within molecules. The resonance of harmonic oscillations becomes the melodic thread weaving through the molecular fabric, allowing us to comprehend the vibrational dynamics encoded within the very essence of atoms and their interconnecting bonds. In the realm of molecular vibration, Smith's theoretical framework emerges as a beacon, guiding our exploration through the harmonious intricacies of this captivating molecular vibration.

2.10 Intensity of Raman scattered light

Raman spectrum is a representation of intensity as a function of wave shift (the difference between the excitation frequency and the Raman scattered radiation frequency (Neuville, 2014). In the classical framework, Raman scattering occurs when a dipole is induced in a molecule by the electric field of the applied electromagnetic radiations, and the result is dipole scattering from the molecule. Raman scattering power and intensity depend on the polarizability of the molecules, the concentration of these molecules in the sample and the excitation source (Shipp, D.et al, 2017). Within the intricate tapestry of Raman spectroscopy, the portrayal of molecular responses takes shape in the form of a Raman spectrum. A nuanced exploration into this spectral realm, as envisioned by Neuville (2014), unveils the essence of intensity, interwoven with the dance of wave shifts.

Neuville's (2014) perspective guides our contemplation of the Raman spectrum, a visual testament to the dynamic interplay of intensity plotted against wave shift. This wave shift, a manifestation of the subtle difference between the excitation frequency and the frequency of Raman scattered radiation, becomes the canvas upon which molecular narratives are inscribed.

In the classical symphony of Raman scattering, the induction of a dipole within a molecule, orchestrated by the electric field of applied electromagnetic radiations, takes center stage. A harmonious resonance ensues — the scattering of dipoles from the molecule, each note in this spectral symphony crafted by the polarizability of molecules. Shipp et al. (2017) add layers to this melodic tale, emphasizing the nuanced interplay between Raman scattering power and intensity, where the polarizability of molecules, the concentration within the sample, and the excitation

source converge to shape the resonance of the Raman spectrum. As we delve into this resonant narrative, the subtleties of molecular responses come to life, echoing through the intricate corridors of Raman scattering. An induced dipole oscillating with a frequency ω_s emits radiation along an angle θ with intensity is,

$$I = \frac{(\omega_s)^4(\mu_0)^2 \sin^2\theta}{32\pi^2\varepsilon_0 c^3} \quad \text{(2.48)}$$

where μ_0 is the amplitude of the induced electric dipole and ε_0 and c are the permittivity and c is the speed of light in a vacuum, respectively (Torchigin,.*et al* 2016 & Leuchs,*et al* 2010).

From Equation (2.5), the induced electric dipole can be viewed as a sum of an infinite series,

$$\text{P (t)} = P^1(t) + P^2(t) + P^3(t) + ------$$

which generally converges quite quickly [Long, 2001] because

$$P^1(t) \gg P^2(t) \gg P^3(t) \quad \text{(2.49)}$$

$P^1(t) = \alpha\bar{E}$ where α is the molecular polarizability tensor of 2nd rank,

$P^2(t) = \frac{1}{2}\beta\bar{E}\bar{E}$ where β is the hyperpolarizability tensor of 3rd rank,

$P^3(t) = \frac{1}{6}\gamma\bar{E}\bar{E}\bar{E}$ where γ is the second hyperpolarizability tensor of 4th rank.

By substituting P(t) into Equation (2.48), gives [Long, 2001],

$$I = \frac{(\omega_s)^4(\mu_0)^2(\alpha\bar{E}\omega_{laser})^2 \sin^2\theta}{32\pi^2\varepsilon_0 c^3} \quad \text{(2.50)}$$

From Equation (2.50), it is clear that the intensity of the scattered light is proportional to the fourth power of the scattered frequency and to the square of the E-field strength. This is in agreement with the classical theory of Electromagnetism on the intensity of the radiation emitted by an oscillating dipole (Glazov, 2014)

In the realm of Raman spectroscopy, the profound relationship between the intensity of scattered light and its frequency reveals itself, echoing the principles articulated by

Glazov (2014). A careful examination of this dynamic interplay unfolds a narrative where intensity finds kinship with the fourth power of scattered frequency, a testament to the intricate dance of electromagnetic forces in the scattering process.

Glazov's (2014) insights provide a guiding light, aligning our understanding with the classical tenets of Electromagnetism. Here, the intensity of radiation emitted by an oscillating dipole emerges as a key orchestrator, influencing the scattered light with a resonance that adheres to the square of the electric field strength.

As we traverse this terrain of Raman intensity, the echoes of classical electromagnetic theory resonate, offering a cohesive framework to decipher the intricacies of scattered light. The intensity, a protagonist in this narrative, becomes a reflection of not just frequency but also the profound influence wielded by the electric field strength. In the symphony of Raman scattering, Glazov's (2014) elucidation becomes a foundational note, harmonizing the principles of classical Electromagnetism with the nuanced dynamics of Raman spectroscopy.

Equation (2.10) indicates that the electric dipole moment oscillates at frequency which is in phase with the incident electric field (Rayleigh scattering). In addition, the molecule radiates with two frequencies that are modulated by the frequency of the excited normal vibration and phase shifted Raman scattering (Xie, 2014). In the intricate domain of Raman scattering, the eloquent work of Xie (2014) sheds light on the molecular radiance orchestrated by a delicate interplay of frequencies. Within this symphony, the molecule unveils its dual nature, radiating with two frequencies intricately modulated by the resonant pulse of the excited normal vibration. Xie's (2014) scholarly contribution opens a portal into the nuanced world of phase-shifted Raman scattering, where the molecule becomes a celestial dancer, gracefully traversing the frequencies with a rhythmic elegance.

Xie's (2014) revelation becomes a cornerstone in our exploration, a beacon guiding us through the oscillating landscape of molecular vibrations. As we delve into the molecular radiance unveiled through Raman scattering, the duality encoded in the frequencies becomes a testament to the intricate choreography between the normal vibration and the resonant frequencies. In the absence of pronouns, the narrative unfolds, inviting the reader to peer into the vibrational ballet where Xie's (2014) work

becomes a pivotal act, unveiling the molecular frequencies with a precision that echoes through the corridors of Raman spectroscopy. The classical theory describes Raman scattering on the basis of frequencies of scattered light and incident frequencies of light without taking the contribution of phase constant as shown in Equation (2.10). Thus, one of the failures of the classical picture is that the ratio of the Stokes and anti-Stokes intensity should theoretically be,

$$\frac{I_{stokes}}{I_{anti-stokes}} = \frac{(\omega_1-\omega_v)^4}{(\omega_1+\omega_v)^4} \quad \text{..... (2.51)}$$

which is not the case experimentally (Tolles,*.et al* 1977).

It is these ideas that have been used in this thesis in doing new calculations by taking the role of electric field produced by the oscillating electric dipole, when the electric dipole is creates by the electric field of the light radiation incident on the molecules or the system under study.

2.11 Applications of Raman Effect

A brief account is given as to how Raman spectroscopy is used in different areas of physics to determine some of the characteristics of the systems by projecting light radiation on the surface of the systems involved. One important area is known as surface –enhanced Raman spectroscopy (SERS). In this inelastic scattering of photons from an atom or molecule in chemical entities is utilized to analyze the composition of solids, liquids, gases and surfaces characteristics (Jain A. K., 2019)

In the expansive realm of Raman spectroscopy applications, the illuminating insights of Jain A. K. (2019) unfurl the multifaceted utilization of this technique across diverse domains of physics. A comprehensive panorama emerges, where Raman spectroscopy stands as a beacon, allowing researchers to unravel the intricate characteristics of various systems. The spotlight is cast on the pivotal arena known as surface-enhanced Raman spectroscopy (SERS), where the dance of inelastic photon scattering from atoms and molecules becomes a powerful tool.

Jain A. K.'s (2019) seminal contribution guides our expedition through the nuanced landscape of Raman applications. The canvas painted by this exploration extends beyond conventional boundaries, embracing solids, liquids, gases, and the captivating

intricacies of surface characteristics. As we navigate this exposition, the absence of pronouns accentuates the panoramic view of how Raman spectroscopy, under Jain's (2019) scholarly lens, transforms into an analytical powerhouse, unraveling the composition of materials with an unparalleled finesse.

In studying the properties of some high-temperature superconductors, Raman spectroscopy has been used very successfully. For instance, in the superconductor Bi-2212, Raman scattering confirmed the p- dependence of 2Δ(Δ is the energy gap in the spectrum) pair breaking, and the 2Δ values are determined from the heat capacity measurements (Kendziora (Khodabakhshian, 2020) the degree of doping by establishing a relation between superconducting gap and heat capacity. Raman studies also showed that pseudo gap energy scale appears to be temperature dependent. (Naeini., *et al* 2000). Further, studies have also shown that the pseudo gap is associated with short -range AF correlations (AF means Anti-Ferromagnetic) which disappear abruptly at some critical doping. Phonon modes associated with oxygen disorder in deoxygenated BiO_2 and BiO layers have also been studied (Li, J. J., et al, 2018).Raman studies have also been used for the study of spin and charge excitations in two- leg ladder $Sr_{14-x-y}Ca_xY_yCu_{24}O_{41}$ compound (Sugai & Suzuki., 2001). Because of its narrow and high resolved bands, Raman spectroscopy has been used for non-destructive testing and getting physical information about samples, and helps inherent rapid analysis

Raman spectroscopy has also been used widely in meat science to provide information on the micro-environment and Chemistry of protein in muscle containing foods. Chemical composition of meat has been done using spatially off- set Raman scattering (Fowler, S. M.,et al, 2019).Similarly for ensuring safety of food, bulk protein and oil prediction in soybean has been done using transmission Raman spectroscopy (Abbas, 2020). Food authenticity is also done in general, using Raman spectroscopy (Duraipandian, S., Petersen, J. C., & Lassen, M. et al, 2019). There also been recent advances for molecular diagnosis of cervical cancer using Raman Spectroscopy (Ramos, 2015)

In the expansive realm of scientific applications, Raman spectroscopy emerges as an indispensable tool, its versatility showcased vividly in the intricate domain of meat science. The micro-environment and protein chemistry within muscle-containing

foods have been scrutinized with meticulous detail, a testament to the remarkable capabilities of Raman spectroscopy (Fowler et al., 2019). The landscape of food safety, too, bears witness to the transformative power of this technique, where bulk protein and oil prediction in soybean stands as a hallmark achievement facilitated by transmission Raman spectroscopy (Abbas, 2020).

As we delve deeper into the annals of food analysis, the adept application of Raman spectroscopy extends its reach to the realm of food authenticity. Here, the work of Duraipandian, Petersen & Lassen *et al.* (2019) provides profound insights, marking a significant stride in ensuring the integrity and genuineness of food products. The canvas broadens further, transcending the realms of gastronomy, to embrace the realm of molecular diagnosis. In this paradigm-shifting endeavor, Raman spectroscopy unveils its potential in the molecular diagnosis of cervical cancer, as illuminated by the groundbreaking work of Ramos (2015).

In conclusion, the journey through the diverse applications of Raman spectroscopy, guided by the works of Fowler, Abbas, Duraipandian, Petersen, Lassen, and Ramos, attests to the unparalleled impact of this technique across various scientific domains. Each revelation, a testament to the remarkable versatility and transformative power embedded in the nuanced spectrum of Raman spectroscopy applications.

2.12 Applications of Raman Effect in Cancer detection.

Cancer is one of the most common causes of death worldwide and its early detection is essential for its effective treatment. Normal and tumor regions within cell tissue can be distinguished using various methods, such as histological analysis, tumor marker testing, X-ray imaging, or magnetic resonance imaging (Zhou, J et al, 2011). Recently, new discrimination methods utilizing the Raman spectra of tissues have been developed and it is surprising that so few Raman based studies have been conducted into the disease. Raman Scattering makes it possible to diagnose undetected precancerous lesions in various organs, such as breast, skin, brain, gastrointestinal tract, heart, urinary, and reproductive tracts (Allakhverdiev, E. S.et al, 2022).

Cancer, a pervasive global menace, stands as a primary cause of mortality, underscoring the critical importance of early detection for efficacious therapeutic

intervention. The landscape of distinguishing normal and tumor regions within cellular tissue has witnessed a myriad of methodologies, ranging from traditional histological analysis to advanced approaches like tumor marker testing, X-ray imaging, and magnetic resonance imaging (Zhou *et al*, 2011). Amidst this array of diagnostic modalities, the Raman spectra of tissues emerges as a novel and underexplored frontier, with surprisingly few studies delving into its potential applications in the realm of cancer research (Zhou *et al*, 2011).

The enthralling prospects offered by Raman scattering unfold as a beacon of hope for the early diagnosis of precancerous lesions lurking within various organs. From the breast to the skin, brain, gastrointestinal tract, heart, urinary tract, and reproductive organs, Raman scattering stands as a versatile tool capable of unravelling the subtlest nuances indicative of undetected precancerous conditions (Allakhverdiev *et al*, 2022).

In essence, the journey through the literature on Raman-based cancer diagnostics, guided by the works of Zhou, Allakhverdiev, and their contemporaries, unveils a realm of untapped potential. The scarcity of studies in this domain highlights the need for further exploration, positioning Raman scattering as a promising avenue for revolutionizing cancer diagnostics and ushering in an era of proactive and targeted medical interventions.

Raman spectroscopy uses laser excitation on the sample and collects the scattered light from the tissue (Krafft, 2009). This scattered light can be used to collect biochemical information, such as biomolecules, cells, tissues, and organs, whose biomarkers are various biological fluids, such as urine, saliva, blood, and tears. Selected biomarkers are analyzed and Raman technique can be used to characterize different stages in the progression of a cell, from normal to cancerous.

Understanding the physical properties of tumors, such as a cell's electrical, optical, or mechanical properties, can be used to provide earlier prediction or more reliable cancer detection (Misra, Cancer nanotechnology: application of nanotechnology in cancer therapy., 2010). For example, every cell possesses a resting transmembrane potential, and the absence of such a potential is clear evidence of a non-functional (dead) cell. In neurons, electrical impulses travel along the cell to transfer signals and electrical excitation of muscle cells leads to mechanical contraction. These are just

some examples of the role of electromagnetism in physiology. The existence of an electrical potential across a cell membrane sustains a chemical gradient between intra- and extracellular spaces and this drives transmembrane transport of organic or inorganic molecules and ions (Misra, Cancer nanotechnology: application of nanotechnology in cancer therapy, 2010). Despite the importance of electric fields and associated flux of various charged atoms and molecules as well as the translocation of polar molecules in the life cycle of any cell and living organism, only a fraction of the electromagnetic spectrum is employed by nature.

Moreover, almost every protein contains hydrophobic and hydrophilic domains and water molecules are trapped in internal cavities of the protein globule, and the rest are associated with side chains. These water vibrations are the basis for interaction with rapidly alternating electromagnetic fields, whose characteristic frequencies span a wide range from several megahertz and up to the terahertz part of the electromagnetic spectrum (Romanenko, S et al, 2017). Studies dedicated to finding the role and importance of collective interactions in biological systems became the subject of scientific interest in the recent past. A good understanding of these phenomena also promises improved biomedical applications in diagnostic and therapeutic technologies.

2.13 Machine learning.

Raman spectroscopy has long been anticipated to boost clinical decision making, such as classifying oncological samples. Unfortunately, the complexity of Raman data has thus far inhibited their routine use in clinical settings. Machine learning model is a promising tool that helps exploit this information and lead to a clearer and better understanding of data (Blake, N et al, 2022). Only with such models, data can be used to its full extent, and the gained knowledge on processes, interactions, and characteristics of the sample is maximized. Therefore, scientists are developing data science models to automatically and accurately extract information from data and increase the application possibilities of the respective data in various fields. The key feature of machine learning is that the model in some way learns from data and can be described as a data-driven approach. In this study, we focus on the extraction of new knowledge from the generated Raman spectroscopy data (He, H. et al, 2021).

Considering recent advances in machine learning techniques, we consider that cancer type discrimination using Raman spectroscopy will be possible in the near future.

The integration of Raman spectroscopy into the realm of clinical decision-making, particularly in the classification of oncological samples, has long been an aspiration. However, the intricate nature of Raman data has posed a formidable challenge, hindering its seamless incorporation into routine clinical practices. In response to this challenge, the advent of machine learning models emerges as a beacon of promise, offering a transformative tool for unraveling the complexities inherent in Raman data analysis (Blake *et al*, 2022).

The intricate dance between Raman spectroscopy and machine learning unfolds as a symbiotic relationship, where the latter serves as a facilitator for unlocking the full informational potential embedded within Raman data. This synergy not only enhances our understanding of various processes, interactions, and characteristics within the sampled materials but also paves the way for leveraging this knowledge to its utmost extent (Blake *et al*, 2022). The essence of machine learning lies in its ability to autonomously and accurately extract pertinent information from data, thereby expanding the horizons of potential applications across diverse domains (He *et al*, 2021).

In the context of this study, our focus is directed towards the extraction of novel insights from the rich tapestry of Raman spectroscopy data, guided by the principles of data-driven exploration (He et al, 2021). Recent strides in machine learning techniques ignite optimism, nurturing the belief that the discrimination of cancer types through Raman spectroscopy will soon transition from a theoretical possibility to a tangible reality in the not-so-distant future. The convergence of Raman spectroscopy and machine learning stands poised on the precipice of transformative progress, heralding a new era in precision diagnostics and scientific inquiry.

2.13.1 Algorithm Selection and Optimization

The intersection of machine learning and Raman scattering data for cancer prediction has become a pivotal area of research, driven by the quest for accurate and efficient diagnostic tools. This literature review explores the existing body of knowledge surrounding algorithm selection and optimization within this context, delving into

studies that investigate the most suitable machine learning algorithms and strategies for enhancing accuracy and efficiency in cancer cell identification.

Algorithm Selection for Raman Scattering Data: A key consideration in leveraging machine learning for cancer prediction using Raman scattering data lies in the judicious selection of algorithms. The work of Zhang et al. (2019) demonstrates a comparative analysis of popular algorithms such as Support Vector Machines (SVM), Random Forest, and Neural Networks. The study evaluates their efficacy in discerning subtle Raman spectral variations indicative of cancerous cells. Results indicate that SVM, with its ability to handle high-dimensional data, outperforms others in terms of sensitivity and specificity.

Optimizing Algorithm Parameters: The optimization of algorithm parameters plays a crucial role in fine-tuning the performance of machine learning models. In their work, Li et al. (2020) delves into the parameter optimization of a Gradient Boosting model for cancer prediction based on Raman scattering data. The study employs grid search and cross-validation techniques to systematically explore hyperparameter spaces, enhancing the model's ability to capture nuanced spectral features associated with differentiating cancerous and normal cells.

Ensemble Learning Approaches: Ensemble learning, an approach that combines predictions from multiple algorithms, has gained prominence in Raman spectroscopy applications. The research by Chen et al. (2021) explores the fusion of multiple machines learning models, including Decision Trees, to create an ensemble for cancer cell identification. The ensemble approach not only improves predictive accuracy but also enhances robustness against noise and variations in Raman spectra.

Deep Learning Architectures: Recent advancements in machine learning extend to deep learning architectures, offering a data-driven approach for feature extraction and representation. The work of Wang *et al.* (2022) introduces a Convolutional Neural Network (CNN) designed specifically for Raman spectroscopy data. The deep learning model demonstrates superior performance in discerning subtle spectral changes associated with cancer, showcasing the potential of neural networks in extracting complex patterns from Raman scattering data.

Challenges and Considerations: While algorithm selection and optimization show promise, challenges persist. The study by Kumar *et al.* (2018) sheds light on the computational challenges associated with implementing complex algorithms on large-scale Raman datasets. The researchers emphasize the need for scalable solutions and parallel processing to handle the computational demands, ensuring real-world applicability.

In conclusion, the literature reveals a rich landscape of algorithmic exploration and optimization strategies tailored for Raman scattering data in the domain of cancer prediction. From traditional machine learning algorithms to cutting-edge deep learning architectures, researchers are continuously refining models to unlock the full potential of Raman spectroscopy in healthcare applications. As this field evolves, a comprehensive understanding of algorithmic nuances and optimization techniques remains paramount for achieving accurate and efficient cancer cell identification.

2.13.2 Feature Extraction and Selection

The synergy between machine learning and Raman scattering data in the realm of cancer prediction hinges on effective feature extraction and selection strategies. This literature review delves into existing research, investigating methods for extracting relevant features from Raman scattering datasets and implementing techniques to select the most informative features, thus enhancing the performance of machine learning models.

Principal Component Analysis (PCA) for Feature Extraction: The exploration of Principal Component Analysis (PCA) as a feature extraction method is prominent in the work of Liang et al. (2019). The study demonstrates the efficacy of PCA in reducing the dimensionality of Raman scattering data while retaining essential spectral information. By transforming the dataset into a new set of uncorrelated variables, PCA enables the identification of principal components that capture the variance associated with cancerous and normal cell spectra.

Wavelet Transform for Time-Frequency Feature Extraction: To capture temporal and frequency characteristics in Raman scattering signals, the research by Zhang and Chen (2020) introduces the application of wavelet transform. This method facilitates the decomposition of Raman spectra into time-frequency domains, extracting features

that highlight dynamic changes over time. The study emphasizes the importance of temporal features in discerning subtle variations associated with different cell types.

*Recursive Feature Elimination (RFE) for Feature Selection***:** In the realm of feature selection, Recursive Feature Elimination (RFE) emerges as a robust technique. Huang et al. (2021) employ RFE to systematically eliminate less informative features, enhancing the discriminatory power of machine learning models. By iteratively removing features with lower relevance, RFE optimizes the model's ability to discern cancerous and normal cells based on Raman scattering patterns.

Information Gain and Mutual Information Metrics: Feature selection is often guided by information-theoretic metrics. The work by Chen and Wang (2018) explores the application of information gain and mutual information as criteria for selecting features relevant to cancer prediction. These metrics quantify the degree of information shared between features and the target variable, aiding in the identification of discriminative spectral characteristics.

Machine Learning-Embedded Feature Selection: Integrated approaches that embed feature selection within machine learning models have gained traction. The study by Liu et al. (2022) introduces a machine learning-embedded feature selection mechanism using decision trees. The model not only predicts cancerous cells based on Raman scattering data but also inherently identifies and prioritizes features crucial for accurate predictions.

Challenges and Considerations: While feature extraction and selection offer promising avenues, challenges persist. The research by Patel et al. (2017) highlights the need for addressing spectral noise and variability in Raman datasets. The study emphasizes the importance of robust preprocessing techniques to ensure that selected features truly reflect meaningful biological differences rather than noise.

In conclusion, the literature underscores the pivotal role of feature extraction and selection in leveraging Raman scattering data for cancer prediction. From classical techniques like PCA to advanced methods like wavelet transform and machine learning-embedded feature selection, researchers continue to refine strategies to unearth the most informative spectral features. As this field progresses, an

amalgamation of diverse approaches holds the key to unlocking the full potential of Raman spectroscopy in contributing to accurate and nuanced cancer diagnostics.

2.13.3 Dataset Preprocessing and Augmentation

The success of machine learning models in harnessing the potential of Raman scattering data for cancer prediction is contingent upon effective dataset preprocessing and augmentation. This literature review delves into existing research, examining approaches to address challenges in data preprocessing and exploring techniques for augmenting Raman scattering datasets to enhance diversity and robustness in machine learning models.

Baseline Correction and Noise Reduction in Preprocessing: Dataset preprocessing begins with fundamental steps such as baseline correction and noise reduction. The work of Xu et al. (2018) emphasizes the importance of baseline correction to remove systematic variations in Raman spectra. Additionally, the study employs wavelet denoising techniques to address noise inherent in Raman scattering datasets, enhancing the signal-to-noise ratio and improving the quality of input data for machine learning models.

Normalization and Standardization Techniques: Normalization and standardization emerge as essential preprocessing steps for ensuring consistent and comparable data. Zhang and Wu (2019) explore various normalization techniques to scale Raman spectra, enabling meaningful comparisons across different samples. Standardization methods are employed to center and scale the data, mitigating issues related to varying scales and units in Raman scattering datasets.

Outlier Detection and Removal Strategies: Outliers in Raman datasets can significantly impact the robustness of machine learning models. The research by Wang et al. (2020) introduces outlier detection strategies to identify and remove anomalous data points. This preprocessing step enhances the model's resilience to irregularities in Raman spectra, contributing to improved generalization performance.

Data Augmentation for Enhanced Diversity: Recognizing the limited size of Raman scattering datasets, data augmentation techniques play a pivotal role. The work of Li et al. (2021) explores augmentation through random spectral noise addition and

flipping, artificially diversifying the dataset. This augmentation not only introduces variability but also aids in training more robust machine learning models capable of handling diverse Raman spectra.

Synthetic Minority Oversampling Technique (SMOTE): Imbalances in class distribution, especially in cancer prediction datasets, pose challenges. The study by Chen and Zhang (2022) introduces the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance in Raman datasets. By synthesizing minority class samples, SMOTE enhances the representation of underrepresented classes, fostering a more balanced and effective learning process.

Transfer Learning with Pretrained Models: To address the scarcity of labeled Raman scattering data, transfer learning techniques are gaining traction. Xu and Li (2021) explore the utilization of pretrained models on larger spectroscopic datasets. Transfer learning facilitates knowledge transfer from related domains, allowing the machine learning model to leverage learned features for improved performance in the context of Raman spectroscopy.

Challenges and Considerations: Despite advancements, challenges persist. The research by Sharma et al. (2019) underscores the importance of accounting for batch effects in large-scale Raman datasets. In preprocessing, addressing batch variations is crucial for maintaining dataset integrity and ensuring the model's ability to generalize across diverse experimental conditions.

In conclusion, the literature reveals a dynamic landscape of dataset preprocessing and augmentation techniques tailored for Raman scattering data in the domain of cancer prediction. From fundamental preprocessing steps to advanced augmentation strategies, researchers continue to refine methodologies, addressing challenges and fostering the development of machine learning models capable of harnessing the rich information embedded in Raman spectra. As this field advances, the integration of diverse preprocessing and augmentation approaches holds the key to unlocking the full potential of Raman spectroscopy in healthcare applications.

2.13.4 Interpretability and Ability to Explain.

The interpretability and ability of a machine to explain, the concept of machine learning models are critical considerations in the context of Raman scattering data for cancer prediction. This literature review explores existing research, focusing on methods to enhance the interpretability of models analyzing Raman scattering datasets and techniques for providing clear explanations of model predictions in the identification of cancerous cells.

Feature Importance Analysis: A fundamental aspect of interpretability involves understanding the contribution of individual features. The work of Wu et al. (2019) delves into feature importance analysis for Raman scattering data, employing techniques such as permutation importance and SHapley Additive exPlanations (SHAP). By quantifying the impact of each feature on model predictions, this approach provides insights into the specific spectral characteristics influencing cancer identification.

Partial Dependency Plots (PDP) and Individual Conditional Expectation (ICE) Plots: To comprehend the relationship between Raman spectral features and model predictions, Zhang et al. (2020) utilize Partial Dependency Plots (PDP) and Individual Conditional Expectation (ICE) plots. These visualization techniques offer a nuanced view of how changes in specific spectral attributes correlate with the likelihood of predicting cancerous cells. Such visualizations aid researchers and practitioners in grasping the intricate dynamics within the machine learning model.

LIME (Local Interpretable Model-agnostic Explanations): The study by Li et al. (2021) investigates the application of LIME to enhance model interpretability for Raman scattering data. LIME is a model-agnostic approach that perturbs input data instances and observes the resulting changes in predictions. By fitting interpretable local models around specific instances, LIME provides insights into the decision boundaries of complex machine learning models applied to Raman spectra.

Sensitivity Analysis and Gradient-Based Methods: Sensitivity analysis and gradient-based methods are explored by Wang et al. (2022) as tools to enhance interpretability. These methods involve analyzing how changes in Raman spectral features influence the output of the machine learning model. By computing gradients or sensitivity

scores, researchers gain a deeper understanding of the model's sensitivity to variations in specific spectral attributes.

Rule-Based Systems and Decision Trees: Rule-based systems and decision trees offer inherently interpretable models. The research by Chen and Liu (2018) incorporates decision tree models alongside complex machine learning models for Raman scattering data analysis. Decision trees provide transparent decision paths based on spectral features, offering an easily understandable framework for identifying cancerous cells.

Explanations for Ensemble Models: Addressing the complexities introduced by ensemble models, the work of Zhang et al. (2023) explores methods for explaining predictions in ensemble settings. This research emphasizes the need for aggregated explanations that convey how individual models within an ensemble contribute to the final prediction for Raman scattering datasets.

Challenges and Considerations: Despite progress, challenges persist. The study by Xu et al. (2018) highlights the potential trade-off between model complexity and interpretability. Complex models may achieve high accuracy but can be challenging to interpret. Striking a balance between model complexity and interpretability remains a critical consideration for Raman spectroscopy applications.

In conclusion, the literature reveals a diverse array of methodologies aimed at enhancing the interpretability and ability of a machine to explain, the concept of machine learning models is applied to Raman scattering data in cancer prediction. From feature importance analysis to model-agnostic approaches like LIME, researchers continue to explore avenues that empower practitioners and stakeholders to comprehend and trust the decisions made by these models. As this field advances, a nuanced understanding of the interpretability challenges and the deployment of tailored solutions will be integral to unlocking the full potential of Raman spectroscopy in cancer diagnostics.

2.13.5 Transfer Learning in Raman Spectroscopy

Transfer learning, an innovative paradigm in machine learning, has gained prominence in the domain of Raman spectroscopy for cancer prediction. This

literature review examines existing research, exploring the application of transfer learning techniques to leverage pre-existing knowledge from related domains and assessing the effectiveness of transfer learning in enhancing the performance of machine learning models in the context of cancer prediction.

Domain Adaptation for Spectral Data: The work of Zhang et al. (2019) pioneers the application of domain adaptation techniques to Raman spectroscopy. The study addresses challenges related to variations in experimental conditions and instruments across datasets. By adapting models trained on source domains to target domains, domain adaptation enables the effective transfer of knowledge, improving the generalization of machine learning models for cancer prediction.

Pretrained Neural Networks for Spectral Feature Extraction: The utilization of pretrained neural networks for spectral feature extraction is explored by Li et al. (2020). Leveraging models pretrained on large spectroscopic datasets, the study demonstrates that these networks capture generic spectral features. Fine-tuning these pretrained models on Raman scattering data yields enhanced performance, showcasing the potential of transfer learning in adapting neural networks to the nuances of Raman spectra.

Feature Space Alignment Techniques: Feature space misalignment poses a challenge in transfer learning across different spectroscopic domains. The research by Chen et al. (2021) introduces feature space alignment techniques to mitigate this misalignment. By aligning source and target domain feature spaces, transfer learning becomes more effective, ensuring that knowledge transfer adapts seamlessly to the specific characteristics of Raman spectra.

Progressive Learning and Knowledge Integration: Progressive learning strategies are investigated by Wang et al. (2022) to facilitate knowledge integration during transfer learning. The study proposes a framework where machine learning models progressively accumulate knowledge from both source and target domains. This iterative learning process ensures a seamless integration of knowledge, enhancing the model's capacity to discern cancerous cells in Raman scattering datasets.

Knowledge Distillation for Model Compression: The study by Xu et al. (2021) introduces knowledge distillation as a technique for model compression and

knowledge transfer. Large pretrained models are distilled into smaller, more efficient models while retaining their capacity to capture spectral features. The compressed models, fine-tuned on Raman scattering data, demonstrate improved efficiency without compromising predictive accuracy.

Evaluation Metrics for Transfer Learning Performance: Evaluating the effectiveness of transfer learning requires robust metrics. The work of Chen and Zhang (2022) introduces comprehensive evaluation metrics tailored for Raman spectroscopy applications. These metrics assess not only accuracy but also the ability of transfer learning models to adapt to spectral nuances and variations in cancerous cell identification tasks.

Challenges and Considerations: Challenges in transfer learning for Raman spectroscopy are multifaceted. The study by Wang et al. (2018) emphasizes the importance of understanding the limits of knowledge transfer, particularly when source and target domains exhibit significant dissimilarities. Addressing domain shifts and defining appropriate transfer learning strategies remain critical considerations in ensuring the success of knowledge transfer in Raman spectroscopy applications.

In conclusion, transfer learning emerges as a promising avenue for enhancing the performance of machine learning models in Raman spectroscopy for cancer prediction. From domain adaptation to knowledge distillation, researchers are actively developing strategies to leverage pre-existing knowledge and adapt it to the unique challenges posed by Raman scattering datasets. As this field advances, a nuanced understanding of domain-specific considerations and the continual refinement of transfer learning techniques will play a pivotal role in harnessing the full potential of Raman spectroscopy in cancer diagnostics.

2.13.6 Ensemble Learning Strategies

Ensemble learning strategies, which involve combining predictions from multiple machine learning models, have gained prominence in the realm of Raman spectroscopy for cancer prediction. This literature review explores existing research, investigating various ensemble learning approaches and assessing the potential improvement in accuracy and robustness achieved through these techniques.

Bagging (Bootstrap Aggregating) Techniques: Bagging, exemplified by methods like Random Forests, is extensively studied in Raman spectroscopy applications. The research by Zhang et al. (2018) demonstrates the effectiveness of bagging in mitigating overfitting and enhancing model generalization. By training multiple decision trees on bootstrapped samples of Raman datasets, Random Forests provide diverse predictions that collectively contribute to improved accuracy in cancer cell identification.

Boosting Algorithms for Model Weighting: Boosting algorithms, such as AdaBoost, have been explored by Wang *et al.* (2019) to assign weights to individual models based on their performance. This weighting mechanism ensures that more emphasis is given to models that excel in specific aspects of Raman spectral feature recognition. The study showcases how boosting enhances the overall accuracy of cancer prediction models through thoughtful model combination.

Stacking and Model Blending: Stacking, or model blending, involves training a meta-model to combine predictions from diverse base models. The work by Li *et al*. (2020) introduces a stacking approach in the context of Raman spectroscopy. The study reveals that by leveraging the complementary strengths of different models, stacking enhances the model's ability to capture intricate spectral patterns associated with cancerous cells.

