Non-invasive Diabetes Detection by Raman Spectrometer

BY

Parawee Tangkiatphaibun 62011185
Pholchanok Udomtanasub 62011205

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF BACHELOR OF
ENGINEERING IN BIOMEDICAL ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2022

SCHOOL OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
PROJECT CERTIFICATE

Project Title Non-invasive Diabetes Detection by Raman Spectrometer

Student Name Parawee Tangkiatphaibun 62011185
Pholchanok Udomtanasub 62011205

Degree Bachelor of Engineering in Biomedical Engineering

Project Advisor Signed: ______________________________
(Assoc. Prof. Dr. Wibool Piyawattanametha)

Committee Signed: ______________________________
(Assoc. Prof. Dr. Matthew Paul Gleeson)

Committee Signed: ______________________________
(Asst. Prof. Dr. Pimkhuan Hannanta-an)

Committee Signed: ______________________________
(Asst. Prof. Dr. Jamie Alexander O'Reilly)

Committee Signed: ______________________________
(Dr. May Phu Paing)

Head of Department Signed: ______________________________
(Assoc. Prof. Dr. Sarinporn Visitsattapongse)

Project Title	Non-invasive Diabetes Detection by Raman Spectrometer
Student Name	Parawee Tangkiatphaibun 62011185
	Pholchanok Udomtanasub 62011205
Degree	Bachelor of Engineering in Biomedical Engineering
Project Advisor	Wibool Piyawattanametha
Academic Years	2022

ABSTRACT

Non-communicable diseases (NCDs) are non-infectious diseases that do not spread from person to person and are caused by undesired behaviors such as excessive alcohol use, smoking, chronic stress, and the consumption of foods rich in sugar, sodium, or cholesterol. The majority of Non-Communicable diseases are chronic obstructive pulmonary disease, diabetes (type 2), hypertension, etc., with diabetes being one of the diseases that has been an issue in the previous five years. According to the International Diabetes Federation (IDF), the number of diabetes patients worldwide in 2017 was approximately 425 million, with an estimated increase to 629 million by 2045.

Diabetes is a condition caused by a malfunction of the insulin enzyme, which causes abnormally high blood glucose levels. It can develop complications such as heart disease, chronic renal disease, nerve damage, and other foot, vision, and hearing issues. Due to the nature of diabetes, blood glucose testing can be used to both diagnose and monitor it. Blood glucose testing is most reliable and accurate when performed with a blood sugar meter, which entails pricking the patient's finger (invasive approach). It is a painful procedure that carries an infection risk. Several non-invasive blood glucose testing technologies are currently available. However, most of them are erroneous. Therefore, it is preferable to use non-invasive glucose monitoring methods that are dependable and accurate. We believe Raman spectroscopy to be the solution to this issue.

Raman spectroscopy is a technique for altering the laser's wavelength through interaction with a sample. A spectrometer can record the shifts, which offer information about the substance. Costing at least $12,000 US, Raman spectrometers are prohibitively expensive for the majority of institutions and universities. We will instead construct the Raman DIY, which costs less than $3,000 US. In this senior project, we aimed to build a Raman spectrometer to measure glucose on SERS substrates and calibrate the results using a benchmark experiment conducted by one of our team members at MSU. Unfortunately, we were unable to obtain satisfactory results as the

system failed to align properly, and we were unable to identify the fingerprint of any substrate or glucose powder due to limitations. Despite the lack of success in obtaining results, we learned valuable lessons about the challenges of building and aligning a Raman spectrometer. We also gained insight into the importance of proper calibration and the need for careful consideration of experimental limitations.

In conclusion, while we were unable to achieve our desired outcome, the project provided a valuable learning experience that will benefit us in our future endeavors in scientific research and development. We hope that our experience will inspire future students to pursue similar projects, learn from our challenges and mistakes, and continue to explore the exciting possibilities of Raman spectroscopy in the field of biomedical research.

ACKNOWLEDGEMENTS

We would like to express my deepest gratitude to my advisors, Assoc. Prof. Wibool Piyawattanametha and Asst. Prof. Zhen Qiu, for their unwavering support, guidance, and expertise throughout this research project. Their invaluable insights and feedback have been critical to the success of this work. We would also like to thank our senior Ph.D. students, Pasin Kuncharin and Aniwat Juhong, for their collaboration and helpful discussions during this research. Also, our senior M.D. students Preeyarat Rithcharung and Nitipon Navaitthiporn's contributions have been invaluable to the project's success.

Parawee Tangkiatphaibun
Pholchanok Udomtanasub

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Symbols/Abbreviations	**Terms**
a-D-glucose	alpha-D-Glucose
A1C	Hemoglobin A1C
AI	Artificial Intelligence
ANN	Artificial neural network
BUN	Blood Urea Nitrogen
CCD	Charged Coupled Device
CCH	Ethynyl radical
CH	Hydrocarbons
CLS	Classical Least Squares
CNN	Convolutional Neural Network
CMOS	Complementary Metal Oxide Semiconductor
CV	Cross Validated
d	Line spacing
D-glucose	Dextrorotatory glucose
DIY	Do It Yourself
DRS	Diffuse Reflectance Spectroscopy
dL	Deciliter
Dm	Diameter
FNN	Feedforward Neural Network
GAN	Generative Adversarial Network
GFR	Glomerular Filtration Rate
ICG	Indocyanine green
InGaAs	Indium Gallium Arsenide
InSb	Indium antimonide
IR	Infrared
KCl	Potassium Chloride
kNN	K-Nearest Neighbor
OCH	2-Oxoquinoline
OCT	Optical Coherence Tomography
OLS	Ordinary Least Squares
MAE	Mean Absolute Error
Mg	Milligram
MIR	Mid-infrared
MIT	Massachusetts Institute of Technology
mL	Milliliter

Mm	Millimeter
mM	Millimole
mW	Milliwatt
n	Diffraction order
NaCl	Sodium chloride
NIR	Near-infrared
Nm	Nanometer
NPs	Nanoparticles
P	Predicted
PAS	Photoacoustic spectroscopy
PCR	Principal Component Regression
PLS	Partial Least Square
ResNet	Residual Network
RMSECV	Root Mean Standard Error for Calibration
ROI	Region of Interest
SERs	Surface Enhanced Raman scattering
SiO2	Silicon Dioxide
SNR	Signal to Noise Ratio
uM	Micromole
U-Net	Convolutional networks for biomedical image
UV	Ultraviolet
USD	US Dollar

CHAPTER 1

INTRODUCTION

The purpose of this chapter is to provide an overview of core ideas as well as to contextualize the inspiration for the research. After that, a discussion of the motivations for the examination of the project and the objectives, followed by an overview of the project. Finally, an overview of the dissertation is given on a per-chapter basis.

1.1 Diabetes

Diabetes, a metabolic disorder characterized by a rise in blood glucose concentration, is one of the most prevalent diseases. The International Diabetes Federation estimated that 382 million people had diabetes in 2013 [1]. There are four distinct types of diabetes, with the first being an autoimmune disorder. The immune system attacks the pancreas, where insulin is produced; 10% of the population has this form of diabetes. The second type of diabetes is insulin resistance, in which the body's resistance to insulin causes sugar to become trapped in blood vessels. Gestational Diabetes is the third type of diabetes, characterized by high blood pressure during pregnancy. The placenta produces hormones that inhibit insulin. Prediabetes is the final type of diabetes. If your blood sugar levels are above average but not high enough to be diagnosed with type 2 diabetes, you have prediabetes. General diabetes symptoms include increased appetite, thirst, weight loss, frequent urination, impaired vision, extreme fatigue, and non-healing wounds [2].

Glucose concentration in the blood can be used to monitor diabetes. Currently, the standard method for measuring blood glucose is minimally invasive blood withdrawal for lab analysis or any biosensor devices for glucose detection [3]. This procedure can be painful or stressful for patients with hemophobia. Now, many people are interested in noninvasive glucose detection because it provides a painless and convenient method for blood glucose monitoring.

1.2 Raman spectroscopy

Raman spectroscopy is a method for altering the wavelength of a laser by interacting with a material, a process known as Raman scattering. Raman scattering is an optical process in which incoming excitation light interacts with a sample, resulting in lower-energy scattered light due to the vibrational modes of the chemical bonds in the sample [4]. A spectrometer can then record the shifts, which provides information about the material. Raman is therefore suitable for noninvasive blood glucose detection using the finger.

However, commercial Raman spectroscopy is too expensive to study and take advantage of non-invasive blood glucose testing. According to cost, we use the Raman DIY model instead. Also, since the Raman spectroscopy principle is more observable, we can study it in greater detail. The Raman DIY utilizes the same principle as commercial Raman spectroscopy, but with more accessible, adjustable components that we can use and at a much lower cost, with costs below 3,000 USD compared to the starting price of 12,000 USD for commercial-grade Raman spectroscopy [5] [6].

1.3 Objective

The detection of clinically-relevant analytes in the blood tissue matrix of human subjects using near-infrared Raman spectroscopy is the long-term objective of this project. The shorter-term research focus of this project is intended to develop a non-invasive blood glucose detector for the treatment of diabetes. Using Raman spectroscopy, we can examine the Raman spectrum from a fingertip to determine if glucose is present in the bloodstream. This optical method offers non-contact, painless, and simultaneous analysis of numerous analytes without the need for any sample preparation. We anticipated finding glucose-specific Raman peak signals that correspond to glucose concentration. During the experiment, we collected multiple samples of the Raman spectrum. We can predict glucose concentration [7] using mathematical methods such as linear regression or PLS regression analysis. is on glucose readings for diabetes patients.

CHAPTER 2

THEORY

This chapter will provide some context for the themes of glucose, creatinine, and urea, which will be discussed throughout the rest of this thesis. Glucose is one of the three blood analytes of clinical importance that will be investigated. It has been determined that concentrations in the millimolar range (mM), which are within the physiological range, are ideal for noninvasive optical technologies. This article provides a concise summary of existing noninvasive optical techniques, with an emphasis on the benefits of each method individually. The previous research that has been done on this subject is reviewed, including the creation of a sensitive instrument, the analysis of chemical concentrations in used human serum and whole blood, and the transcutaneous investigation with the participation of human volunteers. The algorithm of machine learning and deep learning are used for calculation and can be of benefit in the diagnosis of diabetes.

2.1 Blood analytes identified in this dissertation

The diagnostic value of blood analyte levels cannot be overstated. The four most common processes in clinical chemistry are the collection of patient blood, the centrifugation of blood to obtain serum or plasma, the inclusion of specific reagents for chemicals whose amounts are of interest, and the measurement of those concentrations using spectrophotometry. Despite the fact that these tests may be used to detect a broad variety of drugs and that they have advanced to the point where they can be done fast and properly on little volumes of blood, they frequently need to go through a number of processes in a lab before they can be examined. Therefore, from a practical aspect, doctors cannot get fast outcomes using these methods. It would be highly useful to develop a technique to quantify clinically significant chemicals that does not require reagents and does not necessitate drawing blood, as existing clinical chemistry procedures still include blood collection. Analyte concentrations in the blood can be determined with noninvasive blood analysis without requiring direct contact with the blood sample or the test subject. Optical technology is the most applicable modality to consider. Thanks to optical technology, samples may be measured without ever being handled. There is no need for sample preparation, and a wide range of analytes may be examined all at once. In the following discussion, we will utilize the analytes glucose, creatinine, and urea as examples.

2.1.1 Glucose

Each and every one of your cells relies on the carbohydrate glucose for its major source of energy. It can also enter the body after being digested alongside other large carbohydrate molecules or after being absorbed directly in the form of glucose [8]. Although insulin plays the most important role, other hormones also contribute to regulating its levels. When glucose levels in the blood rise, the pancreas produces insulin to lower them. Through this mechanism, glucose may enter cells. Without adequate insulin, glucose builds up outside the cells and into the bloodstream, where it triggers cellular catabolism (the breakdown of fat) [9].

According to the majority, diabetes mellitus is the most prevalent glucose-related condition. Type 1 diabetes (also known as insulin-dependent diabetes) and type 2 diabetes (also known as adult-onset diabetes) are the two most prevalent forms of the disease. Insulin secretion is nonexistent in type 1 diabetes, in which the immune system attacks and destroys the pancreatic beta-cell islets. Individuals with type 1 diabetes must take exogenous insulin daily. Over ninety percent of all cases of diabetes are type 2, in which the patient is able to produce insulin but is unable to utilize it effectively. This is because insulin no longer regulates blood sugar levels effectively in individuals with type 2 diabetes. Diabetes may not be fatal on its own, but if left untreated, it can lead to a variety of severe complications, including heart disease and stroke, vision loss, renal failure, nerve damage, and even amputation. These issues have had a significant economic impact, placing a strain on available funds. With vigilant surveillance and management of blood glucose levels, the risk of complications can be minimized. As a consequence, the American Diabetes Association urges all diabetics to conduct comprehensive and routine glucose monitoring. Individuals with type 2 diabetes should independently monitor their blood sugar levels at least once per day. Nonetheless, a number of studies and articles have argued that more frequent measurement could be advantageous. Patients with type 1 diabetes cannot produce insulin, so they must rely on self-injections three or four times per day (often before each meal and before bed) to maintain normal blood sugar levels. Normal glucose concentrations in the blood or plasma of healthy humans range between 45 and 180 mg/dL (2.5 and 10.0 mM) [10]. In addition to the age, gender, and duration of time since the subject's last meal, additional factors affect the glucose concentration. If one's blood glucose level is consistently above the normal range, they have hyperglycemia, and if their blood glucose level is consistently below the normal range, they have hypoglycemia. A concentration of 1000 mg/dL (56 mM) has been reported. Enzymatic technologies are utilized in clinical chemistry to generate the most accurate glucose levels available. The reaction of f3-D-glucose with glucose oxidase produces gluconic acid and hydrogen peroxide, which can be used as a surrogate for glucose concentration. After a reaction between hydrogen peroxide and an oxygen acceptor catalyzed by peroxidase has been completed, the color produced by the reaction is detected using spectrophotometry. A affordable price is one of the most alluring aspects of the glucose oxidase method. Using glucose oxidase and oxygen electrodes is a further beneficial technique. Using an oxygen-sensing electrode, the reaction between glucose and oxygen is monitored, and the hydrogen peroxide it produces is neutralized with ethanol and iodide. It is possible to determine the glucose concentration with high precision if the metabolic rate is first determined. Due to its precision, linearity, and absence of significant interferences, this method has been extensively adopted in the context of reference methods, with some modifications. Each of the two possible anomeric forms of glucose in aqueous solution has its own Raman spectrum. Use caution when producing in vitro samples or comparing them to the regression vector when employing the balanced form. Both a-D-glucose and its anomeric balanced form are depicted in Figure 2-1 at room temperature. Combining a- and J3-D-glucose produces the anomeric balanced form. When immersed in a water solution, the a shape doesn't instantaneously transform into the balanced form.

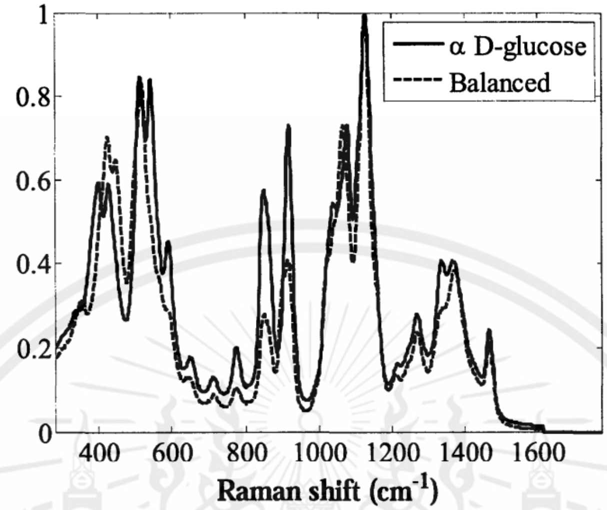


Figure 2-1 a-D-glucose (solid) and anomeric balanced D-glucose (dashed) in DI water Raman spectra. Shih, W.-C. "Quantitative Biological Raman Spectroscopy" in Quantitative Biological Raman spectroscopy for non-invasive blood analysis, vol.12, 2007.

2.1.2 Testing and Diagnosis of Diabetes

a. The A1C test measures glycated hemoglobin. The patient is not required to fast before undergoing this blood test. It is able to provide your average blood sugar level for the past three months. It measures the amount of glucose-attached hemoglobin in the blood, a protein found in red blood cells and responsible for transporting oxygen throughout the body. If you have a greater quantity of sugar-bound hemoglobin, your blood sugar levels will also be elevated. If your A1C level is greater than 6.5% in at least two separate examinations, you have diabetes. Levels of A1C between 5.7% and 6.4% are considered indicative of prediabetes. A score of less than 5.7 is considered average [11].
b. A measurement of the fasting blood glucose level. A blood sample will be obtained after the patient has fasted for at least eight hours (overnight). A fasting blood glucose level of 100 mg/dL or less is considered to be within the normal range. While fasting, a blood glucose level between 100 and 125 mg/dL is considered prediabetes. If your blood sugar levels on two separate occasions are greater than 126 mg/dL, diabetes should be considered [12].
c. Evaluation of one's tolerance to glucose. In order to complete this test, you will need to abstain from food and water for the whole preceding night. After that, you are given a sugary drink to consume, and your doctor will check your blood sugar every 20 minutes

for the next two hours. Less than 140 milligrams of glucose per deciliter is considered to be normal. When the test is repeated after two hours, diabetes is diagnosed when the result is more than 200 mg/dL. A blood sugar level that falls between 140 to 199 mg/dL is considered to be prediabetes [13].

d. Random blood draws. A sample of blood will be taken on an arbitrary basis. Diabetes is diagnosed when a person has a blood sugar level of 200 milligrams per deciliter (mg/dL) or above, regardless of the time of their most recent meal [14].

2.1.3 Creatinine

The waste product creatinine is a result of muscle action and is subsequently transported throughout the body via the circulation. The blood creatinine concentration of a healthy adult male is 1 mg/dL (88 uM). Serum creatinine levels are used alongside demographic information like age, gender, and race to determine a patient's glomerular filtration rate (GFR), a proxy for the kidneys' ability to filter blood and generate urine. The glomerular filtration rate (GFR) quantifies how well one's kidneys filter blood. Kidney health may be estimated from serum creatinine concentrations. The National Kidney Foundation recommends monitoring serum creatinine levels to assess kidney function. A great deal of variation in this measurement, however, can be attributed to the fact that various laboratories use different techniques. Creatinine was found in the water and its Raman spectra is shown in Figure 2-2 [15].

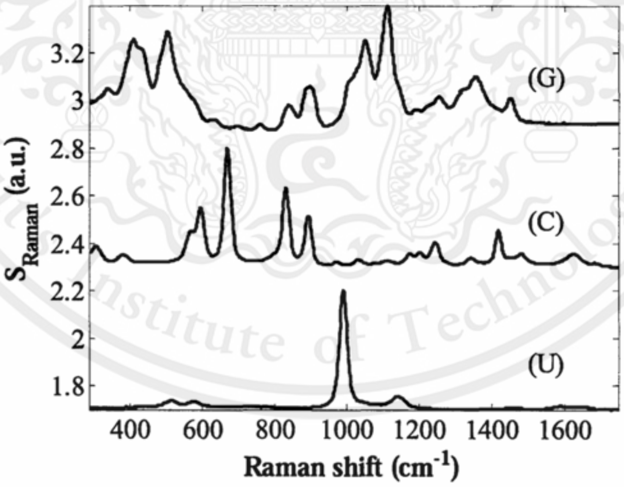


Figure 2-2 Raman spectra of D-glucose (G), creatine (C), and urea (U) with their anomeric excesses removed. Quoted from "Quantitative Biological Raman Spectroscopy" by Shih, W.-C. in Quantitative Biological Raman Spectroscopy for Non-Invasive Blood Analysis. 2007 Essays, Volume 12.

2.1.4 Urea

Urea is produced in the liver by a process called deamination, which breaks down protein (removal of nitrogen from amino acids). The measurement of plasma urea is the form of renal function testing that is performed most commonly. This is due to the fact that urea does not spend a significant amount of time circulating in the bloodstream; rather, it is processed by the kidneys before being eliminated in the urine. The rate of filtration as well as its efficacy both decrease when kidney function declines, leading to an increase in the concentration of urea in the blood. Urea concentration is a useful tool for screening for renal disorders and monitoring the course of these conditions Henry et al.[10]. Throughout the course of history, the nitrogen content of urea was typically reported rather than the real urea concentration. Because of this, measurements of urea levels in the blood are typically referred to as "blood urea nitrogen" concentrations (BUN). Multiplying a blood urea nitrogen concentration by the conversion factor 2.14 yields a urea concentration as the output of the conversion. Plasma urea nitrogen concentrations typically range from 8 to 23 mg/dL (2.9 to 8.2 mM), while urine urea nitrogen concentrations typically range from 60 to 90 mg/dL (21.4 to 32.1 mM). The presence of low quantities of urea in urine may be an indication of malnutrition as well as dysfunction in the kidneys, whereas high levels may be an indication of excessive consumption of protein. Figure 2-2 depicts the Raman spectrum of urea that was measured when it was dissolved in water.

2.2 An analysis of the non-invasive optical methods that are currently in use for glucose detection.

In the following, we will provide a basic overview of direct and indirect optical techniques that have been used for non-invasive measurements of glucose and that have been published in the literature that has been peer-reviewed during the previous two decades. These techniques were developed. Direct methods rely on glucose's unique inherent molecular features, such as optical absorption, Raman scattering, and optical rotation. These are all examples of direct methods. The index of refraction and the scattering coefficient are two examples of indirect techniques. These approaches are based on changes that hyperglycemia has caused to occur in physiological or physical factors. Omar Khalil [16, 17] has penned two in-depth analyses that cover the time period from 1989 to 2003. Our objective here is not to carry out a comprehensive analysis of the relevant literature; rather, we want to acquaint the reader with each approach. The most important fundamentals and characteristics of each method will be covered first, followed by a review of one or more representative studies. These studies will be chosen based on their significance to the industry as well as the presence of adequate data to back up their claims. There are several different types of study designs represented in the cited literature. Overall, it is absolutely necessary that the glucose concentrations in the samples used for calibration span a wide range of values. This criteria can be accomplished for in vitro research either by the diversity of the original sample or by adding a glucose stock solution to the samples. Both methods are acceptable. For in vivo research, one of three different experimental protocols is commonly followed: oral glucose or meal tolerance [17], time-randomized, or glucose clamping [18]. Patients suffering from diabetes have taken part in a number of the studies in order to broaden the scope of the range of glucose change. There is a possibility that spurious correlations may have an effect on the calibration results that are based on glucose tolerance protocols [19]. This is a cause for concern. Since more stringent experimental designs can be implemented, the time-randomized and glucose clamping methods are the superior options. On the other hand, they tend to be substantially more expensive and demand a higher level of compliance from the test respondents. In the interest of brevity, the outcomes of a selection of in vitro and in vivo research that made use of Near-infrared Range (NIR) absorption or Raman spectroscopy, the methods that were utilized the majority of the time, are documented in Tables 2-1 and 2-2. Estimates of the amount of error are presented in the tables with either the letters "CV" or "P" in parenthesis. These letters stand for cross-validated and predicted results, respectively.

2.2.1 Direct approaches

2.2.1.1 Absorption spectroscopy in the Near-infrared Range (NIR)

NIR absorption spectrum analysis is one of the most desirable techniques for in vivo glucose monitoring due to the low cost of the equipment required, the high SNR of the data, and the extent to which NIR light penetrates biological tissue. This is because, at different wavelengths within the NIR spectral range of 800-2500 nm (12500-4000 cm-1), NIR light can penetrate biological tissue to depths of millimeters to centimeters. This is a consequence of the widespread use of NIR absorption spectroscopy. Molecules undergoing IR-active overtone or combination vibrational transitions will absorb light at these wavelengths during their transition. Changes in the dipole moment of molecules are associated with visible infrared spectrum transitions. Longer-wavelength mid- to far-infrared light cannot penetrate tissue as deeply as NIR light. This is because NIR light has shorter wavelengths. This is because the transition between fundamentals and overtones or combinations is much stronger than the transition between overtones and combinations. Due to factors such as the absorption of light by hemoglobin and the dispersion of light, shorter wavelengths provide less penetrability than longer wavelengths. This makes the near-infrared (NIR) spectrum a natural fit for more in-depth studies of living organisms and their tissues. The near-infrared (NIR) spectrum has been studied for its possible use in glucose monitoring applications without requiring invasive procedures. Overtones of OH (10654 cm-1) and CH (8881 cm-1) can be found in the region with shorter wavelengths (14286-7300 cm-1). A combined band of OH and CH atoms (6510 cm-1) and a CH overtone (5924 cm-1) make up the spectral region 6500-5500 cm-1. CCH and OCH deformation bands may be seen at 4423 and 4300 cm-1, respectively, while CH stretch combination bands can be found in the longer wavelength region (5000-4000 cm-1). These ring deformation bands may exhibit greater selectivity for glucose than other regions. Water, fat, and hemoglobin are all potential interferents that must be taken into account when using several wavelengths [16].

The absorption rate of glucose in healthy individuals is many orders of magnitude lower than the absorption rate of glucose by water, the predominant background absorber in tissue.

Overtone and combination bands in molecules are often rather broad, leading to a great deal of spectral overlap. This means that multivariate calibration is necessary for the extraction of quantitative analyte-specific information. A tungsten-halogen lamp is often utilized as the broadband light source in NIR absorption apparatus, together with intermediate optic elements, a Fourier-Transform spectrometer, and an InSb detector. When a Fourier-transform spectrometer is not available, a grating and an array of InGaAs photodiodes can be used instead. Recent work by Olesbergs [20] demonstrated that the signal-to-noise ratio (SNR) may be significantly improved by employing a tunable diode laser rather than lamp illumination. Both the transmission and reflectance modes have been successfully implemented, with fiberoptic probes being widely utilized. The reflection mode [21] is advantageous since it just requires one end of the instrument to function. If the source and detector are on the same side of the measurement location, then the results will be the same. This simplifies the development of optical probes and expands the range of tissues that may be studied. The sample shape and available space are both considerations since the tissue site needs to be positioned so that it is "sandwiched" between the source and the detector for the transmission model [22] to operate. However, the optical path length may be more accurately estimated while operating in the transmission mode as opposed to the reflection mode. Possible improvement in accuracy as a result of this. In early in vitro investigations, glucose-adulterated blood or plasma samples were used to demonstrate NIR absorption spectroscopy's [23, 24] capacity to detect glucose. The findings were significantly better when plasma was used instead of blood in the second attempt at the experiment. This was probably due to the higher absorption and scattering loss provided by red blood cells. After ten minutes, the glucose concentration in the optically probed volume mirrors the plasma blood glucose concentration, as determined by Marbach et al.[21] in their research of in vivo diffuse reflectance spectra recorded from human inner lips. This delay has significant implications for the design of a non-invasive technique, as the spectroscopic signal is generated in large part by glucose molecules existing in the tissue interstitial fluid. Long-term accuracy of an NIR calibration method was analyzed by Samann et al.[25], and the resulting wide error range highlighted the need for highly stable equipment and algorithms that can account for variations in patient physiology. Maruo et al.[26] used a novel calibration model that was simulated computationally to provide accurate predictions regarding the concentration of glucose within a few hours of completing the calibration phase. Spectral residuals equivalent to the glucose net analyte signal were found by Olesberg et al.[27] by removing main components from spectra recorded during a hyperglycemic phase that were taken during a fasting condition. A technique called glucose clamping allowed us to achieve our goal. Because of this, it is reasonable to assume that the spectral signature of glucose is reflected in the NIR data. Table 2-1 below summarizes the results of these and additional well chosen NIR investigations.

Table 2-1 Glucose measurements using NIR absorption spectroscopy						
In vitro						
Author	**Spectral range [cm^{-1}]**	**Mode**	**Sample**	**# Sample**	**Protocol**	**Approx. Error [mM]**
Haaland et al.[11]	6600-4250	transmission	whole blood	various number from 4 individuals	spiked	2(CV)
Small et al [12]	5000-4000	transmission	bovine plasma	69	spiked	0.4-0.5 (P)
In vivo						
Author	**Spectral range [cm^{-1}]**	**Mode**	**Sample**	**# Sample**	**Protocol**	**Approx. Error [mM]**
Robinson et al.[22]	6600-4350	transmission	fingertip	1 diabetic	tolerance	1.1 (CV)
Marbach et al.[21]	9000-5500	reflection	inner lip	133	time-randomized	2.5-3 (CV)
Burmeister et al.[28]	7000-5000	transmission	tongue	5 diabetic	time-randomized	>3 (P)
Samann et al.[25]	12500-7407	reflection	fingertip	10 diabetic		3.1-35.9 (P)
Maruo et al.[29]	6667-5556	reflection	forearm	2 healthy	tolerance	1-2 (CV)
Maruo et al.[26]	6579-5882	reflection	forearm	5 healthy, 8 ICU	tolerance	1.5 (P)
Olesberg et al.[27]	5000-4000	transmission	rat back	1	clamp	2.2 (P)

Table 2-1 Glucose measurements using NIR absorption spectroscopy by Shih, W.-C. “Quantitative Biological Raman Spectroscopy” *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

2.2.2 Absorption spectroscopy within the mid-infrared (MIR) spectrum

To reduce spectral overlap when measuring the fundamental vibrations of glucose, extended wavelength mid-infrared radiation in the range of 2.5-25 gm (4000-400 cm-1) can be utilized. MIR tissue absorption spectra exhibit discrete peaks, thereby enhancing molecular specificity. However, water absorbs many orders of magnitude more light in this spectral range than in the NIR region. As a result, the extent to which light can penetrate biological tissue is drastically diminished and is on the order of 10 to 100 pm. Therefore, it is challenging to obtain a representative sample of glucose-containing fluid in in vivo applications. The fundamental components of a MIR absorption instrument are identical to those of an NIR absorption instrument; however, the light source, optics, and detector have been modified for optimal performance in the mid-infrared spectrum. The preponderance of reported MIR research has been conducted in vitro, including studies employing desiccated samples [30] to reduce water absorption. Despite the fact that MIR absorption spectroscopy can extract clinical information from blood or serum [31], the technique's penetration depth is so severely restricted that it can only be used for measurements in the superficial layer of tissue. Heise et al.[32] attempted to measure glucose in the buccal mucosa and determined that there is no conclusive evidence that glucose can be detected.

2.2.2.1 Near-Infrared (NIR)

Using any particular wavelength of light, Raman spectroscopy permits the investigation of the fundamental vibrational states of molecules. During spontaneous Stokes Raman scattering, the wavelength of the light scattered from a molecule increases. This change in wavelength corresponds to vibrational transitions in the molecule, which are represented by the difference in energy. Raman scattering and infrared absorption each have their own set of selection criteria, but they provide identical information about molecules. Infrared-active vibrational transitions have an effect on the molecule's dipole, whereas Raman-active vibrational transitions have an effect on the molecule's polarizability. More comparisons can be made between NIR absorption spectroscopy and Raman spectroscopy in terms of the data properties. The NIR spectroscopy data have a high SNR, but their features are broad and difficult to distinguish. Raman spectroscopy can produce data with strong spectral features, but feeble signals result in data with a low signal-to-noise ratio. Nonresonant spontaneous Raman scattering will be the subject of the subsequent discussion. Since Raman shifts are independent of excitation wavelength, NIR radiation (typically 785 or 885 nm) is chosen for deeper penetration into biological tissue and to eradicate the laser-induced fluorescence background. This enables the use of NIR radiation. Despite this, NIR tissue fluorescence is still several orders of magnitude greater than physiological glucose Raman signals. Despite the fact that Raman spectra have sharp, distinct peaks, the background fluorescence signal is often a significant drawback of this technique due to the spectrum variation it exhibits and the associated detector noise. Even though Raman spectra have clear, distinct peaks, this is the case. Multivariate calibration is required in order to obtain precise information about the glucose concentration. An NIR Raman instrument consists of a laser source, optic elements for light delivery and collection, and one of the following: a Fourier-Transform spectrometer with an InGaAs detector, a grating with a photodiode array, or a grating with a cooled CCD detector. These components are contained within an enclosure that contains the laser source. Due to the inherent weakness of the Raman signal, collection optics are typically designed with great care to account for its characteristics. The vast majority of the time, a microscope objective is used to collect light, but a paraboloidal mirror has also been used to enhance the quantity of light collected [33]. Both filtered blood serum and blood serum, as well as whole blood, have been used for in vitro measurements [33, 34]. Rohleder et al. [35] used ultrafiltration to remove macromolecules that produce a high Raman background and therefore impair measurement accuracy. This resulted in a substantial increase in the precision of serum measurements. Although the results from whole blood have a larger margin of error than those from filtered or unfiltered serum, they are clinically acceptable [29]. Lambert et al. [36] conducted measurements in human aqueous humor to simulate eye measurements. Concerns regarding laser dosimetry may render impossible certain in vivo applications. To date, only two groups [37, 38] have reported conducting effective in vivo experiments on human participants. Enejder et al.[37] from our lab published glucose concentration measurements in 17 healthy volunteers who were administered an oral glucose tolerance protocol. These individuals were not diabetic. As evidenced by the findings based on a single volunteer and multiple participants, a glucose-specific calibration model was almost certainly developed. Chaiken et al. [38] were able to acquire complete blood Raman spectra in vivo and describe their findings by utilizing tissue modulation. In the subsequent stages, glucose concentrations were determined by analyzing a particular spectral range of the complete blood spectrum. When applied to data from other sources, a calibration model derived from a single individual was capable of producing accurate predictions.

In the following Table 2-2, you will find a summary of these and a few additional selected Raman spectroscopy investigations. As a result of the vast majority of published accounts making use of the same spectral range (-300-1800 cm'), we will not be reporting on specific ranges.

Table 2-2 Glucose measurements using NIR absorption spectroscopy

In vitro

Author	**Excitation [nm]**	**Sample**	**# Sample**	**Protocol**	**Approx. Error [mM]**
Berger et al.[39]	830	serum	66	No further sample preparation	1.5 (CV) 1.7 (P)
Qu et al.[34]	785	serum	24	ultrafiltrated	0.38 (P)
Enejder et al.[33]	830	whole blood	31	pre-selected for hyperglycemia	1.2 (CV)
Rohleder et al.[35]	785	serum	247	ultrafiltrated	0.4 (P)
Pelletier et al.[40]	785	Aqueous humor	17	Measured within contact lens	1-1.5 (P)

In vivo

Author	**Excitation [nm]**	**Sample**	**# Sample**	**Protocol**	**Approx. Error [mM]**
Enejder et al. [37]	830	forearm	17 healthy	tolerance	0.7-1.5 (CV)
Chaiken et al. [38]	785	fingertip	25 diabetics	time-randomized	1.2 (P)

Table 2-2 Glucose measurements using NIR absorption spectroscopy. Source: Shih, W.-C. "Quantitative Biological Raman Spectroscopy" *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

2.2.2.2 Activity in the optical spectrum and polarimetry

Glucose is a chiral molecule, which means that when it comes into contact with light, it causes a rotation in the polarization of that light [29]. Polarimetry is a method that may be utilized to determine the angle of rotation, and it is connected to the concentration of glucose in the blood. Polarimetry is often carried out at a single wavelength, which removes the necessity for calibrating several variables [30]. It is possible that the analysis will become more difficult if one is working with biological tissue that is both birefringent and opaque [31].