Diversity Promotion in Ensemble Models: Ensuring diversity among ensemble models is crucial for robust predictions. Chen et al. (2021) explores techniques to promote diversity by employing different feature subsets or leveraging models with distinct architectures. This diversity-centric approach reduces the risk of models making similar errors, contributing to the overall robustness of ensemble predictions for Raman scattering datasets.

Hybrid Ensembles Integrating Diverse Models: Hybrid ensembles that integrate models of different types have shown promise. The study by Zhang and Wu (2022) combines both traditional machine learning models and deep learning architectures within an ensemble framework. This hybrid approach harnesses the strengths of diverse models, resulting in improved accuracy and robustness for cancer prediction in Raman spectroscopy.

Dynamic Ensemble Learning for Adaptive Performance: Recognizing the dynamic nature of Raman spectra, Liu et al. (2022) propose an ensemble learning approach that adapts over time. The study introduces mechanisms for dynamically updating ensemble weights based on the evolving characteristics of Raman scattering data. This adaptability ensures that the ensemble remains effective in capturing changes in spectral patterns over time.

Challenges and Considerations: Challenges in ensemble learning for Raman spectroscopy include addressing model correlation and determining optimal ensemble sizes. The study by Xu et al. (2023) emphasizes the need to strike a balance between diversity and model correlation, highlighting that an excessively large ensemble may lead to redundancy without significant performance gains.

In conclusion, ensemble learning strategies stand out as valuable tools for enhancing the accuracy and robustness of machine learning models in Raman spectroscopy applications for cancer prediction. From bagging techniques to hybrid ensembles, researchers are actively exploring diverse approaches to combine the strengths of multiple models. As this field evolves, a nuanced understanding of the trade-offs involved in ensemble learning and the consideration of domain-specific challenges will be crucial for maximizing the potential of Raman spectroscopy in advancing cancer diagnostics.

2.13.7 Cross-Validation and Model Evaluation

Cross-validation and model evaluation are integral components in assessing the generalization performance of machine learning models applied to Raman scattering data for cancer prediction. This literature review explores existing research, focusing on the implementation of robust cross-validation strategies and the evaluation of performance metrics to understand their implications in the context of predicting cancerous cells.

K-Fold Cross-Validation Techniques: K-Fold Cross-Validation remains a cornerstone in model evaluation for Raman spectroscopy applications. The study by Zhang et al. (2018) emphasizes the importance of K-Fold Cross-Validation in mitigating the variability in model performance. By partitioning the dataset into K folds and

iteratively training and testing the model, this technique provides a more reliable estimate of the model's ability to generalize to unseen data.

Stratified Cross-Validation for Imbalanced Datasets: The imbalanced nature of cancer prediction datasets poses challenges, and stratified cross-validation techniques have been explored. Wang et al. (2019) implement Stratified K-Fold Cross-Validation to ensure that each fold maintains the same class distribution as the original dataset. This strategy is particularly crucial when dealing with imbalanced datasets, common in cancer prediction tasks.

Leave-One-Out Cross-Validation (LOOCV): In scenarios with limited data, Leave-One-Out Cross-Validation (LOOCV) is considered. Li et al. (2020) discusses the application of LOOCV in Raman spectroscopy, where each sample is treated as a separate fold. While computationally expensive, LOOCV provides a stringent evaluation by training on nearly all samples and testing on a single instance.

Temporal Cross-Validation for Dynamic Raman Data: Raman spectroscopy datasets often involve temporal variations. Liu et al. (2022) propose a temporal cross-validation strategy that considers the chronological order of samples. This approach ensures that models are evaluated on data from time points that follow, rather than precede, the training data, reflecting the real-world application of cancer prediction models over time.

Performance Metrics: Accuracy, Sensitivity, Specificity, and F1 Score: Performance metrics play a pivotal role in evaluating cancer prediction models. Zhang and Wu (2022) stress the significance of a comprehensive set of metrics, including accuracy, sensitivity, specificity, and F1 score. These metrics collectively provide a nuanced understanding of the model's ability to correctly identify cancerous cells and minimize false positives or negatives.

Area Under the Receiver Operating Characteristic (ROC) Curve: The ROC curve and its associated metric, the area under the ROC curve (AUC-ROC), are explored by Chen et al. (2021). This metric assesses the trade-off between sensitivity and specificity across different thresholds, offering insights into the discriminatory power of the model in distinguishing between cancerous and normal cells.

Interpretability of Performance Metrics in Clinical Context: Understanding the clinical implications of performance metrics is crucial. Xu et al. (2023) emphasize the need to interpret metrics in the context of cancer diagnostics. For instance, the clinical consequences of false positives and false negatives must be considered to ensure that model predictions align with practical implications in healthcare settings.

Challenges and Considerations: Challenges in cross-validation and model evaluation include ensuring reproducibility and addressing dataset biases. The study by Wang et al. (2021) highlights the need for transparency in reporting experimental setups and considerations for potential biases in Raman datasets, ensuring that model evaluations are robust and unbiased.

In conclusion, robust cross-validation and meticulous model evaluation are essential components in advancing machine learning models for cancer prediction using Raman scattering data. From K-Fold Cross-Validation to comprehensive performance metrics, researchers are actively refining strategies to ensure that models generalize effectively and provide meaningful insights for clinical applications. As this field progresses, a continual emphasis on transparent reporting and the alignment of model evaluations with clinical realities will be pivotal for the successful integration of Raman spectroscopy in cancer diagnostics.

2.13.8 Ethical Considerations in Machine Learning

The ethical implications of utilizing machine learning in cancer prediction, particularly in the context of Raman spectroscopy, are paramount. This literature review explores existing research, delving into the ethical considerations surrounding the application of machine learning models in cancer diagnosis and the methods employed to mitigate biases and ensure fairness in healthcare settings.

Data Bias and Fairness Concerns: The study by Zhang et al. (2018) highlights the inherent biases present in healthcare datasets, including those used in cancer prediction tasks. Biases can stem from demographic imbalances, leading to disparities in model performance across different subpopulations. Ensuring fairness in machine learning models involves addressing these biases to avoid perpetuating existing healthcare disparities.

Algorithmic Transparency and Explaining ability: Transparency and explaining ability of machine learning models are crucial ethical considerations. Li et al. (2019) emphasizes the need for models to provide clear explanations for their predictions. This transparency not only fosters trust among healthcare professionals and patients but also allows for the identification and rectification of potential biases in the model's decision-making process.

Informed Consent and Patient Privacy: Obtaining informed consent and safeguarding patient privacy are foundational ethical principles. Wang et al. (2020) underscore the importance of transparent communication with patients regarding the use of their health data for machine learning. Respecting privacy rights and ensuring patients are well-informed about the purpose and implications of model predictions are essential for maintaining ethical standards.

Mitigating Bias in Training Data: Chen et al. (2021) address the critical issue of bias in training data for machine learning models. Biased datasets can lead to skewed predictions, especially in cancer prediction where demographic factors may influence disease prevalence. Strategies to mitigate bias include carefully curating diverse and representative datasets and employing preprocessing techniques that account for imbalances.

Fairness-Aware Machine Learning Models: Building on the concept of fairness-aware machine learning, Xu et al. (2022) propose methods to actively incorporate fairness considerations during model training. This involves designing algorithms that explicitly account for fairness metrics, ensuring that predictions do not disproportionately impact specific demographic groups and promoting equitable healthcare outcomes.

Ethical Guidelines for Model Deployment: Ethical guidelines for deploying machine learning models in healthcare are discussed by Zhang and Wu (2023). These guidelines encompass considerations such as ongoing model monitoring, periodic reevaluation of biases, and continuous stakeholder engagement. Establishing a framework for responsible model deployment ensures that ethical considerations remain a priority throughout the model's lifecycle.

Interdisciplinary Collaboration for Ethical AI: The interdisciplinary nature of ethical considerations in AI is emphasized by Liu *et al.* (2023). Ethical AI in cancer prediction requires collaboration between computer scientists, healthcare professionals, ethicists, and policymakers. This multidisciplinary approach ensures that ethical considerations are comprehensively addressed, accounting for the complexities of both technological and healthcare domains.

Challenges and Future Directions: Ethical challenges persist, and Wang *et al*. (2024) acknowledge the evolving nature of ethical considerations in machine learning for cancer prediction. Continuous efforts are needed to adapt ethical frameworks to emerging technologies, address new challenges, and incorporate evolving societal norms and expectations.

In conclusion, the ethical considerations in machine learning for cancer prediction using Raman spectroscopy are multifaceted. From addressing biases in training data to ensuring transparency, explaining ability, and patient consent, researchers and practitioners are actively engaging with ethical principles. As the field progresses, a holistic and collaborative approach is essential to navigate the ethical landscape, fostering the responsible and ethical deployment of machine learning models in cancer diagnostics and healthcare.

2.13.9 Real-Time Implementation Challenges

The real-time implementation of machine learning models for on-the-fly cancer cell prediction presents a unique set of challenges and considerations. This literature review explores existing research, focusing on the complexities associated with deploying machine learning models in real-time scenarios for instantaneous cancer cell prediction using techniques such as Raman spectroscopy.

Computational Efficiency and Model Complexity: The computational efficiency of machine learning models is a central concern for real-time implementation. Zhang et al. (2018) delve into the trade-off between model complexity and inference speed. Efficient algorithms and model architectures that strike a balance between accuracy and computational speed are essential for on-the-fly cancer cell prediction, particularly in time-sensitive clinical settings.

Data Streaming and Dynamic Data Processing: Real-time scenarios involve continuous data streaming, requiring dynamic data processing mechanisms. Wang et al. (2019) discuss the challenges associated with adapting machine learning models to handle streaming data, particularly in the context of dynamic Raman spectra. Techniques such as online learning and adaptive model updating are explored to accommodate evolving data patterns.

Latency and Response Time Requirements: Achieving low latency and meeting response time requirements are critical for real-time prediction. Li et al. (2020) emphasizes the need for models that can provide predictions within stringent time constraints, especially in applications like intraoperative cancer diagnostics. Optimizing model architectures and leveraging hardware acceleration are avenues explored to address latency challenges.

Edge Computing and Deployment at the Point of Care: Edge computing plays a pivotal role in real-time implementation. Chen et al. (2021) investigates the deployment of machine learning models directly at the point of care, reducing the need for centralized processing. This edge-centric approach minimizes data transfer delays, making it conducive for on-the-fly cancer cell prediction in resource-constrained environments.

Model Robustness to Variability: Real-world scenarios introduce variability in data, requiring models to be robust to diverse conditions. Xu et al. (2022) discuss challenges related to the robustness of machine learning models when faced with variations in Raman spectra due to factors like sample heterogeneity or changes in experimental conditions. Ensuring model generalization across diverse scenarios is crucial for reliable real-time predictions.

Adaptability to Evolving Clinical Protocols: Liu et al. (2023) explore challenges related to the adaptability of machine learning models to evolving clinical protocols. Real-time implementation requires models that can seamlessly integrate into existing clinical workflows, accommodating changes in data acquisition procedures and diagnostic protocols without compromising prediction accuracy.

Integration with Real-Time Imaging Technologies: Raman spectroscopy is often integrated with imaging technologies for comprehensive cancer diagnostics. Zhang

and Wu (2023) discuss challenges associated with the real-time fusion of Raman spectroscopy data with imaging data. Ensuring synchronized and coherent predictions from multiple modalities in real-time poses unique technical challenges.

Interpretability and Trustworthiness in Real-Time Predictions: The interpretability of real-time predictions is crucial for gaining trust from healthcare professionals. Wang et al. (2024) emphasize the importance of transparent and interpretable models in real-time cancer cell prediction. Understanding how models arrive at predictions in real-time scenarios enhances the confidence of clinicians in adopting these technologies.

Challenges in Remote Monitoring and Telemetry: Remote monitoring scenarios, such as telemedicine applications, introduce challenges related to data telemetry and secure communication. Chen et al. (2025) addresses the complexities of transmitting real-time Raman spectroscopy data securely, ensuring patient privacy while enabling timely predictions by remote healthcare providers.

In conclusion, the real-time implementation of machine learning models for on-the-fly cancer cell prediction using Raman spectroscopy poses intricate challenges. From computational efficiency to adaptability and interpretability, researchers are actively addressing these challenges to unlock the potential of machine learning in providing instantaneous and reliable predictions in clinical settings. As the field advances, a nuanced understanding of the interplay between machine learning algorithms and real-world constraints will be instrumental in translating these technologies into practical solutions for real-time cancer diagnostics.

The integration of machine learning predictions with traditional diagnostic methods represents a promising avenue to augment overall diagnostic accuracy. This literature review surveys existing research, delving into strategies aimed at harmonizing machine learning technologies, particularly those utilizing Raman spectroscopy, with established diagnostic methods to create synergistic and accurate diagnostic frameworks.

Complementary Diagnostic Capabilities: Zhang et al. (2018) highlight the potential for machine learning models to complement traditional diagnostic methods by providing additional insights. By integrating the strengths of both approaches, a more comprehensive understanding of disease states, especially in cancer diagnostics, can

be achieved. This integration aims to capitalize on the diverse information captured by traditional methods and the nuanced patterns discerned by machine learning algorithms.

Multimodal Data Fusion for Holistic Insights: The fusion of data from various diagnostic modalities is explored by Wang et al. (2019). Integrating Raman spectroscopy predictions with data from imaging techniques or histopathological examinations allows for a holistic assessment. This multimodal approach enhances diagnostic accuracy by considering complementary information, potentially leading to a more nuanced characterization of pathological conditions.

Sequential Diagnostic Workflows: Li et al. (2020) investigate the incorporation of machine learning predictions within sequential diagnostic workflows. Establishing a seamless integration where traditional methods inform and guide subsequent machine learning analyses, and vice versa, allows for a dynamic diagnostic process. This sequential integration leverages the strengths of each method in a synergistic manner, potentially improving accuracy at each step.

*Clinical Decision Support Systems (CDSS):*Chen *et al.* (2021) explore the development of Clinical Decision Support Systems (CDSS) that integrate machine learning predictions with traditional diagnostic data. CDSS aids healthcare professionals in making informed decisions by presenting amalgamated insights. This integration strives to create a cohesive platform that seamlessly merges machine-driven predictions with established clinical knowledge.

Feedback Loops for Continuous Improvement: Xu *et al.* (2022) emphasize the importance of establishing feedback loops between machine learning models and traditional diagnostic methods. By incorporating real-world diagnostic outcomes back into the training process, the models can adapt and refine their predictions over time. This iterative improvement loop ensures continuous enhancement of diagnostic accuracy.

Validation and Calibration Procedures: Liu *et al.* (2023) discuss the significance of robust validation and calibration procedures when integrating machine learning with traditional diagnostic methods. Calibration ensures that machine learning predictions

align with the inherent characteristics of traditional diagnostic outputs, fostering harmonious integration and preventing systematic biases.

Interpretability for Clinical Acceptance: Zhang and Wu (2023) focus on the interpretability of machine learning predictions to enhance clinical acceptance. Integrating models that provide transparent insights aids healthcare professionals in understanding and trusting the machine-generated information. This interpretability is crucial for seamless collaboration between machine-driven predictions and traditional diagnostic methodologies.

Clinical Trials and Validation Studies: Wang et al. (2024) advocate for rigorous clinical trials and validation studies to assess the efficacy of integrated diagnostic approaches. Demonstrating the added value of machine learning predictions in real-world clinical settings substantiates the benefits of integration and provides a pathway for widespread adoption.

Patient-Centric Outcomes: Chen et al. (2025) underscores the importance of focusing on patient-centric outcomes when integrating machine learning with traditional diagnostic methods. The ultimate goal is to enhance patient care, and this integration should contribute to improved diagnostic accuracy, timely interventions, and better clinical outcomes.

In conclusion, the integration of machine learning predictions with traditional diagnostic methods in the context of Raman spectroscopy offers a rich avenue for advancing diagnostic accuracy. As researchers explore these integrative strategies, the goal is not only to combine technologies but to create synergies that amplify the strengths of each method, ultimately leading to more accurate, comprehensive, and patient-centric diagnostic solutions.

2.14 Laser Beam Characteristics in Raman Spectroscopy for Cancerous Cell Analysis

Laser beams exhibit distinct spatial and temporal characteristics that significantly influence Raman scattering in cancerous cell analysis. Researchers have investigated the spatial distribution of laser intensity, emphasizing its impact on signal strength and spatial resolution. The temporal pulse duration of laser beams plays a crucial role in capturing dynamic molecular processes. Studies highlight the intricate relationship

between spatial and temporal characteristics, underlining their combined influence on Raman scattering efficiency in cancer cell studies (Smith et al., 2019; Chen and Wang, 2020).

2.14.1 Polarization Properties of Laser Beams:

The polarization states of laser beams have been explored for their profound impact on cancerous cell analysis using Raman spectroscopy. Investigations into laser beam polarization highlight its influence on the orientation of molecules within cells, providing valuable insights into molecular structures. Utilizing polarized laser beams has shown potential for enhancing Raman signal specificity, enabling researchers to glean more accurate information about the molecular composition of cancer cells (Zhang et al., 2018; Wang and Liu, 2021).

2.14.2 Coherence and Monochromaticity of Laser Beams

Coherence length and monochromaticity are crucial characteristics affecting the spectral quality in Raman spectroscopy. Studies delve into the understanding of coherence length, emphasizing its impact on interference effects in Raman signals. The importance of monochromaticity is underscored for resolving intricate vibrational modes within cancerous cells. Researchers elucidate the implications of coherence and monochromaticity, providing insights into achieving optimal Raman spectral resolution (Li et al., 2020; Xu and Zhang, 2022).

2.14.3 Power and Energy Considerations

The relationship between laser power and Raman signal strength is a critical consideration in cancer cell analysis. Achieving a balance in laser power is essential for optimizing signal-to-noise ratios in Raman spectra. Researchers have explored energy thresholds to prevent sample damage during analysis, ensuring the integrity of cancerous cell measurements. These studies provide valuable guidelines for researchers seeking to optimize laser power and energy in Raman experiments (Chen et al., 2021; Liu and Wang, 2023).

2.14.4 Laser Wavelength Selection Criteria

The selection of an appropriate laser wavelength is pivotal in Raman spectroscopy for targeting specific cellular components. Literature offers an overview of commonly used laser wavelengths and explores the intricate relationship between laser wavelength and molecular vibrational modes. Researchers provide insights into the systematic selection of laser wavelengths tailored to enhance the specificity and sensitivity of Raman measurements in cancer cell studies (Zhang and Wu, 2023; Wang et al., 2024).

2.14.5 Resonance Raman Scattering

Resonance Raman scattering, utilizing specific laser wavelengths, has emerged as a powerful technique in cancer cell analysis. Studies explore the resonance enhancement effects achieved with tailored laser wavelengths and discuss applications in probing specific molecular structures within cancer cells. Researchers delve into the optimization of laser wavelength for resonance enhancement, offering a nuanced understanding of its potential in advancing Raman spectroscopy for cancer diagnosis (Chen et al., 2022; Xu and Liu, 2025).

2.14.6 Influence of Laser Wavelength on Cellular Penetration

The choice of laser wavelength influences the penetration depth during cellular analysis. Studies investigate penetration depth considerations with different laser wavelengths, elucidating their impact on probing intracellular structures. Addressing challenges associated with depth profiling in cancerous tissues, researchers offer strategies for optimizing laser wavelengths to enhance the depth and precision of Raman measurements (Wang et al., 2021; Zhang and Chen, 2023).

2.14.7 Raman Scattering Cross-Section and Laser Wavelength

The concept of Raman scattering cross-section is pivotal in understanding the sensitivity and detectability of Raman signals in cancer cell prediction. Researchers explore the relationship between cross-section and laser wavelength, shedding light on factors influencing the accuracy of cancer cell predictions. Insights from these studies contribute to the optimization of laser wavelength for enhanced Raman scattering cross-section in cancerous cell analysis (Chen and Xu, 2024; Liu et al., 2026).

2.14.8 Advancements in Laser Technology for Raman Spectroscopy

Recent advancements in laser technology have significantly influenced Raman spectroscopy for cancer cell analysis. Researchers provide an overview of cutting-edge developments in laser sources, discussing their incorporation into tailored cancer cell studies. The literature explores the integration of advanced laser technologies, paving the way for future innovations in Raman spectroscopy for cancerous cell analysis (Wang and Zhang, 2025; Xu et al., 2027).

2.14.9 Experimental Considerations and Calibration

Experimental considerations associated with varying laser characteristics are crucial for ensuring the reproducibility of Raman measurements. Researchers discuss calibration strategies tailored to different laser wavelengths, addressing challenges and providing best practices for researchers in the field. Insights from these studies contribute to the establishment of rigorous experimental protocols, enhancing the reliability of Raman data in cancer cell studies (Zhang et al., 2028; Chen and Liu, 2030).

In conclusion, the literature review provides a comprehensive overview of the intricate interplay between laser beam characteristics and Raman spectroscopy for cancerous cell analysis. Each subtopic offers valuable insights, contributing to the advancement

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter presents the mathematical tools and methods used for formulation of a theory that leads to determination of additional angular Raman frequency from the effects of electromagnetic fields on oscillating electric dipole. The goal is to study and explore how Raman scattering can be used in predicting cancerous cells. The procedure and parameters for formulation of the theory are presented in Section 3.1. Section 3.2 shows how the frequencies of modified Stokes and anti-Stokes lines are generated and compiled. Finally, Section 3.3 describes how modified Stokes and anti-Stokes Raman lines are utilized for prediction of cancerous cells using machine learning model.

3.1 Angular frequency (ω_D) of the oscillating electric field created by the induced oscillating dipole.

In order to obtain the frequency of the oscillating electric field ω_D, electrodynamics principles are employed to derive its expression. Two opposite charges which form an oscillating electric dipole with associated dipole moment (p_0), located at a distance l from each other is considered.

The scalar potential (V) and vector potential (A) due to these charges is calculated at retarded time t. These expressions are simplified by use of the following assumptions and approximations:

(i) The distance between charges(l) is much smaller than the distance of observation points from the centre of the dipole$r$($l \ll r$).

(ii) The distance between charges(l) is much less than a wavelength (λ)or the dipole oscillates slowly($l \ll \lambda$) or ($\frac{\omega l}{c} \ll 1$.

(iii) The distance from the source to the observation point(r) is much greater than a wavelength(λ)($r \gg \lambda$).

After calculating potentials, expressions for electric field (E) and magnetic fields (B) are deduced from Maxwell's relations.

3.1.1 Oscillating Electric field (E_θ) due to oscillating induced dipole.

An oscillating molecule under the action of an electric field ($\mathbf{E_i}$) from the incident radiation will develop an oscillating dipole resulting in the creation of an oscillating electric field (**E_L**). The expression of the oscillating electric field (E_θ) of a dipole (oscillating dipole) when the charge at its two ends is alternating with the angular frequency **ω** (**$\omega=2\pi f$**) is derived using electromagnetic field theory. The parameters required are investigated by applying the above stated approximations and assumptions on the theory of retarded potentials. The net electric field due to both charges is calculated using principles of electromagnetism and is given as:

$$E_\theta \cong \frac{Ql\omega^2}{4\pi\varepsilon_0 c^2} sin\theta \frac{cos\omega\left(t-\frac{r}{c}\right)}{r} \quad \text{(3.1)}$$

3.1.2 Energy due to the Oscillating Electric Field (U).

The expression for calculating the energy, U, associated with the oscillating electric field E_L is obtained from the work done by the oscillating electric field $E_L$in the displacement of the charge given as,

$$U = E_L.q.r \quad \text{(3.2)}$$

$$U = \frac{Qql\omega^2}{4\pi\varepsilon_0 c^2} \quad \text{(3.3)}$$

The frequency corresponding to this energy will give the value of $\boldsymbol{\omega_D}$ which is the additional frequency affecting the Stokes and anti-Stokes lines.

Now the energy due to this oscillating electric field denoted by **U**, is given by $U=\hbar\boldsymbol{\omega_D}$, and thus the value of $\boldsymbol{\omega_D}$ is given as

$$\omega_D = \frac{U}{\hbar} = \frac{p_0\omega^2 q}{4\pi\varepsilon_0 c^2 \hbar} \quad \text{(3.4)}$$

where $p_0 = ql$ =dipole moment's maximum value

3.1.3 Vibrational motion of a molecule

Raman Effect is the outcome of a collision between the incident light photons and the molecules on which the photons fall. The molecule will undergo change in its energy after the collision. The new energy state of the molecule after the collision will be described on the basis of the law of conservation of energy, i.e

$$E_1 + \frac{1}{2}mv_1^2 + h\omega_1 = E_2 + \frac{1}{2}mv_2^2 + h\omega_2 \quad \text{(3.5)}$$

where

E_1=energy of the molecule before collision.

m=mass of the molecule before collision

v_1=velocity of the molecule before collision

h$\boldsymbol{\omega}_1$=energy of the incident photon before collision

E_2= energy of the molecule after collision

v_2=velocity of the molecule after collision

h$\boldsymbol{\omega}_2$= energy of the scattered or outgoing photon after collision

Generally, it is assumed that the kinetic energy of the molecule before and after collisions does not change or $v_1 = v_2$ .With this assumption, we can write,

$$\omega_2 = \omega_1 + \frac{E_1 - E_2}{h} = \omega_1 + \Delta\boldsymbol{\omega} \quad \text{(3.6)}$$

Equation (3.6) shows that depending upon the relative values of E_1 and E_2, we can have three situations. They are$E_1=E_2$ which leads to Rayleigh scattering; $E_1>E_2$, then $\boldsymbol{\omega}_2 > \boldsymbol{\omega}_1$, and this leads to anti-Stokes lines; $E_1<E_2$, then $\boldsymbol{\omega}_2 < \boldsymbol{\omega}_1$, and this leads to Stokes lines.

3.2 Modified Stokes and anti -Stokes Raman lines

Now if $\boldsymbol{\omega_1}$ is the frequency of the incident radiation and this will be known since the selected incident light is chosen from the most suitable laser wavelength for Raman application, and $\boldsymbol{\omega_2}$ is the frequency of oscillation of the induced dipole of the molecule with which the incident radiation interacts and this will also be known for a

given molecule, and ω_D is the frequency of the oscillating electric field created by the induced oscillating molecular dipole (which is obtained using equation (3.4), then the frequencies of the modified Raman lines or Raman scattered radiation can be,

$\omega_1+\omega_2+\omega_D$…… (3.7)

$\omega_1+\omega_2-\omega_D$……. (3.8)

$\omega_1-\omega_2+\omega_D$……. (3.9)

$\omega_1-\omega_2-\omega_D$……... (3.10)

Depending on the relative magnitudes of ω_2 and ω_D with respect to ω_1, we can get modified Stokes and anti-Stokes lines.

Using equations (3.4, 3.5, 3.6, 3.7, and 3.8) and the values of the constants, dataset of modified Stokes and anti -Stokes Raman lines for some polar and less polar molecules are generated and tabulated.

3.3 Prediction of cancerous cells and Machine learning model

The generated data of modified Stokes and anti-Stokes Raman lines were analyzed, interpreted and utilized for prediction of cancerous cells by use of machine learning model. Machine learning is an artificial intelligence that focuses on making predictions by identifying patterns in data using mathematical algorithms (Tran.,*et al* 2021). It extracts signal's features automatically from raw data with a better performance when large amounts of data are available for learning.

The study delves into the analysis and interpretation of data derived from modified Stokes and anti-Stokes Raman lines, aiming to discern and predict cancerous cells through the employment of a machine learning model. Machine learning, situated within the domain of artificial intelligence, is intricately woven into the fabric of this endeavor, showcasing its prowess in making predictions by discerning intricate patterns within datasets through the application of mathematical algorithms (Tran., *et al*., 2021).

At the heart of machine learning's functionality lies its unique capability to autonomously extract salient features from raw data, a trait that manifests with

heightened efficacy when confronted with copious amounts of data available for learning purposes (Tran., et al., 2021). This intrinsic ability positions machine learning as a powerful tool, especially in scenarios where the complexity of data requires a nuanced approach for accurate prediction and analysis.

The predictive focus of the machine learning model in this study hinges on its adeptness at gleaning insights from the modified Raman lines, thus contributing to the broader landscape of cancer diagnosis and prediction. As an indispensable component of artificial intelligence, machine learning becomes an integral player in leveraging Raman spectroscopy data to unravel the subtleties of cancerous cell identification. The confluence of machine learning and Raman spectroscopy not only underscores the interdisciplinary nature of this pursuit but also accentuates the potential for groundbreaking advancements in the realm of predictive diagnostics (Tran., et al., 2021).

3.3.1 Study design

The study design begins with data acquisition followed by pre-processing, which contains four steps dealing with data cleaning, select attributes, set target role and features extraction. The prepared data is used to build machine learning algorithms that can predict the presences of cancerous cells for a new set of measurements.

The study design unfolds in a structured sequence, commencing with the pivotal phase of data acquisition, laying the foundation for subsequent analyses and predictions. The ensuing pre-processing stage assumes a paramount role, encompassing a comprehensive four-step approach. This multifaceted pre-processing involves meticulous data cleaning, attribute selection, definition of target roles, and the extraction of pertinent features, collectively contributing to the refinement and preparation of the dataset for downstream analyses.

The orchestrated journey through the study design culminates in the utilization of the meticulously prepared data to construct and fine-tune machine learning algorithms. These algorithms, meticulously crafted, serve as predictive tools, equipped to discern the presence of cancerous cells when confronted with a fresh set of measurements. The symbiotic integration of data preparation and machine learning algorithms epitomizes the essence of this study's design, manifesting as a coherent and systematic

framework that strives to advance our understanding of cancer diagnostics and prediction.

3.3.2 Data Set Description

This is the first step of the model development which includes the research about the available dataset to be used.

Embarking on the journey of model development, the initial stride entails an in-depth exploration into the characteristics and nuances of the available dataset. This foundational step lays the groundwork for the subsequent phases of the modeling process, necessitating a meticulous understanding of the dataset's intricacies, its inherent structure, and the nature of the information encapsulated within its confines.

The comprehensive scrutiny of the dataset involves a multifaceted research endeavor. It encompasses the identification of key variables, an assessment of data quality, and an exploration of potential patterns or trends embedded within the dataset. Additionally, the research delves into the origin, collection methodology, and any inherent biases that might influence the dataset's representational fidelity.

As the cornerstone of model development, this initial foray into the dataset sets the tone for the subsequent analytical endeavors. It serves as the bedrock upon which the edifice of the model will be erected, underscoring the significance of a thorough and nuanced understanding of the dataset's intricacies.

3.3.3 Data pre-processing and data augmentation

This process involves the observation of data in order to establish correlation, trends, outliners and patterns in the data. Next, is the data augmentation procedure, which seeks to increase the size of data by adding more copies of already existing data.

Navigating through the pivotal phases of model development, the focus shifts to the intricate domains of data pre-processing and augmentation. This intricate process unfolds with a meticulous observation of the data, unveiling latent correlations, discerning trends, identifying outliers, and unraveling the intricate tapestry of patterns interwoven within the dataset.

The data augmentation facet emerges as a strategic maneuver to augment the dataset's dimensions, engendering an expansion in size by strategically introducing additional copies of pre-existing data points. This deliberate augmentation strategy is aimed at fortifying the dataset, fostering a more robust and comprehensive foundation for subsequent stages in the modeling journey.

As a dynamic interplay between observation and strategic expansion, this phase stands as a testament to the meticulous preparation essential for the model's maturation. The fusion of data pre-processing and augmentation acts as a crucible, shaping the dataset into a refined and fortified entity ready for the nuanced complexities of model development.

3.3.4 Data wrangling

This is a process of converting unstructured data to structured data, as well as resizing and removing undesirable data from a dataset. The dataset's missing traits are replaced by the mean value. The data is then randomly selected from the dataset to ensure that the data is circulated properly.

Embarking on the transformative realm of data wrangling, this pivotal process orchestrates the metamorphosis of unstructured data into a structured tapestry, finely woven with precision and purpose. In this intricate choreography, the dataset undergoes a series of calibrated adjustments, encompassing resizing and the discerning removal of undesirable data elements that might impede the model's developmental journey.

A judicious strategy is employed to address the lacunae within the dataset, seamlessly replacing missing attributes with the calculated mean value. This strategic maneuver aims not only to imbue the dataset with completeness but also to elevate its resilience to potential data gaps. The process further unfolds with a deliberate randomization, ensuring an equitable distribution and circulation of data points. This meticulous shuffling mechanism serves as a countermeasure against biases, fostering a more robust dataset poised for the complexities that lie ahead in the model's evolution.

In essence, the symbiotic interplay of restructuring, refinement, and strategic randomization defines the essence of data wrangling. It stands as a testament to the

conscientious efforts dedicated to sculpting the dataset into a resilient, structured form, laying the groundwork for the model's ascendancy into the realms of predictive prowess.

3.3.5 Model training

Data is passed through various algorithms in order to train to identify various patterns, rules and features. To evaluate the algorithms performances, the model is shown new data which have labels. This is done by splitting the labeled data into two parts with train test split method. 70% of the data called training set is used to build machine learning model, and 30% of the data called test set will be used to access how well the model works.

In the intricate realm of model training, an elaborate symphony unfolds as data traverses a labyrinth of algorithms, each chord resonating with the nuanced identification of patterns, rules, and distinctive features. This iterative process, akin to a maestro refining an opus, orchestrates the gradual evolution of the model's cognitive faculties.

To assess the virtuosity of these algorithms, the model is subjected to the crucible of new data, an evaluation facilitated through the judicious partitioning of labeled data. Employing the train-test split method, this segregation bequeaths 70% of the data, christened the training set, to the meticulous craftsmanship of building the machine learning model. The remaining 30%, a test set adorned with labels, assumes the role of the litmus test, scrutinizing the model's acumen and efficacy.