2.2.3 Indirect approaches

The presence of physical barriers, which are responsible for variations in the refractive index of the surrounding tissue, is the primary factor that contributes to the scattering of light in biological tissue. This behavior can be observed being carried out in a variety of different ways. Within the framework of the diffusion approximation [32], the decreased scattering coefficient [33] may be approximatively estimated as a function of the number density and diameter of spherical scatterers, in addition to the difference in refractive index between the scatterers and the medium that surrounds them. There is a difference in the refractive indices of the scatterers and the medium that is around them. It is well known that the index difference that exists between the extracellular fluid and structural scatterers found in tissues, such as cell membranes and protein matrix, will change in response to changes in the osmolyte concentration of the tissue. This is only one of many different scenarios that may play out if anything like this were to take place. On the other hand, the mechanism that underlies this phenomenon is not completely known. As a direct consequence of these discrepancies, it is possible that observable differences in the tissue scattering coefficient may become apparent. When examining the effects of changing the refractive index, glucose has a significantly more significant influence than potassium chloride (KC1), sodium chloride (NaC1), or urea [34, 35].

2.2.3.1 Diffuse Reflectance Spectroscopy (DRS)

In contrast to reflectance-mode absorption spectroscopy, diffuse reflectance spectroscopy (DRS) determines the bulk absorption and scattering coefficients by fitting the spectrum to a specific model, as opposed to analyzing the frequency response of the light to extract molecule information [36]. This replaces the traditional method of extracting molecular data. A typical steady-state DRS configuration [37] would include a CCD-based grating spectrometer, intermediate optics, and spatially separated delivery-collection optical fiber sensors. Additionally, research has been conducted on frequency-based methodologies based on a less concrete theoretical framework [38]. Research [37,38] has demonstrated that the scattering coefficient for tissue transport is correlated with glucose levels.

2.2.3.2 Optical Coherence Tomography (OCT)

Using optical coherence tomography (OCT) [35, 39-41], the glucose and other analyte concentrations in living tissue have been detected. OCT, which is based on interferometry, provides the appropriate range (in millimeters) and depth resolution (in micrometers) for the study of tissue reflectance at a localized level. The components of an optical coherence tomography (OCT) system include a broadband light source, intermediate optics, a Michaelson interferometer, fiber-optical instruments, and a detector. Larin et al.[40] discovered a correlation between the slope of the OCT signal versus depth and the glucose content in 15 healthy individuals. It has been hypothesized that glucose-induced regional changes in the index of refraction are associated with the slope of the OCT signal.

2.2.4 Other approaches

Using the body's naturally emitted infrared radiation as the energy source, the fundamental absorption bands of glucose at 10 um are measured to determine thermal emission. For IR absorption spectroscopy, comparable detection equipment is utilized. Malchoff et al.[42] described the evaluation of a prototype that measures infrared emission from the tympanic membrane. Photoacoustic spectroscopy (PAS) is an alternative method for detecting absorption in liquids and gases, in addition to variations in the refractive index. The sample is agitated by employing nano-picosecond laser pulses. Light absorption and consequent localized heating generate ultrasonic waves that can be detected with piezoelectric transducers. PAS has been used to measure in vivo glucose concentrations, but no advantage has been demonstrated over NIR absorption spectroscopy. Indirectly, PAS can be used to detect changes in the refractive index.

2.3 Previous studies conducted in the Spectroscopy Laboratory at MIT

2.3.1 In vitro investigations

Massachusetts Institute of Technology(MIT) research group's initial investigations with Raman spectroscopy focused on human blood as one of the first biological tissues of its kind [44]. Berger et al.[24] performed concentration measurements in biological medium after collecting serum and whole blood samples from 69 individuals over the course of 7 weeks. They reported measuring a large number of analytes in these samples. A number of different analytes, including glucose, uric acid, cholesterol, triglycerides, albumin, total protein, and hematocrit, were evaluated. While compared to the errors that occurred when measuring the overall blood volume, the errors that occurred when calculating serum concentrations were significantly less. This data was interpreted to suggest that the turbidity of whole blood contributed to signal levels that were lower than expected Shih et al.[15]. As a consequence of this, a tool was developed to boost the capacity for signal collecting by more than a factor of four [45]. A second investigation, this one conducted by Enejder et al.[21] using whole blood, revealed that it is possible to measure many analytes with a single sample of blood. As a standard part of their clinical diagnosis, thirty-one participants contributed full blood samples for the purpose of the trial. We obtained 30 spectra from each sample over the course of 5 minutes, with each one lasting exactly 10 seconds. In order to determine the concentrations of the eight reference analytes, the procedures that are standard practice in clinical labs were applied to each sample. By employing PLS and cross validation, it was possible to establish a correlation between the Raman spectra and the relevant reference concentrations. The outcomes of a PLS leave-one-out cross validation test that was performed on the full blood dataset are presented in Table 2-3. The usual range for adult males living in the United States is provided for each analyte. This is done in order to provide the reader with a better understanding of the significance of the RMSECV documents. 46 With the exception of total cholesterol, every test parameter reveals a substantial correlation between the predicted and the reference concentrations, with r2 values of 0.93 or above. The only exception is total cholesterol. The only variable that did not connect substantially with the outcome is total cholesterol (0.66). A adequate correlation between the reference and the measured concentration levels is indicated when the r 2 value is more than 0.9. This value may be thought of as a rule of thumb. Keep in mind that the square of the correlation coefficient between two vectors is the number that is referred to as the r2 value. In this particular scenario, both the original concentration, which will be referred to as the "reference concentration," and the forecast concentration, which will be referred to as the "projected concentration," will serve as vectors.

PLS and leave-one-out cross validation were applied to develop a one-of-a-kind calibration model [47,48]. The data from each participant was put to use in this process. The amounts of glucose in the subjects' blood ranged from 68 to 223 milligrams per deciliter. A volunteer is seen in Figure 2-3 sitting with his arm extended toward the gadget. For the aim of conducting an analysis of the data, Raman spectra were chosen that fell anywhere between 1545 and 355 cm-1. In the in vivo spectra, there were distinct Raman signals that were placed on top of a background that was quite uniform and wide. The process by which the broad backdrop is gradually erased is referred described as fluorescence photobleaching. In the following step, a 13-point Savitzky-Golay filter was applied to each 3-minute sample in order to smooth out the spectra and increase the signal-to-noise ratio.

Table 2-3 Cross-validated results of calibration on eight analytes.

Analyte (units)	**Cross-Validated Error (RMSECV)**	**Normal Range (adult males)**	**r^2**
Glucose (mg/dL)*	21	45-180	0.97
Urea (mg/dL)	4.9	17-50	0.94
Cholesterol (mg/dL)	30	150-250	0.66
Triglycerides (mg/dL)	38	10-190	0.92
Total Protein (mg/dL)	0.31	6-8.3	0.94
Albumin (g/dL)	0.11	3.2-4.5	0.94
Hemoglobin (g/dL)	0.66	13-17.5	0.94
Hematocrit (%)	1.7	35.9-50.4	0.94

*(For glucose: 1mM = 18 mg/dL)

Table 2-3 Glucose measurements using NIR absorption spectroscopy. Source: Shih, W.-C. “Quantitative Biological Raman Spectroscopy” *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

2.3.2 In vivo studies

Enejder et al.[26] carried out a transcutaneous investigation using a glucose challenge test on 17 healthy volunteers who were not diabetics. In order to do an in-depth analysis of each individual, we employed PLS in conjunction with leave-one-out cross validation. Several different grouping algorithms were used, and the information about glucose's spectrum was ultimately stored in the b vector that was produced.

2.3.2.1 Methods and experimental protocols

In the course of this research project, a series of spectra were acquired on the forearms of human participants in combination with a technique that measured their oral glucose tolerance. During this test, the subject drinks a solution that has a high concentration of glucose. As a result, their blood glucose levels rise to more than double the amount seen when the subject has been fasting. Throughout the length of the operation, which ranged from one to two and a half hours for each participant, Raman spectra and reference glucose concentrations from blood samples were monitored at various intervals. The reference measurement for the blood analysis that was performed using finger sticks was given by a Hemocue glucose analyzer.

Figure 2-3 A volunteer was found seated next to the optical table with his forearm pressed on the equipment. Source: Shih, W.-C. "Quantitative Biological Raman Spectroscopy" *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

PLS was used in conjunction with leave-one-out cross validation[47,48] to construct an individual calibration model from the data collected from each participant in the study (ref). The levels of glucose in the blood of each and every volunteer ranged anywhere from 68 to 223 mg/dL. The volunteer shown in Figure 2-3 is seated with his arm positioned such that it is facing the instrument. For the purpose of data analysis, Raman spectra were chosen to fall within the range of 1545-355 cm^{-1}. The spectra that were acquired in vivo had a vast and diffuse background that was superposed with tiny and distinct Raman features. Fluorescence photobleaching is the process that describes the broad backgrounds gradual deterioration over time. In order to get a higher signal-to-noise ratio, the spectra of each 3-minute sample were then smoothed using a 13-point Savitzky-Golay filter.

2.3.2.2 Results and discussion

Each volunteer's combined background and Raman spectra were analyzed using PLS with start leaving cross validation. By combining 8 elements from each respondent, a b vector was constructed. Results showed an MAE of 7.8% and a r^2 of 0.83 on average. The mean absolute error (MAE) was 16.9% when data from all 17 subjects was combined. Positive and encouraging results were found when multiple glucose spectral signatures were discovered inside the regression vectors. These signatures offered supporting evidence that glucose was indeed measured.

2.4 Artificial intelligence

Raman spectroscopy is a type of spectroscopy that can be used to conduct an indirect analysis of the vibrational states of different substances. Information on the vibrational states of the sample can be used to determine the spectroscopic imprint of the sample. Then, these spectroscopic signatures can be used in a variety of application situations to determine the chemical composition of a sample without modifying it, or to predict a sample property such as a patient's health condition. There is a wide variety of application scenarios, ranging from biological diagnostics to material science-related issues, of which these two examples only represent a small fraction. However, because the Raman signal is quite weak and Raman spectroscopy lacks identifiers, the data obtained from Raman spectroscopy does not specifically target anything. Consequently, analyzing Raman spectra is a challenging task that necessitates the application of chemometric models based on the concept of machine learning. Deep learning, a subset of representation learning techniques, has been demonstrated to be highly effective in the analysis of Raman spectra and photonic data in general within the field of data science. Within the scope of this paper, we will discuss recent advancements made in deep learning algorithms for Raman spectroscopy, as well as the current implementation challenges. Luo et al.[41]. In nearly all of the most significant fields of scientific research, the application of deep learning algorithms has met with tremendous success. Deep learning exists within the subfield of representation learning within the discipline of artificial intelligence (AI). There are always three layers in artificial neural networks (ANNs): an input layer, a hidden layer, and an output layer. The methodologies of deep learning are a subset of ANNs. The majority of the time, feedforward neural networks, also known as FNNs, serve as the foundation of deep learning algorithms. After the neurons in the network's hidden levels have used their weights to process the data, the processed data is returned to the input layer via the output layer. This occurs once the neurons in the concealed levels have processed the data. Data are initially transmitted into the network by the input layer, which is responsible for this action. Backpropagation and gradient-based optimization are two prevalent techniques for adjusting the weights and bias of a neural network. Backpropagation is a specific instance of backpropagation. As long as sufficient vast volumes of labeled training data are available, the structural foundation of a deep learning network makes it feasible for the network to represent increasingly complex functions. This functionality can be obtained by incorporating additional network entities and layers [42].

2.4.1 Convolutional Neural Networks (CNN)

CNN's convolutional layer is the most important part of the network. In addition to fully-connected layers, a CNN will frequently utilize bulk normalization layers, pooling layers, and pooling layers. The input of one convolutional layer is convolved by the kernels of that layer, and the result is then used as the input for the subsequent layer. While the weights of each layer are being modified, the convolution kernel of each layer is being trained; consequently, the kernel-generated feature maps are being updated. In addition, pooling layers are utilized to subsample data in order to reduce the data's dimensions and the computational complexity. The two most prevalent forms of pooling algorithms are maximal pooling and average pooling. A completely connected layer is typically located at the conclusion of a CNN. This layer links every neuron of the preceding layer to the output. The typical organizational structure of a CNN model is illustrated in Figure 2-4. It is essential to remember that CNNs are distinguished by two distinct characteristics: parameter sharing and local connection. These concepts reduce the total number of parameters, resulting in more efficient calculations, as stated by Luo et al. [41].

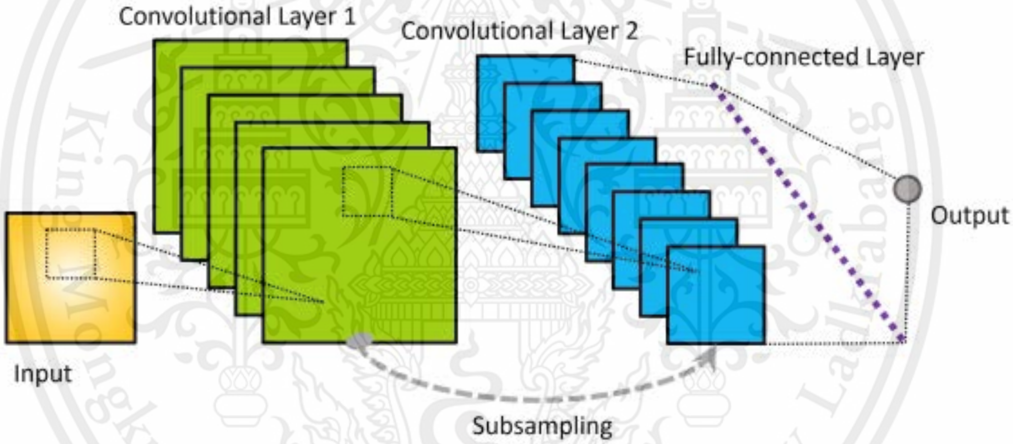


Figure 2-4 demonstrates CNN's organizational structure. On the left, the yellow square represents input data, while the gray circle represents output data. The green and blue squares, respectively, represent the first and second convolutional layers, and a pooling phase (represented by a grey arrow with dashes) subsamples between them. Before emission, a fully-connected layer is used at the end of the CNN (shown by the purple dotted line). Source: R. Luo, J. Popp, & T. Bocklitz Deep Learning for Raman Spectroscopy: A Review 2022. *Analytica*, *3*(3), 287–301.

2.4.2 Residual Network (ResNet)

ResNet's residual block, which is depicted in Figure 2-5, is the network's fundamental building element. The residual block uses a shortcut to skip between layers. This design can avoid the vanishing gradient problem, which could prevent the neural network from learning further during the training process. According to network length, the most frequently employed ResNets are ResNet-50, ResNet-101, and ResNet-152, all of which can be classified as CNN variants.

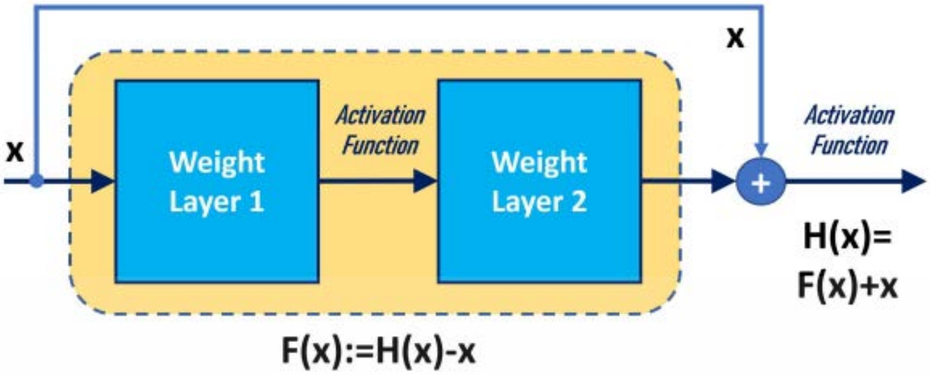


Figure 2-5 Blockage. There is a shortcut from x to H. This network architecture provides skip connections, which carry gradient information between layers and avoid the vanishing gradient problem. Source: R. Luo, J. Popp, & T. Bocklitz Deep Learning for Raman Spectroscopy: A Review 2022. *Analytica*, *3*(3), 287–301.

2.4.3 Autoencoder

Figure 2-6 illustrates how an autoencoder, which uses a bottleneck structure, can be used to achieve unsupervised learning. An autoencoder consists of two primary components: an encoder and a decoder, which are intended for data input and output, respectively. A feature representation of the input data is formed as the data pass through a bottleneck on their way from the input to the output. The U-Net, which was initially for the purpose of biomedical image segmentation [44], is an example of an important application that makes use of an autoencoder structure that was generated by means of a CNN.

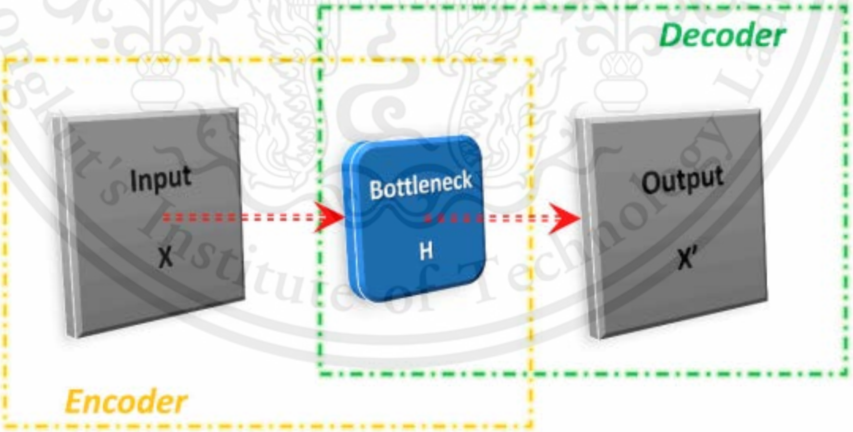


Figure 2-6 Autoencoder bottlenecks. The yellow encoder maps the input X to the bottleneck H, while the green decoder maps H to X'. The encoder analyzes X and creates H (containing key information); next, the decoder translates H into the desired output X' by R. Luo, J. Popp, & T. Bocklitz Deep Learning for Raman Spectroscopy: A Review 2022. *Analytica*, *3*(3), 287–301.

2.4.4 Generative Adversarial Network (GAN)

Goodfellow et al. [45] introduced a very intriguing deep learning architecture dubbed Generative Adversarial Network (GAN). Goodfellow et al.[46] introduced a very intriguing deep learning architecture they dubbed GAN, which is comprised of a generator G and a discriminator D. Goodfellow et al. introduced the GAN architecture for deep learning, which they deemed to be quite intriguing. During the training phase, G's objective is to increase the likelihood that D will commit an error, whereas D's objective is to distinguish between genuine and produced data instances. Using the minimax loss, which is technically a minimax two-player game, training is conducted in a sequential fashion. As a result, G is able to generate data that derives from an approximated distribution that is comparable to the distribution of the input data. In the interim, D endeavors to distinguish between actual photos in the training D.

2.4.5 Artificial Neural Networks(ANN)

Neural networks have emerged as one of the most extensively utilized systems for pattern recognition and have discovered a broad range of uses across the board in the field of biomedicine. This is due in large part to the fact that neural networks are able to learn from their own mistakes. An artificial neutral network, also known as an ANN, is a network that consists of several layers and nodes, each of which is responsible for performing a basic operation on a number of inputs in order to produce a single output. This allows the network to function as a single entity. Via weighted connections, the signal is sent from the input layer of the standard format known as a multilayer perceptron (see Figure 2-7), which is a multilayer perceptron, to the layer of hidden nodes. The processing of the signal is under the purview of this layer. When dealing with spectral applications, there is often just one input node for each every spectral signal or portion of the spectrum. Analysis of spectra is another name for the field of spectral applications. Signals from all of the input nodes are to be collected by each concealed node, which is accountable for doing so. After that, these signals are combined together in the form of a weighted total, and then they are sent through a non-linear "activation function," which is often a smoothed step function or sigmoid.This has the effect of applying a soft threshold to the total inputs, with the weights determining which spectral features each hidden node responds to in the most favorable manner. This has the effect of applying a soft threshold to the whole inputs. The signals coming from the hidden nodes are merged in the form of a weighted sum by the output nodes, and the output nodes may or may not apply a non-linear adjustment to the combined signal. In classification applications, there is typically one output node corresponding to each class, and maybe one belonging to a "unknown" class. In certain cases, there may also be a "unknown" class. In addition to that, there is a possibility of one. The output node that generates the highest output value determines which class a new input vector belongs to, and that class is the one to which the new input vector is assigned. It is a typical practice for computer systems to utilize the value of a single output node as the dependent variable in regression studies [47].

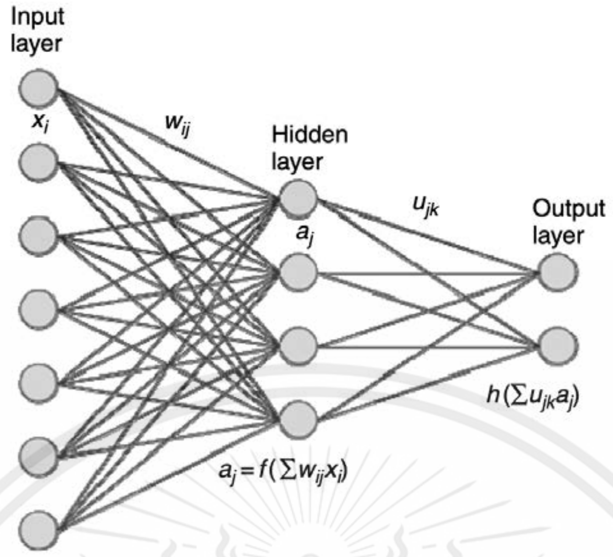


Figure 2-7 Schematic of ANN with seven input, four hidden, and two output nodes. Input values are xi, hidden layer weights are wij, hidden layer outputs are aj, output layer weights are ujk, and hidden and output layer transfer functions are f and h. Source: Timothy, M.D. *"Non-linear methods for the analysis of metabolic profiles(n.d)"* in Elsevier B.V. 2007, 7, 213.

2.4.6 K-Nearest Neighbour classification

The K-Nearest Neighbor (kNN) rule is a technique to classification that is often considered to be the most straightforward of all supervised classification methods. In order to classify test samples, k nearest neighbors in the multidimensional parameter space are used, and each sample is then assigned to the class that occurs the most frequently among those k neighbors. The approach, despite its apparent simplicity, has a good theoretical basis in non-parametric density estimation and has the ability to frequently outperform methods that are far more complex. The sole need for the approach is the selection of k, which represents the number of neighbors that will be taken into account throughout the classification process. Smaller values of k will result in the selection of the training points that are the most capable of estimating the right classification at the test point. These will be the points that are closest to the test point. This estimate, on the other hand, will be susceptible to severe statistical fluctuations due to the limited number of participants. On the other hand, large values of k can minimize statistical mistakes, but they also let distant points to participate to the classification, which may have the effect of smoothing out some of the features of the class distributions. k is often determined to be the value that results in the least amount of classification error based on some independent validation data or by following techniques that involve cross-validation. Similar to the k-means algorithm, the kNN algorithm just generates a classification for each item without any related visualisation or explanation; as a result, its usage as a standalone tool is rather uncommon. The fact that it performed so well in competitive classification, on the other hand, has promoted its usage as a benchmark against which other methods might be evaluated. These aspects are exemplified by the vast majority of metabolic profile studies that make use of the method [48].

2.4.7 Applications for Raman Spectroscopy

In the discipline of Raman spectroscopy, traditional approaches of machine learning have been utilized to a significant degree. In the vast majority of circumstances, you will be required to carry out processes like as data pre-processing, feature extraction (or feature selection), and data modeling. Deep learning, on the other hand, allows for the possibility that the work of such complicated phases may be performed entirely by a single neural network, assuming that there is adequate training data. This, of course, is something that can only be said if there are sufficient facts to begin with. The four essential aspects of deep learning applications for Raman spectroscopy are pre-processing, classification, regression, and highlighting, and they are depicted in Figure 2-8. On the basis of the many kinds of outputs that the program creates, these components can be subdivided into more components later on. After training the model using a Raman spectrum as an input, a pre-processing model will split the input into several portions and will generally output a specified area of interest (ROI) of the 1D spectral data. This occurs after the model has been trained with the spectrum. A label is what comes out of a categorization model. The output of a regression model is either a number or a probabilistic value. A highlighting model will segment the input into several segments and will typically produce a particular region of interest (ROI) of the 1D spectral data. Luo et al.[41].

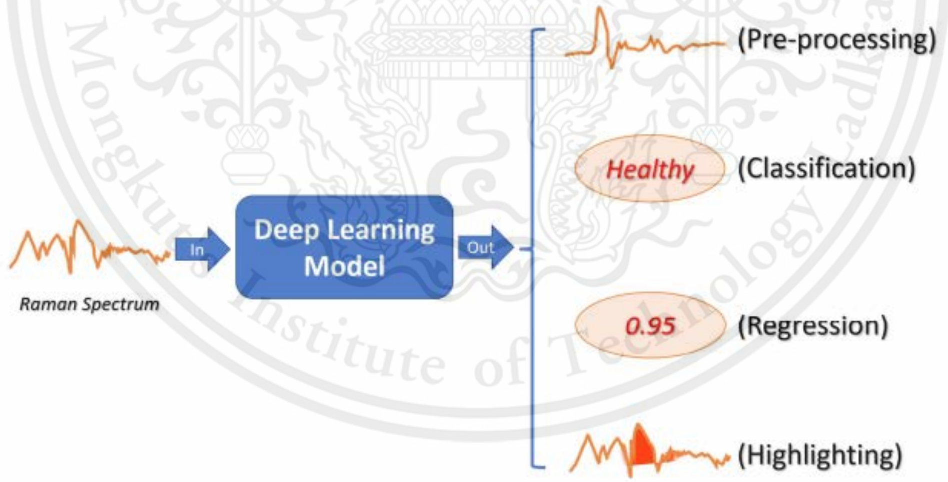


Figure 2-8 Four Raman applications that leverage machine learning. Models for preprocessing, classification, regression, and highlighting based on output. The output of a pre-processing model is another Raman spectrum, the output of a classification model is a label, the output of a regression model is a number or probability value, and the output of a highlighting model is the spectral ROI of the input spectrum. Source: R. Luo, J. Popp, & T. Bocklitz Deep Learning for Raman Spectroscopy: A Review 2022. *Analytica*, *3*(3), 287–301

2.5 Summary

Optical detection is the modality of choice for the creation of a totally non-invasive technology for blood analysis, which offers the highest potential for success in terms of how well it might be implemented. Due of its unrivaled molecular selectivity, the Raman spectroscopy technique stands out as one of the most promising optical approaches. The review that is contained inside this chapter goes through the therapeutically important blood analytes that are presented in this thesis. Since their concentrations are under the millimolar (mM) range, they are suitable for use in optical technologies that do not need the invasive penetration of the human body. In order to provide the reader with the information necessary to assess the calibration findings in the part that will follow, a comprehensive overview of multivariate calibration is presented. This is done for the aim of giving the reader with the knowledge required. Methods such as NIR and MIR absorption spectroscopy, NIR Raman spectroscopy, polarimetry, diffuse reflectance spectroscopy, optical coherence tomography, thermal emission spectroscopy, and photoacoustic spectroscopy were discussed in this portion of the article. The information that the reader needs to analyze effectively will be provided to them. A review of the previous work carried out at the Spectroscopy Laboratory at MIT was carried out. This review included measurements of numerous blood analytes in discarded human serum, whole blood, and human individuals utilizing PLS with leave-one-out cross validation. This study was one of the accomplishments that had been accomplished in the field of NIR Raman spectroscopy previously. In comparison to methods that rely on subjectivity for the pre-processing phase, deep learning makes it feasible to simplify the pre-processing stage and even makes it possible to build an automatic pre-processing solution. This is the case since deep learning can learn from examples. The application of a solution is now feasible as a result of the fact that deep learning makes this capability conceivable. Some deep learning models incorporate all of the pre-processing processes into a single network, along with the eventual goal, which may be classification or regression. Because of this, the model is able to learn more efficiently. This can be accomplished in a roundabout fashion.

CHAPTER 3

METHODOLOGY

3.1 Raman spectroscopy

An inelastic scattering phenomenon known as Raman scattering is utilized in the process of Raman spectroscopy, which measures the fundamental vibrational states of molecules. The phenomenon of Raman scattering, which was found by Raman and Krishna [49], results from interactions between photons and molecules. According to the principles of classical physics, the contact can be seen as a disruption of the electronic cloud that is already existing in the molecule. Rayleigh scattering is named after Lord Rayleigh, who discovered that the great majority of photons that are dispersed as a result of light being scattered from a molecule are elastically scattered. While elastic scattering is taking place, there is no frequency shift that takes place since the scattered photon has the same amount of energy as the photon that was fed into the system. Photons, on the other hand, have the power to either transfer energy to molecular vibrations or absorb energy from them through a process known as Raman scattering. This occurs when two photons collide with one another. An incident photon with the energy hv_L, where h is Planck's constant and v_L is the frequency of the excitation laser, can cause the excitement of a molecule into a virtual state that has a lower energy need than an electronic transition. This can be caused by the molecule being exposed to an excitation laser with a frequency of v_L.This process, which is known as Stokes-Raman scattering [50], involves the creation of a new photon that is then scattered from the "virtual state." Light that has been scattered by a Stokes-Raman scattering process will have an energy equal to $h(v_L + v_R)$, and its frequency will be $v_R < v_L$. Figure 3-1 is a schematic representation of the Rayleigh process, as well as the Stokes Raman process and the anti-Stokes Raman process. As demonstrated in Figure 3-2 for acetaminophen powder in a quartz cuvette, a Raman spectrum is made up of a scattering intensity vs. energy plot. Each peak is associated with a specific Raman shift derived from the hv_L value of the incident light energy. The Raman shift in wavenumbers (cm^{-1}), also known as the energy difference between the initial and final vibrational states, v, can be computed using the formula $v = (v_L - v_R)/c$, where c refers to the speed of light. Raman shifts are consistent and unaffected by the frequency of the excitation laser used in the experiment. This affords the flexibility to select a laser wavelength that is appropriate for the particular purpose at hand.

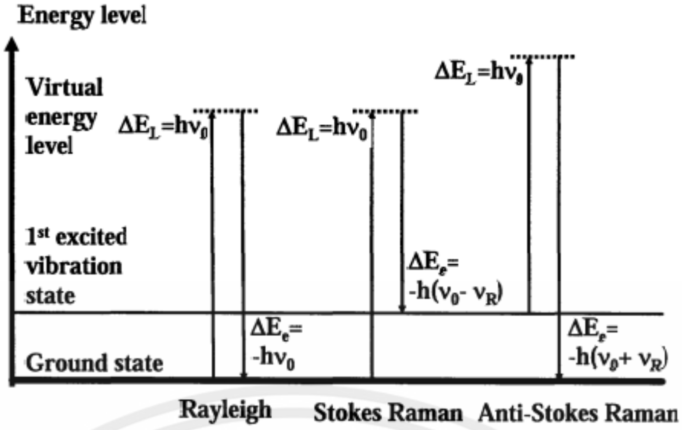


Figure 3-1 A diagram depicting the energies involved in Rayleigh, Stokes Raman, and anti-Stokes Raman scattering by Shih, W.-C. "Quantitative Biological Raman Spectroscopy" *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

Individual molecule vibrations also affect the absorption of infrared radiation. Despite the fact that Raman spectroscopy investigates vibrational transitions indirectly via scattering, the energy range of the Raman shift is identical to that of IR absorption, and in many instances, the same energies are observed. Raman and infrared absorption each have their own set of criteria for selecting molecules, but the information they provide is identical. IR absorption can be used to measure the vibrational frequencies responsible for a change in a molecule's permanent dipole. The Raman scattering technique entails the measurement of vibrational frequencies that induce a polarizability shift.

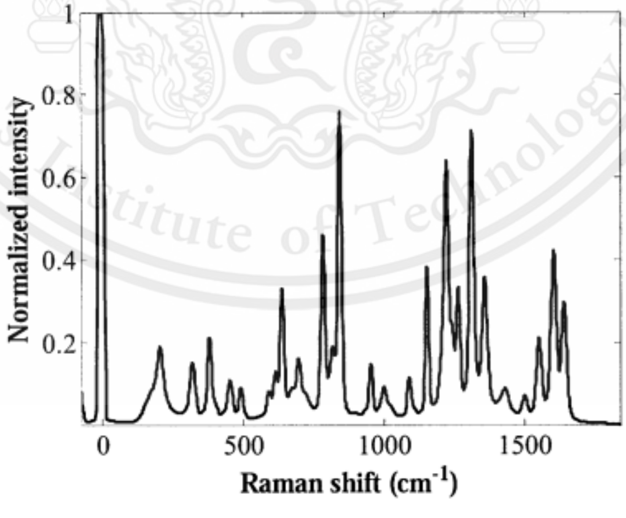


Figure 3-2 A Raman spectrum is created by plotting the scattered intensity vs the energy. As an example, this figure takes the readings from a quartz cuvette containing acetaminophen powder by Shih, W.-C. "Quantitative Biological Raman Spectroscopy" *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

Near-infrared (NIR) absorption spectroscopy is a significant method in the context of the development of quantitative Raman spectroscopy. NIR is founded on overtone and combination bands of mid-infrared transitions. These transitions are considerably weaker than fundamental transitions, and quantum mechanics deems them "prohibited." Near-infrared (NIR) absorption spectroscopy is a significant method in the context of the development of quantitative Raman spectroscopy. NIR is founded on overtone and combination bands of mid-infrared transitions. These transitions are considerably weaker than fundamental transitions, and quantum mechanics deems them "prohibited."

3.2 Biological factors

3.2.1 Using near-infrared radiation

As Raman shifts are independent of excitation wavelength when using near-infrared radiation, this technique allows the excitation wavelength to be tailored to the needs of a variety of applications. Due to its low-energy optical radiation, profound penetration depth, and reduced background fluorescence, NIR excitation is advantageous for the investigation of biological samples. These three benefits justify using NIR excitation as the method of investigation. Utilizing an excitation wavelength from the NIR spectrum prevents the potentially hazardous ionization of sample components. It is feasible to conduct optical sampling at greater distances, on the order of one millimeter, because the NIR region is devoid of significant absorbers. The comparatively low fluorescence background associated with NIR excitation enables the recovery of a Raman signal that is one order of magnitude weaker. As a result, we have chosen an excitation wavelength of 830 nm in order to fully exploit the "diagnostic window" depicted in Figure 3-3 while maintaining an acceptable quantum efficiency in the silicon-encased charge coupled device (CCD) detector.

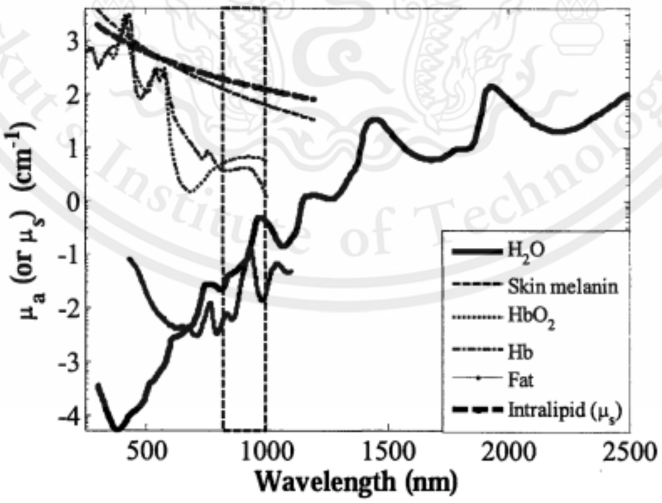


Figure 3-3 A number of different substances' absorption spectra, including those of water, skin melanin, hemoglobin, and fat. The scattering spectrum of 10% Intralipid, which is a lipid emulsion, is also shown. This lipid emulsion is exhibited because it is widely used to replicate the scattering that happens in tissue. Source: http://omlc.ogi.edu/spectra/index.html.