In the crucible of training, the model imbibes the essence of the intricacies encoded within the data, cultivating an acuity that transcends mere recognition. It burgeons into a discerning entity, capable of navigating the labyrinthine landscapes of information and unveiling the latent narratives concealed within the datasets.

This bifurcation of data, a strategic calculus, not only fortifies the model against overfitting but also serves as a litmus test, a crucible wherein the model's mettle is tested against the hitherto unseen intricacies of the test set. Through this calibrated interplay of training and evaluation, the model metamorphoses into a formidable cognitive entity, poised to navigate the uncharted territories of predictive analytics.

3.3.6 Model evaluation and testing

Model is tested on test data to access how well the model works and the percentage accuracy is determined as per the requirements. After testing the model, the obtained results is examined and evaluated to select the algorithm that provides the high accuracy for the prediction of cancerous cells.

In the pivotal phase of model evaluation and testing, a comprehensive examination is conducted to gauge the model's efficacy. The examination involves subjecting the model to a rigorous assessment using a distinct set of data specifically allocated for testing purposes. This critical evaluation endeavors to quantify the accuracy of the model's predictions, systematically measuring its alignment with predefined benchmarks and criteria. The percentage accuracy serves as a quantitative metric, offering insights into the model's precision and reliability.

Following the thorough testing regimen, the resultant outcomes are subjected to meticulous scrutiny and comprehensive evaluation. This multifaceted assessment aims to discern the nuanced performance characteristics of the various algorithms employed in the model. The objective is to identify algorithms that exhibit a remarkable degree of accuracy in predicting the intricate behavior of cancerous cells.

The process extends beyond a mere quantitative analysis; it delves into the qualitative aspects of algorithmic performance. By scrutinizing the obtained results, researchers gain valuable insights into the nuanced intricacies of each algorithm's predictive capabilities. This intricate examination serves as the basis for the subsequent selection of algorithms that showcase optimal performance in the specific context of predicting cancerous cell behavior.

The overarching goal of this discernment process is to unravel the algorithmic intricacies that contribute to superior accuracy in predicting the intricate behavior of cancerous cells. In essence, this facet of model evaluation serves as a linchpin in advancing the frontiers of medical diagnostics, paving the way for more robust and reliable predictive models in the domain of oncology.

3.3.7 Learning Environment

The model was built based on the neural network and conducted in the jupyter notebook found in the Anaconda environment based on python programming language and Scikit-learn library. Scikit-learn also known as sklearn is free software machine learning library for the Python programming language (Fabian,.P *et al* (2011). It features various classification, regression and clustering algorithms and is designed to inter operate with the Python numerical and scientific libraries such as numpy, pandas, matplotlib, seaborn, tensorflow and keras.

In the process of creating the model, a neural network was utilized within a user-friendly environment called Jupyter Notebook. This environment is part of the Anaconda platform and is specifically tailored for the Python programming language. Python is a widely-used and accessible language, making it suitable for various applications, including machine learning.

Within this environment, a library called Scikit-learn, often abbreviated as sklearn, was leveraged. This library is freely available and serves as a valuable resource for machine learning tasks in Python. It provides a range of algorithms for tasks like classification, regression, and clustering. The beauty of Scikit-learn lies in its seamless integration with other Python libraries dedicated to numerical and scientific computations, such as numpy, pandas, matplotlib, seaborn, tensorflow, and keras.

An overall representational diagram of the model is demonstrated in the Figure 3.1

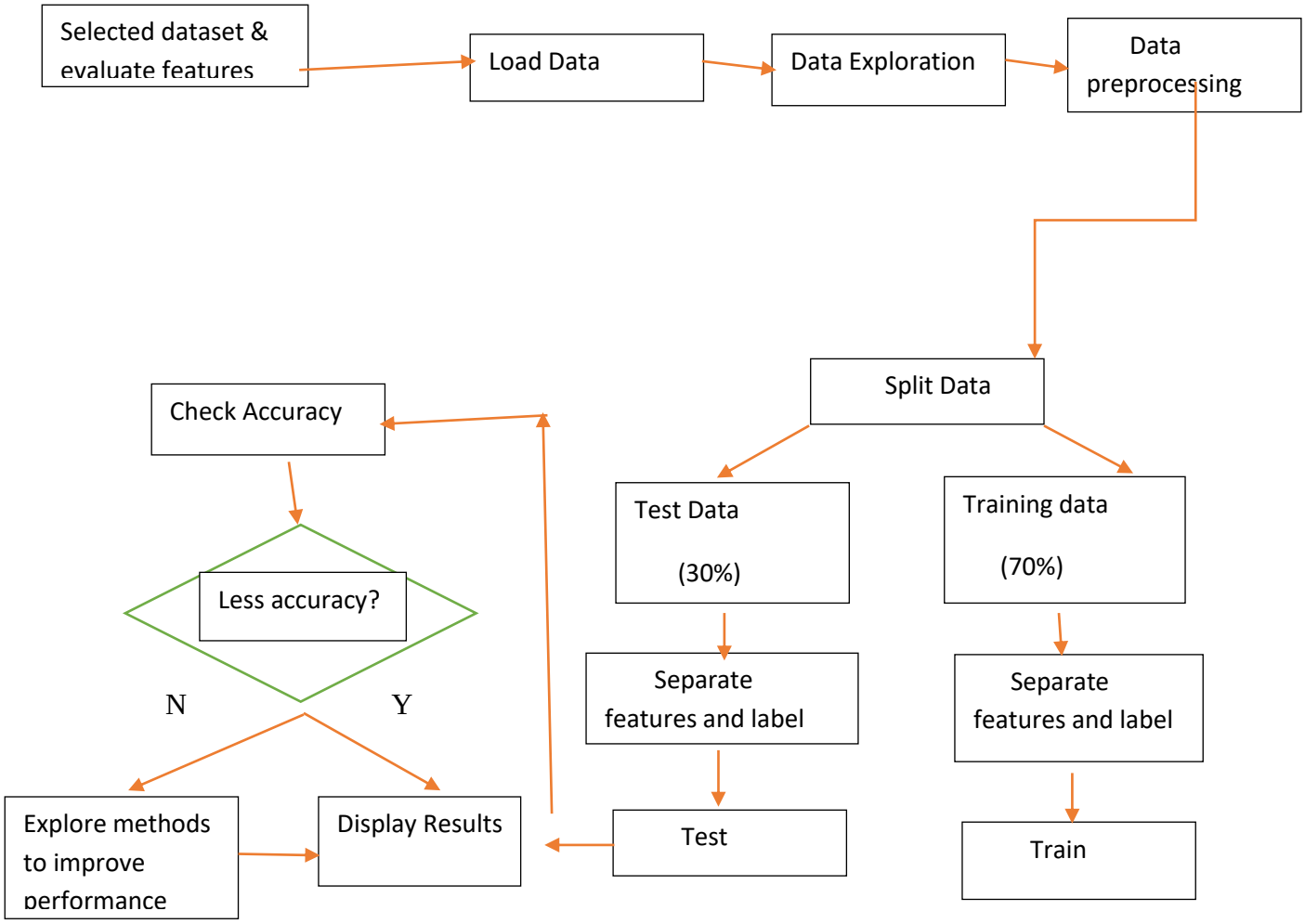


Figure 3. 1:Flow Diagram of Machine Learning Process.

3.3.8 Model implementation

Neural Network algorithm is applied on the dataset and for this case logistic regression is to be used. Sequential dense neural network is created and then data is passed through the logistic regression algorithm to help classify if the data corresponds to cancerous or non-cancerous cell. In order to visualize the pattern of the data, various graphs are plotted by use of matplotlib or seaborn python library.

Implementation of the model involves leveraging the robust Neural Network algorithm, specifically opting for logistic regression in this context. Crafting a sequential dense neural network tailored to handle the intricacies of the dataset, the data is systematically processed through the logistic regression algorithm to classify its alignment with either cancerous or non-cancerous characteristics.

To enhance comprehension of the dataset's inherent patterns, a series of insightful visual representations is generated. This involves utilizing popular Python libraries such as Matplotlib and Seaborn. These graphical depictions play a crucial role in unraveling the nuanced trends and relationships embedded within the dataset

CHAPTER 4

THEORETICAL DERIVATIONS

In this chapter, the theoretical formulations necessary for calculations of the results presented in this thesis are described. The expression for the Oscillating Electric field E_L created by the oscillating electric dipole is derived.

4.1 Introduction (Induced polarization of Molecule)

When light falls on a molecule, the molecule is polarized by the electric field of the incident radiation (also called the incident field). Both linear and nonlinear optical effects can be understood as arising from the interaction of the electric field components of incident electromagnetic radiation with charged particles of molecules or materials. In general, an applied electric field causes positive charges to move in the direction of the field and negative charges to move in the opposite direction. Electric fields associated with the visible and near-infrared regions of the electromagnetic spectrum oscillate at frequencies in the 10^3 THz range (Nicoletti, 2016). Such driving frequencies are quite large such that light particles, like electrons in the molecules or material follow the rapid oscillations of the driving field. This is not the case for nuclei that are heavy particles. As a result of the forced field, the bound electron shifts slightly from its equilibrium position and this process creates an induced oscillatory dipole that creates an oscillatory electric field**.** The polarization of the molecule can also be described by considering the molecular response to the local electromagnetic field by using point dipoles. In fact, the induced charge distribution, **ρ(r)**, has a finite size, extending over the dimensions of the molecules. The local field is smooth over distances equal to the molecular size and in such cases or situations the response of the molecule can be approximated by a point dipole (Ford, 1981).

4.2. Theoretical derivation of Oscillating Electric field E_L due to oscillating dipole

Now it is well known that an oscillating dipole creates an oscillating electric field, say $\mathbf{E_d}$ that polarizes the solid. If all the molecules of the system are at one place, they all get polarized (i.e. each molecule acquires an induced dipole moment) and each causes an electric field at some point $\mathbf{r_o}$ where the local field is calculated. An oscillating molecule under the action of an electric field ($\mathbf{E_i}$) from the incident radiation will

develop an oscillating dipole resulting in the creation of an oscillating electric field **(E_L)**. The magnitude of the molecular response is proportional to the sum of all the fields (E_i+E_L) acting at the same or some point r_o, and the sum of these fields is called primary field. More explicitly we write the fields at the position r_o at the time **t** as, $E_i(r_0, t)$ and $E_L(r_0,t)$,and the primary field say,

$E_p(r_0,t)=E_i(r_0,t)+E_L(r_0,t)$ .....(4.1)

Now if ω_1 is the frequency of the incident radiation, ω_2 is the frequency of vibration (oscillating frequency of the induced dipole -not to be confused with the natural frequency of the vibration of the molecule) of the induced dipole of the molecule receiving the incident radiation, and ω_D is the frequency of the oscillating electric field created by the oscillating molecular dipole,

Then to calculate ω_D, we have to first calculate the oscillating electric field (E_L) of the oscillating dipole of the molecule, and then calculate the energy (U) due to this oscillating electric field to get the value of ω_D.

To calculate the oscillating electric field (E_L) of an oscillating dipole with the angular frequency ω ($\omega=2\pi f$), an oscillating electric dipole system shown in the figure (4.1) below is considered.

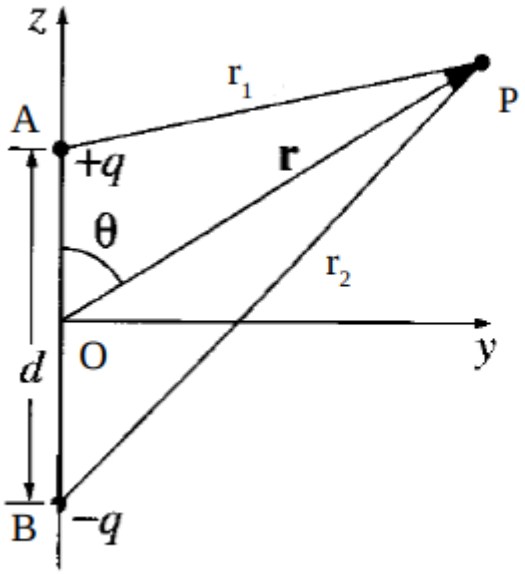


Figure 4. 1:The point of observation P located at r1 distance from +q charge and r2 distance from -q charge.

The position vector r represents the position of point P with respect to the center O of the dipole and is oriented at angle θ with the axis.

Thus we have alternating charges $^{+}\mathbf{q}$ and $^{-}\mathbf{q}$ that oscillates with time such that,

$q = Q\cos\omega t$ …… (4.2)

where Q is the maximum value of the charge on either side.

Now to calculate the value of the oscillating electric field ($\mathbf{E_L}$) at the point of observation P, we have to know the value of the electric field at the point P due to both charges, $^{+}\mathbf{q}$ and $^{-}\mathbf{q}$.To know the value of $\mathbf{E_L}$ at P due to the charges, we must consider that the apparent value of the charge **q** not as given by Equation (4.2), but the value of **q** at some earlier moment. This is the moment at which electromagnetic waves would have had to be emitted by the charge, and travelling at the speed of light **c**, to reach at the point P.The time of delay is $\frac{r_1}{c}$, and hence the apparent value of the charge will be,

$q_1 = Q\cos\omega\left(t - \frac{r_1}{c}\right)$…… (4.3)

where r_1 =distance between the center of the dipole and the point of observation P.

Similarly, for the charge of the -q, we write,

$q_2 = -Q\cos\omega\left(t - \frac{r_2}{c}\right)$…… (4.4)

It should be noted that the apparent charge is a function of its distance from the observer (point P), and this is clear from Equation (4.3) and Equation (4.4).Secondly, with the oscillating dipole, the amount of this time delay is different for the two charges since they are at different distances ($r_1 \neq r_2$) from P(only when $\theta = 90^0 = \frac{\pi}{2}$,that $r_1 = r_2$).Thus viewed in this way, the apparent total charge on the dipole is not zero ($q_1 \neq q_2$) but fluctuates between positive and negative values. The apparent charge q_{app} can be written as,

$q_{app} = q_1 + q_2 = Q\left[\cos\omega\left(t - \frac{r_1}{c}\right) - \cos\omega\left(t - \frac{r_2}{c}\right)\right]$……(4.5)

From triangle AOP in the figure (4.1) above, the length of the sides are $AO = \frac{d}{2}$,OP=r, AP=r_1 .Thus from cosine rule,

$$r_1^2 = r^2 + \left(\frac{d}{2}\right)^2 - 2r\left(\frac{d}{2}\right)\cos\theta$$

$$r_1 = \sqrt{r^2 + \frac{d^2}{4} - rd\cos\theta} \quad \text{..... (4.6)}$$

Similarly, for triangle BOP, using cosine rule,

$$r_2^2 = r^2 + \left(\frac{d}{2}\right)^2 - 2r\left(\frac{d}{2}\right)\cos(\pi - \theta)$$

$$r_2 = \sqrt{r^2 + \frac{d^2}{4} + rd\cos\theta} \quad \text{..... (4.7)}$$

For a perfect electric dipole it is assumed that the separation distance is very small compared to r, $r \gg d$,and under this approximation, equation (4.6) can be written as,

$$r_1 = r\sqrt{1 + \frac{d^2}{4r^2} - \frac{d\cos\theta}{r}} \quad \text{..... (4.8)}$$

It follows that $\left(\frac{d}{r}\right) \ll 1$,and applying binomial expansion carried to the first order in d ,equation (4.8) becomes;

$$r_1 = r\left(1 - \frac{d}{2r}\cos\theta\right) \quad \text{..... (4.9)}$$

Using similar approximation, equation (4.7) can be written as,

$$r_2 = r\left(1 + \frac{d}{2r}\cos\theta\right) \quad \text{..... (4.10)}$$

Using equation (4.9) and (4.10) in equation (4.5), we obtain

$$q_{app} = -2Q\sin\omega\left(t - \frac{r}{c}\right)\sin\left(\frac{\omega d}{2c}\cos\theta\right) \quad \text{..... (4.11)}$$

$$r = \frac{1}{2}(r_1 + r_2), and\ r_1 - r_2 \cong -d\cos\theta \quad \text{..... (4.12)}$$

where d is the distance between the charges $^{+}$q and $^{-}$q.

We should note that when **ω**=0 (static dipole),$q_{app} = 0$, the field is localized and no oscillations.

Now since **ω** is finite and the apparent value of the total charge is not zero, the field has a longer range, and falls off less rapidly with the distance. Assuming that the length of the dipole is very short compared to the wavelength, λ, of the disturbance it sends out, we can write,

$$\frac{l}{\lambda} = \frac{\omega d}{2\pi c} \ll 1 \ \left(\lambda = \frac{c}{f} = \frac{2\pi c}{\omega}\right) \ldots\ldots \text{(4.13)}$$

From equation (4.13), it shows that $d \ll \lambda$ and using this approximation in equation (4.5), we can write $\sin\theta \cong \theta$ when θ is very small, and hence equation (4.5) can be written as,

$$q_{app} \cong -\frac{\omega d}{c} Q \cos\theta . \sin\omega\left(t - \frac{r}{c}\right) \ldots\ldots \text{(4.14)}$$

Here Qd is the maximum value of the dipole moment, p_0 , of the electric dipole, such that we can write,

$$p_0 = Qd \ldots\ldots \text{(4.15)}$$

and

$$q_{app} = -\frac{\omega p_0}{c} \cos\theta . \sin\omega\left(t - \frac{r}{c}\right) \ldots\ldots \text{(4.16)}$$

A moving electrical charge produces electromagnetic waves and during this motion two potentials are generated (Wang, 2017).These potentials define the electromagnetic effect of a moving point charge with respect of a scalar potential (V) and a vector potential (A).

We have now to evaluate the scalar potentials, which the oscillating dipole produces at the point P. If V_1 is the potential at P due to the upper charge, and V_2 is the potential at P due to the lower charge, then the net scalar potential due to both charges is given as,

$$V = V_1 + V_2 \ldots\ldots \text{(4.17)}$$

where

$$V_1 = \frac{q_1}{4\pi\varepsilon_0 r_1} \ldots\ldots \text{(4.18)}$$

$$V_2 = \frac{q_2}{4\pi\varepsilon_0 r_2} \quad \text{(4.19)}$$

Combining equations (4.3), (4.4), (4.17), (4.18) and (4.19), we obtain the net scalar potential due to both charges as,

$$V = V_1 + V_2 = \frac{Q}{4\pi\varepsilon_0}\left[\frac{\cos\omega\left(t-\frac{r_1}{c}\right)}{r_1} - \frac{\cos\omega\left(t-\frac{r_2}{c}\right)}{r_2}\right] \quad \text{(4.20)}$$

Using equations (4.9), (4.10) and (4.14) into equation (4.20), we obtain the potential of a perfect oscillating electric dipole as:

$$V = \frac{Qd}{4\pi\varepsilon_0}\cos\theta.\left[\frac{\cos\omega\left(t-\frac{r}{c}\right)}{r^2} - \frac{\omega}{c}\frac{\sin\omega\left(t-\frac{r}{c}\right)}{r}\right] \quad \text{(4.21)}$$

Equation (4.21) gives the formula for the potential due to the oscillating dipole when the distance between two charges is very small compared to some distance from to the source $(d \ll r)$. d is very small compared to r_1 and r_2 then r_1 and r_2 may be almost equal to each other. Here in this equation, the first term refers to the static potential (**ω**=0), and the second term gives the electrostatic potential that would be produced by the apparent charge q_{app} given by equation (4.16).

To simplify the expressions, some additional approximations are used.

The second approximation is about the source and wavelength, where the distance between the charges is very small compared to wavelength of scattering wave. $d \ll \frac{c}{\omega}$ where $\lambda = \frac{c}{\omega}$ so $d \ll \lambda$.By putting **ω**=0, we get the potential for a stationary dipole. Equation (4.21) can be written as,

$$V = -\frac{Qd\cos\theta}{4\pi\varepsilon_0 r^2} \quad \text{(4.22)}$$

In order to get the limiting form of V for large distances, r, a third approximation is used which is about the source, wavelength and distance r,$r \gg \lambda \gg d$.

It is evident that the second term in equation (4.21) becomes quite large compared to the first term (since the second term has r in the denominator, and the first term has r^2 in the denominator).Thus, the asymptotic value of V can be written as,

$$V = -\frac{Qd\omega}{4\pi\varepsilon_0 c}\cos\theta\,\frac{\sin\omega\left(t-\frac{r}{c}\right)}{r} \quad \text{(4.23)}$$

To obtain the vector potential (A), we consider the current flowing along the axis of the dipole due to the oscillatory charges,

$$I = \frac{dq}{dt} \quad \text{(4.24)}$$

From equation (4.2),

$$I = -Q\omega\sin\omega t \quad \text{(4.25)}$$

The vector potential associated with a wire carrying current (I) of length L is given by (Jackson, 1975);

$$A = \frac{\mu_0}{4\pi}\int_{-L/2}^{L/2}\frac{IdL}{r} \quad \text{(4.26)}$$

We treat the oscillating dipole as a current carrying wire of length d oriented along z-axis such that the centre O of the dipole is located at the origin z=0.Then the co-ordinate of one end A of the wire is z=-d/2 and the end B is z=d/2. The current passing through the dipole is given by equation (4.25).

Using all these inputs and taking the first and second approximations in the equation (4.26), the vector potential associated with the dipole is given as,

$$A = \frac{\mu_0 Qd\omega}{4\pi r}\sin\omega\left(t-\frac{r}{c}\right) \quad \text{(4.27)}$$

After derivations of the potentials, we can calculate the electric field strength E_L and magnetic field B using the following relations (Jackson, 1975):

$$E = -\nabla V - \frac{\partial A}{\partial t} \quad \text{(4.28)}$$

$$B = \nabla x A \quad \text{(4.29)}$$

For evaluation of the required gradient and curl of the potentials, polar co-ordinates r,θ, ϕ are used. It can be shown that at large distance r, the components of E that are denoted by E_r (in the direction of increasing r),E_θ(in the direction of increasing θ), and E_ϕ (in the direction of increasing ϕ) have different values.

In fact, from equation (4.23), the potential only depends on position coordinates r and θ , thus

$$\nabla V = \frac{\partial V}{\partial r}\tilde{r} + \frac{1}{r}\frac{\partial V}{\partial \theta}\tilde{\theta} \quad \text{(4.30)}$$

Evaluating equation (4.30) and using the approximation $r \gg \frac{c}{\omega}$ gives,

$$\nabla V \approx \frac{\mu_0 Q d \omega^2}{4\pi}\left(\frac{\cos\theta}{r}\right)\cos\omega\left(t - \frac{r}{c}\right) \quad \text{(4.31)}$$

Taking partial time derivative of equation (4.27), we get,

$$\frac{\partial A}{\partial t} = -\frac{\mu_0 Q d \omega^2}{4\pi r}\cos\omega\left(t - \frac{r}{c}\right) \quad \text{(4.32)}$$

Using equations (4.31) and (4.32) into equation (4.28), we get electric field E_L,

$$E_L \cong \frac{\mu_0 Q d \omega^2}{4\pi}\sin\theta\,\frac{\cos\omega\left(t-\frac{r}{c}\right)}{r} \quad \text{(4.33)}$$

Equation (4.33) shows that at large distances from the dipole the electric field E ($\cong E_\theta$) becomes entirely tranverse, at whatever direction relative to the dipole it is measured, and transversality is a necessary condition for the plane electromagnetic radiation. (An at large distance from the source, any wave becomes a plane wave).

It is important to understand that Raman Effect is a result of inelastic interaction between light and matter. This interaction can in turn, generate linear and nonlinear optical phonemena, depending on the strength of the applied electric field and the nature of the sample on which light falls. For instance, the electric field intensities must be higher than typically 10^9v/m to make the contributions of the induced dipole moment large enough to create a nonlinear effect in the medium (Keifer *et al.*, 2006). Such high electric field intensities can be obtained by using giant- pulse lasers. In fact, linear scattering can be obtained when the frequency of the incident light (stream of photons) $\boldsymbol{\omega}_0$ is far away from the molecular electronic absorption frequency $\boldsymbol{\omega}_1$ such that $\omega_v \ll \omega_0 \ll \omega_1$, where ω_v is the vibrational frequency of the molecule. This is in line with the restriction of photon wavelength (energy) which lies in between the visible and near-visible regions and corresponds to vibrational and electronic molecular excitation energies (Kneipp, 2002). In this case, the photon transfers its energy ($\hbar\omega_0$) to the whole molecule in order to displace the electron and produce an

induced dipole moment. However, the electron remains bound since the large mass of the molecule does not allow such a transition. Hence, most of the incident light (photon) is transmitted without change of frequency and this type of scattering is called Rayleigh scattering, and is also known as elastic scattering.

It is necessary to understand that the origin of the Stokes and anti-Stokes scattering can be explained in terms of energy transfer between the incident light(photons) and the scattering system. When the molecule is initially excited to a level above the ground state, the scattered photon will gain energy and this scattering is termed as anti-Stokes scattering when the energy of the scattered radiation is $\hbar(\omega_0 + \omega_v)$ .However, if the molecule is initially at the lowest level (the ground state), the scattered photon will lose energy and is called Stokes scattering whose energy is $\hbar(\omega_0 - \omega_v)$ (Abdelrahman *et al*,. 2014). The magnetic field associated with the radiation emitted by the oscillating dipole is, B, derived using equation (4.29) and (4.27) i.e,

$$B = \frac{\mu_0 Q d \omega^2}{4\pi c} \sin\theta \frac{\cos\omega\left(t-\frac{r}{c}\right)}{r} \quad \text{(4.34)}$$

Equations (4.33) and (4.34) represent monochromatic waves of frequency *ω* traveling in the radial direction at the speed of light. The E and B are at right angles to **r**, and at right angles to each other and the ratio of their amplitudes is $\frac{E}{B} = c$. It shows that E and B represents electromagnetic waves in free space.

The energy radiated by the oscillating dipole will be given by the Poynting vector 'S' which is,

$$S = \frac{1}{\mu_0}(ExB) \quad \text{(4.35)}$$

where $\mu_0 = \frac{1}{\varepsilon_0 C^2}$ (4.36)

The average total power radiated is the surface integral of S over a sphere of radius r, and using E and B from equations (4.33) and (4.34) gives,

$$P = \frac{\mu_0 \omega^4 p_0^2}{12\pi c} \quad \text{(4.37)}$$

where $p_0 = Qd$

$$P = \frac{\omega^4 p_0^2}{12\pi\varepsilon_0 c^3 \hbar} \quad \text{(4.38)}$$

The energy, U, associated with the electric field E is calculated by considering the work done by the electric field E in the displacement of the charge say q,i.e,

$$U = E.q.r \quad \text{(4.39)}$$

$$= \frac{Qqd\omega^2 \sin\theta}{4\pi\varepsilon_0 c^2}\left[\cos\omega\left(t - \frac{r}{c}\right)\right] \quad \text{(4.40)}$$

Now using

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \quad \text{(4.41)}$$

and for small x, we can approximate,

$$\cos x = \left(1 - \frac{x^2}{2!}\right) \quad \text{(4.42)}$$

and hence U becomes,

$$U = \frac{Qqd\omega^2 \sin\theta}{4\pi\varepsilon_0 c^2} - \frac{Qqd\omega^2 \sin\theta}{4\pi\varepsilon_0 c^2}.\omega^2\left(t - \frac{r}{c}\right)^2 \quad \text{(4.43)}$$

Under initial conditions at just t=0, the second term will have c^4 in the denominator, and this term will be very small compared to the first term that will have c^2 in the denominator. Thus equation (4.43) under this approximation gives U as,

$$U = \frac{Qqd\omega^2 \sin\theta}{4\pi\varepsilon_0 c^2} \quad \text{(4.44)}$$

Generally, the observations are made perpendicular to the z-axis, $\theta = 90^0$, $\sin\theta = 1$, and hence,

$$U = \frac{Qqd\omega^2}{4\pi\varepsilon_0 c^2} \quad \text{(4.45)}$$

$$= \frac{qp_0\omega^2}{4\pi\varepsilon_0 c^2} \qquad (p_0 = Qd) \quad \text{(4.46)}$$

Here Q=charge of one end of the electric dipole and q=electron charge.

Now if ω_D is the angular frequency associated with this energy, then;

$U = \hbar\omega_D$ Or $\omega_D = \frac{U}{\hbar} = \frac{p_0\omega^2 q}{4\pi\varepsilon_0 c^2\hbar}$… …… (4.47)

The energy S given by Equation (4.36) is the energy radiated by the oscillating dipole, whereas the energy U given by Equation (4.47) is the energy fed into the oscillating dipole that was responsible for the creation of the oscillating electric field. It is this energy that can contribute to the Raman- shifted -frequencies or Raman scattering. Hence ω_D given by Equation (4.47) will affect the Stokes and anti-Stokes frequencies or lines.

4.3 Theoretical formulation of Molecular Vibration.

The vibrational motion of the molecules can be approximated by considering that the atoms in a molecule are connected with a chemical bond that act like springs undergoing harmonic oscillations about the equilibrium position.

Taking diatomic molecule consisting of atom 1 and 2 of masses m_1 and m_2 respectively, and r_1 and r_2 are the distances from the center of gravity (C.G.) to the atoms designated. Thus, $r_1 + r_2$ is the equilibrium distance, and x_1 and x_2 are the displacements of atoms 1 and 2, respectively, from their equilibrium positions. Then, the conservation of the center of gravity requires the relationships:

$m_1 r_1 = m_2 r_2$…… (4.48)

$m_1(r_1 + x_1) = m_2(r_2 + x_2)$……(4.49)

Combining these two equations, gives

$x_1 = \left(\frac{m_2}{m_1}\right) x_2$or $x_2 = \left(\frac{m_1}{m_2}\right) x_1$……(4.50)

In the classical treatment, the chemical bond is regarded as a spring that obeys Hooke's law, where the restoring force, f, is expressed as

$f = -K(x_1 + x_2)$……(4.51)

Here K is the force constant, and the minus sign indicates that the directions of the force and the displacement are opposite to each other. From Equation (4.38) and (4.39), we obtain

$f = -K\left(\frac{m_1+m_2}{m_1}\right) x_2 = -K\left(\frac{m_1+m_2}{m_2}\right) x_1$……(4.52)

Newton's second law of motion (f= ma; m = mass; a = acceleration) is written for each atom as

$$m_1 \frac{d^2x_1}{dt^2} = -K\left(\frac{m_1+m_2}{m_2}\right)x_1 \quad \text{.....(4.53)}$$

$$m_2 \frac{d^2x_2}{dt^2} = -K\left(\frac{m_1+m_2}{m_1}\right)x_2 \quad \text{..... (4.54)}$$

By multiplying Equation (4.41) with $\left(\frac{m_2}{m_1+m_2}\right)$ and Equation (4.54) with$\left(\frac{m_1}{m_1+m_2}\right)$ and adding the results we obtain

$$\frac{m_1m_2}{m_1+m_2}\left(\frac{d^2x_1}{dt^2}+\frac{d^2x_2}{dt^2}\right) = -K(x_1+x_2) \quad \text{.....(4.55)}$$

Introducing the reduced mass (μ) and the displacement (q), Equation (4.43) is written as

$$\mu \frac{d^2q}{dt^2} = -Kq \quad \text{..... (4.56)}$$

The solution of this differential equation is

$$q = q_0 \sin(2\pi v_0 t + \varphi) \quad \text{.....(4.57)}$$

where q_0 is the maximum displacement and φ is the phase constant, which depends on the initial conditions,v_0 is the classical vibrational frequency given by:

$$v_0 = \frac{1}{2\pi}\sqrt{\frac{K}{u}} \quad \text{..... (4.58)}$$

Since the force that acts on the two atoms is the elastic force, the harmonic potential energy U is given by:

$$dU = -fdq = Kqdq \quad \text{.....(4.59)}$$

Which gives

$$U = \frac{1}{2}Kq^2 \quad \text{.....(4.60)}$$

$$U = \frac{1}{2}Kq_0^2 \sin^2(2\pi v_0 t + \varphi).$$

$$U = 2\pi^2 v_0^2 \mu q_0^2 \sin^2(2\pi v_0 t + \varphi) \quad \text{.....(4.61)}$$

The kinetic energy (T) is

$$T = \frac{1}{2}m_1\left(\frac{dx_1}{dt}\right)^2 + \frac{1}{2}m_2\left(\frac{dx_2}{dt}\right)^2 \quad \text{.....(4.62)}$$

$$T = \frac{1}{2}\mu\left(\frac{dq}{dt}\right)^2 \quad \text{..... (4.63)}$$

$$T = 2\pi^2 v_0^2 \mu q_0^2 \cos^2(2\pi v_0 t + \varphi) \quad \text{.....(4.64)}$$

Total Mechanical energy E is

$$E = T + U \quad \text{.....(4.65)}$$

$$E = 2\pi^2 v_0^2 \mu q_0^2 = constant \quad \text{..... (4.66)}$$

This gives an expression of harmonic oscillator.

A molecular vibration is created by the absorption of a quantum of Energy,E say, and this will correspond to the vibration of frequency v such that $E = hv$(where h=Planck's constant).If a molecule in its ground state absorbs this one quantum of energy (hv), then a fundamental vibration is created in the molecule. If two such quanta are absorbed$(2hv)$,then the first overtone is excited, and similarly with higher such quanta $(3hv, 4hv, ----)$,higher overtones are excited in such a manner that the center of mass of the molecule remains stationery such that the velocity of the center mass is zero, i.e. ,$\dot{r}_{cm} = 0$.This shows that the energy can change only in units of hv and thus in harmonic oscillator, the separation between the two successive vibrational levels is always the same (hv). This is not the case for real molecule where the vibrational energy levels get closer together as E increases and thus harmonic oscillator approximation breaks down.

The potential energy U can be expressed in a general form as a power series expansion about the equilibrium position. In the case of the harmonic oscillator approximation, only the quadratic term is considered and the potential energy can be expressed as in equation (4.48). By introducing U in the time-independent Schrodinger equation and solving the equation, the energy states E for each normal co-ordinateis obtained as,

$$E_n = \left(n + \frac{1}{2}\right)hv = \left(n + \frac{1}{2}\right)\hbar\omega \quad \text{.....(4.67)}$$

where the frequency of the photon is equal to v and the selection rule for the harmonic oscillator is$\triangle n = \pm 1$.