The figure 3-3 shows the spectra of many endogenous tissue absorbers. Water, melanin, hemoglobin, and fat are endogenous tissue absorbers. 10% scattering Intralipid is a lipid emulsion used to mimic tissue dispersion. The diagnostic window contains NIR minima. Figure's dashed rectangle shows this area.

3.2.2 A Signal Observed in the Context of Biological Raman Spectra

In the case of biological samples, the Raman spectra are typically accompanied by a substantial background. Fluorescence is applied quite frequently to determine the source of the background signal, particularly in circumstances that involve the utilization of UV/visible laser stimulation. Several components present in biological tissue are fluorescent. The fluorescence backdrop is affected by macromolecules such as proteins and lipids, and these molecules have a contribution to make Qu et al.[34] . When triggered by UV-visible light, the skin's autofluorescence can be utilized to diagnose disease states including psoriasis [51] and diabetes. This technique has been employed for decades. In the absence of an enzyme, glycation is the bonding of a single sugar molecule, also including glucose, to a protein or lipid molecule.Under these conditions, changes in the autofluorescence of collagen due to glycation were detected. Glycation causes changes in the autofluorescence of collagen [52]. In addition, in order to diagnose dysplasia, the autofluorescence that is produced by the components of epithelial tissue is utilized in our laboratory as a diagnostic tool [53]. The presence of the backdrop, in addition to the shot noise that is produced by the backdrop, is a factor that works against the overall capability of detection. In addition, background variation makes it impossible to continue with the multivariate analysis. Fluorescence photobleaching is the word that is being used to characterize the background deterioration that was seen in the human study. This deterioration was seen as a result of the human inquiry Zeng et al.[54, 55]. modeled the signal degradation from skin under UV-visible light excitation using a double-exponential function, with the time constants being attributed to varying photobleaching rates of distinct fluorophores in the stratum corneum and dermis. This function was used to mimic the signal degradation from epidermis when stimulated by UV-visible light. Jongen and Sterenborg [56] contend that the turbidity of tissue modifies the decay properties of a single fluorophore such that it no longer follows a single exponential decay curve. This means that a singular exponential cannot be used to characterize the decay curve.The conclusions of this study run counter to the findings of the earlier study. As a result of this, the decay profile does not need to be given by extra exponentials with a variety of different time constants. The physical justification for this argument is that the fluorescence signal that is measured for a turbid medium that is composed of multiple layers is the sum of the contributions that are made by each layer. This is the fluorescence signal that is measured for a turbid medium that is composed of multiple layers. Fluorescence emanating from a layer that is deeper will have a less intense appearance and will photobleach at a more leisurely pace as a result of the decreased laser intensity. This is because the fluorescence is originating from a layer that is deeper. As a direct consequence of this, the proportional contribution of fluorescence coming from deeper layers will show up as lower levels of signal. that lose their strength more slowly, while the surface layers have stronger signals that lose their strength more quickly. This demonstrates the significant impact that the optical qualities of the sample can have on the behavior of light as it is seen. Even though implicit multivariate calibration approaches have the potential to mitigate the negative impacts of the background to a certain amount, their effectiveness will never be fully restored. Therefore, it is preferable to either lessen the backdrop during the data collecting phase or eliminate it in a way that does not result in the introduction of artifacts. Polynomial fit is the foundation of the vast majority of the background removal algorithms described in the research literature. Due to the fact that the backdrop has very little structure, it is possible to suffice with a slowly varying low-order polynomial to approximate the background [57-59] According to the findings of these authors, a polynomial of the fifth order is the most successful strategy for fitting the background.

3.2.3 Heterogeneities in human skin

If the sample is continuously stirred, it is usually safe to assume that the analyte concentration is uniform throughout the sample. This is especially true for liquid samples like blood serum or even whole blood. Heterogeneity is an important consideration with regard to biological tissue, particularly human skin. Using confocal Raman spectroscopy, researchers were able to investigate the in-depth morphological structures as well as the molecular components of skin[60]. The epidermis and the dermis are the two primary layers that make up the skin's layered structure. The epidermis is the topmost layer of skin, and it is made up of several layers within itself, including the stratum corneum, stratum lucidum, and stratum granulosum. Keratin, which makes up around 65% of the stratum corneum, is the human epidermis' most abundant protein and is considered to be its primary ingredient. In addition to being a layered tissue, the dermis is predominately made up of collagen and elastin. The dermis contains blood capillaries; hence, this layer is selected for visual examination because of its location. Interstitial fluid is typically located at the interface between the epidermis and the dermis. Even so, it has been proposed that the vast majority of glucose molecules sampled using a non-invasive optical methodology are found in the extracellular fluid.

3.3 Methods of calibration using several variables

The explicit and the implicit methods of calibrating are the two categories that are available. The explicit approaches make use of the spectra of the constituent components, which can either be measured or estimated. The ordinary least squares (OLS) method and the classical least squares method (CLS) are two examples. Methods that are explicit produce more transparent models and findings that are simpler to interpret. In biological applications, it may be difficult to generate precisely controlled experimental conditions, spectra of high quality, and precise concentration measurements of each component. When not all of the spectra of the various components can be determined, implicit calibration methods are frequently employed. Because they can operate under conditions in which the quantity of spectra used for calibrating is smaller than the number of wavelengths sampled, factor analysis techniques such as principal component regression (PCR) [61] and partial least squares (PLS) [62] are commonly employed. The most prevalent of these are PCR and PLS. For example, a calibration set could include 30 spectra, with each spectrum containing 500 data points (wavelengths). In contrast to explicit approaches, the efficacy of implicit methods cannot be readily assessed using standard statistical criteria such as "fitness." According to the research, seven spurious effects, including system drift and co-variations between constituents, can be misinterpreted as originating from the target analyte. As a result of this circumstance, hybrid methods that combine aspects of explicit and implicit strategies to accomplish superior performance have emerged as a viable alternative.

3.4 SERs

As Fleischmann and his colleagues were researching pyridine that had been adsorbed onto an electrochemically roughened silver surface in 1973, they came upon the SERs effect and made the finding that led to its identification. They were successful in seeing the SERs effect despite the fact that they were unable to precisely define what it was that was taking place to create the enhanced Raman signal. This led to the realization that bringing molecules into proximity (through processes such as adsorption) with arrangements of nanoparticles (NPs), or metal surfaces that have nanoscale roughness, caused an enhancement of their Raman scattering intensity that was disproportional to the concentration of molecules that were present. This was realized as a result of the fact that bringing molecules into proximity (through processes such as adsorption). It is known that two processes, namely electromagnetic enhancement and chemical enhancement, contribute to the enhancement. Although the precise mechanism that underlies the SERs effect is still the subject of much discussion to this day, it is known that these processes contribute to the enhancement. During electromagnetic enhancement, plasmon oscillations that are perpendicular to the metal surface induce enhancement twice: first, of the light that was incident (excitation), and second, of the light that was Raman scattered. This results in an increase in the intensity of the electromagnetic field. The substrates used for SERs can be modified to maximize the localized surface plasmons, and the "hotspots" that result from the tendency of these plasmons to concentrate are referred to as "hotspots" in the industry. Figure 3-4 presents an instance of the SERs effect, in which the "hotspots" are depicted as being located at the very edges of the surface[63].

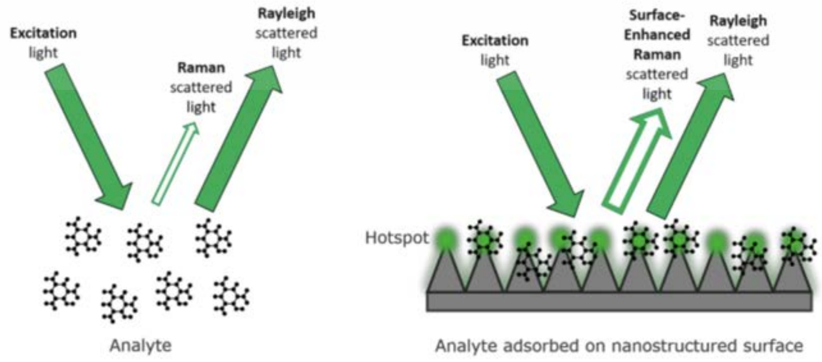


Figure 3-4 Raman scattering enhancement from a SERS substrate. Source: Granite, S. Optimisation of SERS for Glucose Sensing: Raman spectroscopy *in Edinburgh Instruments.*

Because inter- and intra- molecular charge transfers are what generate chemical enhancement, the enhancement is highest for molecules that are adsorbed on the metal surface. The enhancement factor for the electromagnetic contribution is on the order of 104, while the enhancement factor for the chemical contribution is 102. When these two factors are combined, the result is a 10°-fold boost in intensity of SERs in comparison to standard Raman scattering. Due to the fact that the plasmon frequencies of gold and silver lie within the range of the most frequent excitation wavelengths used for Raman spectroscopy (visible and NIR), these two metals are typically utilized as substrates for SERs experiments.

Researchers have developed many different types of SERs substrates since the discovery of SERs; nevertheless, the majority of these substrates may be divided into two categories: colloidal nanoparticles and nano-patterned surfaces.

In the case of colloids, the enhancement factor can be affected by a number of the properties of the colloids. In this application note, the aggregation of the NPs as well as their size are investigated with the goal of improving the sensitivity of SERs to detect glucose. The sensing of glucose is essential for the research of diabetes as well as the investigation of the influence of meals on blood glucose levels in persons who are healthy as well as those who have diabetes. In order to improve and evaluate the glucose sensing capabilities of our DIY Raman microscope, a 785 nm laser was utilized. The augmentation was accomplished with the utilization of gold nanoparticles that were obtained from ONSPEC LITE (SERs Substrate).

3.5 Instrumentation

In order to acquire the Raman Spectrum and a Raman peak of the desired quality. In this experiment, the do-it-yourself Raman spectroscopy will be constructed and utilized. Because we are using a laser on the human body as a fingertip, we are utilizing a high power laser diode that has a wavelength of 785 nm (L785P090). This is because it required the right range of wavelength and power in order to reach the area of the bloodstream and interstitial fluid where glucose is present. A dichroic mirror is traversed by the laser as it travels (FF801-Di02). After being reflected back from the Sample dichroic mirror, the laser will travel through it and then be routed via a filter (BLP01-830R) in order to isolate light noise from incident light. After that, the light will travel through various components of the instrument known as a spectrometer, which begins with the entrance slit. A collimation lens will assist in making the light beam uniform while simultaneously increasing the amount of light that will be gathered in the detector. And a diffraction grating that has a rate of 1200 l/min, which acts as the diffraction of light by reflecting a beam of light into a detector that we will be employing, which will be a Complementary Metal Oxide Semiconductor sensor or CMOS camera (BFLY-PGE-31S4M/C-C 1/1.8" Blackfly® PoE GigE Monochrome). Which has a quantum efficiency (QE) that is appropriate to receive our goal wavelength for Raman scattering, which is above 785 nanometers and is approximately 40% QE. After we have been successful in obtaining the findings in the form of intensity and wavelength peak bands, we then produce them using the software that comes with the camera and is known as Spinnaker. Processing the light intensity into a graph may be accomplished with MATLAB. Following that, we include it into our deep learning CNN model as the dataset.

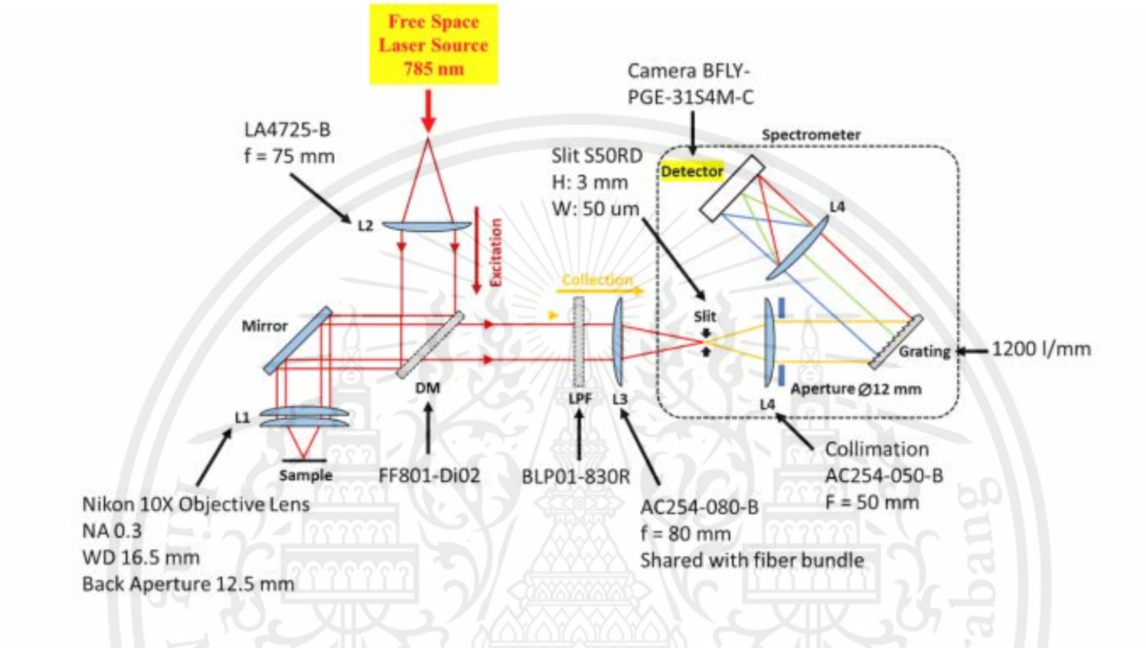


Figure 3-5. The 2D diagram of Raman optical spectroscopy

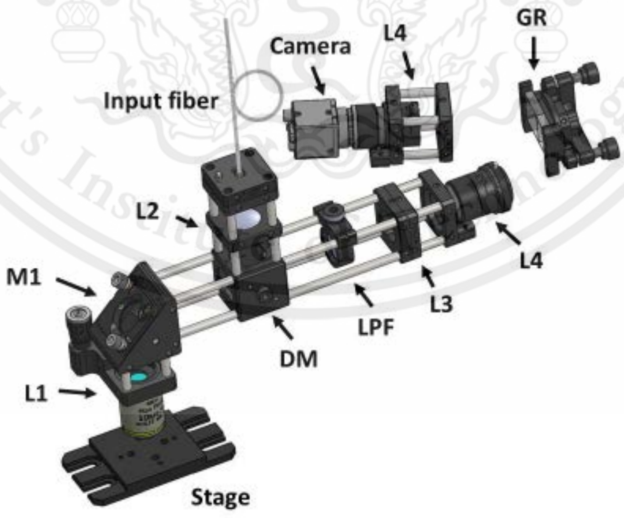


Figure 3-6. The 3D Design of Raman spectroscopy

3.6 Collection of data

Because the Raman Spectrum produced by Raman scattering is a very low and weak signal, a significant amount of integration time is required during the process of picture collecting in order to be able to read any relevant and discernible light bands produced by the Raman scattering region. Because of time and resource constraints, the only available information that can be gathered for this project is a picture of the detector. This is the only piece of data that can be obtained. Therefore, we will carry out the experiment in order to collect some kind of light pattern that will simulate the actual Raman spectrum picture.