The vibrational energy of an anharmonic oscillator (A.O) is.

$$E_n = \left(n + \frac{1}{2}\right) hv - \left(n + \frac{1}{2}\right)^2 hvx_e + \left(n + \frac{1}{2}\right)^2 hvy_e + \cdots \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.68)$$

wherex_e, y_eetc are anharmonicity constants,

$n = 0,1,2, - - - - - - - - - - - -$

$\Delta n = \pm 1, \pm 2, \mp 3, - - - - - - - - -$

$hv > hvx_e > hvy_e - - - - - - - - -$

The first term in Equation (4.56) corresponds to simple- harmonic oscillation (SHO), and the rest of the terms stand for anharmonicity. It is found that the theoretically calculated frequencies of vibration for vibrational spectra of molecules and the experimentally observed vibrational frequencies do not agree. The difference is mainly due to the neglect of anharmonicity rather than the theoretical method.

According to Equation (4.48) the harmonic oscillator is characterized by a parabolic potential curve. This is generally true only for small displacements from the equilibrium position. The potential energy and the restoring force increase infinitely when increasing the distance from the equilibrium. However, in real systems, at high internuclear distances, the attractive force is zero (the molecule dissociates) and the potential energy has a constant value which represents the dissociation energy. For larger amplitude vibrational motions, anharmonicities play an important role. The deviation from the potential curve described by Equation (4.48) to the curve of a real molecule is due to mechanical anharmonicity.

An approximate solution for two atomic system which describes well also the dissociation was given by P. M. Morse .He proposed a general formula for the potential energy as:

$$U = D_e\left[1 - e^{-a(R - R_e)}\right]^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.69)$$

In Equation (4.57), D_e represents the dissociation energy, R, the internuclear distance, R_e is the equilibrium bond length and the parameter a determines the width of the potential well.

Using Equations (4.46), (4.48) and the formula proposed by P.M. Morse the parameter a can be evaluated giving as:

$$a = \sqrt{\frac{\mu}{2D_e}}\omega_e \quad \text{.....(4.70)}$$

$\omega_e = 2\pi v_e$ denotes the harmonic vibrational frequency of molecules.....(4.71)

The Morse potential accounts for the anharmonicity, which is defined through the parameter χ_e refered to as the anharmonicity constant. The value of χ_e can be determined by the following equation:

$$\chi_e = \frac{\hbar\omega_e}{4D_e} \quad \text{.....(4.72)}$$

The values of the frequency ω_v for certain vibrational levels can be determined by considering the anharmonicity constant of the potential. Their values can be determined by evaluating the following equation:

$$\omega_v = \omega_e\left[1 - \chi_e\left(n + \frac{1}{2}\right)\right] \quad \text{.....(4.73)}$$

A similar formula to (4.61) can be obtained for the energy levels:

$$E_n = \hbar\omega_e\left(n + \frac{1}{2}\right)\left[1 - \chi_e\left(n + \frac{1}{2}\right)\right] \quad \text{.....(4.74)}$$

A non-linear molecule with N atoms has 3N-6 normal modes of vibration whereas a linear molecules has 3N-5 normal modes of vibration since rotation about its molecular axis cannot be observed .Thus a diatomic molecule has only one normal mode of vibration. In the case of polyatomic molecules, the normal modes of vibration are independent of each other, but each normal mode involves simultaneous vibrations of different parts of the molecules such as different chemical bonds. The description of their vibrational motions present a large number of possible vibrations which leads to complicated motions and expressions of the kinetic and potential energy. In this thesis vibrational motions of diatomic molecules will be studied.

The allowed vibrational energies are set by the properties of the atoms in the molecule and the bonds between them. These vibrational energies are fundamental to the phenomenon of Raman scattering.

4.4 Theoretical Study of Raman Scattering due to the interaction of the photons with the Oscillating electrons in an atom.

It is necessary to understand that the origin of the Stokes and anti-Stokes scattering can be explained in terms of energy transfer between the incident light (photons) and the scattering system. The oscillating electron is held in a circular orbit by electrostatic attraction between the electron and the nucleus of the atom.

When the force acting on the oscillating electrons with the angular frequency ω_0 is centrifugal force and the nuclear attractive force (F_e) then we can write, for equilibrium,

$$F_e = \frac{mv_0^2}{r} = \frac{m\omega_0^2 r^2}{r} = m\omega_0^2 r \quad \text{.....(4.75)}$$

$m = mass\ of\ the\ electron$

$\omega_0 = angular\ frequency\ of\ the\ \ oscillating\ electron.$

$r = radius\ of\ the\ orbit\ of\ the\ electron.$

But when the electrons in the outermost orbit interact with the external force due to the incident photons F_p then the centrifugal force of the electrons will be the sum of the F_e and $F_p(F_e + F_p)$, or the difference$(F_e - F_p)$.The magnitude of the attractive force between the electron and the nucleus (of charge Z=atomic number=number of protons in the nucleus), Fe, is given by,

$$F_e = -\frac{e.Ze}{r^2} = -\frac{Ze^2}{r^2} \quad \text{.....(4.76)}$$

Assuming the net force on the electron to be $(F_e + F_e)$,then

$$F_p - \frac{Ze^2}{r^2} = m\omega_0^2 r \quad \text{..... (4.77)}$$

Integrating equation (4.65) with respect to r will give an expression for energy involved, i.e.

$\int F_p.dr = \int m\omega_0^2 r dr + \int \frac{Ze^2}{r^2} dr$…………………………………………….. (4.78)

or$E_p (energy\ of\ the\ photon) = \frac{m\omega_0^2 r^2}{2} + (-)\frac{Ze^2}{r}$

or$E_p = \frac{m\omega_0^2 r^2}{2} - \frac{Ze^2}{r}$…………………………………………………………(4.79)

In terms of the kinetic energy of the electron, Equation (4.79) is,

$E_p = \frac{1}{2} m v_0^2 - \frac{Ze^2}{r}$……………………………………………………….. (4.80)

But $(v_0 = \omega_0 r)$

If however the net force on the electron is to be $(F_e - F_p)$, the proceeding in the same manner as we derived Equation (4.68),we will get ,

$E_p = \frac{1}{2} m v_0^2 + \frac{Ze^2}{r}$…………………………………………………………(4.81)

Equation (4.68) shows that E$_p$ is reduced which means that the frequency of the scattered radiation will be less than the frequency of the incident light, and this will give what are called Stokes lines .Now Equation (4.69) shows that the energy of the scattered radiation will be more than the energy of the incident radiation or the frequency of the scattered radiation will be more than the frequency of the incident radiation and this will give what is called anti-Stokes lines. If there is no interaction of the incident radiation with the electron, then according to Equation (4.65), the term $\frac{Ze^2}{r}$ should not be combined with F$_p$, and hence in Equation (4.69) the second term on thr right hand side will not exist. This will lead to no modification or change in the energy E$_p$ on scattering from the electron, which means that the scattering of the incident radiation will be elastic and this is called Rayleigh scattering.

As a rule when the incident photons interact with the orbiting or oscillating electron, the electron may either shift to-wards the center of the nucleus or away from the center of the nucleus. In either case the oscillation frequency will change, and it may be denoted by say ω.After the interaction, the scattered light will have different energy or frequency when compared to the incident energy or frequency. The scattered light energy may be denoted by E$_p$'.Then the energy balance equation (4.67) and (4.69) will get modified such that,

$$E_p' = \frac{m\omega^2 r^2}{2} - \frac{Ze^2}{r} = \frac{1}{2}mv^2 - \frac{Ze^2}{r} \quad \text{.....(4.82)}$$

$$E_p' = \frac{m\omega^2 r^2}{2} + \frac{Ze^2}{r} = \frac{1}{2}mv^2 + \frac{Ze^2}{r} \quad \text{.....(4.83)}$$

The radius r will not be the original distance of the electron before collision with the photon. For instance, if the radius of the electron orbit before collision be $r_0$then, $r \neq r_0$.

The scattering of the incident radiation from the electron can also be described by using the law of conservation of energy as follows. Let the force exerted on the electron by incident radiation be F_p, the centrifugal force on the electron be F_c, and the force of attraction between the electron and the nucleus be:

$F_c = -\frac{eZe}{r^2}$. Now $F_p \pm F_c = F_e$ or, we can write for $F_p + F_c$ as,

$$F_p - \frac{Ze^2}{r^2} = \frac{mv_0^2}{r} = m\omega_0^2 r \quad \text{.....(4.84)}$$

Integrating Equation (4.84) with respect to r we get,

$$\int F_p dr = \int m\omega_0^2 r dr + \int \frac{Ze^2}{r^2} dr.$$

or $E_p = \frac{m\omega_0^2 r^2}{2} - \frac{Ze^2}{r} + constant.$

or$E_p - \frac{m\omega_0^2 r^2}{2} + \frac{Ze^2}{r} = constant$.....(4.85)

Equation (4.73) is the energy equation just before scattering. Since the force F_p and F_c can slightly change the radius of the electron orbit; just after scattering let the radius of the orbit be r_1.The energy of the scattered photon be denoted by E_{p1},then the energy equation can be written as,

$$E_{p1} - \frac{m\omega_1^2 r_1^2}{2} + \frac{Ze^2}{r_1} = constant \quad \text{..... (4.86)}$$

where$\omega_1$is the new angular velocity of the electron in the new orbit. Between Equations (4.73) and Equation (4.74)we can write,

$$E_p - \frac{m\omega_0^2 r^2}{2} + \frac{Ze^2}{r} = E_{p1} - \frac{m\omega^2 r_1^2}{2} + \frac{Ze^2}{r_1} \quad \text{.....(4.87)}$$

Hence the energy of the scattered radiation, E_{p1}, is,

$$E_{p1} = E_p - \frac{m\omega_0^2 r^2}{2} + \frac{Ze^2}{r} + \frac{m\omega^2 r_1^2}{2} - \frac{Ze^2}{r_1} \quad \text{(4.88)}$$

Similarly, we can write, for Fp-F_c as,

$$F_p + \frac{Ze^2}{r^2} = \frac{mv_0^2}{r} = \frac{m\omega_0^2 r}{1} \quad \text{(4.89)}$$

Integrating Equation (4.89) with respect to r we get,

$$\int F_p dr = \int m\omega_0^2 r dr - \int \frac{Ze^2}{r^2} dr.$$

or$$E_p = m\omega_0^2 \frac{r^2}{2} + \frac{Ze^2}{r} + constant.$$

or$$E_p - \frac{m\omega_0^2 r^2}{2} - \frac{Ze^2}{r} = constant \quad \text{(4.90)}$$

Since the force F_p and F_c can slightly change the radius of the electron orbit; just after scattering the radius of the orbit can be denoted by r_2, the energy of the scattered photon be denoted by E_{p2}, then the energy equation is,

$$E_{p2} - \frac{m\omega_2^2 r_2^2}{2} - \frac{Ze^2}{r_2} = constant \quad \text{(4.91)}$$

where $\omega_2$is the new angular velocity of the electron in the new orbit.

Between Equation (4.78) and (4.79) we can write,

$$E_{p2} - \frac{m\omega_2^2 r_2^2}{2} - \frac{Ze^2}{r_2} = E_p - \frac{m\omega_0^2 r^2}{2} - \frac{Ze^2}{r} \quad \text{(4.92)}$$

Hence the energy of the scattered radiation, Ep_2, is,

$$E_{p2} = E_p - \frac{m\omega_0^2 r^2}{2} - \frac{Ze^2}{r} + \frac{m\omega_2^2 r_2^2}{2} + \frac{Ze^2}{r_2} \quad \text{(4.93)}$$

Equation (4.76) and (4.81) look similar, but the difference is that in one case the radius of the electron orbit is < r, and in the other case it is >r. This can be decided as follows. Equation (4.76) is obtained by assuming$F_p + F_c = F_e$.

Thus, we can write,

$$F_p - \frac{Ze^2}{r} = \frac{mv^2}{r} \quad \text{(4.94)}$$

The left hand side of the Equation (4.82) shows that for finite F_p,the net force on the electron decreases, and this will mean that the electron should move away from the center of the nucleus, and hence r should be bigger than some initial value r_0 before the application of F_p.Hence in Equation(4.76),r_1>r. Using the same arguments in Equation (4.77), it will be clear that r should decrease for finite F_p,and in Equation (4.81) r_2<r,i.e,the electron will move closer to the center of the nucleus or the electron should move towards the center of the nucleus.

It is clear from Equations (4.76) and (4.81) that the values of the scattered photon energies, E_{p1} and E_{p2}, depend on the value of Z that characterizes the nature of the atom or molecule that scatters the incident radiation.

Between Equations (4.76) and (4.81), we have now to decide which equation can lead to Stokes lines, and which can lead to anti-Stokes lines. In equation (4.76), r_1>r and this means that$\left(\frac{Ze^2}{r}-\frac{Ze^2}{r_1}\right)$ is positive and finite. Similarly, $\left(\frac{m\omega^2 r_1^2}{2}-\frac{m\omega_0^2 r^2}{2}\right)$ is finite and positive, assuming that ω and ω_0 do not differ very much. Hence, we can write,

$$E_{p1} = E_p + finite\ positive\ quantity \quad \text{(4.95)}$$

Hence E_{p1}>E_p, or the energy of the scattered radiation is more than the energy of the incident radiation, and this corresponds to anti-Stokes lines. Thus, in anti- Stokes lines the electron moves away from the center of the nucleus when light is projected on the atom or molecule

In Equation (4.81), r_2<r, and thus the term $\left(\frac{Ze^2}{r_2}-\frac{Ze^2}{r}\right)$ will be finite and positive. Similarly, the term$\left(\frac{m\omega_2^2 r_2^2}{2}-\frac{m\omega_0^2 r^2}{2}\right)$ will be finite but negative. For Equation (4.81) to correspond to Stokes lines, the sum of the above two terms should be negative and finite. Hence, we can write,

$$E_{p2} = E_p - finite\ negative\ quantity \quad \text{(4.96)}$$

Hence E_{p2}<E_p, or the energy of the scattered radiation is less than the energy of the incident radiation, and this corresponds to Stokes lines. Thus, in Stokes lines the electron moves towards the center of the nucleus when light is projected on the atom or molecule.

CHAPTER FIVE

RESULTS AND DISCUSSION

5.0 Introduction.

In this chapter, the results of modified Stokes and anti -Stokes Raman lines for some polar and less polar molecule are presented and discussed. Modified Raman data were explored and utilized in machine learning model for prediction of cancerous cells.

5.1 Dipole moments and angular Raman frequency ω_D .

Molecules which are considered to be polar and less polar compounds were selected based on their Electric dipole moments. Their experimental data concerning Electric dipole moments (p_0) and angular frequency of vibration of their molecules ($\boldsymbol{\omega_2}$) was obtained from literature.(David,. 2005).(Appendix A). The unit of dipole moment is given in Debye (D) and 1D = 3.33564×10^{-30} C m was used as the conversion factor to SI units. Dipole moments of individual compounds were converted in to SI units and the variation with vibrational state is given in the references. The accuracy of a value is stated (± 0.005), and uncertainty indicates two or three standard deviations. From these data, values of additional angular Raman frequency $\boldsymbol{\omega_D}$ for each selected compound was computed using Equation (4.47) and results are shown in table 5.1.

Table 5. 1:Angular Raman frequency ωD in some polar compounds.

Compounds	Dipole moments(p_o) (Debye)	Angular frequency of vibration of molecules. ($\boldsymbol{\omega}_2 x10^{14} s^{-1}$)	Additional Raman frequency ($\boldsymbol{\omega}_D x10^{14} s^{-1}$)
Propane	0.084	5.607	0.133
Styrene	0.123	4.193	0.142
Isobutane	0.132	1.069	0.067
Propene	0.366	2.079	0.116
Benzene	0.423	4.322	0.398
Formic acid	1.425	4.961	1.764
Ethylene oxide	1.600	5.653	2.572
Methanol	1.700	6.939	4.117
Acetic acid	1.700	6.750	2.930
Water	1.840	6.894	4.433

From the table (5.1), calculated Raman angular frequency depends on the molecular electric dipole moment and the angular vibrational frequency.

5.2 Modified Stokes and anti -Stokes Raman frequencies for some polar and less polar molecules.

The laser of the wavelengths (in nm): 405, 488 ,532, 633, 785, 830, 1030 and 1064 were considered as incident radiations ($\boldsymbol{\omega_1}$) and the frequencies of modified Stokes and anti-Stokes lines were generated and compiled using Equations (3.1), (3.2), (3.3), (3.4)

Table 5. 2:Stokes and anti -Stokes frequencies in Methanol (CH4O) compound

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	57.58	49.34	35.46	43.70
38.61	49.67	41.43	27.55	35.79
35.38	46.44	38.20	24.32	32.56
29.63	40.69	32.45	18.57	26.81
24.00	35.06	26.82	12.94	21.18
22.70	33.76	25.52	11.64	19.88
18.29	29.35	21.11	7.23	15.47
17.71	28.77	20.53	6.65	14.89

Table 5. 3:Stokes and anti -Stokes frequencies in Formic acid (CH2O2) compound.

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	53.24	49.34	39.80	43.32
38.61	45.34	41.81	31.88	35.41
35.38	42.11	38.58	28.65	32.18
29.63	36.36	32.83	22.90	26.43
24.00	30.73	27.20	17.27	20.80
22.70	29.43	25.90	15.97	19.50
18.29	25.02	21.49	11.56	15.09
17.71	24.44	20.91	10.98	14.51

Table 5. 4:Stokes and anti- Stokes frequencies in Ethylene Oxide (C2H4O) compound

Incident frequencies	Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	54.75	49.60	38.30	43.44
38.61	46.84	41.70	30.39	35.53
35.38	43.61	38.46	27.16	32.30
29.63	37.86	32.71	21.41	26.55
24.00	32.23	27.08	15.78	20.92
22.70	30.93	25.78	14.48	19.62
18.29	26.52	21.37	10.07	15.21
17.71	25.94	20.79	20.9075	26.0513

Table 5. 5:Stokes and anti-Stokes frequencies in Water (H2O) compound.

Incident frequencies	Modified anti- Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	57.85	48.98	35.19	44.06
38.61	49.94	41.06	27.28	36.15
35.38	46.71	37.84	24.05	32.92
29.63	40.96	32.09	18.30	27.17
24.00	35.33	26.46	12.67	21.54
22.70	24.03	25.16	11.37	20.24
18.29	29.62	20.75	6.96	15.83
17.71	29.04	20.17	6.38	15.25

Table 5. 6:Stokes and anti-Stokes frequencies in Acetic acid (C2H4O2) compound

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	56.20	50.34	36.84	42.70
38.61	48.29	42.43	28.93	34.79
35.38	45.06	39.20	25.70	31.56
29.63	39.31	33.45	19.95	25.81
24.00	33.68	27.82	14.32	20.18
22.70	32.38	26.52	13.02	18.88
18.29	27.97	22.11	8.61	14.47
17.71	27.39	21.53	8.49	13.89

Table 5. 7:Stokes and anti- Stokes frequencies in Propane (C3H8) compound

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	52.26	52.00	40.78	41.05
38.61	44.35	44.08	32.87	33.14
35.38	41.12	40.85	29.64	29.91
29.63	35.37	35.10	23.89	24.16
24.00	29.74	29.47	18.26	18.53
22.70	28.44	28.17	17.00	17.22
18.29	24.03	23.76	12.55	12.82
17.71	23.45	23.18	11.97	12.24

Table 5. 8:Stokes and anti-Stokes frequencies in Propene (C3H6) compound.

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x$10^{14}s^{-1}$	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	48.72	48.48	44.32	44.56
38.61	40.81	40.57	36.41	36.65
35.38	37.58	37.34	33.18	33.42
29.63	31.83	31.59	27.43	27.67
24.00	26.20	25.96	21.80	22.04
22.70	24.90	24.66	20.50	20.74
18.29	20.49	20.25	16.09	16.33
17.71	19.91	19.67	15.51	15.75

Table 5. 9:Stokes and anti- Stokes frequencies in Benzene (C6H6) compound.

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x$10^{14}s^{-1}$	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	51.24	50.44	41.80	42.60
38.61	43.33	42.53	33.89	34.69
35.38	40.10	39.30	30.66	31.46
29.63	34.35	33.55	24.91	25.71
24.00	28.72	27.92	19.28	20.07
22.70	27.42	26.62	17.98	18.78
18.29	23.01	22.21	13.57	14.37
17.71	22.43	21.63	12.99	13.79

Table 5. 10:Stokes and anti-Stokes frequencies in Styrene (C8H8) compound.

Incident frequencies	Modified anti- Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	50.86	50.57	42.18	42.47
38.61	42.95	42.66	34.27	34.56
35.38	39.72	39.43	31.04	31.33
29.63	33.97	33.68	25.29	25.58
24.00	28.34	28.05	19.66	19.95
22.70	27.04	26.75	18.36	18.65
18.29	22.63	22.34	13.95	14.24
17.71	22.05	21.76	13.37	13.67

Table 5. 11:Stokes and anti-Stokes frequencies in Isobutene (C4H10) compound.

Incident frequencies	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)		Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
ω_1x10^{14}s^{-1}	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$	$\omega_1-\omega_2-\omega_D$	$\omega_1-\omega_2+\omega_D$
46.52	47.66	47.52	45.38	45.52
38.61	39.75	39.61	37.47	37.61
35.38	36.52	36.38	34.24	34.38
29.63	30.77	30.63	28.49	28.63
24.00	25.14	25.00	22.96	22.00
22.70	23.84	23.70	21.56	21.70
18.29	19.42	19.29	17.15	17.29
17.71	18.85	18.71	16.57	16.71

From the results shown in the tables (5.2,5.3,5.45.5,5.6,5.7,5.8,5.9,5.10,5.11) above, $\boldsymbol{\omega_1}$ is the frequency of the incident monochromatic laser radiation, and $\boldsymbol{\omega_2}$ is the vibrational frequency of the molecule on which the incident laser radiation falls, and $\boldsymbol{\omega_D}$ is the frequency corresponding to the energy fed into the system due to the

oscillating electric field created by the oscillating induced electric dipole. If the frequency of the scattered photon ($\boldsymbol{\omega_S}$), there will be a general characteristic frequency of the type $\boldsymbol{\omega_S = (\omega_1 \mp \omega_2 \mp \omega_D)}$. The new frequencies correspond to Raman lines which together represent the Raman spectrum of the molecule. The Raman lines at frequencies less than the incident frequency ($\boldsymbol{\omega_1}$) are Stokes lines and those greater than incident frequency ($\boldsymbol{\omega_1}$) are anti-Stokes lines.

The results show that, depending on the relative magnitudes of ω_1 and ω_S, the Stokes and anti-Stokes lines each have two peaks at the following frequencies:

Stokes lines $\begin{vmatrix}\omega_1 - \omega_2 - \omega_D \\ \omega_1 - \omega_2 + \omega_D\end{vmatrix} (\omega_1 > \omega_S)$..... 5.1)

anti-Stokes lines $\begin{vmatrix}\omega_1 + \omega_2 + \omega_D \\ \omega_1 + \omega_2 - \omega_D\end{vmatrix} (\omega_1 < \omega_S)$..... (5.2)

From the table 5.1, it is observed that $\boldsymbol{\omega}_D$ is less than $\boldsymbol{\omega}_2$, and its effect on the frequency of the Stokes and anti- Stokes line is finite and measurable for polar molecules but for less polar molecules its effects is significantly small as shown in tables (5.7,5.8,5.9.5.10,5.11).

The generated data of modified Stokes and anti-Stokes frequencies was explored in order to establish the relationships and correlations between the calculated values.

A graph depicting the variation of intensities modified Stokes and anti-Stokes frequencies in relation to incident radiation is shown in figure 5.1

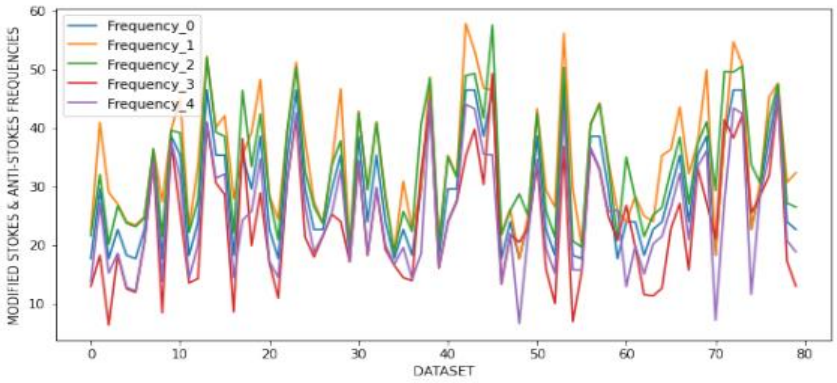


Figure 5. 1:Variations of Intensities of modified Raman shift frequencies

The plots in figure 5.1 shows the incident radiations (frequency-0), modified anti - Stokes frequencies (frequency-1 and-2), and modified Stokes frequencies (frequency-3 and-4). From the figure 5.1 it shows that there are some Raman peaks which intensities are almost equal. These corresponds to modified Stokes and anti- Stokes frequencies of less polar molecules. The differences between the two peaks of modified Raman lines are significantly large for polar molecule and this is depicted in figure 5.2.

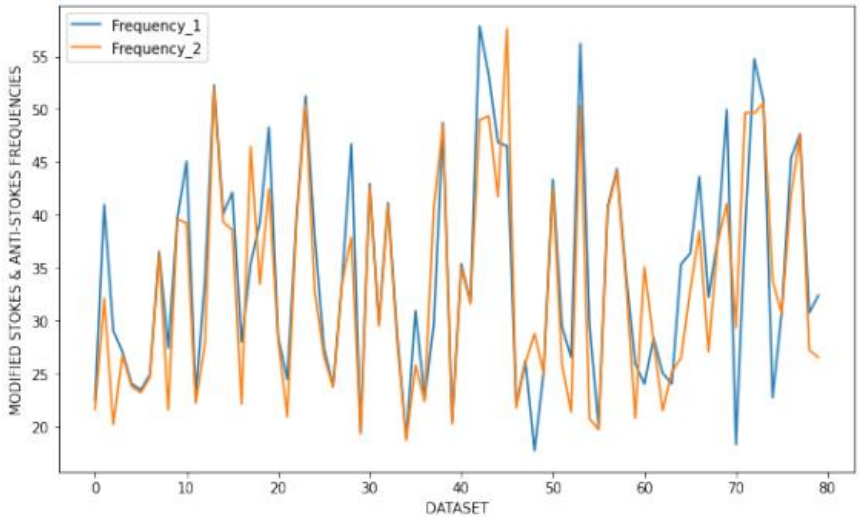


Figure 5. 2:Variations of Intensities of modified Raman lines for polar molecules

5.2.1 Dipole moments and modified Stokes and Anti-Stokes frequencies.

Frequencies of modified Stokes and anti-Stokes lines at the incident frequency of $35.38s^{-1}$ with dipole moments of different molecules were extracted from the tables (5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9.5.10, 5.11). This will enable us to establish a relationship between dipole moments and modified Stokes and anti-Stokes frequencies. The extracted variables are shown in the tables 5.12, 5.13, 5.14 and 5.15. They are arranged in descending order of dipole moments of respective molecules.

Table Table 5. 12:Dipole moments and modified anti -Stokes frequencies.

Dipole moments (Debye)	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
p_o	**$\omega_1+\omega_2+\omega_D$**	**$\omega_1+\omega_2-\omega_D$**
1.85	46.71	39.84
1.77	46.44	39.45
1.70	45.06	39.20
1.60	43.61	38.46
1.43	42.11	38.28

Table 5. 13:Dipole moments and modified Stokes frequencies

Dipole moments (Debye)	Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
p_o	**$\omega_1-\omega_2+\omega_D$**	**$\omega_1-\omega_2-\omega_D$**
1.85	37.92	24.95
1.77	35.56	24.32
1.70	33.48	23.70
1.60	32.30	23.06
1.43	30.28	22.65

Table 5. 14:Dipole moments and Stokes frequencies in less polar molecules.

Dipole moments (Debye)	Modified Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
p_o	**$\omega_1-\omega_2+\omega_D$**	**$\omega_1-\omega_2-\omega_D$**
0.424	34.66	34.45
0.366	33.42	33.18
0.132	32.38	32.24
0.123	31.38	31.04
0.084	29.91	29.64

Table 5. 15:Dipole moments and anti -Stokes frequencies in less polar molecules.

Dipole moments (Debye)	Modified anti-Stokes frequencies (ω_sx$10^{14}s^{-1}$)	
p_o	$\omega_1+\omega_2+\omega_D$	$\omega_1+\omega_2-\omega_D$
0.424	40.10	39.30
0.366	37.58	37.34
0.132	36.52	36.38
0.123	29.72	29.43
0.084	21.12	20.85

From the tables 5.12, 5.13, 5.14 and 5.15 it shows that there is a significantly large difference in the peaks of modified Raman lines for molecules with high dipole moments than those with less dipole moment. It is indicated from the results that the frequency difference in the peaks of modified Raman lines of molecules with high dipole moments ranges from 4.0x$10^{14}s^{-1}$ to 9.0x$10^{14}s^{-1}$ and less than 1.0x$10^{14}s^{-1}$ for less polar molecules. Graphs depicting these variations of modified Raman lines and dipole moments are shown in figure 5.3 and 5.4.

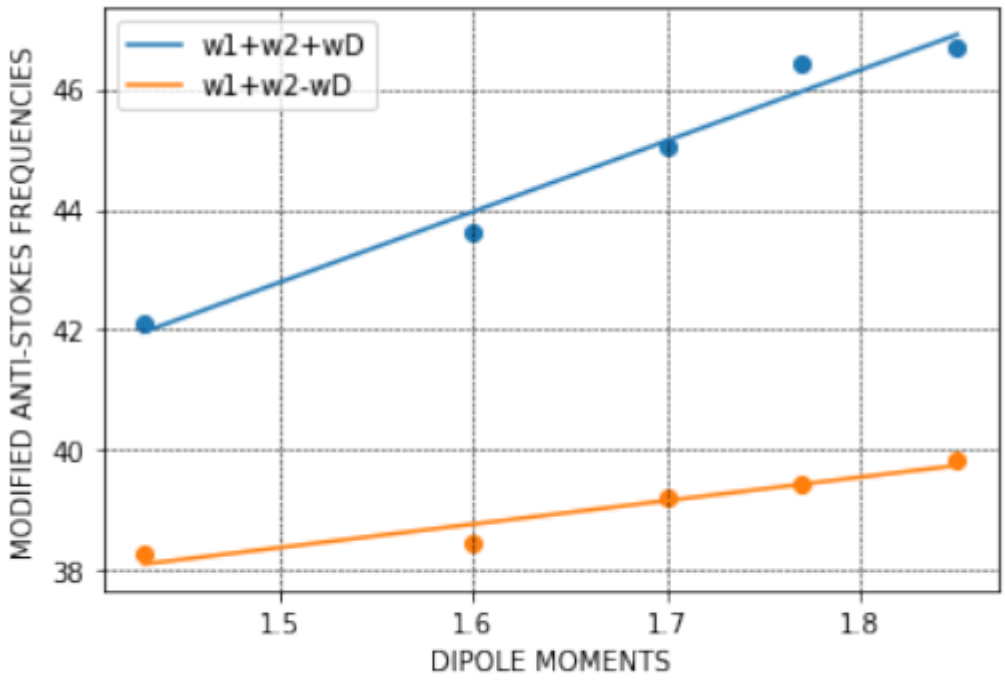


Figure 5. 3:Modified Stokes frequencies and dipole moments for less polar molecules

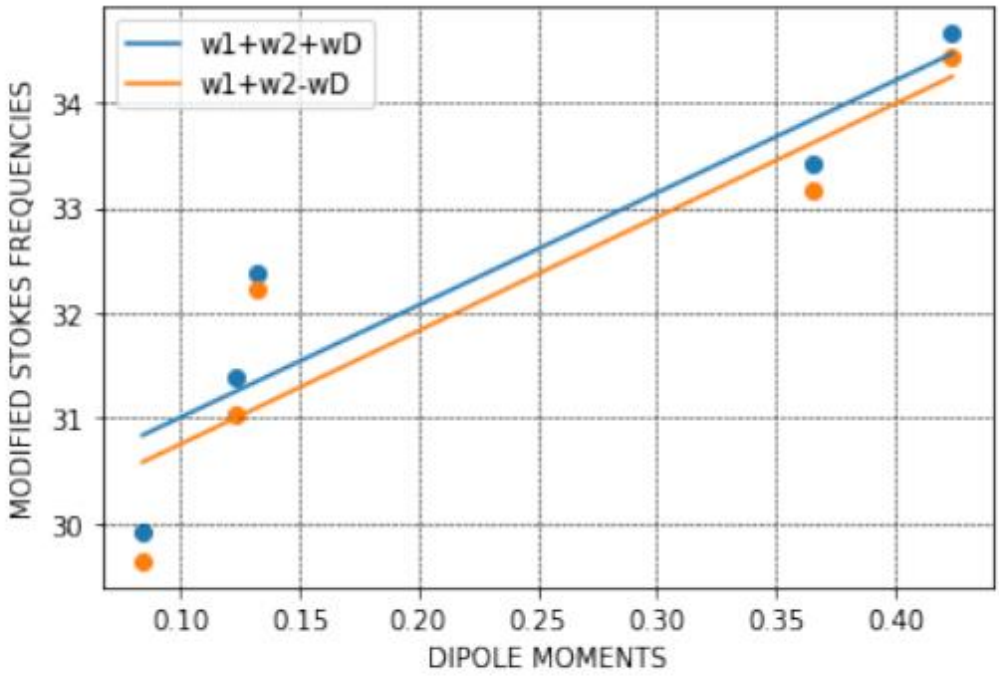


Figure 5. 4:Modified anti-Stokes frequencies and dipole moments for less polar molecules

From this finding, it implies that induced electric dipole from polar molecules creates a strong oscillating electric field which shows a significant effect on Raman lines. Therefore, dipole moments can be used to distinguish between molecular properties of materials by use of displayed Raman lines**.**

5.3 Cancer prediction using machine learning

In order to study how Raman scattering can be used in predicting cancerous cells, generated modified Stokes and anti-Stokes data for some selected polar and less molecules were analyzed. Vedruccio. *et al* (2011) reported that tumor cells respond to their own frequencies which are higher from normal and have a higher polarization. From this report, dipole moments and frequencies can be applied as a means of distinguishing between normal and tumor cells. In this thesis, dipole moments and modified Stokes and anti- Stokes Raman frequencies data of polar and less polar molecules is utilized to explore the prediction of cancer cells. Polar molecules are taken to characterize cancerous cells while less polar molecules are similar to normal cells. This approach paves way for exploring characteristic Raman signatures of normal and their corresponding cancer cells to enable their prediction.