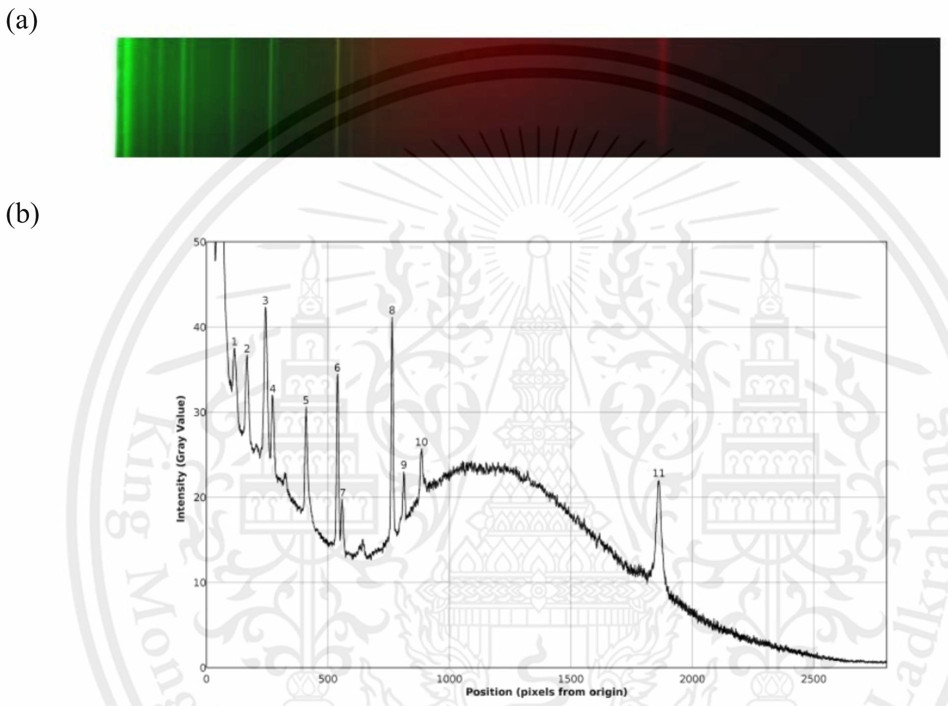
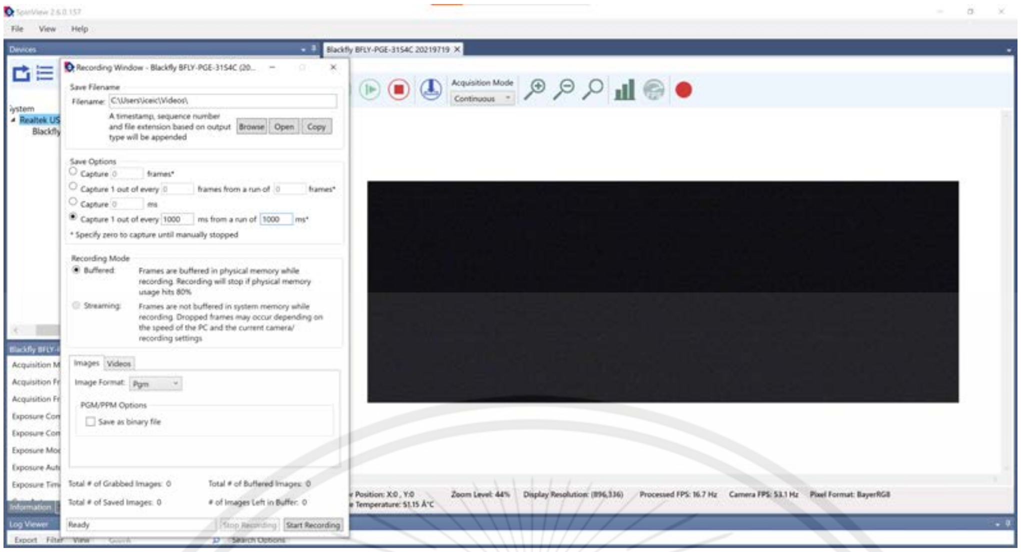


Figure 3-7. Example of the Raman spectrum that is found in reference paper material representing the final destination of DIY Raman result. (a) Digital picture of the Raman spectrum of pure benzene. (b) The plot of the Raman spectrum of pure benzene from figure(a). Source: contributors, P. L. (n.d.). A homemade cost effective Raman spectrometer with high performance *in Public Lab* Available: https://publiclab.org/notes/emontoya57/12-16-2015/a-homemade-cost-effective-raman-spectrometer-with-high-performance [Accessed Oct 25, 2022]

The program known as spinnaker is already pre-installed on the camera that we are using. This we may utilize to change any and all parameters as well as record an image or video of the light being caught.



Recording Window - Blackfly BFLY-PGE-3154C (20...

Save Filename

Filename: C:\Users\iceic\Videos\

A timestamp, sequence number and file extension based on output type will be appended

Save Options

○ Capture 0 frames*

○ Capture 1 out of every 0 frames from a run of 0 frames*

○ Capture 0 ms

● Capture 1 out of every 1000 ms from a run of 1000 ms*

* Specify zero to capture until manually stopped

Recording Mode

● Buffered: Frames are buffered in physical memory while recording. Recording will stop if physical memory usage hits 80%

○ Streaming: Frames are not buffered in system memory while recording. Dropped frames may occur depending on the speed of the PC and the current camera/ recording settings

Image Format: Pgm

PGM/PPM Options

☐ Save as binary file

Total # of Grabbed Images: 0

Total # of Buffered Images: 0

Total # of Saved Images: 0

# of Images Left in Buffer: 0

Ready

Position: X:0 , Y:0

Temperature: 51.15 Â°C

Zoom Level: 44%

Display Resolution: (896,336)

Processed FPS: 16.7 Hz

Camera FPS: 53.1 Hz

Pixel Format: BayerRG8

Figure 3-8 Interface of Spinnaker

3.7 Deep Learning development

3.7.1 Convolutional Neural Network (CNN).

As our point of reference, we will be using the article titled "Application of convolutional neural networks for categorization of adult mosquitoes in the field" [65]. They make an effort to segregate a large number of mosquitoes, but the ones that we choose to study are referred to as Aedes mosquitoes and Culex mosquitoes. They train the program using 823 photographs, and it functions effectively, as demonstrated by the fact that it achieves accuracies of 100% and 90%, respectively, for the categorization of Aedes and Culex mosquitoes. As a result, this model may be dealt with by separating the curves that represent the peak intensities.

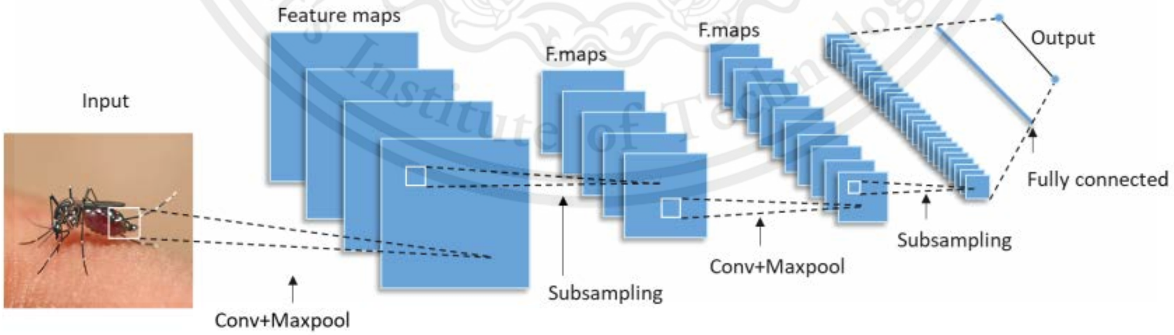
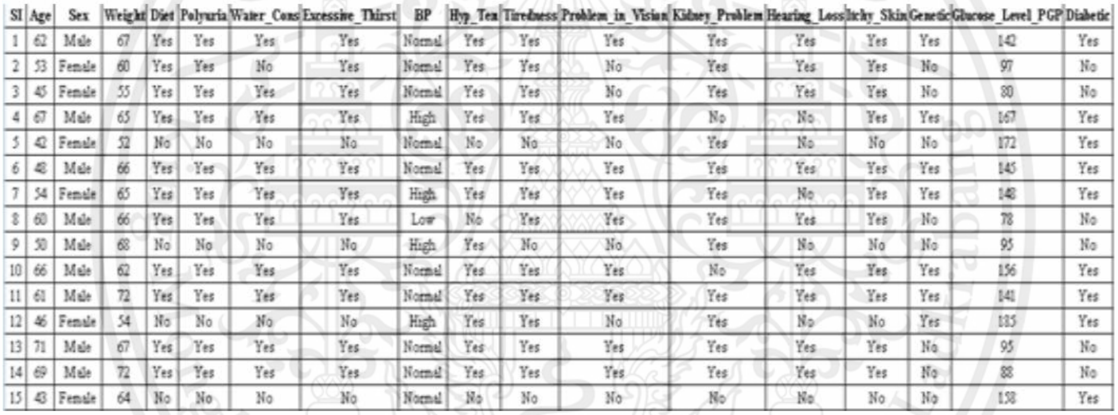


Figure 3-9 The CNN model that we got, can detect the Aedes mosquito and Culex Mosquito separately. D. Motta, A. Á. B. Santos, I. Winkler, B. A. S. Machado, D. A. D. I. Pereira, A. M. Cavalcanti, E. O. L. Fonseca, F. Kirchner, and R. Badaró, “Application of convolutional neural networks for classification of Adult Mosquitoes in the field,” *PLOS ONE*, 14-Jun-2019.

3.7.2 K-nearest neighbors algorithm(KNN).

We make use of the reference entitled "K-Nearest Neighbor (KNN) Learning based Diabetes Mellitus Prediction and Analysis for eHealth Services" [66]. It focuses particularly on predicting and analyzing diabetes mellitus, a condition that refers to a set of metabolic disorders that are characterized by a high blood sugar level for an extended period of time. Diabetes mellitus refers to a chronic condition that refers to a set of metabolic disorders. KNN is a simple machine learning method that may be utilized for the purpose of developing disease risk prediction models on the basis of relevant health data. In order to accomplish this objective, they provide a learning-based prediction model called an Optimum K-Nearest Neighbor (Opt-KNN) model that is based on a patient's routine characteristics across a number of variables. This technique determines the optimal number of neighbors to use in order to get the lowest possible error rate and provide improved prediction results in the final model. The more accurate forecast indicates that we will utilize the KNN to collaborate on the development of a telemedicine application in order to get the data and make the process more efficient.



SI	Age	Sex	Weight	Diet	Polyuria	Water_Cons	Excessive_Thirst	BP	Hyp_Ten	Tiredness	Problem_in_Vision	Kidney_Problem	Hearing_Loss	Itchy_Skin	Genetic	Glucose_Level_PGP	Diabetic
1	62	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	142	Yes
2	53	Female	60	Yes	Yes	No	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	97	No
3	45	Female	55	Yes	Yes	Yes	Yes	Normal	Yes	Yes	No	Yes	Yes	Yes	No	80	No
4	67	Male	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	No	No	Yes	Yes	167	Yes
5	42	Female	52	No	No	No	No	Normal	No	No	No	Yes	No	No	No	172	Yes
6	48	Male	66	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	145	Yes
7	54	Female	65	Yes	Yes	Yes	Yes	High	Yes	Yes	Yes	Yes	No	Yes	Yes	148	Yes
8	60	Male	66	Yes	Yes	Yes	Yes	Low	No	Yes	Yes	Yes	Yes	Yes	No	78	No
9	50	Male	68	No	No	No	No	High	Yes	No	No	Yes	No	No	No	95	No
10	66	Male	62	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	No	Yes	Yes	Yes	156	Yes
11	61	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	141	Yes
12	46	Female	54	No	No	No	No	High	Yes	Yes	No	Yes	No	No	Yes	135	Yes
13	71	Male	67	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	95	No
14	69	Male	72	Yes	Yes	Yes	Yes	Normal	Yes	Yes	Yes	Yes	Yes	Yes	No	88	No
15	43	Female	64	No	No	No	No	Normal	No	No	No	No	No	No	No	158	Yes

Figure 3-10. Example of the dataset for the KNN model. Source: I. Sarker, M. Faruque, H. Alqahtani, and A. Kalim, “K-nearest neighbor learning based diabetes mellitus prediction and analysis for eHealth Services,” *ICST Transactions on Scalable Information Systems*, p. 162737, 2018.

3.9 Summary

In this chapter, the high-level requirements and design of a system that utilizes a Raman spectrometer for non-invasive blood glucose monitoring were discussed. The first part of the chapter focused on providing a more in-depth description of Raman spectroscopy's components and alignments, as well as our difficulty factors and calibration methodologies. After that, the proposed solution was dissected in detail in the following sections: our substrate, instrumentation and design, data collecting, and deep learning development. In Chapter 4, which details the implementation of the blood glucose level analysis by deep learning, which led to the analysis of non-invasive diabetes detection utilizing SERs substrate, the outcomes of the low-cost DIY Raman spectrometer are discussed in additional detail.

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

In this chapter, we present the comprehensive results of the optimization process, which include representative Raman spectra of various parameters, as well as the signal-to-noise ratio and detection limit of the method. We also describe the optimization of parameters to achieve the best performance of our DIY Raman spectrometer. Despite our best efforts, the results of our system were not satisfactory for detecting Raman scattering signals. Therefore, we will provide a detailed discussion of each parameter and component, highlighting their contributions to the system's limitations and challenges. These results provide important insights into the performance of the detection system and the potential difficulties of using Raman spectroscopy for glucose detection.

4.2 The DIY Raman spectroscopy core system design

The system used in this study consists of two main parts: the Raman part and the spectrometer part, as shown in Figure 4-1. Originally, we planned to use a fiber laser but later decided to use a laser diode instead. The Raman optical part consists of five components. The laser diode we used required collimation, which we achieved using an aspherical lens (A230TM-B). However, due to limitations in the experiment, we needed to extend the beam size, so we added a beam expander. Our dichroic mirror was an 801 nm edge BrightLine single-edge standard epi-fluorescence dichroic beamsplitter (FF801-Di02), which reflects wavelengths between 450 nm and 790 nm and transmits wavelengths between 813 nm and 1100 nm. We later replaced this mirror with an ultrafast-enhanced silver mirror (UM10-AG) with a coating range of 750 nm to 1000 nm. The objective lens we used was a 10X Olympus Plan Achromat Objective (RMS10X), which we later changed to a 20X Olympus Plan Achromat Objective (RMS20X) to increase the laser power at the sample. Despite these modifications, we encountered some limitations in the experiment, particularly concerning the beam size and detector sensitivity, which affected the quality of our results. The spectrometer part of our system includes a slit, a diffraction grating, a collimating lens, a focusing lens, and a detector. We added the long pass filter BLP01-830R-25, which has an average passband transmission range of 93% at 853.2–1600 nm. Before the beam comes through the slit, we use an achromatic doublet (AC254-080-B) for focusing the lens. The slit is S50K - Ø1" mounted slit, 50 ± 3 µm wide, and the beam collides with AC254-050-B before passing through the grating. The detector (BFLY-PGE-31S4M/C-C) then picks up the Raman scattering after it diffracts on the reflective diffraction grating (GR25-1208). However, we have encountered many unexpected variables from the alignment and calibration process. Therefore, we have four changing concept or the component of the systems design in total.

4.3 The system design, alignment and, calibration.

As the Raman spectroscopy consists of two main parts. The Raman system parts can be further subdivided into 2 parts as will be described below.

4.3.1 Laser Source or Excitation source of Raman scattering.

We are using a laser diode as the nature of light to diverge over a certain distance. The divergent laser light can decrease the working distance and decrease the output of laser power which can further decrease the Raman scattering chance to occur. With this the laser beam is needed to be collimated to improve beam size and power. This setup can help ensure that the beam is collimated to have a longer working distance over the optical system to decrease the signal loss as result for undesirable focusing length.

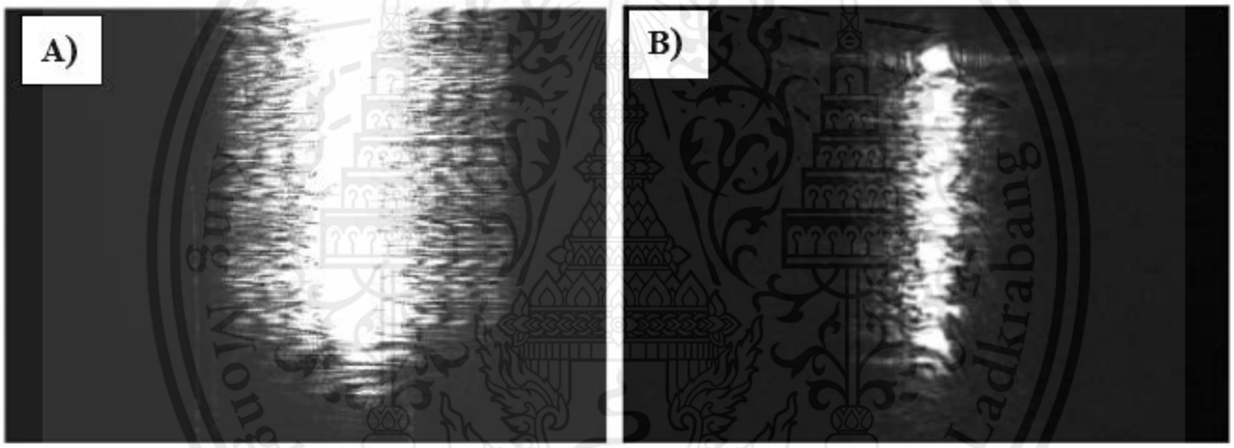


Figure 4-1. A) Laser source without collimation calibration. B) Laser source with collimation calibration.

beam divergence formula

$$\Theta = \frac{Dm}{f}$$

Θ = Divergence Angle of a laser diode.
f = Focal Length of a collimating lens.
Ø = Beam Diameter, D_m = Collimated Beam Radius = Ø/2

Beam collimation output is using a lens with focal length *f*, then the result will be a beam with a radius $D_m = \Theta/2 \times f$.

Beam expander magnification formula

$$M = \frac{f2}{f1}$$

M = beam expansion ratio or magnification of input beam.
$f1$ = Focal Length of an input lens.
$f2$ = Focal Length of an output lens.

Formula for calculating the beam expansion ratio or magnification is the same for both Keplerian and Galilean designs. The beam expansion ratio equals the focal length of the output lens divided by the focal length of the input lens. Lens 1 the input lens of incident light, whose focal length is $f1$. The second lens (Lens 2) with a focal length of $f2$ and the beam expansion ratio (M).