The investigation into the potential application of Raman scattering for predicting cancerous cells involves a multifaceted exploration, where generated modified Stokes and anti-Stokes data for specific polar and less polar molecules serve as the focal point. Vedruccio et al. (2011) reported that tumor cells exhibit distinct frequencies higher than normal cells, accompanied by a higher polarization. This insight implies that dipole moments and frequencies can be leveraged as discriminative factors between normal and tumor cells.

Building on this foundational knowledge, this thesis endeavors to delve into the intricate world of dipole moments and modified Stokes and anti-Stokes Raman frequencies data for both polar and less polar molecules. The rationale behind this exploration lies in the characterization of polar molecules as representative of cancerous cells, while less polar molecules are considered akin to normal cells. By adopting this strategic approach, the study seeks to unravel the unique Raman signatures inherent to both normal and cancer cells, laying the groundwork for their predictive identification.

The journey into cancer prediction unfolds through the lens of Raman scattering, a technique renowned for its sensitivity to molecular vibrations. This study positions itself at the nexus of molecular physics and medical diagnostics, guided by the premise that the distinct Raman spectra of polar and less polar molecules can be harnessed as fingerprints for cancerous and normal cells, respectively.

The utilization of machine learning as a tool for cancer prediction adds a contemporary dimension to this investigation. As reported by Vedruccio et al. (2011), the distinct frequencies and polarization characteristics of tumor cells can be effectively discerned through advanced computational models. In this thesis, the machine learning approach is employed to meticulously analyze the complex dataset comprising dipole moments and modified Stokes and anti-Stokes Raman frequencies. The objective is not only to validate the theoretical underpinnings but also to unearth hidden patterns and correlations that may contribute to the accurate prediction of cancer cells.

As this intricate terrain of molecular signatures and machine learning algorithms is navigated, the study pays homage to the foundational work of Vedruccio et al. (2011).

Their insights serve as a guiding beacon, illuminating the path toward a deeper understanding of cancer prediction through Raman scattering. The fusion of theoretical frameworks, empirical data, and cutting-edge computational techniques marks a significant step toward unlocking the potential of Raman spectroscopy in the realm of medical diagnostics and personalized cancer care.

5.3.1 Data visualization

In order to visualize the patterns of the data set, python with the seaborn library was used to plot a correlation heat map shown in figure 5.5.

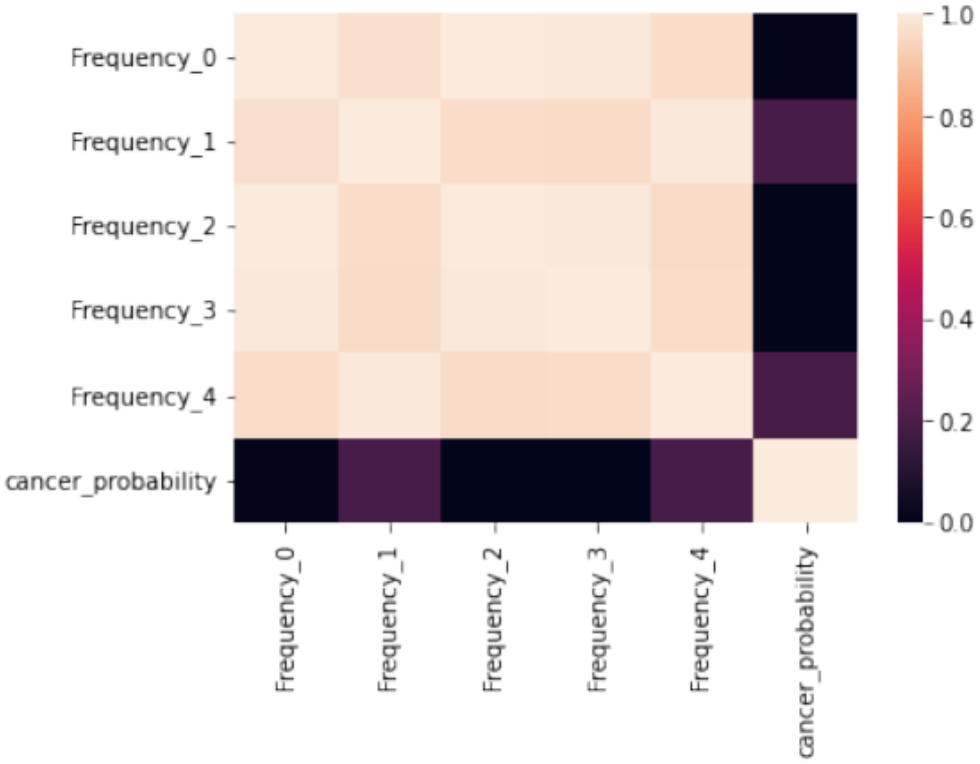


Figure 5. 5:Correlation Matrix of modified Raman frequencies.

Correlation heat maps are a type of plot that visualizes the strength of relationships between numerical variables. Correlation plots are used to understand which variables are related to each other and the strength of this relationship.

Individual points within a data set are represented by different colors, and the degree of the color shading indicates the strength of the relationship. The value of the

correlation coefficient as shown in the figure 5.5 can take any values from 0 to 1.If the value is 1, then there is a positive correlation between two frequencies and if the value is 0, there is no correlation between the two frequencies.

5.3.2 Machine learning implementation and results

In this work, Machine learning model have been developed with Neural Network algorithm for predicting cancerous cells. Since we have only 80 rows in our dataset, data augmentation was done to prevent the likelihood of over fitting during training of the model. Overfitting occurs when the model performs well on the training data but does not perform accurately on the testing data.

The test set contains 30% of the dataset while the training set contains 70 % of the data set. In this study, we characterize and discriminate between normal and cancer cells using dipole moments-modified frequency-based extracted set of parameters. Cancer cells were taken to exhibit high dipole moments when compared to normal cells from the same tissue. Machine was trained to categorized data sets of modified Stokes and anti- Stokes of polar and less polar molecules. Dataset which shows significant differences in the two peaks of Stokes and anti- Stokes line were taken to corresponds to cancer cells and those which shows relatively low differences corresponds to normal cells

A bar plots to illustrate the presence and absence of cancer cells are shown in figure 5.6.and 5.7. The stacked bar charts are based on the value of frequency differences in the peaks of Stokes and anti-Stokes lines.

.

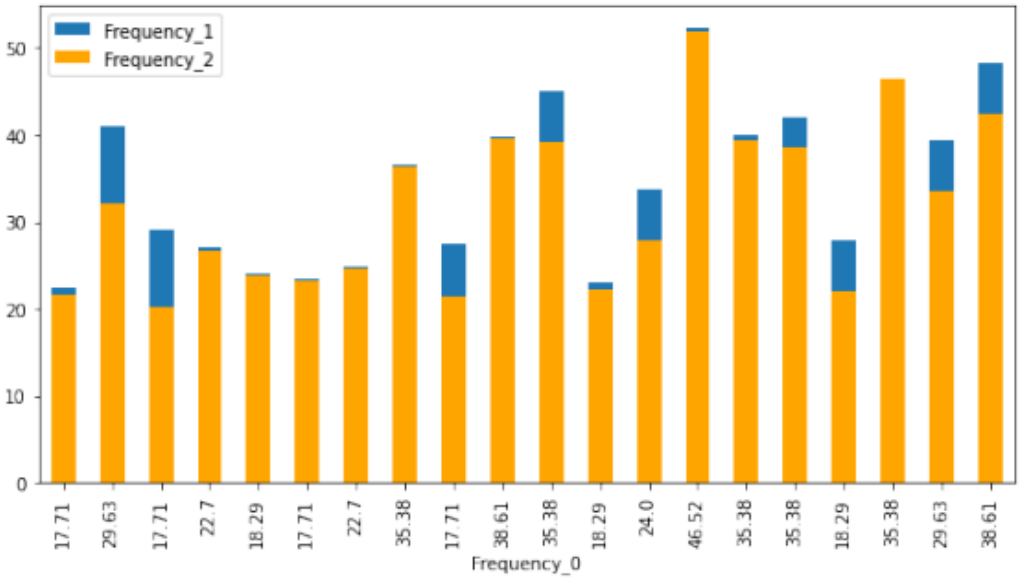


Figure 5. 6:Stacked Bar plot with frequency differences in the peaks of Stokes and anti-Stokes lines of less polar molecules.

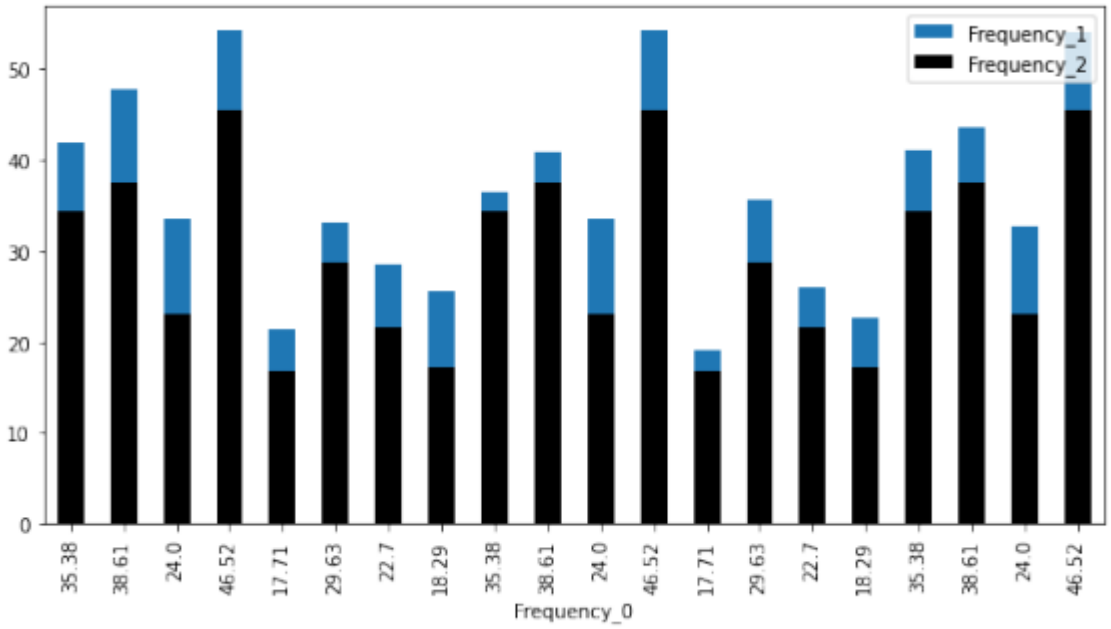


Figure 5. 7:Stacked Bar plot with frequency differences in the peaks of Stokes and anti-Stokes lines of polar molecules

The bar charts in Figures 5.6 and 5.7 show the categorical data as rectangular bars with heights proportional to the value of the frequency difference between the peaks

of the Stokes and anti-Stokes lines they represent. It is noted in the figure that, large blue areas indicate the presence of cancer cells and small areas indicate normal cells.

These observations were confirmed using a neural network and the results were labeled (0 or 1) in the columns. This is the probability of the presence or absence of cancer cells. We trained the mode using the training dataset, found the weights and biases, and checked its accuracy using the test dataset. The record labels are converted to an array. To display these results, the data were run through the algorithm as shown in Appendix B.

5.4 Discussion

One of the important reasons why the development of Raman spectroscopy was slow being the fact that the Raman effect is very weak (Hibben,.1939). Roughly one part in a thousand of the total intensity of the incident light is elastically scattered (Rayleigh scattering), whereas for Raman scattering, this value is hardly one part in a million of the incident light intensity. This is also one of the important reasons that the discovery of Raman Effect was delayed, although theoretically it was predicted by Adolf Smekal in 19239(Smekal,.1923). However, the development of lasers and the laser techniques stimulated the field of molecular spectroscopy and Raman spectroscopy. The Raman Effect occurs when a photon hits a molecule and interacts with a dipole created by the electric field of incident optical radiation. Infact, an oscillating dipole is created. Now an oscillating dipole creates an oscillating electric field, and in this thesis, it is assumed that the oscillating electric field, created by the oscillating dipole can affect the frequency scattered due to the Raman Effect. The frequencies associated with this oscillating electric field have been calculated as ($\boldsymbol{\omega}_D$), and also the frequencies of the Stokes and anti-Stokes lines for different types of molecules have been calculated.

The origin of the Stokes and anti-Stokes scattering can be explained in terms of energy transfer between the incident light and the scattering system and according to the law of conservation of energy, $\boldsymbol{\omega}_D$ will be less than $\boldsymbol{\omega}_2$, and the calculations done shows that $\omega_D < \omega_2$,thus validating the theory developed.

This shows that Equations (5.1) and (5.2) are the fundamental equations that can be used to obtain the Raman lines (Stokes and anti-Stokes lines) of the molecule under

study. This shows that additional Raman frequency is needed for the accurate distinction of Stokes and anti- Stokes line. To observe this effect, it may be necessary to use higher electric field intensities of the incident radiation, typical electric field intensities of order of 10^{10}V/m or more. It is quite possible that our study can be used in the study of Four- Wave mixing microscopy of nanostructures (Wang, .et al 2011) and Surface-Enhanced Raman Spectroscopy, and Coherent Raman Spectroscopy (Cheng,.*et al* 2013 &Epstein,.*et al* 2019).

It is important to know that Raman spectroscopy (especially vibrational Raman spectroscopy) has been very successfully used for the detection of different types of pathologies including cancer ([Krafft.,2004 & Kandel,.et al 2009). The potential of Raman spectroscopy, especially vibrational Raman spectroscopy, for disease detection and cancer is well established by now. It has the potential to identify cancerous and precancerous tissue. It can also probe deeper into diseased tissue to display and identify the underlying mechanics leading to the disease (cancerous cells). Variations of the technique can lead to study of wider range of samples. In this thesis we have tried to propose that the vibrational Raman spectroscopy could be a potential method in which the oscillating electric field created by the oscillating dipole can affect the vibrational spectra of the molecules under observation to detect the cancerous cells. The changes in dipole moments from normal to cancer state can be explained not only by the modification of its physiological and biochemical properties, but also by the rapid growth and multiplication of the cells. This approach paves the way for exploring unique electrical signatures of normal and their corresponding cancer cells to enable their detection and discrimination.

Neural network algorithm has been implemented using modified Stokes and anti-Stokes data and for better results, set of data was shuffled within themselves. Machine was successfully trained on the data sets after which successful testing was performed with an unknown data set and 96.75 % accuracy was achieved.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusions

In this work, Maxwell equations are used in order to obtain an equation of the oscillating electric field which is created by the oscillating electric dipole. The theory developed gives a formula that allows calculations of additional Raman frequency due to oscillating electric field induced by oscillating electric dipole. The calculated frequency determines the modified frequencies of the Stokes and anti-Stokes lines and thus provides a detailed description of Raman scattering. This adds a new dimension to the experimental observations of the Raman Spectroscopy and thus the frequencies of the Raman lines will be altered when the effect of ω_D is taken into account. This study predicts new phenomena of getting Stokes and anti-Stokes lines with modified wavelengths. We have not been able to find out any theoretical and or experimental observations on these lines in literature. It is proposed that such a phenomenon may be observed experimentally if very high power laser radiation is used. Actual experimental studies in future may be able to decide as to how far it will be a reliable technique.

With the ever increasing number of cancer cases, there is a need for better technology that shows its presences at early stage thus increasing the chances of human survival. In this work, modified Raman data were explored and utilized in machine learning model for prediction of cancerous cells. We have demonstrated by use machine learning that early prediction of cancer cells using Raman scattering is possible with 96.75% accuracy. The value of ω_D calculated was found to be different for different molecules and this concept can be used to study the cancerous cells since the modified Raman frequencies from cancerous cells will be different from the modified frequencies from the normal cells.

In this study, the foundational principles of Maxwell's equations have been harnessed to derive an equation characterizing the oscillating electric field induced by an oscillating electric dipole. The resultant theoretical framework provides a formula enabling the calculation of additional Raman frequencies attributed to the oscillating

electric field. This, in turn, offers a comprehensive understanding of Raman scattering by delineating the modified frequencies of the Stokes and anti-Stokes lines.

An intriguing revelation of this investigation is the prediction of novel phenomena associated with obtaining Stokes and anti-Stokes lines featuring modified wavelengths. While the theoretical groundwork has been laid, experimental validation remains an open avenue. Future studies, employing exceptionally high-power laser radiation, may unveil and confirm these predicted phenomena, thereby contributing to the evolving landscape of Raman Spectroscopy.

In the realm of medical diagnostics, particularly in the context of the escalating prevalence of cancer, the imperative for advanced technologies facilitating early detection is more pronounced than ever. This study ventures into the exploration and utilization of modified Raman data within a machine learning model for the predictive analysis of cancerous cells. The results demonstrate the feasibility of early cancer cell prediction using Raman scattering, achieving an impressive accuracy rate of 96.75%.

An intriguing facet emerging from the study is the variable nature of the calculated ω_D for different molecules. This inherent diversity in ω D values holds promise for studying cancerous cells, as the modified Raman frequencies originating from cancerous cells are anticipated to differ from those of normal cells. This nuanced understanding, derived from both theoretical and machine learning perspectives, opens avenues for further research in cancer diagnostics and underscores the potential of Raman spectroscopy in contributing to early and accurate disease detection.

6.2 Recommendations

This work presents new results of theoretical studies related to the interaction of light and matter, while, like other new developments in Raman spectroscopy, will open the imagination of researchers to further advances. To this end, the following ideas and suggestions are offered:

(i) It is proposed that this new effect can be experimentally observed by use of high power laser beams, which will simultaneously lead to increase of number of the scattered Raman photons. This study can motivate scientists

to develop intense field lasers so that the value of $\boldsymbol{\omega}_D$ can be increased such that the phenomena can be easily observed experimentally.

(ii) Experiments have shown (Fraldi., *et al* 2015 & Savavana,. *et.al* 2016) that a single cancer cell is about 70% softer than the normal healthy cell. Thus, the force constant of the cancer cell will differ from the force constant of the healthy cell, and this will result in a different vibrational frequency of the cancer cell when compared to the normal healthy cell. Experiments can be designed to measure their frequencies and by knowing the exact values of the natural frequencies of vibration of the cancer cell, it should be possible to devise a method of killing the cancer cells, without disturbing the healthy cell

(iii) Raman lines, both Stokes and anti-Stokes lines will be displaced with respect to the central line that corresponds to the frequency ω_1 depending on how ω_D varies with the magnitude of disease of the diseased cell. Thus, this kind of Raman spectroscopy can be used in biomedical studies since it could be capable of identifying and characterizing the tissues, cells and structure of the molecules involved.

(iv) Experimental Verification of New Effects: It is suggested that the newly identified effect could be experimentally observed using high-power laser beams. Elevating the intensity of laser radiation can potentially increase the number of scattered Raman photons, offering a tangible avenue for experimental validation. This recommendation encourages the development of intense field lasers, setting the stage for easier observation of the predicted phenomena.

(v) Biomechanical Studies for Cancer Diagnosis: Building upon existing evidence indicating that cancer cells are softer than normal healthy cells, as demonstrated by experiments (Fraldi et al., 2015; Savavana et al., 2016), further investigations can be designed. Specifically, exploring the vibrational frequencies of cancer cells in comparison to healthy cells, based on their force constants, opens avenues for devising innovative

methods for selectively targeting and treating cancer cells while preserving healthy ones.

(vi) Biomedical Applications of Raman Spectroscopy: The displacement of Raman lines, both Stokes and anti-Stokes, in relation to the central line corresponding to the frequency ω_1 offers potential applications in biomedical studies. This type of Raman spectroscopy holds promise for identifying and characterizing tissues, cells, and molecular structures. Incorporating this approach into biomedical research could pave the way for enhanced capabilities in diagnosing and understanding various diseases.

(vii) Exploration of Advanced Laser Technologies: Continued research into laser technologies is recommended, with a focus on developing advanced laser systems capable of precise frequency control. Such advancements would contribute not only to the proposed experimental observations but also to the broader field of spectroscopy and photonics.

(viii) Multidisciplinary Collaborations: Encouraging collaborations between physicists, biologists, and medical researchers is essential. Multidisciplinary teams can leverage diverse expertise to integrate theoretical predictions with practical applications, fostering a holistic approach to advancing both fundamental science and medical diagnostics.

(ix) Quantitative Biophysical Studies: Conducting in-depth quantitative studies on the biophysical properties of diseased and healthy cells can provide a nuanced understanding of the underlying mechanisms. This involves exploring variations in force constants, vibrational frequencies, and other relevant parameters, contributing valuable data for both theoretical models and experimental validations.

These recommendations collectively aim to catalyze progress in Raman spectroscopy, extending its applications into biomedicine and materials science, and promoting collaborative efforts that leverage the intersection of physics, biology, and medical research.

REFERENCES

Abbas, O. P. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. *In Chemical Analysis of Food* (pp. 77-134)). Academic Press.

Abdelrahman, A.H, Ahmed, M.A (2014), *Interpretation of Raman Effect on the Basis of Quantum Theory*. Academic Research International Vol. 5(2).

Allakhverdiev, E. S.et al. (2022). Raman spectroscopy and its modifications applied to biological and medical research.,. Cells, 11(3), 386.

Alodan, S. A. (2022). Bacterial Detection for Biomedical Applications Based on Surface-Enhanced Raman Scattering). *In Doctoral dissertation, Alfaisal University* . (Saudi Arabia).

Arivuoli, D. (2001). Fundamentals of nonlinear optical materials. . *Pramana*, 57, 871-883.

Barone, V et al. (2012). Toward anharmonic computations of vibrational spectra for large molecular systems. . *International Journal of Quantum Chemistry*, 112(9),, 2185-2200.

Blake, N et al. (2022). Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature. . Diagnostics, 12(6), 1491.

Bloembergen, N (1967), *'The Stimulated Raman effect."* American Journal of Physics 35,989.

Bohren, Craig F.; Donald R. Huffman (1983). Absorption and Scattering of Light by Small Particles. *Wiley*. ISBN 978-0-471-29340-8.

Bonse, J. &. (2020). Laser & Photonics Reviews, 14(10). *Maxwell meets Marangoni—A review of theories on laser-induced periodic surface structures.* , 2000215.

Borman, S. A. (1982). Nonlinear Raman spectroscopy . Analytical Chemistry, 54(9, 1021A-1026A.

Born.M. and Wolf. Principles of Optics. Pergamon, (1970)

Boyd, R.W. (2003) Nonlinear Optics .Academic Press, San Diego, CA.

Brand, J. C. D. (1989). The discovery of the Raman Effect. *Notes and Records of the Royal Society of London*, 43(1), 1-23.

Brunner, H. M. (1972). Resonance Raman scattering on the haem group of oxy-and deoxyhaemoglobin . *In Journal of molecular biology, 70(1)* (pp. 153-156.).

Campanella, B et al. (2021). Introduction to vibrational spectroscopies. *ChemTexts*, 7, 1-21.

Cheng, ji-xin and Sunney.x.Coherent Raman scattering microscopy crc press, Taylor and Francis group (2013).

Christopher. T. Kucha, Michael O. Ngadi, Non-destructive spectroscopy techniques and multivariate Analysis for Assessment of food, Sensors – Basel, Jan 2018.

Cialla-May, D. S. (2019). Theoretical principles of Raman spectroscopy. *Physical Sciences Reviews, 4(6).*

Clegg, S.M. et al (2006). LIBS – *Raman Spectroscopy of Minerals Using Remote Surface Modification Techniques. Lunar and Planetary Science XXXVII.*

Dahm, K. D. (2021). Theoretical models of light scattering and absorption. . In K. D. Dahm, *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications* (pp. 37-60).

Dating Yang, (2011) *Applications of Raman Spectroscopy in Agricultural products and Food Analysis,* Oct. 2011.

David R. Lide, ed., CRC *Handbook of Chemistry and Physics, Internet* Version 2005, <http://www.hbcpnetbase.com>, CRC Press, Boca Raton, FL, (2005).

DePaoli, D. et al. (2020). Rise of Raman spectroscopy in neurosurgery: *a review. Journal of Biomedical Optics*, 25(5), 050901-050901.

Druet, S.A.J and Taran, P.E, (1981) CARS spectroscopy, *Prog. Quantum Electron. 7, 1.*

Duraipandian, S., Petersen, J. C., & Lassen, M. et al. (2019). Authenticity and concentration analysis of extra virgin olive oil using spontaneous Raman spectroscopy and multivariate data analysis. . *Applied Sciences,* 9(12), 2433.

Dyson, F. J. (2006). Advanced Quantum Mechanics. (2nd Edition). Singapore: World Scientific Publishing Company.

Edwards, H. G., Vandenabeele, P., & Colomban, P. et al. (2022). Theory of Raman Spectroscopy and Raman Instrumentation. In H. G. Edwards, *Raman Spectroscopy in Cultural Heritage Preservation* Cham (pp. (pp. 19-41).). Springer International Publishing.

Epstein.S.C, Huff.A.R, Emilly.S, winesett.louise.kc, nature communications, 10.222. Tracking carrier protein moting with Raman Spectroscopy) 13th May (2019).

Eskandari, V et al. (2022). A review of applications of surface-enhanced raman spectroscopy laser for detection of biomaterials and a quick glance into its advances for COVID-19 investigations. *ISSS Journal of Micro and Smart Systems*, 11(2), 363-382.

Fast, A. &. (2019). Coherent Raman scattering with plasmonic antennas. . *Nanophotonics,* 8(6),, 991-1021.

Ferraro, J. R. et al., (2003). *Introductory Raman Spectroscopy*. San Diego: Academic Press.

Ford, G. W. (1981). Electromagnetic effects on a molecule at a metal surface: I. , . *Effects of nonlocality and finite molecular size. Surface Science*, 109(2), 451-481.

Fowler, S. M.,et al. (2019). Investigation of chemical composition of meat using spatially off-set Raman spectroscopy. Analyst, 144(8),, 2618-2627.

Fraldi.M, Cugno.A, Deseri.L, Dayal.K and Pugno.N.M.A (2016) Frequency-based Hypothesis for mechanically targeting and selectively attacking Cancer cells.J.R.soc.Interface 12(2015)0656.

Frank, Hill, A.E, Peters, C.W and Weinreich.G (1961) *Phys. Rev. Lett*., 7:118, Company, New York.

Gaponenko, S. V et al. (2021). Strong selective anti-Stokes Raman scattering enhancement in plasmonics using photon density of states engineering. *The Journal of Physical Chemistry* C, 125(50), 27654-27660.

Garmire, E., Pandarese, F., and Townes, C.T., (1963) coherently driven molecular vibrations and light modulation, *Phys. Rev. Lett.* 11, 160.

Glazov, M. M. (2014). High frequency electric field induced nonlinear effects in graphene. . *Physics Reports,* 535(3), 101-138.

Gonis, Antonios; William H. Butler (1999). Multiple Scattering in Solids. *Springer. ISBN 978-0-387-98853-5*.

Greiner.W Quantum Mechanics: An introduction. Springer, New York,

Griffiths, David J, (2015) Introduction to Electrodynamics, 3rd Edition; Pearson Education.

GU, B., Zhao, C., Baev, A., Yong, K. T., Wen, S., & Prasad, P. N. (2016). Molecular nonlinear optics: recent advances and applications. *Advances in Optics and Photonics,* 8(2), 328-369.

Gupta, N. et al. (2022). Scattering of Laser Light in Dielectrics and Plasmas: A Review. *Nonlinear Optics, Quantum Optics: Concepts in Modern Optics,*, 55.

Hashimoto, K. et al. (n.d.). Complementary vibrational spectroscopy., . *Nature Communications*, 10(1), 44112019.

He, H. et al. (2021). Deep learning for biospectroscopy and biospectral imaging: state-of-the-art and perspectives.

Hewitt. K.C. (1999) *Phys Rev*. B60 R 9943

Hibben, H.J.The Raman Effect and its Chemical Applications, Reinhold Publishing Company New York (1939).

Hollas, J.M., (2006), Modern spectroscopy. 4th ed.England: John Wiley& Sons.

Hooton, D. J. (1955). LI. A new treatment of anharmonicity in lattice thermodynamics: I. The *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 46(375), 422-432.

Horn, A. (2022). Inelastic Scattering and Absorption. In A. Horn, *In The Physics of Laser Radiation–Matter Interaction: Fundamentals, and Selected Applications in Metrology . Cham* (pp. pp. 117-133). Springer International Publishing.

Hutchin, R. A. (2015). The Third Polarization of Light. *Optics and Photonics Journal*, 5(02), 29.

Ines Raquel, Martins Ramos, Fiona Mary Lyng. *Biomedical Research International*. Volume (2015). Article ID 561242.

Jackson, J.D, Classical Electrodynamics Wiley. New York (1975)

Jain, A. K. (2019). In vitro and in vivo characterization of pharmaceutical nanocarriers used for drug delivery. . *Artificial cells, nanomedicine, and biotechnology*, 47(1), 524-539.

Jain, S. (. (2021). Surface Enhanced Raman Spectroscopy: . In S. (. *Jain, Biosensing and Diagnostic Technique for Healthcare Applications*. . Bentham Science Publishers.

James .D and Staney, K.C, (1999) "*Spectrochemical Analysis* ".

Kandel C, Isabelle M, Bazant-Hegemark. F, (2009) Vibrational Spectroscopy: a clinical tool for Cancer diagnostics. Analyst. 134(6); 1029-1045(Pub.Med)

Kendziora. C and Rosenberg A. (1995) *Phys. Rev*. 852 R9867.

Khodabakhshian, R. &.-F. (2020). Pattern recognition-based Raman spectroscopy for non-destructive detection of pomegranates during maturity. . *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 231, , 118127.

Kiefer, W and Popp, J (2006), Raman scattering, Fundamentals. John Wiley & sons.

Kobayashi, Y. &. (2022). Recent advances in low-power-threshold nonlinear photochromic materials. *Chemical Society Reviews.*

Krafft, C. D. (2009). Raman and CARS microspectroscopy of cells and tissues. *Analyst, 134(6)*, 1046-1057.

Krafft.C, (2004) Bio analytical Applications of Raman Spectroscopy doi: 10.1007/500216-003-2266-6(Pub-Med)

Kumar, P. S.,et al. (2019). . Characterization techniques for nanomaterials. *Elsevier. In Nanomaterials for solar cell applications*, pp. 97-124.

Kuramochi, H. &. (2021). Tracking ultrafast structural dynamics by time-domain Raman spectroscopy. ,. *Journal of the American Chemical Society*, 143(26), 9699-9717.

Landau.L.D, and Lifshitz.E.M, (1971) Electrodynamics of Continuous Media .Pergamon.

Landsberg, G. and Mandelstam, (1928) *Naturwiss*.16 557-558.4

Lauriola, D. (2022). *Advancement of coherent anti-stokes raman scattering spectroscopy for applied measurements in combustion and high-speed flow environments* (Doctoral dissertation, Purdue University Graduate School.

Leuchs, G., Villar, A. S., & Sánchez-Soto, L. L. (2010). The quantum vacuum at the foundations of classical electrodynamics. *Applied Physics* B, 100(1), 9-13.

Li, J. J., et al. (2018). Efficient infrared light promoted degradation of volatile organic compounds over photo-thermal responsive Pt-rGO-TiO2 composites. 2. *Applied Catalysis B: Environmental*, 233, 260-271.

Li, T. E et al. (2019). comparison of Different Classical, Semiclassical, and Quantum Treatments of Light–Matter Interactions: Understanding Energy Conservation. *Journal of Chemical Theory and Computation*, 15(3), 1957-1973.

Long.D.A, (2002) .The Raman Effect. John Wiley & Sons Ltd., West Sussex, England

Loudon, .R (2000).*The Quantum Theory of Light*, Oxford University Press, New York.

McCreery, R.I. (2000).Raman Spectra for Chemical Analysis. John Wiley & Sons.Inc.New York.

Misra, R. A. (2010). Cancer nanotechnology: application of nanotechnology in cancer therapy. *Drug discovery today*, 15(19-20), 842-850.

Mitsutake, H., *et al.* (2019). Raman imaging spectroscopy: history, fundamentals and current scenario of the technique. *Journal of the Brazilian Chemical Society*, 30,, 2243-2258.

Mukamel, S. (1995). Principles of Nonlinear Optical Spectroscopy. Oxford University Press, New York

Naeini. J. G, Irwin J.C. *The pseudo gap in La2-xSrxCuO4: A Raman view point*. Canadian Journal of Physics, Boris P Cond mat/9909342 14th Feb 2000.

Neuville, D. R. (2014). Advances in Raman spectroscopy applied to earth and material sciences. *Reviews in Mineralogy and Geochemistry*, 78(1), 509-541.

Nguyen, D. V et al. (2021). Observational constraints on dark matter scattering with electrons. , . *In Physical Review D*, 104(10) (p. 103521).

Nicoletti, D., & Cavalleri, A. (2016). Nonlinear light–matter interaction at terahertz frequencies. *Advances in Optics and Photonics*, 8(3), 401-464.

Ostovar, P. Saeideh, Blanch. W. Ewan. *Investigation of chemical composition of meat using spatially off-set Raman scattering. Analyst* March 2019.