4.3.1.1 The first system design

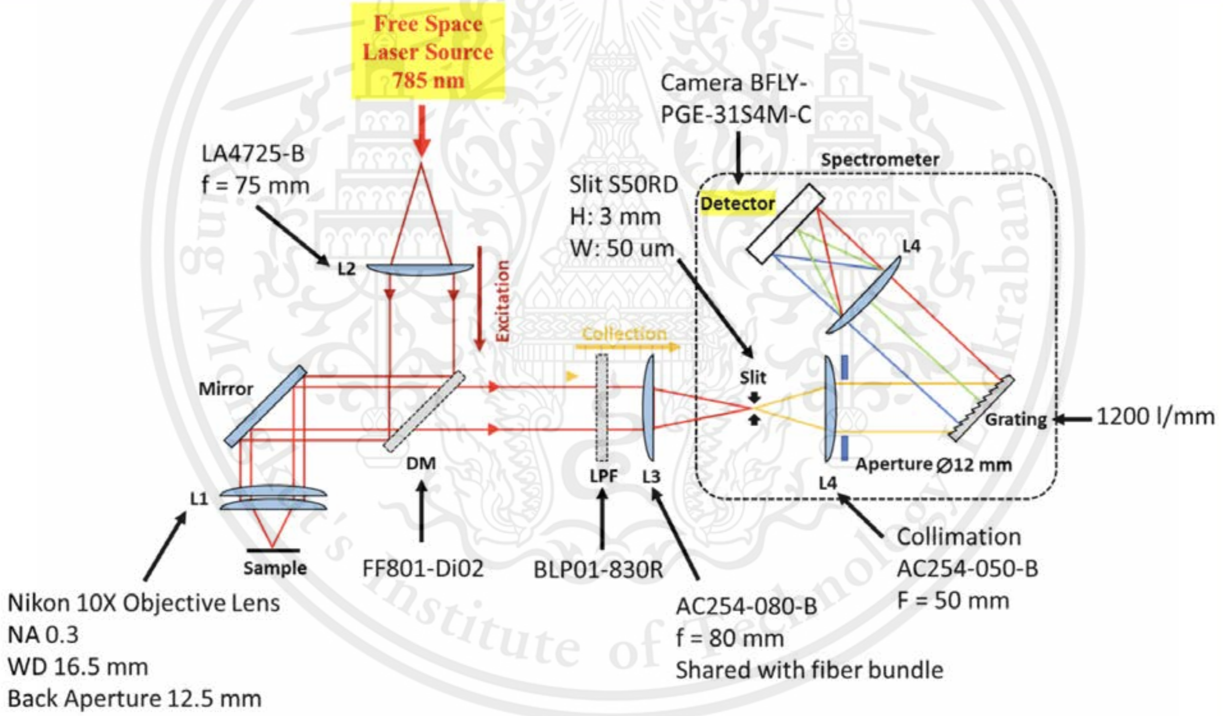


Figure 4-2. 2D Diagram: Design-1

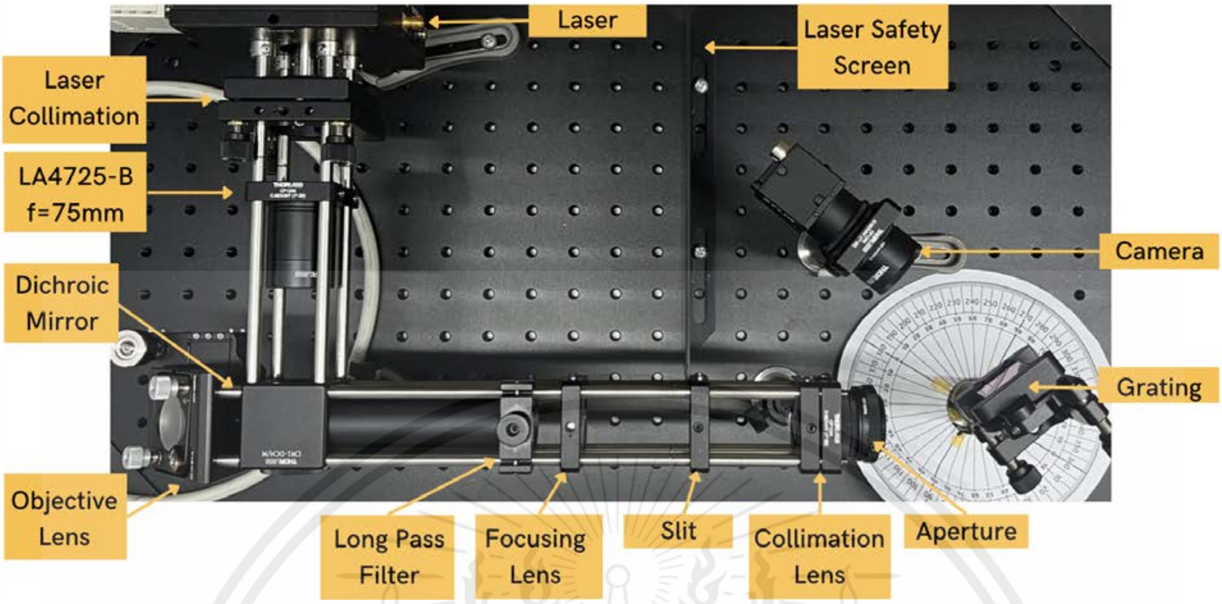


Figure 4-3. System: Design-1

Calculate the beam size of system design-1.

1) LA4725-B lens: focal length of 75 mm
2) L785P090 Laser diode: divergence angle of 17 degree (halves = 8.5 degree)

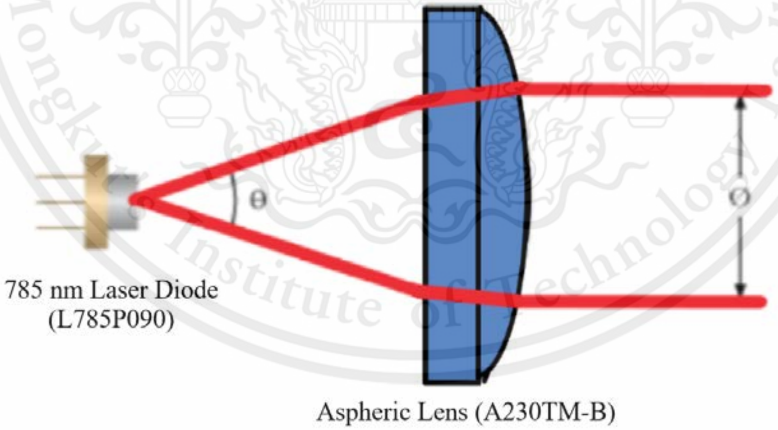


Figure 4-4. Beam size calculation diagram for system-1

$$\tan(8.5°) = \frac{Dm}{75\,mm} = Dm = 11.2\,mm$$

$$Dm = \frac{Ø}{2} = Ø = 11.2x2 = 22.4\,mm$$

From this calculation of a collimated laser with a diameter of 22.4 mm. We have done some measurements with a laser to approximate the real diameter of this collimated beam being around 8 mm in diameter. The results using the LA4725-B lens have not been very useful as the lens itself has too far focal length and too large beam size resulting in power drop and it is almost impossible to collimate a laser light with this LA4725-B lens so we take this problem into consideration to further develop our Raman system.

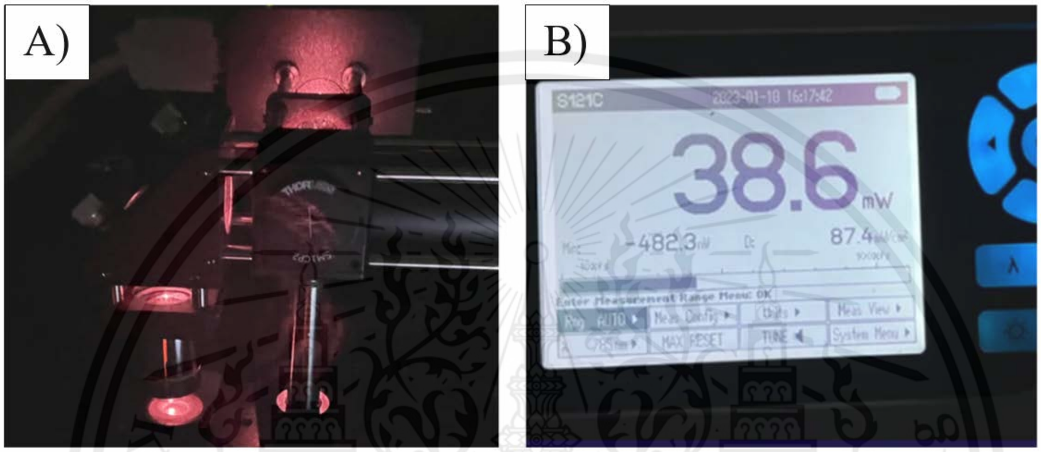


Figure 4-5. The setup of design-1 Raman system A) Overall system with the beam that scattering around the system, B) Result from measuring laser power at objective lens with 38.6 mW from 80 mW output.

4.3.1.2 The second system design

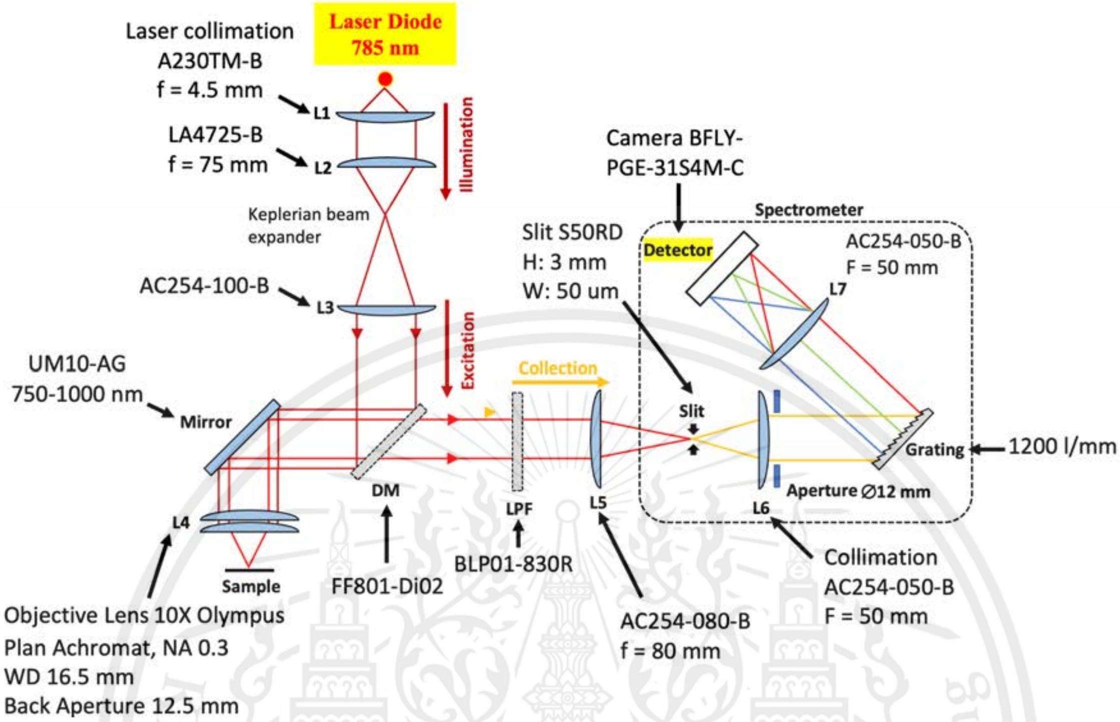


Figure 4-6. 2D Diagram: Design-2

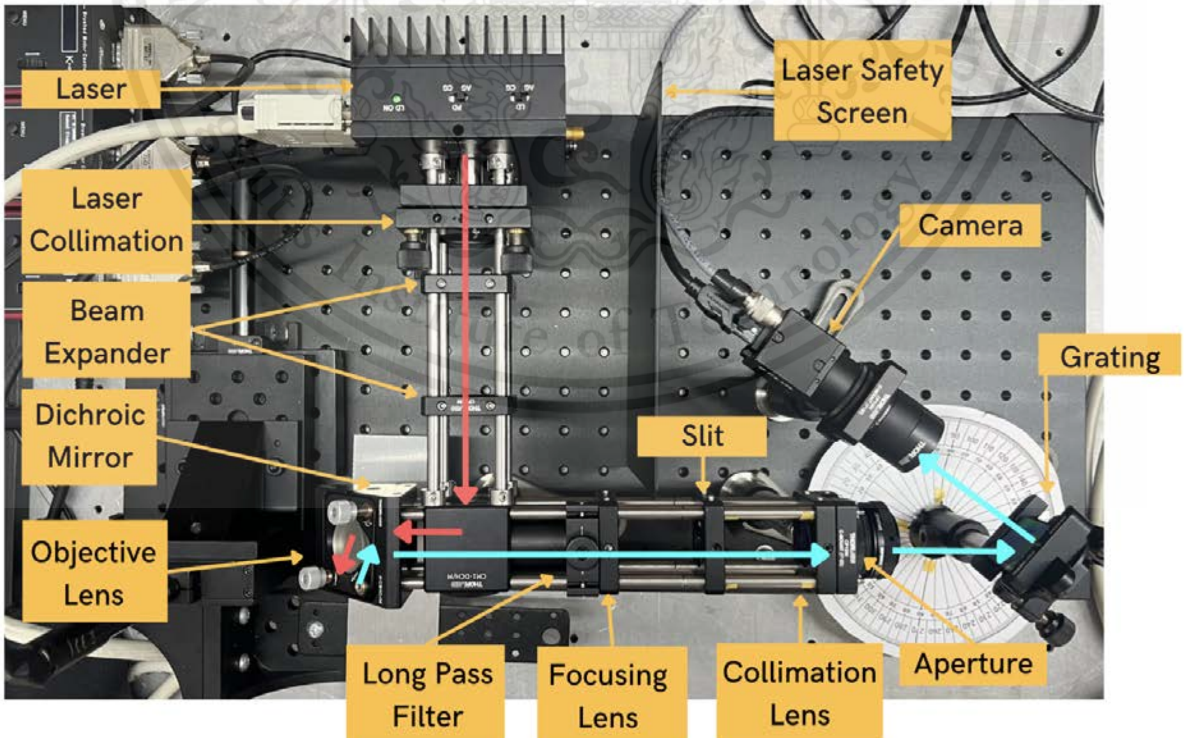


Figure 4-7. System: Design-2

Calculate the beam size of system design-2.

1) A230TM-B lens: focal length of 4.51 mm.
2) L785P090 Laser diode: divergence angle of 17 degree (halves = 8.5 degree).
3) Beam expander first lens LC4210-B: focal length of -25 mm.
4) Beam expander second lens LA4725-B: focal length of 75mm.

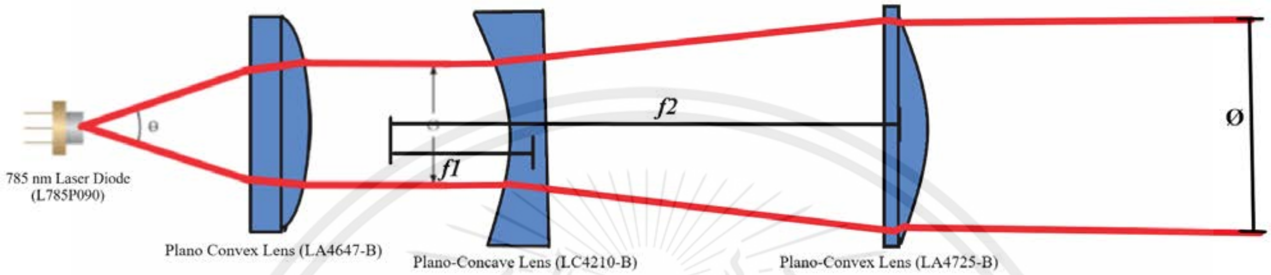


Figure 4-8. Beam size calculation diagram for system-2

$$tan(8.5°) = \frac{Dm}{4.51\,mm} = Dm = 0.647\,mm$$

$$Dm = \frac{Ø}{2} = Ø = 0.647\,x\,2 = 1.348\,mm$$

With beam expander of LC4210-B and LA4725-B

$M = \frac{f2}{f1} = \frac{75\,mm}{25\,mm} = 3$ times of magification = 1.348 mm x 3 = 4.04 mm

Figure 4-9. Laser collimation for system design-2 using a Galilean beam expander.

From Galilean beam expander setup calculations of a collimated laser with a diameter of 4.04 mm We have done some measurements with a laser to approximate the real diameter of this collimated beam being around 4 mm in diameter. The results using the beam expander lens have been better as the lens itself can collect more light resulting in more power increasing at the objective lens, the working distance does go up as the beam is much more collimated using a better collimating lens.

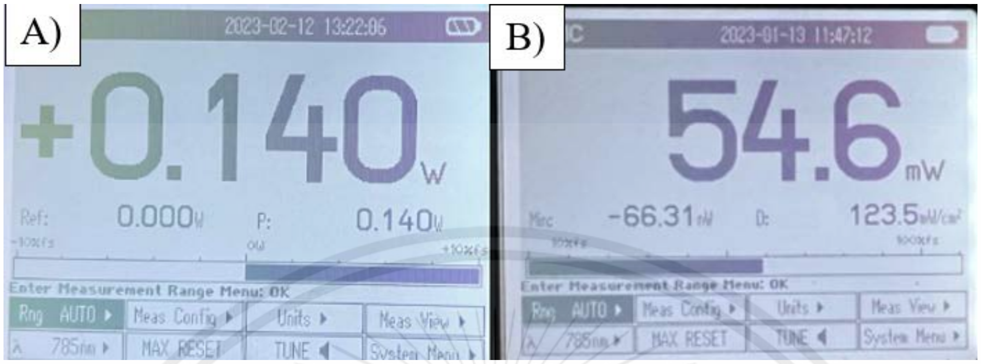


Figure 4-10. A) Laser collimation input power of 140 mW , B) Laser output power of 54.6 mW at the objective lens.