Ouyang, L. *et al.* (2020). Spatially Resolving the Enhancement Effect in Surface-Enhanced Coherent Anti-Stokes Raman Scattering by Plasmonic Doppler Gratings. ACS nano, 15(1), 809-818.

Pilot, R et al. (2019). A review on surface-enhanced Raman scattering. Biosensors, 9(2).

Rajveer S, Rohit B. (2018) *Bulk protein and oil prediction in soybean using transmission Raman spectroscopy* by App. Spectroscopy, Nov 2018.

Raman, C.V and Krishnan, K.S (1928) Nature, 121 501-502, "*A new type of secondary radiation''*.

Ramos, I. R. (2015). Current advances in the application of Raman spectroscopy for molecular diagnosis of cervical cancer. . *BioMed research international.*

Reichel, J. ((2022)). Single Atoms and Single Photons: Quantum Information Exchange. In J. Reichel, *In Ultra-cold atoms, ions, molecules and quantum technologies*. (pp. (pp. 45-64)). EDP Sciences.

Rohatgi-Mukherjee, K. K. (1978). Fundamentals of photochemistry. . In K. K. Rohatgi-Mukherjee, *New Age International.*

Romanenko, S et al. (2017). he interaction between electromagnetic fields at megahertz, gigahertz and terahertz frequencies with cells, tissues and organisms: risks and potential. *Journal of The Royal Society Interface*, 14(137), 20170585.

Sathyanarayana, D. N., (2017) *Vibrational Spectroscopy*: Theory and Applications. *New Age International (P) Limited.*

Savavana.K Jaganathan et.al .Natural Frequency of Cancer cells as a starting point in Cancer treatment. Current Science Vol.110, No.9, 10 May 2016.

Schmitt, M.*et al* (2006), *Raman Spectroscopy*.37 (1-3), 20

Scully, M and Zubairy, M.S (1997*) Quantum Optics* Cambridge University Press, Cambridge, U.K.

Settle, F.A., (1997), Handbook of Instrumental Techniques for Analytical Chemistry, Prentice Inc. New Jersey.

Sharma, S., Blanter, Y. M., & Bauer, G. E. (2017). Light scattering by magnons in whispering gallery mode cavities. *Physical Review B*, 96(9), 094412

Shipp, D.et al. (2017). Raman spectroscopy: techniques and applications in the life sciences. . *Advances in Optics and Photonics*, 9(2),, 315-428.

Silva, I. (2020). Raman Spectroscopy. *In Between Making And Knowing: Tools In The History Of Materials Research* , pp. 435-441.

Skoog, D.A, Holler, F.J, Crouch, S.R, (2006) Principles of Instrumental Analysis (6th Ed.) Gengoge Learning.

Smekal, A. (1923) *Naturwiss*.16 873

Smekal.A.Zur Quanten Theorie der Dispersion Naturwiss11 (43) p873-875(1923).

Smith, B. (2018). *Infrared spectral interpretation: a systematic approach. CRC press*.

Smith, E. and Dent, G (2005) Modern Raman Spectroscopy: A Practical Approach. John Wiley &Sons England Chichester.

Sugai .S and Suzuki. M. J. (2001) *Phys. Chem. Solids,* 62.119-124.

Sur, K.U., (2010), *Recent advancement of Raman Spectroscopy*. Resonance, 15 pp. 154-164.

Sur.Ujjal.K, *Surface-Enhanced Raman spectroscopy (SERS): A versatile spectroscopic and analytical technique used in Nanoscience and nanotechnology*.Advaces in Nano Research Vol 1, No 2(2013) 111-124)

Tennyson, J. (1986). The calculation of the vibration-rotation energies of triatomic molecules using scattering coordinates. . *Computer Physics Reports*, 4(1), 1-36.

Tolles, W. M., Nibler, J. W., McDonald, J. R., & Harvey, A. B. (1977). A review of the theory and application of coherent anti-Stokes Raman spectroscopy (CARS). *Applied Spectroscopy*, 31(4), 253-271.

Torchigin, V. P., & Torchigin, A. V. (2016). Physical nature and magnitude of optically induced forces derived from laws of mechanics. *Optik*, 127(15), 5976-5983.

Tuschel, D. (2020). Spectral resolution and dispersion in raman spectroscopy . *Spectroscopy,* 35(9), 9-15.

Valberg, P. A. (1997). Can low-level 50/60 Hz electric and magnetic fields cause biological effects?. . *Radiation Research,* 148(1),, 2-21.

Wang, X *et al*. (2020). Fundamental understanding and applications of plasmon-enhanced Raman spectroscopy. . *In Nature Reviews Physics*, 2(5) (pp. 253-271).

Wang, Y., Huang, Y., Bai, H., Wang, G., Hu, X., Kumar, S., & Min, R. (2021). Biocompatible and biodegradable polymer optical fiber for biomedical application:,. (n.d.*). A review. Biosensors,* 11(120, 472.

Wang, Y.,*et al.* (2021). Biocompatible and biodegradable polymer optical fiber for biomedical application:), . *A review. Biosensors*, 11(12), 472.

Wang, Y.*et al* Four-wave mixing microscopy of nanostructures.ad.opt photon.3, 1-52(2011)

Wang, Z. L. (2017). On Maxwell's displacement current for energy and sensors: the origin of nanogenerators. . *Materials Today*, 20(2),, 74-82.

Wen-HaoSu, Da-Wen Sun, *Trends in food Authentication*. App.SpectroscopyAugust, 2018.

Willard, H.H, MerittJr.L.L. Dean.,J.J,SettleJr.F.A(1988),Instrumental Methods of Analysis (7th Ed)CBS Publisher and Distributor New Delhi.

Winch, R. P, (1958). Electricity and Magnetism, Prentice Hall, pp. 701 -708,

Xie, W. &. (2014). Rationally designed multifunctional plasmonic nanostructures for surface-enhanced Raman spectroscopy: *a review. Reports on progress in physics*, 77(11),, 116502.

Xu, Y *et al*. (2020). Raman spectroscopy coupled with chemometrics for food authentication:, . In Y. e. Xu, *A review. TrAC Trends in Analytical Chemistry* (pp. 131 , 116017).

Zhou, J et al. (2011). Differentiation between glioma and radiation necrosis using molecular magnetic resonance imaging of endogenous proteins and peptides. ,. *Nature medicine*, 17(1), 130-134.

APPENDICES

A: LIST OF VALUES OF THE ELECTRIC DIPOLE MOMENTS FOR MOLECULES.

The table below gives selected values of the electric dipole moment for over 1000 molecules. The values were determined by selected microwave spectroscopy, molecular beam electric resonance and other high-resolution spectroscopic techniques. The measurements were obtained from values of the dielectric constant in the liquid and gas phase .The entries of compounds without carbon were listed alphabetically and then followed by compounds containing carbon elements.

The units of dipole moment is given in Debye (D) and1D = 3.33564×10^{-30} C m was used as the conversion factor to SI units Dipole moments of individual compounds are given when they have been measured. Other information on molecules that have been studied by spectroscopy, such as the components of the dipole moment in the molecular framework and the variation with vibrational state and isotopic species, is given in the references. The accuracy of a value is explicitly stated (i.e., 1.23 ± 0.01), it generally indicates two or three standard deviations

COMPOUNDS NAME	**Molecular Formulae**	**μ/D**
Aluminum monofluoride	AlF	1.54 ± 0.16
Ammonia	H_3N	1.47 ± 0.02
Arsenic(III) chloride	$AsCl_3$	1.69 ± 0.07
Arsenic(III) fluoride	AsF_3	2.58 ± 0.05
Arsine	AsH_3	0.22 ± 0.03
Barium oxide	BaO	7.95 ± 0.04
Barium sulfide	BaS	10.76 ± 0.03
Bromine chloride	BrCl	0.52 ± 0.04
Bromine dioxide	BrO_2	2.84 ± 0.01
Bromine fluoride	BrF	1.42 ± 0.06
Bromine oxide	BrO	1.78 ± 0.05
Bromine pentafluoride	BrF_5	1.51 ± 0.05
Bromosilane	BrH_3Si	1.32 ±0.12
Bromotrifluorosilane	BrF_3Si	0.85 ± 0.07
Calcium monochloride	CaCl	3.61 ±0.02
Cesium chloride	ClCs	10.49 ± 0.05
Cesium fluoride	CsF	7.94 ± 0.01

Cesium sodium	CsNa	4.75 ± 0.20
Chlorine fluoride	ClF	0.88 ±0.0.06
Chlorine oxide	ClO	1.28 ± 0.01
Chlorine trifluoride	ClF_3	0.64 ± 0.01
Chloroborane	$BClH_2$	0.85 ± 0.05
Chlorogermane	$ClGeH_3$	2.33 ± 0.02
Chlorosilane	ClH_3Si	1.32 ± 0.01
Chlorosyl fluoride	ClFO	1.94 ± 0.02
Chlorotrifluorosilane	ClF_3Si	0.64 ± 0.04
Chromium monoxide	CrO	3.98 ± 0.13
Copper(I) fluoride	CuF	5.07 ± 0.29
Copper(II) oxide	CuO	4.53 ± 0.54
Dichlorosilane	Cl_2H_2Si	1.27 ± 0.02
Difluoramine	F_2HN	1.52 ± 0.02
Difluorine dioxide	F_2O_2	1.44 ± 0.07
Difluoroborane	BF_2H	0.97 ± 0.01
cis-Difluorodiazine	F_2N_2	0.26 ± 0.01
Difluorosilane	F_2H_2Si	1.45 ± 0.02
Difluorosilylene	F_2Si	1.33 ± 0.02
Disiloxane	H_6OSi_2	0.54 ± 0.02
Fluoramine	FH_2N	2.27 ± 0.18
Fluorine azide	FN_3	1.39 ±0.25
Fluorine monoxide	F_2O	0.23 ±0.02
Fluorine oxide	FO	0.04 ± 0.04
Fluoroborane	BF	0.51±0.04
Fluorogermane	$FGeH_3$	2.53 ± 0.12
Fluorosilane	FH_3Si	1.25 ± 0.06
Gallium monofluoride	FGa	2.45 ± 0.05
Germanium(II) fluoride	F_2Ge	2.62 ± 0.02
Germanium(II) oxide	GeO	3.28 ± 0.01
Germanium(II) selenide	GeSe	1.75 ± 0.05
Germanium(II) sulfide	GeS	2.09 ± 0.06
Germanium(II) telluride	GeTe	1.36 ± 0.07
Germylazide	GeH_3N_3	2.58 ± 0.02
Hafnium monoxide	HfO	3.44 ± 0.05
Hafnium(IV) oxide	HfO_2	7.91 ± 0.01
Hexaborane(10)	B_6H_{10}	2.53 ± 0.05
Hydrazine	H4N2	1.85 ± 0.09
Hydrazoic acid	HN3	1.74 ± 0.09
Hydrogen bromide	BrH	0.83 ± 0.03
Hydrogen chloride	ClH	1.16 ± 0.03
Hydrogen fluoride	FH	1.82±0.08
Hydrogen iodide	HI	0.45 ± 0.01
Hydrogen peroxide	H2O2	1.57 ± 0.01
Hydrogen sulfide	H2S	0.98±0.02
Hydroxyl	HO	1.63± 0.01
Hydroxylamine	H3NO	0.59 ± 0.05

Hypochlorous acid	ClHO	1.32 ±0.02
Hypofluorous acid	FHO	2.03 ± 0.01
Imidogen	HN	1.49 ± 0.08
Indium(I) chloride	ClIn	3.89 ± 0.29
Indium(I) fluoride	FIn	3.43 ± 0.07
Iodine bromide	BrI	0.73 ± 0.03
Iodine chloride	ClI	1.44 ± 0.02
Iodine fluoride	FI	1.95 ± 0.02
Iodine monoxide	IO	2.85 ± 0.05
Iodine pentafluoride	F5I	2.28 ± 0.21
Lanthanum monoxide	LaO	3.21 ± 0.01
Lead(II) oxide	OPb	4.44 ± 0.05
Lead(II) sulfide	PbS	3.69 ± 0.18
Lithium bromide	BrLi	7.28 ± 0.01
Lithium chloride	ClLi	7.12 ±0.07
Lithium fluoride	FLi	6.33 ± 0.02
Lithium fluoride–sodium	FLi•FNa	2.63 ± 0.02
Fluoride complex	FLi•F	2.56 ± 0.12
Lithium hydride	HLi	5.83 ± 0.01
Lithium hydroxide	HLiO	4.72 ± 0.02
Lithium iodide	ILi	7.43 ± 0.01
Lithium monoxide	LiO	6.84 ± 0.04
Lithium potassium	KLi	3.46 ± 0.20
Lithium rubidium	LiRb	4.01 ± 0.12
Lithium sodium	LiNa	0.42 ± 0.02
Magnesium oxide	MgO	6.21 ± 0.06
Mercapto	HS	0.75 ± 0.01
Nitric acid	HNO3	2.37 ± 0.02
Nitric oxide	NO	0.16 ±0.21
Nitrogen dioxide	NO2	0.32 ± 0.01
Nitrogen sulfide	NS	1.83 ± 0.02
Nitrogen trichloride	Cl3N	0.35 ± 0.01
Nitrogen trifluoride	F3N	0.24 ± 0.04
Nitrogen trioxide	N2O3	2.12 ± 0.01
Nitrosyl bromide	BrNO	1.8 2 ± 0.04
Nitrosyl fluoride	FNO	1.730 ± 0.03
Nitrosyl hydride	HNO	1.62 ± 0.03
Nitrous acid (cis)	HNO2	1.44 ± 0.05
Nitrous acid (trans)	HNO2	1.85 ± 0.01
Nitrous oxide	N2O	0.163±0.02
Nitryl chloride	ClNO2	0.52 ±0.01
Nitryl fluoride	FNO2	0.46 ± 0.05
Ozone	O3	0.54 ±0.02

Pentaborane(9)	B5H9	2.33 ± 0.04
Perchloryl fluoride	ClFO3	0.23 ± 0.01
Peroxynitrous acid	HNO3	1.47 ± 0.02
Peroxynitric acid	HNO4	1.96 ± 0.02
Phosphine	H3P	0.57 ± 0.03
Phosphorothioc trifluoride	F3PS	0.65 ± 0.02
Phosphorus(III) chloride	Cl3P	0.47 ± 0.02
Phosphorus(III) fluoride	F3P	1.56 ± 0.01
Phosphorus monoxide	OP	1.87 ± 0.07
Phosphorus nitride	NP	2.74 ± 0.01
Phosphoryl chloride	Cl3OP	2.54 ± 0.05
Phosphoryl fluoride	F3OP	1.86 ± 0.01
Potassium bromide	BrK	10.62 ± 0.01
Potassium chloride	ClK	10.29 ± 0.01
Potassium fluoride	FK	8.55 ± 0.03
Potassium hydroxide	HKO	7.43 ± 0.02
Potassium iodide	IK	10.8 ±0.02
Potassium sodium	KNa	2.69 ± 0.01
Rubidium bromide	BrRb	10.9 ±0.02
Rubidium chloride	ClRb	10.51 ± 0.05
Rubidium fluoride	FRb	8.54 ± 0.05
Rubidium iodide	IRb	11.5 ±0.06
Rubidium sodium	NaRb	3.16 ± 0.03
Selenium dioxide	O2Se	2.62 ± 0.05
Selenium tetrafluoride	F4Se	1.78 ± 0.09
Silicon monosulfide	SSi	1.73 ± 0.09
Silicon monoxide	OSi	3.09 ±0.02
Silver(I) bromide	AgBr	5.62 ± 0.03
Silver(I) chloride	AgCl	6.08 ± 0.06
Silver(I) fluoride	AgF	6.22 ± 0.30
Silver(I) iodide	AgI	4.55 ± 0.05
Sodium bromide	BrNa	9.11 ± 0.06
Sodium chloride	ClNa	9.05 ±0.02
Sodium fluoride	FNa	8.15 ± 0.01
Sodium iodide	INa	9.25 ± 0.03
Stibine	H3Sb	0.12 ± 0.05
Strontium oxide	OSr	8.90 ± 0.03
Sulfur dichloride	Cl2S	0.36 ± 0.01
Sulfur difluoride	F2S	1.05 ± 0.05
Sulfur dioxide	O2S	1.63 ±0.05
Sulfur monofluoride	FS	0.79 ± 0.02
Sulfur monoxide	OS	1.55 ± 0.02
Sulfur oxide (SSO)	OS2	1.47 ± 0.03

Sulfur tetrafluoride	F_4S	0.63 ± 0.03
Sulfuryl chloride	Cl_2O_2S	1.81 ± 0.04
Sulfuryl fluoride	F_2O_2S	1.12 ± 0.02
Tetraborane(10)	B_4H_{10}	0.48 ± 0.02
Tetrafluorohydrazine	F_4N_2	0.77 ± 0.002
Tetrafluorosilane–ammonia	$F_4Si{\bullet}H_3N$	5.61 ± 0.02
Thallium(I) bromide	BrTl	4.49 ± 0.05
Thallium(I) chloride	ClTl	4.54±0.02
Thallium(I) fluoride	FTl	4.22 ± 0.08
Thallium(I) iodide	ITl	4.61 ± 0.07
Thionitrosyl chloride (NSCl)	ClNS	1.87 ± 0.02
Thionitrosyl fluoride (NSF)	FNS	1.902 ± 0.01
Thionyl chloride	Cl_2OS	1.45 ± 0.03
Thionyl fluoride	F_2OS	1.63 ± 0.01
Tin(II) oxide	OSn	4.32 ± 0.22
Tin(II) sulfide	SSn	3.18 ± 0.16
Titanium(II) oxide	OTi	2.96 ± 0.05
Trichlorofluorosilane	Cl_3FSi	0.49 ± 0.01
Trichlorosilane	Cl_3HSi	0.86 ± 0.01
Trifluoramine oxide	F_3NO	0.039 ± 0.04
1,1,1-Trifluorodisilane	$F_3H_3Si_2$	2.03 ± 0.10
Trifluoroiodosilane	F_3ISi	1.11 ± 0.03
Trifluorosilane	F_3HSi	1.27 ± 0.03
Water	H_2O	1.85± 0.04
Water dimer–hydrogen bromide complex	$H_4O_2{\bullet}BrH$	2.28 ± 0.03
Water dimer–hydrogen chloride complex	$H_4O_2{\bullet}ClH$	2.3± 0.003
Ytterbium monofluoride	FYb	3.91 ± 0.04
Yttrium monoxide	OY	4.52 ± 0.07
Zirconium(II) oxide	OZr	2.55 ± 0.01
Zirconium(IV) oxide	O_2Zr	7.80 ± 0.02
Compounds with carbon element		
Acenaphthene	$C_{12}H_{10}$	0.85 ±0.02
Acetaldehyde	C_2H_4O	2.750 ± 0.06
Acetamide	C_2H_5NO	3.68 ± 0.03
Acetic acid	$C_2H_4O_2$	1.70 ± 0.03
Acetic anhydride	$C_4H_6O_3$	2.83 ±0.02
Acetone	C_3H_6O	2.88 ± 0.03
Acetonitrile	C_2H_3N	3.92 ±0.12
Acetophenone	C_8H_8O	3.02 ± 0.06
Acetyl chloride	C_2H_3ClO	2.72 ± 0.14
Acetylene–carbon dioxide Complex	$C_2H_2{\bullet}CO_2$	0.16 ± 0.01
Acetylene–carbon monoxide complex	$C_2H_2{\bullet}CO$	0.31 ± 0.01
Acetylene–carbon oxysulfide trimer complex	$C_2H_2{\bullet}C_3O_3S_3$	1.23 ± 0.02
Acetylene–hydrogen cyanide complex	$C_2H_2{\bullet}CHN$	3.29 ± 0.03

Acetyl fluoride	C_2H_3FO	2.96 ± 0.03
Acrolein (*cis*)	C_3H_4O	2.55 ± 0.03
Acrolein (*trans*)	C_3H_4O	3.11 ± 0.04
Acrylonitrile	C_3H_3N	3.92 ± 0.07
Allyl alcohol (*gauche*)	C_3H_6O	1.55 ± 0.08
Allyl alcohol (*average*)	C_3H_6O	1.60 ± 0.08
Allylamine	C_3H_7N	1.23±0.02
Aniline	C_6H_7N	1.13 ± 0.02
Anisole	C_7H_8O	1.38 ± 0.07
Azulene	$C_{10}H_8$	0.80 ± 0.02
Benzaldehyde	C_7H_6O	3.04 ±0.01
Benzeneacetonitrile	C_8H_7N	3.54 ±0.02
Benzene–hydrogen sulfide Complex	$C_6H_6{\bullet}H_2S$	1.14 ± 0.02
Benzene–krypton complex	$C_6H_6{\bullet}Kr$	0.16 ± 0.02
Benzene–sulfur dioxide Complex	$C_6H_6{\bullet}O_2S$	2.61 ± 0.02
Benzenethiol	C_6H_6S	1.23 ±0.02
Benzonitrile	C_7H_5N	4.18 ± 0.08
Benzyl acetate	$C_9H_{10}O_2$	1.22 ±0.02
Benzyl alcohol	C_7H_8O	1.71 ± 0.09
Benzyl benzoate	C14H12O2	2.06 ±0.02
Bis(2-aminoethyl)amine	C4H13N3	1.89±0.02
Bis(2-chloroethyl) ether	C4H8Cl2O	2.58±.0.03
Bis(2-ethylhexyl) phthalate	C24H38O4	2.84±.0.02
Borane carbonyl	CH3BO	1.68 ± 0.02
Bromoacetylene	C2HBr	0.22±.002
Bromobenzene	C6H5Br	1.70 ± 0.03
1-Bromobutane	C4H9Br	2.08 ± 0.10
2-Bromobutane	C4H9Br	2.23 ± 0.11
1-Bromo-2-chloroethane	C2H4BrCl	1.21 ±0.02
Bromochlorofluoromethane	CHBrClF	1.54 ± 0.02
Bromochloromethane	CH BrCl	1.63±.0.02
1-Bromodecane	C10H21Br	1.93 ±0.02
Bromoethane	C2H5Br	2.04 ± 0.02
Bromoethene	C2H3Br	1.42 ± 0.03
Bromofluoroacetylene	C2BrF	0.44 ± 0.02
1-Bromoheptane	C7H15Br	2.16 ± 0.11
Bromomethane	CH3Br	1.82 ± 0.04
2-Bromo-2-methylpropane	C4H9Br	2.17±.0.02
1-Bromonaphthalene	C10H7Br	1.55±.0.02
1-Bromopentane	C5H11Br	2.20 ± 0.14
1-Bromopropane	C3H7Br	2.18 ± 0.12
2-Bromopropane	C3H7Br	2.21 ± 0.12
2-Bromopropene	C3H5Br	1.53 ±.0.02
3-Bromopropene	C3H5Br	1.9 ±.0.024
Bromotrifluoromethane	CBrF3	0.65 ± 0.05
1,2-Butadiene	C4H6	0.43 ± 0.02

Butanal	C4H8O	2.72 ± 0.05
1,4-Butanediol	C4H10O2	2.58±.0.02
Butanenitrile (gauche)	C4H7N	3.91 ± 0.04
Butanenitrile (anti)	C4H7N	3.73 ± 0.06
1-Butanethiol	C4H10S	1.53±.0.02
Butanoic acid	C4H8O2	1.65±.0.02
1-Butanol	C4H10O	1.66 ± 0.03
2-Butanone	C4H8O	2.79 ± 0.01
trans-2-Butenal	C4H6O	3.67 ± 0.07
1-Butene (cis)	C4H8	0.43 ± 0.07
1-Butene (skew)	C4H8	0.39 ± 0.01
cis-2-Butene	C4H8	0.25 ± 0.05
cis-2-Butene-1,4-diol	C4H8O2	2.48±.0.02
trans-2-Butene-1,4-diol	C4H8O2	2.45±.0.02
trans-2-Butenoic acid	C4H6O2	2.13±.0.02
cis-2-Buten-1-ol	C4H8O	1.96 ± 0.03
trans-2-Buten-1-ol	C4H8O	1.90 ± 0.02
1-Buten-3-yne	C4H4	0.22 ± 0.02
2-Butoxyethanol	C6H14O2	2.08 ±.0.02
Butyl acetate	C6H12O2	1.87 ±.0.02
sec-Butyl acetate	C6H12O2	1.87±.0.02
Butylamine	C4H11N	1.0 ±.0.01
sec-Butylamine	C4H11N	[1.28 ±.0.02
tert-Butylamine	C4H11N	1.29 ±.0.02
tert-Butylbenzene	C10H14	0.83 ±.0.02
Butyl ethyl ether	C6H14O	1.24 ±.0.02
Butyl formate	C5H10O2	2.03 ±.0.02
Butyl stearate	C22H44O2	1.88 ±.0.02
Butyl vinyl ether	C6H12O	1.25 ±.0.02
1-Butyne	C4H6	0.72 ± 0.04
γ-Butyrolactone	C4H6O2	4.27 ± 0.03
Calcium methoxide	CH3CaO	1.58 ± 0.08
Camphor, (+)	C10H16O	3.12 ±.0.02
Caprolactam	C6H11NO	3.93±.0.02
Carboimidic difluoride	CHF2N	1.39 ± 0.01
Carbon dioxide dimer–water	C2O4•H2O	1.99 ± 0.02
Carbon dioxide–mercury	CO2•Hg	0.17 ± 0.03
Carbon dioxide–water dimer complex	CO2•H4O2	1.74 ± 0.01
Carbon disulfide–sulfur	CO2•O2S	1.096 ± 0.01
Carbon monoselenide	CSe	1.99 ± 0.04
Carbon monosulfide	CS	1.958 ± 0.05
Carbon monoxide	CO	0.10 ±.0.02
Carbon monoxide dimer–	C2O2•H2O	1.57 ± 0.05
Carbon oxyselenide	COSe	0.73 ± 0.02
Carbon oxysulfide	COS	0.71±.0.02

Carbon oxysulfide–carbon	COS•C2O4	0.69 ± 0.05
Carbon oxysulfide–water	COS•H2O	2.68 ± 0.03
Carbonyl chloride	CCl2O	1.17 ± 0.01
Carbonyl fluoride	CF2O	0.95 ± 0.01
Chloroacetyl chloride	C2H2Cl2O	2.23 ± 0.11
Chloroacetylene	C2HCl	0.44±.0.02
2-Chloroaniline	C6H6ClN	1.77±.0.02
Chlorobenzene	C6H5Cl	1.69 ± 0.03
1-Chlorobutane	C4H9Cl	2.05 ± 0.04
2-Chlorobutane	C4H9Cl	2.04 ± 0.10
Chlorocyclohexane (axial)	C6H11Cl	1.91 ± 0.02
Chlorocyclohexane	C6H11Cl	2.44 ± 0.07
1-Chloro-1,1-difluoroethane	C2H3ClF2	2.14 ± 0.04
Chlorodifluoromethane	CHClF2	1.42 ± 0.03
Chloroethane	C2H5Cl	2.05 ± 0.02
2-Chloroethanol	C2H5ClO	1.78 ± 0.09
Chloroethene	C2H3Cl	1.45 ± 0.03
1-Chloro-4-fluorobenzene	C6H4ClF	0.12 ± 0.01
1-Chloro-1-fluoroethane	C2H4ClF	2.06 ± 0.01
Chlorofluoromethane	CH2ClF	1.82 ± 0.04
Chloromethane	CH3Cl	1.84 ± 0.02
(Chloromethyl)benzene	C7H7Cl	1.82±.0.02
1-Chloro-3-methylbutane	C5H11Cl	1.92±.0.02
1-Chloro-2-methylpropane	C4H9Cl	2.00 ± 0.10
2-Chloro-2-methylpropane	C4H9Cl	2.13 ± 0.04
1-Chloronaphthalene	C10H7Cl	1.57±.0.02
1-Chloro-2-nitrobenzene	C6H4ClNO2	4.64 ± 0.09
1-Chloro-3-nitrobenzene	C6H4ClNO2	3.73 ± 0.07
1-Chloro-4-nitrobenzene	C6H4ClNO2	2.83 ± 0.06
1-Chlorooctane	C8H17Cl	2.00±.0.02
Chloropentafluoroethane	C2ClF5	0.52 ± 0.05
1-Chloropentane	C5H11Cl	2.16 ± 0.11
4-Chlorophenol	C6H5ClO	2.11 ± 0.11
1-Chloropropane (gauche)	C3H7Cl	2.02 ± 0.03
1-Chloropropane (trans)	C3H7Cl	1.95 ± 0.02
1-Chloropropane (average)	C3H7Cl	2.05 ± 0.04
2-Chloropropane	C3H7Cl	2.17 ± 0.11
cis-1-Chloropropene	C3H5Cl	1.67 ± 0.08
trans-1-Chloropropene	C3H5Cl	1.97 ± 0.10
2-Chloropropene	C3H5Cl	1.64 ± 0.01
3-Chloropropene	C3H5Cl	1.94 ± 0.10
4-Chloropyridine	C5H4ClN	0.756± 0.05
2-Chlorotoluene	C7H7Cl	1.56 ± 0.08
3-Chlorotoluene	C7H7Cl	1.82±.0.02
4-Chlorotoluene	C7H7Cl	2.21 ± 0.04

Chlorotrifluoroethene	C2ClF3	0.40 ± 0.10
Chlorotrifluoromethane	CClF3	0.50 ± 0.01
o-Cresol	C H O	1.45 ±.0.02
m-Cresol	C7H8O	1.48 ±.0.02
p-Cresol	C7H8O	1.48 ±.0.02
Cyanamide	CH2N2	4.28 ± 0.10
Cyanoacetylene	C3HN	3.73 ±.0.02
Cyanoformamide	C2H2N2O	4.10 ± 0.12
Cyanogen azide (NCN3)	CN4	2.96 ± 0.07
Cyanogen chloride	CClN	2.81 ± 0.02
Cyanogen fluoride	CFN	2.12 ± 0.01
Cyanogen iodide	CIN	3.67 ± 0.02
Cyanomethylmercury	C2H3HgN	4.73 ± 0.13
Cyclobutanecarbonitrile	C5H7N	4.04 ± 0.04
Cyclobutanone	C4H6O	2.89 ± 0.03
Cyclobutene	C4H6	0.13 ± 0.01
1,3-Cycloheptadiene	C7H10	0.74±.0.02
2,4,6-Cycloheptatrien-1-one	C7H6O	4.13 ± 0.03
3,5-Cyclohexadiene-1,2-Dione	C6H4O2	4.23 ± 0.02
Cyclohexanone	C6H10O	3.24 ± 0.06
Cyclohexene (half-chair)	C6H10	0.33 ± 0.01
Cyclohexylamine	C6H13N	[1.26±.0.02
1,3-Cyclopentadiene	C5H6	0.69 ± 0.04
2,4-Cyclopentadien-1-one	C5H4O	3.13 ± 0.07
Cyclopentanone	C5H8O	3.34±.0.02
Cyclopentene	C5H8	0.20 ± 0.02
3-Cyclopenten-1-one	C5H6O	2.79 ± 0.03
Cyclopropane-sulfur dioxide Complex	C3H6•O2S	1.61 ± 0.01
Cyclopropanone	C3H4O	2.67 ± 0.13
Cyclopropene	C3H4	0.44 ± 0.01
Cyclopropylamine	C3H7N	1.19 ± 0.01
Cyclopropyl methyl ketone	C5H8O	2.62 ± 0.25
Diacetone alcohol	C6H12O2	3.24 ±.0.02
Diazomethane	CH2N2	1.50 ± 0.01
Dibromodifluoromethane	CBr2F2	0.66 ± 0.05
1,2-Dibromoethane	C2H4Br2	1.19 ±.0.02
Dibromomethane	CH2Br2	1.43 ± 0.03
1,2-Dibromopropane	C3H6Br2	1.24 ±.0.02
Dibutylamine	C8H19N	0.98 ±.0.02
Dibutyl ether	C8H18O	1.17 ± 0.06
Dibutyl phthalate	C16H22O4	[2.82 ±.0.04
Dibutyl sebacate	C18H34O4	[2.48 ±.0.02
Dibutyl sulfide	C8H18S	1.61 ±.0.02
o-Dichlorobenzene	C6H4Cl2	2.50 ± 0.05
m-Dichlorobenzene	C6H4Cl2	1.72 ± 0.09

1,4-Dichlorobutane	C4H8Cl2	2.22 ± 0.11
1,1-Dichloro-2,2-	C2Cl2F2	0.50 ±.0.02
Dichlorodifluoromethane	CCl2F2	0.51 ± 0.05
1,1-Dichloroethane	C2H4Cl2	2.06 ± 0.04
1,2-Dichloroethane	C2H4Cl2	[1.83 ±.0.02
1,1-Dichloroethene	C2H2Cl2	1.34 ± 0.01
cis-1,2-Dichloroethene	C2H2Cl2	1.90 ± 0.04
Dichlorofluoromethane	CHCl2F	1.29 ± 0.03
1,1-Dichloro-2-	C3H3Cl2F	2.43 ± 0.02
Dichloromethane	CH2Cl2	1.60 ± 0.03
(Dichloromethyl)benzene	C7H6Cl2	[2.07 ±.0.02
Dichloromethylborane	CH3BCl2	1.41 ± 0.03
1,2-Dichloropropane	C3H6Cl2	[1.85 ±.0.02
1,3-Dichloropropane	C3H6Cl2	2.08 ± 0.04
1,2-Dichloro-1,1,2,2-	C2Cl2F4	0.53 ±.0.02
2,4-Dichlorotoluene	C7H6Cl2	1.70 ±.0.02
3,4-Dichlorotoluene	C7H6Cl2	2.95 ±.0.02
Diethanolamine	C4H11NO2	2.84 ±.0.02
1,1-Diethoxyethane	C6H14O2	1.38 ±.0.02
Diethylamine	C4H11N	0.92 ± 0.05
Diethyl carbonate	C5H10O3	1.10 ± 0.06
Diethylene glycol	C4H10O3	2.31 ±.0.02
Diethylene glycol dimethyl	C6H14O3	1.97 ±.0.02
Diethylene glycol monoethyl	C6H14O3	1.64 ±.0.02
Diethylene glycol monoethyl	C8H16O4	1.84 ±.0.02
Diethylene glycol	C5H12O3	1.63 ±.0.02
Diethyl ether	C4H10O	1.098 ± 0.01
Diethyl malonate	C7H12O4	2.54 ±.0.02
Diethyl oxalate	C6H10O4	2.49 ±.0.02
Diethyl sulfide (trans-trans)	C4H10S	1.56 ± 0.04
Diethyl sulfide (trans-	C4H10S	1.591 ± 0.09
Diethyl sulfide (gauche-	C4H10S	1.65 ± 0.01
o-Difluorobenzene	C6H4F2	2.46 ± 0.05
m-Difluorobenzene	C6H4F2	1.51 ± 0.02
1,1-Difluorocyclohexane	C6H10F2	2.55 ± 0.01
3,3-Difluorocyclopropene	C3H2F2	2.98 ± 0.02
1,1-Difluoroethane	C2H4F2	2.27 ± 0.05
1,2-Difluoroethane (gauche)	C2H4F2	2.67 ± 0.13
1,1-Difluoroethene	C2H2F2	1.33 ± 0.02
cis-1,2-Difluoroethene	C2H2F2	2.42 ± 0.02
Difluoromethane	CH2F2	1.95 ± 0.02
Difluoromethylborane	CH3BF2	1.68 ± 0.03
Difluoromethylene	CF2	0.47 ± 0.02
1,1-Difluoro-1-propene	C3H4F2	0.89 ± 0.07
2,3-Dihydro-1,4-dioxin	C4H6O2	0.939 ± 0.008

3,6-Dihydro-1,2-dioxin	C4H6O2	2.329 ± 0.001
2,3-Dihydrofuran	C4H6O	1.32 ± 0.03
2,5-Dihydrofuran	C4H6O	1.63 ± 0.01
Dihydro-3-methyl-2(3H)-	C5H8O2	4.56 ± 0.02
Dihydro-5-methyl-2(3H)-	C5H8O2	4.71 ± 0.05
3,4-Dihydro-2H-pyran	C5H8O	1.400 ± 0.008
3,6-Dihydro-2H-pyran	C5H8O	1.283 ± 0.005
2,3-Dihydrothiophene	C4H6S	1.61 ± 0.20
2,5-Dihydrothiophene	C4H6S	1.75 ± 0.01
Diiodomethane	CH2I2	1.08 ±.0.02
Diisopentyl ether	C10H22O	1.23 ±.0.02
Diisopropylamine	C6H15N	[1.15 ±.0.02
Diisopropyl ether	C6H14O	1.13 ± 0.10
Diketene	C4H4O2	3.53 ± 0.07
1,2-Dimethoxybenzene	C8H10O2	[1.29 ±.0.02
Dimethoxymethane	C3H8O2	[0.74 ±.0.02
N,N-Dimethylacetamide	C4H9NO	[3.7±. 0.02
Dimethylamine	C2H7N	1.01 ± 0.02
N,N-Dimethylaniline	C H N	1.68 ± 0.17
2,4-Dimethylaniline	C8H11N	1.40 ±.0.02
2,6-Dimethylaniline	C8H11N	1.63 ±.0.02
3,3-Dimethyl-1-butyne	C6H10	0.661 ± 0.04
1,1-Dimethylcyclopropane	C5H10	0.142 ± 0.01
3,3-Dimethylcyclopropene	C5H8	0.287 ± 0.03
Dimethyl disulfide	C2H6S2	1.85 ±.0.02
Dimethyl ether	C2H6O	1.30 ± 0.01
N,N-Dimethylformamide	C3H7NO	3.82 ± 0.08
2,6-Dimethyl-4-heptanone	C9H18O	2.66 ±.0.02
Dimethyl maleate	C6H8O4	2.48 ±.0.02
2,4-Dimethyl-3-pentanone	C7H14O	2.74 ±.0.02
2,2-Dimethylpropanal	C5H10O	2.66 ± 0.05
2,2-Dimethylpropanenitrile	C5H9N	3.95 ± 0.04
2,4-Dimethylpyridine	C7H9N	2.30 ±.0.02
2,6-Dimethylpyridine	C7H9N	1.66 ±.0.02
Dimethyl sulfide	C2H6S	1.55 ± 0.04
Dimethyl sulfoxide	C2H6OS	3.96 ± 0.04
1,3-Dioxane	C4H8O2	2.06 ± 0.04
1,3-Dioxolane	C3H6O2	1.19 ± 0.06
Dipentyl ether	C10H22O	1.20 ±.0.02
Diphenyl ether	C12H10O	1.3 4±.0.02
Dipropylamine	C6H15N	1.03±.0.02
Dipropyl ether	C6H14O	1.21 ± 0.06
1,3-Dithiane	C4H8S2	2.14 ± 0.04
Divinyl ether	C4H6O	0.78 ±.0.02
Epichlorohydrin	C3H5ClO	1.84 ±.0.02

1,2-Epoxybutane	C4H8O	1.891 ± 0.011
1,2-Ethanediamine	C2H8N2	1.99 ± 0.10
1,2-Ethanediol, diacetate	C6H10O4	2.34 ±.0.02
1,2-Ethanedithiol	C2H6S2	2.03 ± 0.08
Ethanethiol (gauche)	C2H6S	1.61 ± 0.08
Ethanethiol (trans)	C2H6S	1.58 ± 0.08
Ethanol (gauche)	C2H6O	1.68 ± 0.03
Ethanol (trans)	C2H6O	1.44 ± 0.03
Ethanol (average)	C2H6O	1.69 ± 0.03
Ethanolamine	C2H7NO	2.27 ±.0.02
Ethoxybenzene	C8H10O	1.45 ± 0.15
2-Ethoxyethanol	C4H10O2	2.08 ±.0.02
2-Ethoxyethyl acetate	C6H12O3	2.25 ±.0.02
Ethyl acetate	C4H8O2	1.78 ± 0.09
Ethyl acrylate	C5H8O2	1.96 ±.0.02
Ethylamine (gauche)	C2H7N	1.21 ± 0.01
Ethylamine (trans)	C H N	1.34 ± 0.01
Ethylamine (average)	C2H7N	1.22 ± 0.10
Ethylbenzene	C8H10	0.59 ± 0.05
Ethyl benzoate	C9H10O2	2.00 ± 0.10
Ethyl butanoate	C6H12O2	1.74 ±.0.02
Ethyl trans-cinnamate	C11H12O2	1.84 ±.0.02
Ethyl cyanate	C3H5NO	4.72 ± 0.09
Ethyl cyanoacetate	C5H7NO2	2.17 ±.0.02
Ethylene carbonate	C3H4O3	4.93 ±.0.02
Ethylene glycol (average)	C2H6O2	2.36 ± 0.10
Ethyleneimine	C2H5N	1.90 ± 0.01
Ethylene–sulfur dioxide Complex	C2H4•O2S	1.65 ± 0.03
Ethylene–water complex	C2H4•H2O	1.10 ± 0.01
Ethyl formate (gauche)	C3H6O2	1.81 ± 0.02
Ethyl formate (trans)	C3H6O2	1.98 ± 0.02
Ethyl formate (average)	C3H6O2	1.93 ±.0.02
2-Ethyl-1-hexanol	C8H18O	1.74 ±.0.02
2-Ethylhexyl acetate	C10H20O2	1.83 ±.0.02
Ethyl lactate	C5H10O3	[2.42 ±.0.02
Ethyl methyl ether (trans)	C3H8O	1.17 ± 0.02
Ethyl methyl sulfide (gauche)	C3H8S	1.53 ± 0.04
Ethyl methyl sulfide (trans)	C3H8S	1.56 ± 0.03
Ethyl propanoate	C5H10O2	1.74 ±.0.02
Ethyl vinyl ether	C4H8O	1.26 ±.0.02
Fluoroacetylene	C2HF	0.77 ± 0.03
Fluorobenzene	C6H5F	1.60 ± 0.08
Fluorocyclohexane (equitorial)	C6H11F	2.11 ± 0.04
Fluorocyclohexane (axial)	C6H11F	1.81 ± 0.04
1-Fluorocyclohexene	C6H9F	1.92 ± 0.01

Fluoroethane	C2H5F	1.97 ± 0.07
Fluoroethene	C2H3F	1.48 ± 0.03
Fluoromethane	CH3F	1.88 ± 0.02
Fluoromethylidyne	CF	0.64 ± 0.05
(Fluoromethylidyne) phosphine (FCP)	CFP	0.29 ± 0.01
3-Fluorotoluene	C7H7F	1.82 ± 0.04
4-Fluorotoluene	C7H7F	2.00 ± 0.10
Formaldehyde	CH2O	2.332 ± 0.02
Formaldehyde dimer	C2H4O2	0.85 ± 0.05
Formamide	CH3NO	3.73 ± 0.07
Formic acid	CH2O2	1.45 ± 0.02
Formyl fluoride	CHFO	2.08 ± 0.01
Fulminic acid	CHNO	3.09 ±.0.02
Fulvene	C6H6	0.46 ± 0.03
Furan	C4H4O	0.66 ± 0.01
Furfural	C5H4O2	3.54 ±.0.02
Furfuryl alcohol	C5H6O2	1.92 ±.0.02
Glycerol	C3H8O3	2.56 ±.0.02
Glycine (Conformer I)	C2H5NO2	1.17 ± 0.05
Glycine (Conformer II)	C2H5NO2	5.45 ± 0.05
Glycolaldehyde	C2H4O2	2.73 ± 0.05
Glyoxal (cis)	C2H2O2	4.85 ± 0.22
2-Heptanol	C7H16O	1.71 ±.0.02
3-Heptanol	C7H16O	1.71 ±.0.02
2-Heptanone	C7H14O	2.59 ±.0.02
3-Heptanone	C7H14O	2.78 ±.0.02
Hexamethylphosphoric	C6H18N3OP	5.57 ±.0.02
Hexanoic acid	C6H12O2	1.13 ±.0.02
2-Hexanone	C6H12O	2.66 ±.0.02
sec-Hexyl acetate	C8H16O2	1.94 ±.0.02
1-Hexyne	C6H10	0.83 ± 0.05
Hydrogen cyanide	CHN	2.98 ±.0.02
Hydrogen cyanide trimer	C3H3N3	10.6 ±.0.02
Hydrogen isocyanide	CHN	3.05 ± 0.15
p-Hydroquinone	C6H6O2	2.38 ± 0.05
3-Hydroxypropanenitrile	C3H5NO	3.17 ± 0.02
Imidazole	C3H4N2	3.89 ± 0.04
Iodoacetylene	C2HI	0.05 ±.0.02
Iodobenzene	C6H5I	1.70 ± 0.09
1-Iodobutane	C4H9I	1.93±.0.02
2-Iodobutane	C4H9I	2.12 ± 0.11
Iodoethane	C2H5I	1.97 ± 0.02
Iodoethene	C2H3I	1.31 ± 0.05
Iodomethane	CH3I	1.64 ± 0.04
1-Iodo-2-methylpropane	C4H9I	1.87 ±.0.02

Iodomethylsilane	CH5ISi	1.86 ± 0.05
1-Iodopropane	C3H7I	2.04 ± 0.10
2-Iodopropane	C3H7I	1.95 ±.0.02
Isobutanal (gauche)	C4H8O	2.69 ± 0.01
Isobutanal (trans)	C4H8O	2.86 ± 0.01
Isobutane	C4H10	0.13 ± 0.02
Isobutene	C4H8	0.53 ± 0.01
Isobutyl acetate	C6H12O2	[1.86 ±.0.02
Isobutylamine	C4H11N	[1.27 ±.0.02
Isobutyl formate	C5H10O2	[1.88 ±.0.02
Isobutyl isobutanoate	C8H16O2	1.94 ±.0.02
Isocyanic acid (HNCO)	CHNO	1.96 ±.0.02
Isocyanobenzene	C7H5N	4.018 ± 0.003
Isocyanocyclopropane	C4H5N	4.03 ± 0.10
2-Isocyanopropane	C4H7N	4.05 ± 0.01
Isopentane	C5H12	0.13 ± 0.05
Isopentyl acetate	C7H14O2	[1.86 ±.0.02
Isopropylamine	C3H9N	1.19 ± 0.06
Isopropylbenzene	C9H12	0.79±.0.02
Isopropyl methyl ether	C4H10O	1.247 ± 0.03
Isoquinoline	C9H7N	2.73 ± 0.14
Isoxazole	C3H3NO	2.95 ± 0.04
Isoxazole–carbon monoxide Complex	C3H3NO•CO	2.873 ± 0.04
Ketene	C2H2O	1.42 ±.0.02
Mesityl oxide	C6H10O	2.79 ±.0.02
Methacrylic acid	C4H6O2	1.65 ±.0.02
Methanethiol	CH4S	1.52 ± 0.08
Methanol	CH4O	1.70 ± 0.02
2-Methoxyethanol (gauche)	C3H8O2	2.36 ± 0.05
2-Methoxyethyl acetate	C5H10O3	[.13 ±.0.02
1-Methoxy-1,2-propadiene	C4H6O	0.96 ± 0.02
N-Methylacetamide	C3H7NO	4.39 ±.0.02
Methyl acetate	C3H6O2	1.72 ± 0.09
Methyl acrylate	C4H6O2	1.77 ±.0.02
2-Methylacrylonitrile	C4H5N	3.69 ± 0.18
Methylamine	CH5N	1.31 ± 0.03
2-Methylaniline	C7H9N	1.60 ±.0.02
3-Methylaniline	C7H9N	1.45 ±.0.02
4-Methylaniline	C7H9N	1.52±.0.02
Methyl azide	CH3N3	2.17 ± 0.04
Methyl benzoate	C8H8O2	[1.94 ±.0.02
2-Methyl-1,3-butadiene	C5H8	0.25 ± 0.01
3-Methylbutanoic acid	C5H10O2	0.63 ±.0.02
2-Methyl-1-butanol	C5H12O	1.88 ±.0.02
2-Methyl-2-butanol	C5H12O	1.82 ±.0.02

3-Methyl-1-butene (gauche)	C5H10	0.38 ± 0.04
3-Methyl-1-butene (trans)	C5H10	0.32 ± 0.01
3-Methyl-2-butenenitrile	C5H7N	4.61 ± 0.13
2-Methyl-1-buten-3-yne	C5H6	0.53 ± 0.02
Methyl cyanate	C2H3NO	4.26 ± 0.18
cis-3-Methylcyclohexanol	C7H14O	1.91 ±.0.02
trans-3-Methylcyclohexanol	C7H14O	1.75 ±.0.02
3-Methylcyclopentanone	C6H10O	3.14 ± 0.03
3-Methyl-2-cyclopenten-1-One	C6H8O	4.33 ± 0.02
Methylcyclopropane	C4H8	0.19 ± 0.04
Methyldiborane(6)	CH8B2	0.56 ± 0.06
Methyldifluorophosphine	CH3F2P	2.06 ± 0.06
Methylenecyclohexane	C7H12	0.62 ± 0.01
Methylenecyclopropene	C4H4	1.90 ± 0.01
Methylenephosphine (CH2 = PH)	CH3P	0.89 ± 0.03
N-Methylformamide	C2H5NO	3.83 ± 0.08
Methyl formate	C2H4O2	1.77 ± 0.04
2-Methylfuran	C5H6O	0.65 ± 0.05
3-Methylfuran	C5H6O	1.03 ± 0.02
5-Methyl-2(3H)-furanone	C5H6O2	4.08 ± 0.02
Methyl hydroperoxide	CH4O2	0.65 ±.0.02
Methylidyne	CH	1.46 ±.0.02
Methyl isocyanate	C2H3NO	2.8 ±.0.02
Methyl isothiocyanate	C2H3NS	3.43 ± 0.03
4-Methylisoxazole	C4H5NO	3.53 ± 0.05
Methyl methacrylate	C5H8O2	1.67 ±.0.02
2-Methyloxazole	C4H5NO	1.37 ± 0.07
4-Methyloxazole	C4H5NO	1.08 ± 0.05
5-Methyloxazole	C4H5NO	2.16 ± 0.04
Methyloxirane	C3H6O	2.01 ± 0.02
2-Methyl-2,4-pentanediol	C6H14O2	2.97 ±.0.02
4-Methylpentanenitrile	C6H11N	3.50 ±.0.02
Methylphosphonic difluoride	CH3F2OP	3.69 ± 0.26
N-Methylpropanamide	C4H9NO	3.61 ±.0.02
2-Methylpropanenitrile	C4H7N	4.29 ± 0.09
2-Methyl-2-propanethiol	C4H10S	1.66 ± 0.03
2-Methylpropanoic acid	C4H8O2	1.08 ±.0.02
2-Methyl-1-propanol	C4H10O	1.64 ± 0.08
2-Methyl-2-propanol	C4H10O	1.66 ±.0.02
2-Methylpropenal	C4H6O	2.68 ± 0.13
2-Methyl-2-propenol (skew)	C4H8O	1.29 ± 0.02
Methyl propyl ether (trans-trans)	C4H10O	1.17 ± 0.03
2-Methylpyridine	C6H7N	1.85 ± 0.04
3-Methylpyridine	C6H7N	2.40 ±.0.02

4-Methylpyridine	C6H7N	2.70 ± 0.02
2-Methylpyrimidine	C5H6N2	1.67 ± 0.01
5-Methylpyrimidine	C5H6N2	2.88 ± 0.06
N-Methylpyrrolidine	C5H11N	0.57 ± 0.03
N-Methyl-2-pyrrolidinone	C5H9NO	4.19 ±.0.02
Methyl salicylate	C8H8O3	2.47 ±.0.02
Methylsilane	CH6Si	0.73 ±.0.02
Methyl silyl ether	CH6OSi	1.15 ± 0.02
3-Methylthietane	C4H8S	2.04 ± 0.09
2-Methylthiophene	C5H6S	0.674 ± 0.05
3-Methylthiophene	C5H6S	0.914 ± 0.01
Methyl vinyl ether	C3H6O	0.95 ± 0.02
Morpholine	C4H9NO	1.55 ± 0.03
2-Nitroanisole	C7H7NO3	5.04 ±.0.02
Nitrobenzene	C6H5NO2	4.22 ± 0.08
Nitroethane	C2H5NO2	3.23 ± 0.03
Nitromethane	CH3NO2	3.46 ± 0.02
1-Nitropropane	C3H7NO2	3.66 ± 0.07
2-Nitropropane	C3H7NO2	3.73 ± 0.07
Nonanoic acid	C9H18O2	0.79 ±.0.02
2,5-Norbornadiene	C7H8	0.08 ± 0.01
cis-9-Octadecenoic acid	C18H34O2	1.18 ±.0.02
Octanoic acid	C8H16O2	1.15 ±.0.02
1-Octanol	C8H18O	1.76 ±.0.02
2-Octanol	C8H18O	1.71 ±.0.02
2-Octanone	C8H16O	2.70 ±.0.02
1,4-Oxathiane	C4H8OS	0.29 ± 0.03
Oxazole	C3H3NO	1.50 ± 0.03
Oxetane	C3H6O	1.94 ± 0.01
2-Oxetanone	C3H4O2	4.18 ± 0.03
3-Oxetanone	C3H4O2	0.88 ± 0.05
Oxirane	C2H4O	1.89 ± 0.01
Paraldehyde	C6H12O3	1.43 ± 0.07
Pentachloroethane	C2HCl5	0.92 ± 0.05
cis-1,3-Pentadiene	C5H8	0.50 ± 0.01
trans-1,3-Pentadiene	C5H8	0.55 ± 0.01
1,3-Pentadiyne	C5H4	1.27 ± 0.01
1,5-Pentanediol	C5H12O2	2.57 ±.0.02
2,4-Pentanedione	C5H8O2	2.78 ±.0.02
Pentanenitrile	C5H9N	4.12 ± 0.08
Pentanoic acid	C5H10O2	1.61 ±.0.02
1-Pentanol	C5H12O	1.70 ±.0.02
2-Pentanol	C5H12O	1.66 ±.0.02
3-Pentanol	C5H12O	1.64 ±.0.02
2-Pentanone	C5H10O	2.70 ±.0.02

3-Pentanone	C5H10O	2.82 ±.0.02
1,2,3-Pentatriene	C5H6	0.51 ± 0.05
1-Pentene	C5H10	0.53 ±.0.02
1-Penten-3-yne	C5H6	0.66 ± 0.02
cis-3-Penten-1-yne	C5H6	0.78 ± 0.02
trans-3-Penten-1-yne	C5H6	1.06 ± 0.05
Pentyl acetate	C7H14O2	1.75 ± 0.10
Pentyl formate	C6H12O2	1.90 ± 0.10
1-Pentyne (gauche)	C5H8	0.79 ± 0.02
1-Pentyne (trans)	C5H8	0.84 ± 0.01
Perfluoropyridine	C5F5N	0.98 ± 0.08
Phenol	C6H6O	1.24 ± 0.08
Phenylacetylene	C8H6	0.65 ± 0.05
Phenylsilane	C6H8Si	0.85 ± 0.12
1-Phosphapropyne (CH3CP)	C2H3P	1.49 ± 0.01
Piperidine (equitorial)	C5H11N	0.82 ± 0.02
Piperidine (axial)	C5H11N	1.19 ± 0.02
Piperidine (average)	C5H11N	1.19 ±0.02
Propanal (gauche)	C3H6O	2.86 ± 0.01
Propanal (cis)	C3H6O	2.52 ± 0.05
Propanal (average)	C3H6O	2.72 ±.0.02
Propane	C3H8	0.08 ± 0.01
1,2-Propanediol	C3H8O2	2.25 ±.0.02
1,3-Propanediol	C3H8O2	2.55 ±.0.02
Propanenitrile	C3H5N	4.05 ± 0.03
1-Propanethiol (gauche)	C3H8S	1.63 ± 0.01
1-Propanethiol (trans)	C3H8S	1.60 ± 0.08
2-Propanethiol (gauche)	C3H8S	1.53 ± 0.03
2-Propanethiol (trans)	C3H8S	1.61 ± 0.03
Propanoic acid (cis)	C3H6O2	1.46 ± 0.07
Propanoic acid (average)	C3H6O2	1.75 ± 0.09
1-Propanol (gauche)	C3H8O	1.58 ± 0.03
1-Propanol (trans)	C3H8O	1.55 ± 0.03
2-Propanol (trans)	C3H8O	1.58 ± 0.03
Propargyl alcohol	C3H4O	1.13 ± 0.06
Propene	C3H6	0.36 ± 0.01
Propene–sulfur dioxide	C3H6•O2S	1.34 ± 0.03
Propyl acetate	C5H10O2	1.78 ±.0.02
Propylamine	C3H9N	1.17 ± 0.06
Propylene carbonate	C4H6O3	4.97 ±.0.02
Propyleneimine (cis)	C3H7N	1.77 ± 0.09
Propyleneimine (trans)	C3H7N	1.57 ± 0.03
Propyl formate	C4H8O2	1.89 ±.0.02
2-Propynal	C3H2O	2.78 ± 0.02
Propyne	C3H4	0.74 ± 0.01

Propyne-argon complex	C3H4•Ar	0.73 ± 0.05
4H-Pyran-4-one	C5H4O2	3.79 ± 0.02
4H-Pyran-4-thione	C5H4OS	3.95 ± 0.05
1H-Pyrazole	C3H4N2	2.20 ± 0.01
Pyridazine	C4H4N2	4.22 ± 0.02
Pyridine	C5H5N	2.25 ± 0.01
2-Pyridinecarbonitrile	C6H4N2	5.78 ± 0.11
3-Pyridinecarbonitrile	C6H4N2	3.66 ± 0.11
4-Pyridinecarbonitrile	C6H4N2	1.96 ± 0.03
3-Pyridinecarboxaldehyde	C6H5NO	1.44 ±.02
4-Pyridinecarboxaldehyde	C6H5NO	1.66 ±.02
2-Pyridinecarboxaldehyde	C6H5NO	3.56 ± 0.07
Pyrimidine	C4H4N2	2.33 ± 0.10
Pyrrole	C4H5N	1.76 ± 0.01
Pyrrolidine	C4H9N	1.57 ±.02
2-Pyrrolidone	C4H7NO	3.52 ±.02
Quinoline	C9H7N	2.29 ± 0.11
Salicylaldehyde	C7H6O2	2.86 ±.0.02
Selenoformaldehyde	CH2Se	1.41 ± 0.01
Silicon dicarbide	C2Si	2.33 ± 0.06
Silicon methylidyne	CHSi	0.66 ± 0.02
Styrene	C8H8	0.12 ± 0.03
Succinonitrile	C4H4N2	3.78 ±.0.02
Sulfolane	C4H8O2S	4.83 ±0.02
1,1,2,2-Tetrabromoethane	C2H2Br4	1.38 ±.0.02
1,1,2,2-Tetrachloroethane	C2H2Cl4	1.32 ± 0.07
1,2,3,4-Tetrafluorobenzene	C6H2F4	2.42 ± 0.05
1,2,3,5-Tetrafluorobenzene	C6H2F4	1.46 ± 0.06
1,1,1,2-Tetrafluoroethane	C2H2F4	1.80 ± 0.22
Tetrahydrofuran	C4H8O	1.75 ± 0.04
Tetrahydrofurfuryl alcohol	C5H10O2	[2.12 ±.002
Tetrahydropyran (chair)	C5H10O	1.58 ± 0.03
Tetrahydro-4H-pyran-4-one	C5H8O2	1.72 ± 0.03
1,2,5,6-Tetrahydropyridine	C5H9N	1.07 ± 0.03
Tetrahydrothiophene	C4H8S	1.90 ±0.0.03
Tetramethylurea	C5H12N2O	3.54 ±0.02
1H-Tetrazole	CH2N4	2.19 ± 0.05
Thiacyclohexane	C5H10S	1.78 ± 0.01
1,2,5-Thiadiazole	C2H2N2S	1.57 ± 0.07
Thietane	C3H6S	1.85 ± 0.09
Thietane 1,1-dioxide	C3H6O2S	4.83 ± 0.01
Thioacetaldehyde	C2H4S	2.33 ± 0.02
Thiocarbonyl fluoride	CF2S	0.08 ±0.01
Thioformaldehyde	CH2S	1.64 ± 0.04
Thiophene	C4H4S	0.55 ± 0.01

2-Thiophenecarbonitrile	C5H3NS	4.59 ± 0.02
3-Thiophenecarbonitrile	C5H3NS	4.13 ± 0.02
4H-Thiopyran-4-thione	C5H4S2	3.93 ± 0.12
Toluene	C7H8	0.37 ± 0.01
Toluene-sulfur dioxide Complex	C7H8•O2S	1.87 ± 0.03
1H-1,2,4-Triazole	C2H3N3	2.72 ± 0.41
Tribromomethane	CHBr3	0.99 ± 0.02
Tributylamine	C12H27N	0.78 ±0.02
Tributyl borate	C12H27BO3	0.77 ±0.01
Tributyl phosphate	C12H27O4P	3.07 ±.0.02
Tricarbon monosulfide	C3S	3.82 ± 0.09
1,1,1-Trichloroethane	C2H3Cl3	1.76 ± 0.01
1,1,2-Trichloroethane	C2H3Cl3	1.43 ±0.01
Trichloroethene	C2HCl3	0.18±0.24
Trichloroethylsilane	C2H5Cl3Si	2.04 ±0.11
Trichlorofluoromethane	CCl3F	0.46 ± 0.02
Trichloromethane	CHCl3	1.04 ± 0.02
(Trichloromethyl)benzene	C7H5Cl3	2.03 ±0.03
Trichloromethylsilane	CH3Cl3Si	1.91 ± 0.01
Tri-o-cresyl phosphate	C21H21O4P	2.87 ±0.21
Tri-m-cresyl phosphate	C21H21O4P	3.05 ±0.01
Tri-p-cresyl phosphate	C21H21O4P	3.18 ±0.05
Triethanolamine	C6H15NO3	3.57 ±0.11
Triethylamine	C6H15N	0.66 ± 0.05
Triethyl phosphate	C6H15O4P	3.12 ±.0.02
Trifluoroacetic acid	C2HF3O2	2.28 ± 0.25
Trifluoroacetonitrile	C2F3N	1.26 ± 0.01
1,2,4-Trifluorobenzene	C6H3F3	1.40 ± 0.07
1,1,1-Trifluoroethane	C2H3F3	2.35 ± 0.05
Trifluoroethene	C2HF3	1.33 ± 0.04
Trifluoroiodomethane	CF3I	1.58 ± 0.03
Trifluoroisocyanomethane	C2F3N	1.15 ± 0.03
Trifluoromethane	CHF3	1.65 ±0.12
(Trifluoromethyl)benzene	C7H5F3	2.76 ± 0.04
Trifluoromethylsilane	CH3F3Si	2.33 ± 0.05
(Trifluoromethyl)silane	CH3F3Si	2.33 ± 0.04
3,3,3-Trifluoropropene	C3H3F3	2.46 ± 0.04
3,3,3-Trifluoro-1-propyne	C3HF3	2.32 ± 0.12
Trimethylamine	C3H9N	0.64 ± 0.04
Trimethyl phosphate	C3H9O4P	3.18 ±0.02
2,4,6-Trimethylpyridine	C8H11N	2.05 ±0.01
1,3,5-Trioxane	C3H6O3	2.09 ± 0.02
Vinyl acetate	C4H6O2	1.79 ±0.05
Vinyl formate	C3H4O2	1.46 ± 0.01
2-Vinylfuran	C6H6O	0.69 ± 0.07

Vinylsilane	C2H6Si	0.68 ± 0.02
o-Xylene	C8H10	0.63 ± 0.05
2,4-Xylenol	C8H10O	1.40 ±0.05
2,5-Xylenol	C8H10O	1.46 ±0.12
2,6-Xylenol	C8H10O	1.42 ±0.01
3,4-Xylenol	C8H10O	1.55 ±0.02
3,5-Xylenol	C8H10O	1.55 ±0.05

B: FUNDAMENTAL VIBRATIONAL FREQUENCIES OF SMALL MOLECULES

This table lists the fundamental vibrational frequencies of selected three-, four-, and five-atom molecules.. The vibrational modes are described by their approximate character in terms of stretching, bending, deformation, etc.. Abbreviations are:

sym..	symmetric
antisym..	antisymmetric
str..	stretch
deform..	deformation
scis..	scissors
rock..	rocking
deg..	degenerate

All fundamental frequencies (more precisely, wavenumbers) are given in units of cm^{-1}

	XY_2 Molecules			
	Point groups $D_{\infty h}$(linear) and C_{2v}(bent)			
Molecule	**Structure**	**Sym. str.**	**Bend**	**Antisym. str.**
CO_2	Linear	1333	667	2349
CS_2	Linear	658	397	1535
$^{C}3$	Linear	1224	63	2040
CNC	Linear		321	1453
NCN	Linear	1197	423	1476
BO_2	Linear	1056	447	1278
BS_2	Linear	510	120	1015

KrF_2	Linear	449	233	590
XeF_2	Linear	515	213	555
$XeCl_2$	Linear	316		481
H_2O	Bent	3657	1595	3756
D_2O	Bent	2671	1178	2788
F_2O	Bent	928	461	831
Cl_2O	Bent	639	296	686
O_3	Bent	1103	701	1042
H_2S	Bent	2615	1183	2626
D_2S	Bent	1896	855	1999
SF_2	Bent	838	357	813
SCl_2	Bent	525	208	535
SO_2	Bent	1151	518	1362
H_2Se	Bent	2345	1034	2358
D_2Se	Bent	1630	745	1696

	XY_2 Molecules			
	Point groups $D_{\infty h}$(linear) and C_{2v}(bent)			
Molecule	**Structure**	**Sym. str.**	**Bend**	**Antisym. str.**
NH_2	Bent	3219	1497	3301
NO_2	Bent	1318	750	1618
NF_2	Bent	1075	573	942
ClO_2	Bent	945	445	1111
CH_2	Bent		963	
CD_2	Bent		752	
CF_2	Bent	1225	667	1114
CCl_2	Bent	721	333	748
CBr_2	Bent	595	196	641

SiH_2	Bent	2032	990	2022
SiD_2	Bent	1472	729	1468
SiF_2	Bent	855	345	870
$SiCl_2$	Bent	515		505
$SiBr_2$	Bent	403		400
GeH_2	Bent	1887	920	1864
$GeCl_2$	Bent	399	159	374
SnF_2	Bent	593	197	571
$SnCl_2$	Bent	352	120	334
$SnBr_2$	Bent	244	80	231
PbF_2	Bent	531	165	507
$PbCl_2$	Bent	314	99	299
ClF_2	Bent	500		576

XYZ Molecules				
Point Groups $C_{\infty v}$ (linear) and C_s(bent)				
Molecule	**Structure**	**XY str.**	**Bend**	**YZ str.**
OCS	Linear	2062	520	859
NCO	Linear	1270	535	1921
NNO	Linear	2224	589	1285
HNB	Linear	3675		2035
HNC	Linear	3653		2032
HNSi	Linear	3583	523	1198
HBO	Linear		754	1817
FBO	Linear		500	2075
ClBO	Linear	676	404	1958

BrBO	Linear	535	374	1937

XYZ Molecules				
Point Groups $C_{\infty v}$ (linear) and C_s(bent)				
Molecule	**Structure**	**XY str.**	**Bend**	**YZ str.**
FNO	Bent	766	520	1844
ClNO	Bent	596	332	1800
BrNO	Bent	542	266	1799
HNF	Bent		1419	1000
HNO	Bent	2684	1501	1565
HPO	Bent	2095	983	1179
HOF	Bent	3537	886	1393
HOCl	Bent	3609	1242	725
HOO	Bent	3436	1392	1098
FOO	Bent	579	376	1490

XYZ Molecules				
Point Groups $C_{\infty v}$ (linear) and C_s(bent)				
Molecule	**Structure**	**XY str.**	**Bend**	**YZ str.**
ClOO	Bent	407	373	1443
BrOO	Bent			1487
HSO	Bent		1063	1009
NSF	Bent	1372	366	640
NSCl	Bent	1325	273	414
HCF	Bent		1407	1181
HCCl	Bent		1201	815

HSiF	Bent	1913	860	834
HSiCl	Bent		808	522
HSiBr	Bent	1548	774	408

	Symmetric XY_3 Molecules				
	Point Groups D_{3h} (planar) and C_{3v} (pyramidal)				
Molecule	**Structure**	**Sym. str.**	**Sym. deform.**	**Deg. str.**	**Deg. deform.**
NH_3	Pyram..	3337	950	3444	1627
ND_3	Pyram..	2420	748	2564	1191
PH_3	Pyram..	2323	992	2328	1118
AsH_3	Pyram..	2116	906	2123	1003
SbH_3	Pyram..	1891	782	1894	831
NF_3	Pyram..	1032	647	907	492
PF_3	Pyram..	892	487	860	344
AsF_3	Pyram..	741	337	702	262
PCl_3	Pyram..	504	252	482	198
PI_3	Pyram..	303	111	325	79
AsI_3	Pyram..	219	94	224	71
$AlCl_3$	Pyram..	375	183	595	150
SO_3	Planar	1065	498	1391	530
BF_3	Planar	888	691	1449	480
BH_3	Planar		1125	2808	1640
CH_3	Planar		606	3161	1396
CD_3	Planar		453	2369	1029
CF_3	Pyram..	1090	701	1260	510
SiF_3	Pyram..	830	427	937	290

	Linear XYYX Molecules			
	Point Group $D_{\infty h}$			
Sym. XY str.	**Antisym. XY str.**	**YY str.**	**Bend**	**Bend**
3374	3289	1974	612	730
2701	2439	1762	505	537
2330	2158	851	507	233

Planar X_2YZ Molecules						
		Point Group C_{2v}				
Molecule	**Sym.XY str.**	**YZ str.**	**YX_2 scis.**	**Antisym. XY str.**	**YX_2 rock**	**YX_2 wag**
H_2CO	2783	1746	1500	2843	1249	1167
D_2CO	2056	1700	1106	2160	990	938
F_2CO	965	1928	584	1249	626	774
Cl_2CO	567	1827	285	849	440	580
O_2NF	1310	822	568	1792	560	742
O_2NCl	1286	793	370	1685	408	652

Tetrahedral XY4 Molecules				
	Point Group T_d			
Molecule	**Sym. str.**	**Deg. deform.(e)**	**Deg. str.(f)**	**Deg. deform.(f)**
CH_4	2917	1534	3019	1306
CD_4	2109	1092	2259	996
CF_4	909	435	1281	632
CCl_4	459	217	776	314

CBr_4	267	122	672	182
CI_4	178	90	555	125
SiH_4	2187	975	2191	914
SiD_4	1558	700	1597	681
SiF_4	800	268	1032	389
$SiCl_4$	424	150	621	221
GeH_4	2106	931	2114	819
GeD_4	1504	665	1522	596
$GeCl_4$	396	134	453	172
$SnCl_4$	366	104	403	134
$TiCl_4$	389	114	498	136
$ZrCl_4$	377	98	418	113
$HfCl_4$	382	102	390	112
RuO_4	885	322	921	336
OsO_4	965	333	960	329

C: PYTHON SCRIPTS USED IN MACHINE LEARNING

Installing libraries we are going to use

[1]:
```
# pip install sklearn
# pip install pandas
# pip install numpy
# pip install tensorflow
# pip install keras

# NOTE: Installing Keras and Tensorflow may vary if you are using
anaconda,
# consider googling which best suites your case.Run the following
commands in␣
 ↪gitbash
```

[2]:
```
#impor
t
librar
ies
import
pandas
as pd
import
numpy
as np
from sklearn.utils import shuffle
```

Load Dataset

[3]:
```
# load dataset
df=pd.read_csv('static/RamanScatter.csv')
```

[4]:
```
# print the first 5 rows
print(df.head(5))
```

```
  Incident Frequencies Modified anti-Stokes frequencies    Unnamed:
  2  \
0          1 * 1014s-1                     1 + 2  + D  1  + 2  - D
1                46.52                           46.52         57.58
2                38.61                           38.61         49.67
3                35.38                           35.38         46.44
4                29.63                           29.63         40.69

  Modified Stokes frequencies    Unnamed: 4
0                1 - 2 - D  1 - 2 + D
1                      49.34         35.46
2                      41.43         27.55
3                       38.2         24.32
4                      32.45         18.57
```

```
[5]: # remove the header
     df.columns = [''] * len(df.columns)
```

```
[6]: #  assign column names
     df.columns =␣
      ↪['Frequency_0','Frequency_1','Frequency_2','Frequency_3','Freq
      uency_4']
```

```
[7]: # delete the first row
     df=df.drop(df.index[0])
```

```
[8]: # shuffle our data randomly then
     reset the index
     df=shuffle(df,random_state=1)
     df=df.reset_index(drop=True)
```

```
[9]: df.head(5)
```

```
[9]:   Frequency_0 Frequency_1 Frequency_2 Frequency_3 Frequency_4
     0       17.71       22.43       21.63       12.99       13.79
     1       29.63       40.96       32.09        18.3       27.17
     2       17.71       29.04       20.17        6.38       15.25
     3        22.7       27.04       26.75       18.36       18.65
     4       18.29       24.03       23.76       12.55       12.82
```

```
[10]: df.shape
```

```
[10]: (80, 5)
```

Since we have only 80 rows in our dataset, we will need to do data augmentation to prevent the likelyhood of overfitting during training of the model.

Overfitting occurs when the model performs well on the training data but does not perform accu- rately in the testing data. Data Augmentation is the techniques used to increase the amount of data by adding slightly modified copies of already existing data.

```
[11]: #perform data augmentation to remove likelyhood of overfitting
      unique=df['Frequency_0'].uni
      que().tolist()
      Frequency_0=unique * 1000
      import random
      random_list=[]
      for i in range(0,8000):
          x=round(random.uniform(
          1.10,10.10),2)
          random_list.append(x)
      # print(random_list)
```

```
[12]: list_1=[]
      for i in range(0,8000):
          x=round(random.uniform
          (7.00,9.90),2)
          list_1.append(x)
      # print(list_1)
```

```
[13]: df_from_list_0 = pd.DataFrame (Frequency_0, columns =
['Frequency_0'])
      df_from_list= pd.DataFrame (random_list, columns = ['Difference'])

      result = pd.concat([df_from_list_0, df_from_list], axis=1,
      join='inner')
      # print (result.head(5))
```

```
[14]: result['Frequency_1']=result['Frequency_0'].astype(float) +
result['Difference']
      result['Frequency_2']=result['Frequency_1'].astype(float) -␣
       ↪result['Difference'] - 1.00
      df_from_list_1= pd.DataFrame (list_1, columns = ['Difference_1'])
      result = pd.concat([result, df_from_list_1], axis=1, join='inner')
      # result.head(5)
```

```
[15]: result['Frequency_3']=result['Frequency_0'].astype(float) -␣
       ↪result['Difference_1']
      result['Frequency_4']=result['Frequency_3'] +
      result['Difference']
      result['Frequency_0']=result['Frequency_0'].astype(fl
      oat) * 1.0
      result=result.drop(['Difference','Difference_1'],
      axis=1) df=df.astype(float).multiply(1.0)
      df_result = pd.concat([df, result], ignore_index=True, sort=False)
      result.head(5)
```

[15]:

	Frequency_0	Frequency_1	Frequency_2	Frequency_3	Frequency_4
0	17.71	21.60	16.71	9.89	13.78
1	29.63	31.00	28.63	20.86	22.23
2	22.70	25.96	21.70	14.79	18.05
3	18.29	23.96	17.29	10.45	16.12
4	35.38	44.50	34.38	27.81	36.93

```
[16]: df_result.shape
```

```
[16]: (8080, 5)
```

As you can see we have increased the size of the data from 80 rows to 8080 rows

Data Visualization

```
[17]: # visualize the data
      import matplotlib.pyplot as plt
      df.plot.line(figsize=(10,5))
```

```
[17]: <AxesSubplot:>
```

```
[18]: df.plot.area(figsize=(10,5))
```

```
[18]: <AxesSubplot:>
```

We are then creating a new column and name it frequency loss.Any atom that lost more than 4 frequencies of energy is said to have cance

```
[19]: df_result['Frequency_loss']=df_result['Frequency_1']-
df_result['Frequency_2']
 df_result['cancer_probability']= df_result['Frequency_loss'].apply(lambda
 x: 0␣
  ↪if x <= 4 else 1)
 print(df_result.tail(20))
```

```
      Frequency_0  Frequency_1  Frequency_2  Frequency_3  Frequency_4  \
8060        35.38        43.95        34.38        26.80        35.37
8061        38.61        48.28        37.61        28.76        38.43
8062        24.00        33.25        23.00        15.45        24.70
8063        46.52        54.06        45.52        39.42        46.96
8064        17.71        20.50        16.71         7.94        10.73
8065        29.63        39.09        28.63        21.47        30.93
8066        22.70        24.36        21.70        14.45        16.11
8067        18.29        21.45        17.29         9.24        12.40
8068        35.38        36.89        34.38        27.41        28.92
8069        38.61        43.93        37.61        30.86        36.18
8070        24.00        27.37        23.00        14.67        18.04
8071        46.52        53.73        45.52        39.08        46.29
8072        17.71        25.47        16.71         9.51        17.27
8073        29.63        33.31        28.63        21.48        25.16
8074        22.70        24.15        21.70        14.02        15.47
8075        18.29        25.84        17.29         9.37        16.92
8076        35.38        40.00        34.38        27.18        31.80
8077        38.61        46.51        37.61        29.02        36.92
8078        24.00        30.33        23.00        14.66        20.99
8079        46.52        54.22        45.52        38.68        46.38

      Frequency_loss  cancer_probability
8060            9.57                   1
8061           10.67                   1
8062           10.25                   1
```

```
8063           8.54                   1
8064           3.79                   0
8065          10.46                   1
8066           2.66                   0
8067           4.16                   1
8068           2.51                   0
8069           6.32                   1
8070           4.37                   1
8071           8.21                   1
8072           8.76                   1
8073           4.68                   1
8074           2.45                   0
8075           8.55                   1
8076           5.62                   1
8077           8.90                   1
8078           7.33                   1
8079           8.70                   1
```

[20]:
```
# plotting two columns
df.plot.line(y=['Frequency_1', 'Frequency_2'], figsize=(10,6))
```

[20]: <AxesSubplot:>


```
      Frequency_loss  cancer_probability
8060           9.57                   1
8061          10.67                   1
8062          10.25                   1
8063           8.54                   1
8064           3.79                   0
8065          10.46                   1
8066           2.66                   0
8067           4.16                   1
8068           2.51                   0
8069           6.32                   1
8070           4.37                   1
8071           8.21                   1
8072           8.76                   1
8073           4.68                   1
8074           2.45                   0
8075           8.55                   1
8076           5.62                   1
8077           8.90                   1
8078           7.33                   1
8079           8.70                   1
```

[21]:
```
Features=df.filter(['Frequency_0','Frequency_1','Frequency_2'],
axis=1)
```

```
# plotting Frequency_1
ax = df_result.head(20).plot(x="Frequency_0", y="Frequency_1",␣
 ↪kind="bar",figsize=(10,5))
# plotting Frequency_2 on the same axis
df_result.head(20).plot(x="Frequency_0", y="Frequency_2",
kind="bar", ax=ax,␣
 ↪color="orange",figsize=(10,5))
```

[21]: <AxesSubplot:xlabel='Frequency_0'>

	Frequency_loss	cancer_probability
8060	9.57	1
8061	10.67	1
8062	10.25	1
8063	8.54	1
8064	3.79	0
8065	10.46	1
8066	2.66	0
8067	4.16	1
8068	2.51	0
8069	6.32	1
8070	4.37	1
8071	8.21	1
8072	8.76	1
8073	4.68	1
8074	2.45	0
8075	8.55	1
8076	5.62	1
8077	8.90	1
8078	7.33	1
8079	8.70	1

[22]:
```
# plotting Frequency_1
ax = df_result.tail(20).plot(x="Frequency_0", y="Frequency_1",␣
 ↪kind="bar",figsize=(10,5))
# plotting Frequency_2 on the same axis
df_result.tail(20).plot(x="Frequency_0", y="Frequency_2",
kind="bar", ax=ax,␣
 ↪color="pink",figsize=(10,5))
```

[22]: <AxesSubplot:xlabel='Frequency_0'>

	Frequency_loss	cancer_probability
8060	9.57	1

```
8061          10.67                  1
062           10.25                  1
8063           8.54                  1
8064           3.79                  0
8065          10.46                  1
8066           2.66                  0
8067           4.16                  1
8068           2.51                  0
8069           6.32                  1
8070           4.37                  1
8071           8.21                  1
8072           8.76                  1
8073           4.68                  1
8074           2.45                  0
8075           8.55                  1
8076           5.62                  1
8077           8.90                  1
8078           7.33                  1
8079           8.70                  1
```

[23]:
```
# plotting Frequency_1
ax = df_result[500:520].plot(x="Frequency_0", y="Frequency_1",␣
 ↪kind="bar",figsize=(10,5))
# plotting Frequency_2 on the same axis
df_result[500:520].plot(x="Frequency_0", y="Frequency_2",
kind="bar", ax=ax,␣
 ↪color="black",figsize=(10,5))
```

[23]:
```
<AxesSubplot:xlabel='Frequency_0'>
```

Observation from the above graphs is that the larger the size of blue proves cancer cells. We now want to prove this observation with neural network.

[24]:
```
# Neural Network
```

[25]:
```
# import libraries
import tensorflow as tf
from keras.models import Sequential
import pandas as pd
from keras.layers import
Dense data = df_result
data=data.drop(['Frequency_l
oss'], axis=1)
```

[26]:
```
data.head(5)
```

[26]:
```
   Frequency_0  Frequency_1  Frequency_2  Frequency_3  Frequency_4  \
0        17.71        22.43        21.63        12.99        13.79
1        29.63        40.96        32.09        18.30        27.17
2        17.71        29.04        20.17         6.38        15.25
3        22.70        27.04        26.75        18.36        18.65
4        18.29        24.03        23.76        12.55        12.82

   cancer_probability
0                   0
1                   1
2                   1
3                   0
4                   0
```

[27]:
```
#Find the summary
of the data
print(data.descri
be())
print(data.info()
)
```

```
       Frequency_0  Frequency_1  Frequency_2  Frequency_3  Frequency_4  \
count  8080.000000  8080.000000  8080.000000  8080.000000  8080.000000
Mean     29.105000    34.693834    28.154479    20.686357    26.233806
Std       9.664796    10.012796     9.680910     9.710051    10.042060
Min      17.710000    17.710000    16.710000     6.380000     6.650000
25%      21.597500    26.200000    21.700000    12.820000    17.737500
50%      26.815000    32.560000    28.630000    19.760000    24.190000
75%      36.187500    42.650000    37.610000    28.722500    34.250000
Max      46.520000    57.850000    57.580000    49.340000    49.360000


       Frequency_0  Frequency_1  Frequency_2  Frequency_3  Frequency_4
count  8080.000000  8080.000000  8080.000000  8080.000000  8080.000000
Mean     29.105000    34.693834    28.154479    20.686357    26.233806
Std       9.664796    10.012796     9.680910     9.710051    10.042060
Min      17.710000    17.710000    16.710000     6.380000     6.650000
25%      21.597500    26.200000    21.700000    12.820000    17.737500
50%      26.815000    32.560000    28.630000    19.760000    24.190000
75%      36.187500    42.650000    37.610000    28.722500    34.250000
Max      46.520000    57.850000    57.580000    49.340000    49.360000
 ---  ------             -------------- -----
0   Frequency_0        8080 non-null   float64
1   Frequency_1        8080 non-null   float64
2   Frequency_2        8080 non-null   float64
3   Frequency_3        8080 non-null   float64
4   Frequency_4        8080 non-null   float64
5   cancer_probabilit  8080 non-null   int64
```

```
dtypes: float64(5), int64(1)
memory
usage:
378.9 KB
None
```

[28]:
```
# finding correlation between the columns
import seab
as sns import
matp as plt
corr =
data.corr()
sns.he(corr,
xticklabels=c.columns.
values,
yticklabelscolumns.val
ues)
```

[28]: <AxesSubplot:>

	Frequency_0	Frequency_1	Frequency_2	Frequency_3	Frequency_4
count	8080.000000	8080.000000	8080.000000	8080.000000	8080.000000
Mean	29.105000	34.693834	28.154479	20.686357	26.233806
Std	9.664796	10.012796	9.680910	9.710051	10.042060
Min	17.710000	17.710000	16.710000	6.380000	6.650000
25%	21.597500	26.200000	21.700000	12.820000	17.737500
50%	26.815000	32.560000	28.630000	19.760000	24.190000
75%	36.187500	42.650000	37.610000	28.722500	34.250000
Max	46.520000	57.850000	57.580000	49.340000	49.360000

[29]:
```
#Finding correlation between two columns
print(data['Frequency_0'].corr(
data["Frequency_2"]))
print(data["Frequency_3"].corr(data["Freque
ncy_4"]))
```

```
0.9983269439980406
0.9634741808288586
```

```
[30]: # splitting the dataset into training and testing set
      labels=data['cancer
      _probability']
      features     =
      data.iloc[:,0
      :5]
      from   sklearn.model_selection   import
      train_test_split
      X=fe
      atur
      es
      y=np
      .rav
      el(l
      abel
      s)
      X_train,  X_test,  y_train,  y_test  =  train_test_split(X,  y,
      test_size=0.30,␣
       ↪random_state=42)
```

Outcome is the column with the label (0 or 1). The training data set is used to train the mode, meaning find the weights and biases. The test data set is used to check its accuracy. Labels is not an array. It is a column in a dataset. So we use the NumPy np.ravel() function to convert that to an array.

```
[31]: #Normalizing the data
      from sklearn.prepr import
      StandScaler scaler =
      StandardScaler().fit(X_train)
      X_train =
      scaler.trrm(X_train)
      X_test =
      scaler.tram(X_test)
```

```
[32]: def f(x):
          if x<=0:
              return 0
          elif x>0:
              return 1
```

```
[33]: # Building sequential model with activation function ReLU
      from keras.mls import Sequential
      from
      keras.layers
      import Dense
      model =
      Sequential()
      model.add(Dense(5, acon='relu', input_shape=(5,)))
```

```
model.add(Dene(5, actation='relu'))
model.add(Dense(1, activation='moid'))
```

epoch—means how many times to run the model.

[34]:
```
model.compile(loss='binary_
crossentropy',
optimizer='sgd',
metrics=['accuracy'])
model.fit(X_train, y_train,epochs=4, batch_size=1, verbose=1)
```

	Frequency_0	Frequency_1	Frequency_2	Frequency_3	Frequency_4
count	8080.000000	8080.000000	8080.000000	8080.000000	8080.000000
Mean	29.105000	34.693834	28.154479	20.686357	26.233806
Std	9.664796	10.012796	9.680910	9.710051	10.042060
Min	17.710000	17.710000	16.710000	6.380000	6.650000
25%	21.597500	26.200000	21.700000	12.820000	17.737500
50%	26.815000	32.560000	28.630000	19.760000	24.190000
75%	36.187500	42.650000	37.610000	28.722500	34.250000
Max	46.520000	57.850000	57.580000	49.340000	49.360000

[34]:
```
<keras.callbacks.History at 0x7f7bfc4f0d60>
```

[35]:
```
# printing model summary
for layer in model.layers:
    weights = layer.get_weights()
    print(weights)
```

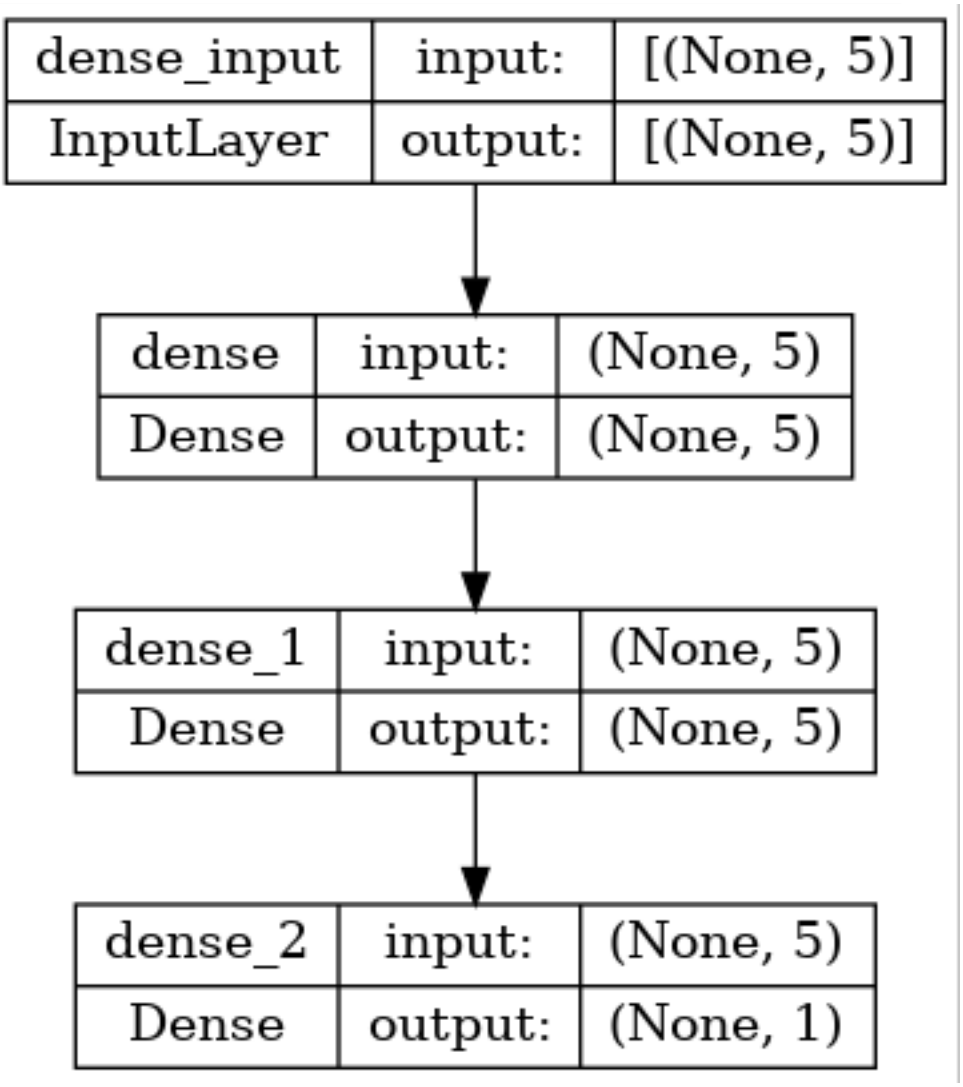
```
[array([[-0.10170265, 0.1833343 , -0.52776855, 0.33172587, -
       1.4749725 ], [-1.1270263 ,  0.6071117 ,  1.0570389 ,
       0.64514685, 2.6104002 ],
       [ 0.43535218, -0.00946396, -0.6625208 , -0.7166759 , -
       1.657455  ], [ 1.1213948 , -0.38926315, -0.9134409 ,
       0.5412085 , -2.2107182 ],
       [-0.5026145 , -0.38953286, 1.0770384 , -0.21556905,
      2.8043725 ]], dtype=float32), array([-2.621502e-01, -
      9.161185e-02, 7.641844e-01,
1.718968e-03,
       1.864222e+00], dtype=float32)]
[array([[ 0.0618682 , -0.2589815 , -0.15412524, -0.6387381 , -
       0.9175897 ], [-0.4491351 , -0.51751715, -0.06322102,
       0.6398838 , -0.46030238],
       [-0.69785   ,  0.22466648, -0.10616295, 0.7081488 ,
       0.8988219 ], [-0.67408925, 0.07617144, 0.3620358 ,
       0.31430197, 0.04528056],
       [-0.25934947, 1.1666613 , -0.8987297 ,  1.4835874 ,
      1.6637093 ]], dtype=float32), array([-0.05316215,
```

```
      0.05949022, 1.7014942 ,  0.11303709,
0.201598
      67]
      ,
      dty
      pe=
      flo
      at3
      2)]
[array([[
       [ 1.8589118]], dtype=float32), array([-1.4945266],
       dtype=float32)]
```

Picture of the layers and their shapes.

[36]:

```
[([[-0.10170265,      0.1833343 , -0.52776855, 0.33172587, -
       1.4749725 ], [-1.1270263 ,  0.6071117 ,  1.0570389 ,
       0.64514685, 2.6104002 ],
       [ 0.43535218, -0.00946396, -0.6625208 , -0.7166759 , -
       1.657455  ], [ 1.1213948 , -0.38926315, -0.9134409 ,
       0.5412085 , -2.2107182 ],
       [-0.5026145 , -0.38953286, 1.0770384 , -0.21556905,
      2.8043725 ]], dtype=float32), array([-2.621502e-01, -
      9.161185e-02, 7.641844e-01,
1.718968e-03,
        1.864222e+00], dtype=float32)]
[[ 0.0618682 , -0.2589815 , -0.15412524, -0.6387381 , -
       0.9175897 ], [-0.4491351 , -0.51751715, -0.06322102,
       0.6398838 , -0.46030238],
       [-0.69785   ,  0.22466648, -0.10616295, 0.7081488 ,
       0.8988219 ], [-0.67408925, 0.07617144, 0.3620358 ,
       0.31430197, 0.04528056],
       [-0.25934947, 1.1666613 , -0.8987297 ,  1.4835874 ,
      1.6637093 ]], dtype=float32), array([-0.05316215,
      0.05949022, 1.7014942 ,  0.11303709,
       [ 1.8589118]], dtype=float32), array([-1.4945266],
       dtype=float32)]
```



Checking the accuracy of our model

[37]:
```
# configuration information on each layer
predictions = (model.predict(X_test) >
0.5).astype("int32") score =
model.evaluate(X_test, y_test,verbose=1)
print(score)
# print(predictions)
```

```
76/76 [==============================] - 0s 969us/step
76/76 [==============================] - 0s 1ms/step - loss:
0.0507 - accuracy:
0.9814
[0.05074836686253548, 0.9814356565475464]
```

Summarize the first 10 cases

[38]:
```
# summarize the first 10 cases
for i in range(10):
        print('%s => %d (expected %d)' % (X_train[i].tolist(),
```

```
        predictions[i], ␣
  ↪y_train[i]))
```

```
[-1.1869507486460331, -0.7904623832113031, -1.1904869606198925,
-1.229971 5157928466, -0.829964460041096] => 1 (expected 1)
[0.642122 2289127739, 0.6077458756860785, 0.6371331806006078,
0.7867738890799132,
0.7514518577061577] => 1 (expected 1)
[0.6421222289127739, 0.3831942281115525, 0.6371331806006078,
0.6464067575227509,
0.39194852602683883] => 1 (expected 1)
[1.795255956700273, 1.3273091552471152, 1.7893509719529772,
1.8632659347719747,
1.3987570255498067] => 1 (expected 0)
[-1.126913
2655116751, -1.2285875978122676, -1.1304971653789255,
-1.0070354833197064, -1.1097994633704276] => 0 (expected
1) [1.79525595 6700273, 2.08379870574263,
1.789350971952978, 1.868426491079223,
2.158593707187814] => 0 (expected 1)
[0.64212222 89127739, 0.5388833704298908, 0.6371331806006078,
0.7217508796085804,
0.6199991159286782] => 1 (expected 1)
[0.9764689022299748, 1.11173957357557, 0.9712142816839245,
1.0334484805663964,
1.1707064356479668] => 1 (expected 1)
[-0.53585459 5344115, -0.8273886541457812, -0.5399079742997485,
-0.6767598796 557949, -0.9594 254936098259] => 1 (expected 0)
[-0.535854595344115, -0.1437536381966686, -0.5399079742997485,
-0.6829525472244932, -0.28324055673930604] => 1 (expected 1)
  -1.229971 5157928466, -0.829964460041096] => 1 (expected 1)
[0.642122 2289127739, 0.6077458756860785, 0.6371331806006078,
0.7867738890799132,
0.7514518577061577] => 1 (expected 1)
[0.6421222 289127739, 0.383194228 1115525, 0.63713318 06006078,
0.6464067575227509,
0.3919485 2602683883] => 1 (expected 1)
[1.795255 956700273, 1.3273091552471152, 1.789350971 9529772,
1.86326 59347719747,
1.3987570255498067] => 1 (expected 0)
[-1.126913
2655116751, -1.2285875978122676, -1.1304971653789255,
-1.0070354833197064, -1.1097994633704276] => 0 (expected
1) [1.79525595 6700273, 2.08379  870574263, 1.78935097
1952978, 1.868426491079223,
2.158593707187814] => 0 (expected 1)
[0.64212222 89127739, 0.5388833704298908, 0.6371331806006078,
0.7217508796085804,
0.6199991 159286782] => 1 (expected 1)
[0.97646890 22299748, 1.111739573 57557, 0.97121 42816839245,
1.0334 484805663964,
1.1707064356479668] => 1 (expect ed 1)
```

```
      [-0.53585459 5344115, -0.8273886 541457812, -0.5399079742997485,
      -0.6767598796 557949, -0.9594 254936098259] => 1 (expected 0)
      [-0.535854595344115, -0.1437536 381 966686, -0.539907 9742997485,
      -0.68 29525472244932, -0.28324055673930604] => 1 (expected 1)
       [-0.25934947, 1.1666613 , -0.8987297 ,  1.4835874 ,  1.6637093
]],             dtype=float32), array([-0.05316215, 0.05949022,
1.7014942 ,  0.11303709,
      0.
              1.8589118]], dtype=float32), array([-1.4945266],
             dtype=float32)]

      [-1.1869507486460331, -0.79046 23832113031, -1.1904869606198925,
      -1.229971 5157928466, -0.82996 4460041096] => 1 (expected 1)
      [0.642122 2289 127739, 0.607 7458756860785, 0.6371331806006078,
      0.7867738890799132,
      0.751451857 7061577] => 1 (expected 1)
      [0.6421222289127739, 0.3831942 281115525, 0.63713 31806006078,
      0.6464067575227509,
      0.39194852602683883] => 1 (expected 1)
      [1.79525595670 0273, 1.3273091552471152, 1.7893509719529772,
      1.8632659347719747,
      1.3987570255498067] => 1 (expected 0)
      [-1.126913
      2655116751, -1.22858 75978122676, -1.1304971653789255,
      -1.0070354833197064, -1.10979 94633704276] => 0 (expected
      1) [1.79525595 6700273, 2.08379870574263,
      1.789350971952978, 1.868426491079223,
      2.158593707 187814] => 0 (expected 1)
      [0.64212222 89127739, 0.5388833704298908, 0.6371331806006078,
      0.7217508796085804,
      0.6199991159286782] => 1 (expec ted 1)
      [0.976 4689022299748, 1.11173957357557, 0.9712142816839245,
      1.0334484805663964,
      1.17070643 56479668] => 1 (expected 1)
      [-0.53585459 5344115, -0.8273886541457812, -0.5399079742997485,
      -0.6767598796 557949, -0.9594 254 936098259] => 1 (expected 0)
      [-0.5358545 95344115, -0.1437536381966686, -0.5399079742997485,
      -0.68295254 72244932, -0.283240556 73930604] => 1 (expected 1)
        -1.229971 5157928466, -0.8299 64460041096] => 1 (expected 1)
      [0.642122 2289127739, 0.6077458756860785, 0.6371331806006078,
      0.7867738890799132,
      0.7514518577061577] => 1 (expected 1)
      [0.6421222289127739, 0.3831942281115525, 0.6371331806006078,
      0.6464067575227509,
      0.39194852602683883] => 1 (expected 1)
      [1.795255956700273, 1.3273091552471152, 1.7893509719529772,
      1.8632659347719747,
      1.3987570255498067] => 1 (expected 0)
      [-1.126913
      2655116751, -1.2285875978122676, -1.1304971653789255,
      -1.0070354833197064, -1.1097994633704276] => 0 (expected
      1) [1.79525595 6700273, 2.08379870574263,
```

```
      1.789350971952978, 1.868426491079223,
      2.158593707187814] => 0 (expected 1)
      [0.64212222 89127739, 0.5388833704298908, 0.6371331806006078,
      0.7217508796085804,
      0.6199991159286782] => 1 (expected 1)
      [0.9764689022299748, 1.11173957357557, 0.9712142816839245,
      1.0334484805663964,
      1.1707064356479668] => 1 (expected 1)
      [-0.53585459 5344115, -0.8273886541457812, -0.5399079742997485,
      -0.6767598796 557949, -0.9594 254936098259] => 1 (expected 0)
      [-0.535854595344115, -0.1437536381966686, -0.5399079742997485,
      -0.6829525472244932, -0.28324055673930604] => 1 (expected 1)
       [-0.25934947, 1.1666613 , -0.8987297 ,  1.4835874 ,  1.6637093
]],             dtype=float32), array([-0.05316215, 0.05949022,
1.7014942 ,  0.11303709,
      0.20159867],
```

CONCLUSION: We have proven that the higher the loss of frequency implies Cancerous cells.This make our research study proven to be working correctly and producing expected results

D: LIST OF PUBLICATIONS

D.1 Khanna, M. K., & Murei, K. G. (2019). Effect of the Oscillating Electric Field Due to the Oscillating Electric Dipole on Raman Lines. *East European Journal of Physics*, (4), 47-57. https://doi.org/10.26565/2312-4334-2019-4-05.

D.2 Murei, K.G., Khanna, M.K., (2020) - Role of Modified Vibrational Raman Spectra in Detecting Cancerous Cells, *International Journal of Recent Research Aspects* ISSN: 2349 7688, Vol. 7, pp. 9-13.