4.3.1.3 The third system design

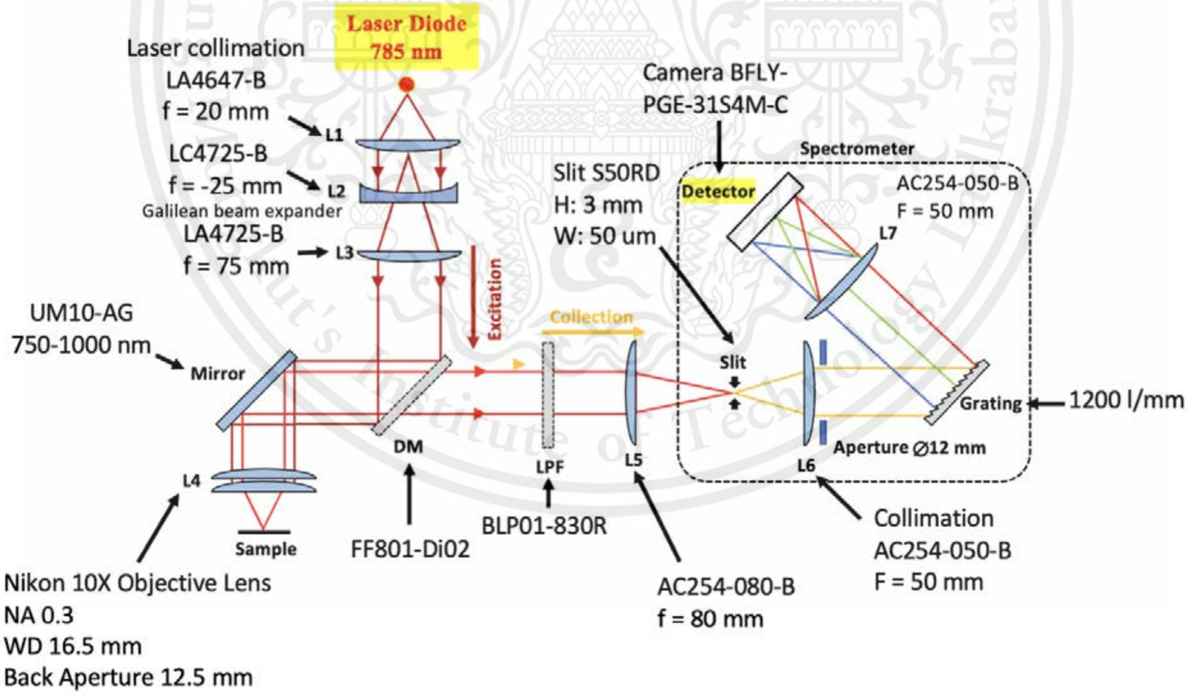


Figure 4-11. 2D Diagram: Design-3

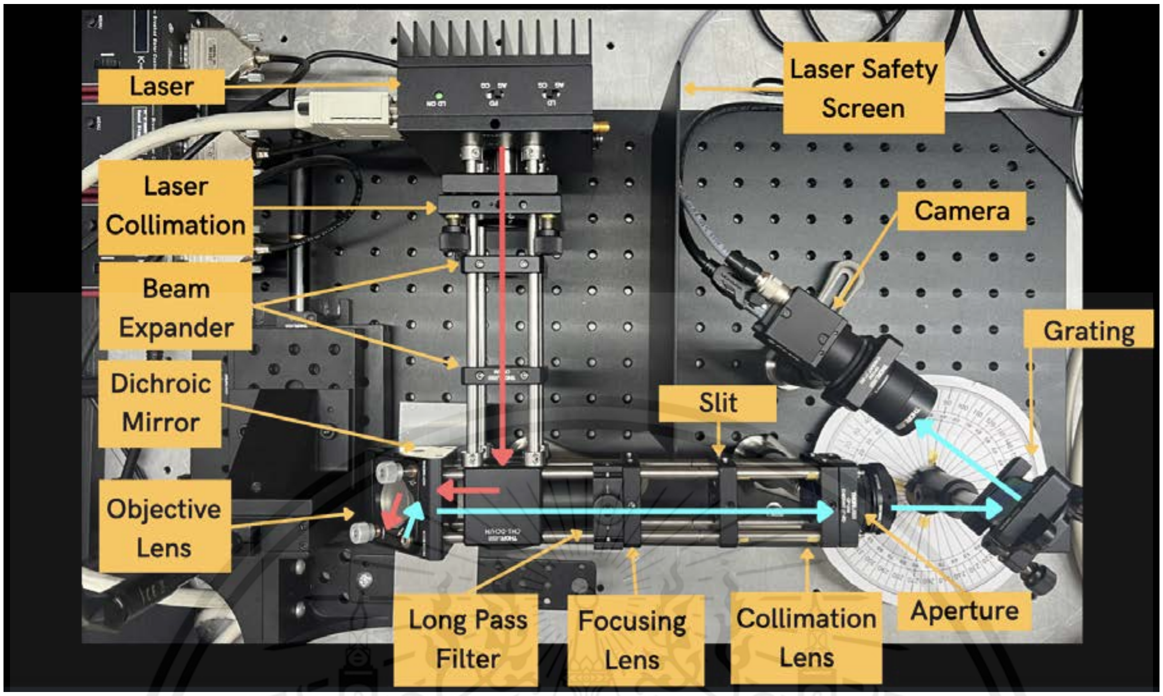


Figure 4-12. System: Design-3

Calculate the beam size of system design-3.

1) A230TM-B lens: focal length of 4.51 mm
2) L785P090 Laser diode: divergence angle of 17 degree (halves = 8.5 degree)
3) Beam expander first lens LA4725-B: focal length of 75 mm
4) Beam expander second lens AC254-100-B: focal length of 100 mm

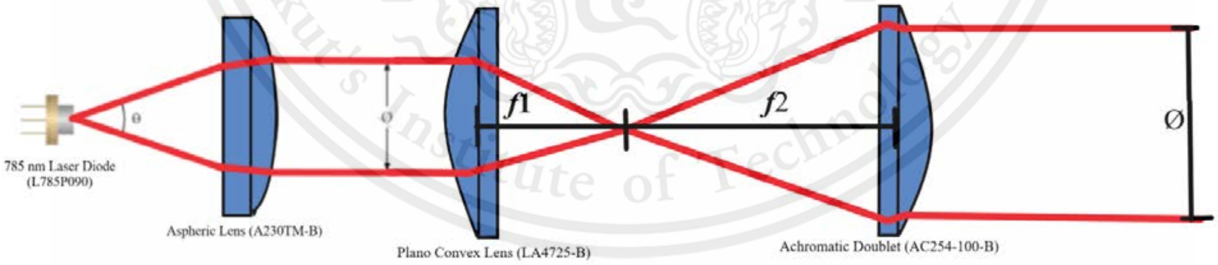


Figure 4-13. Beam size calculation diagram for system-3

$$tan(8.5°) = \frac{Dm}{20\ mm} = Dm = 0.674\ mm$$

$$Dm = \frac{Ø}{2} = Ø = 0.674\ x\ 2 = 1.348\ mm$$

With beam expander of LC4210-B and LA4725-B

$M = \frac{f2}{f1} = \frac{100\ mm}{75\ mm} = 3$ times of magification = 1.348 mm x 1.33 = 1.793 mm

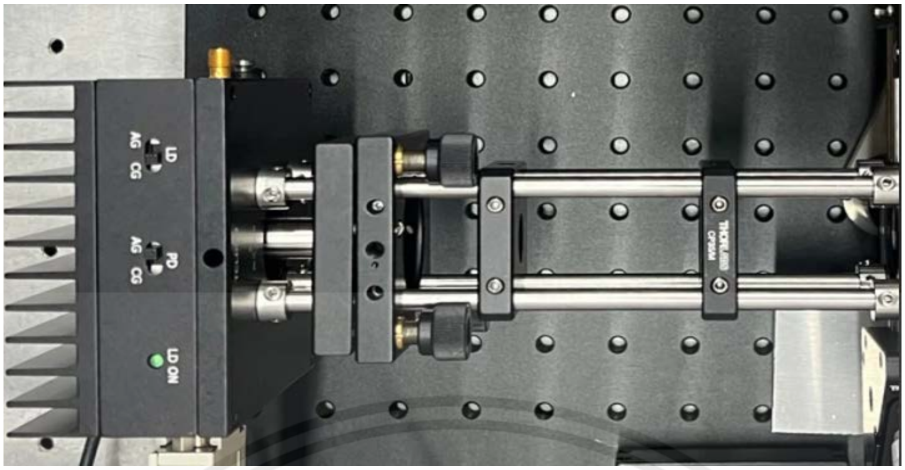


Figure 4-14. Laser collimation for system design-3 using a Keplerian beam expander.

From Keplerian beam expander setup calculation for a collimated laser with a diameter of 1.793. We have done some measurements with a laser to approximate the real diameter of this collimated beam being around 5 mm in diameter. The results using the beam expander lens have a better result compared to galilean as the beam size resulting in power increasing at the objective lens, the working distance does go up as the beam is much more collimated using a better collimating lens. so we take this problem into consideration to further develop our Raman system. To further the proper collimation system we ordered the new lens that is more suitable for our system of AC080-020-B with the new laser of L785H1.

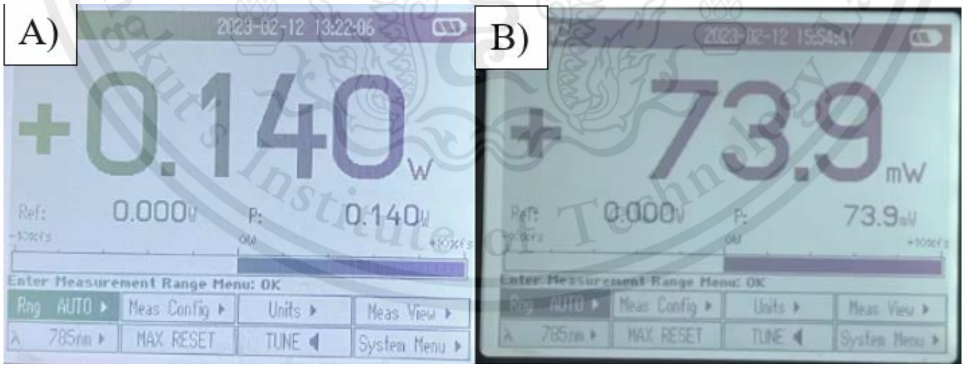


Figure 4-15. A) Laser collimation input power of 140 mW , B) Laser output power of 73.9 mW at the objective lens.

4.3.1.4 The fourth system design

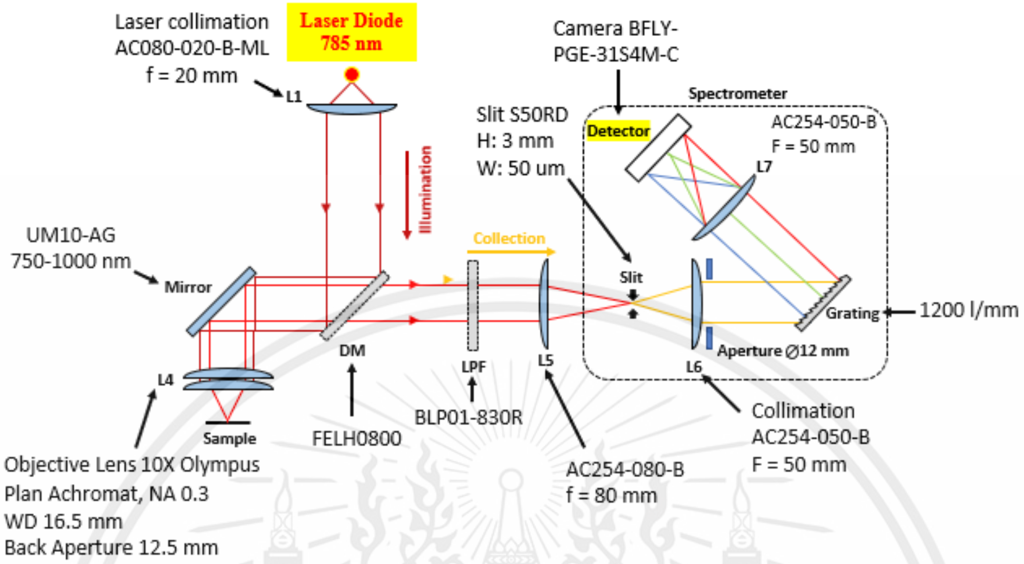


Figure 4-16. 2D Diagram: Design-4

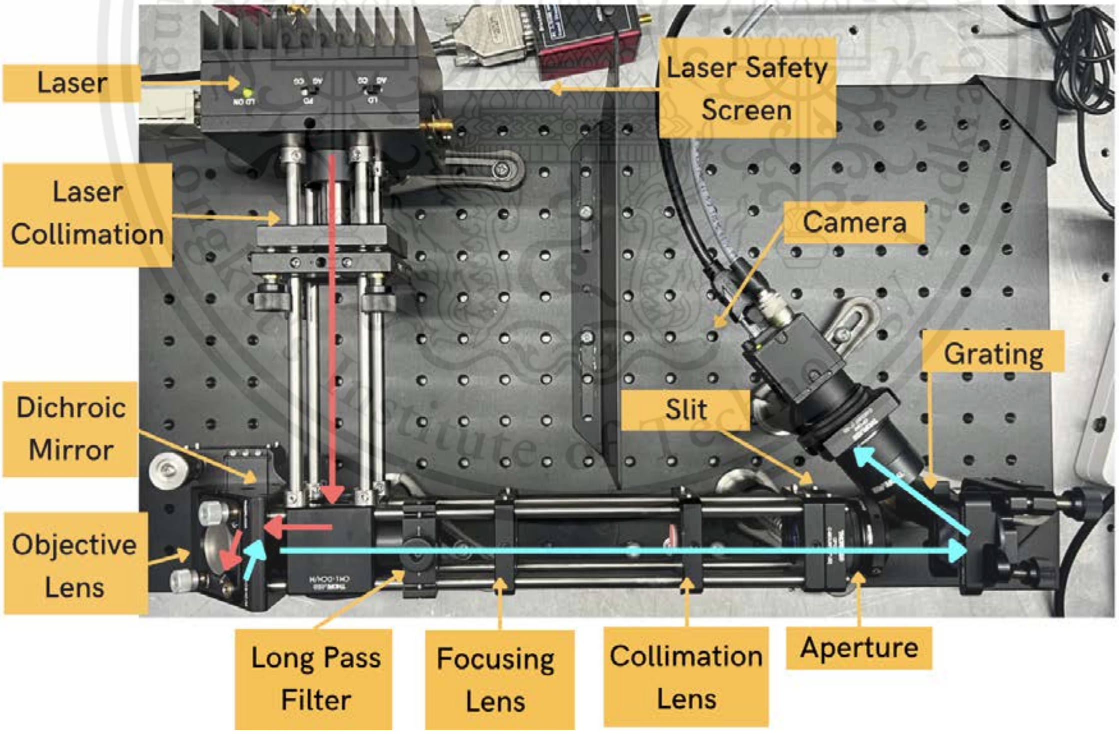


Figure 4-17. System: Design-

Calculate the beam size of system design-4.

1) AC080-020-B lens: focal length of 20 mm
2) L785H1 Laser diode: divergence angle of 16 degree (halves = 8 degree)

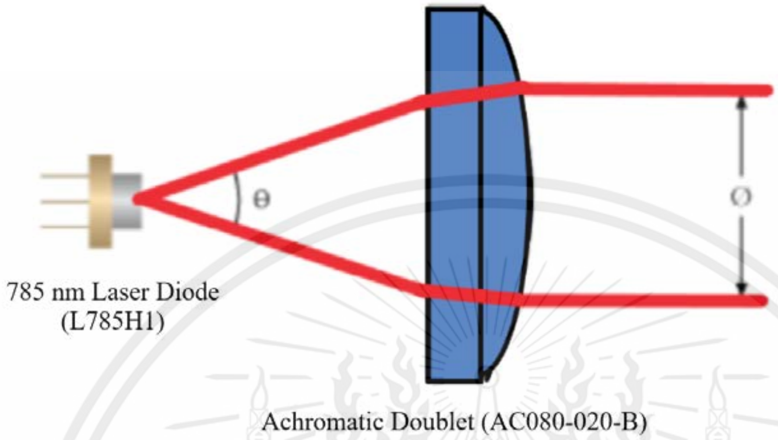


Figure 4-18. Beam size calculation for system-4

$$tan(8°) = \frac{Dm}{20\ mm} = Dm = 2.81\ mm$$

$$Dm = \frac{\varnothing}{2} = \varnothing = 2.81x2 = 5.62\ mm$$

From this calculation of a collimated laser with a diameter of 5.62 mm. We have done some measurements with a laser to approximate the real diameter of this collimated beam being around 6 mm in diameter with a very uniform light beam.

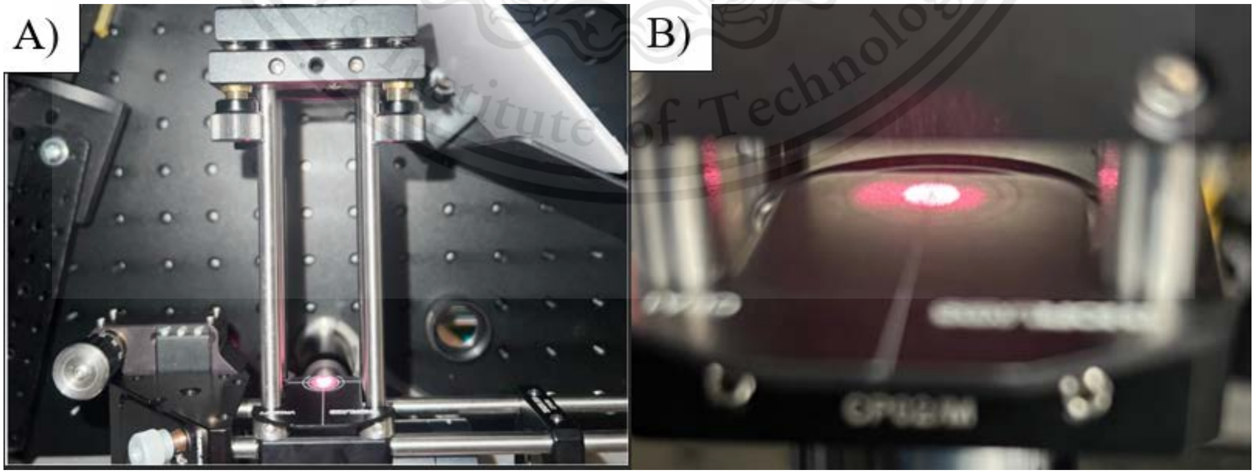


Figure 4-19. A) Laser collimation for system design-4 , B) Using a test target to measure a beam size with the approximation to be around 6 mm.

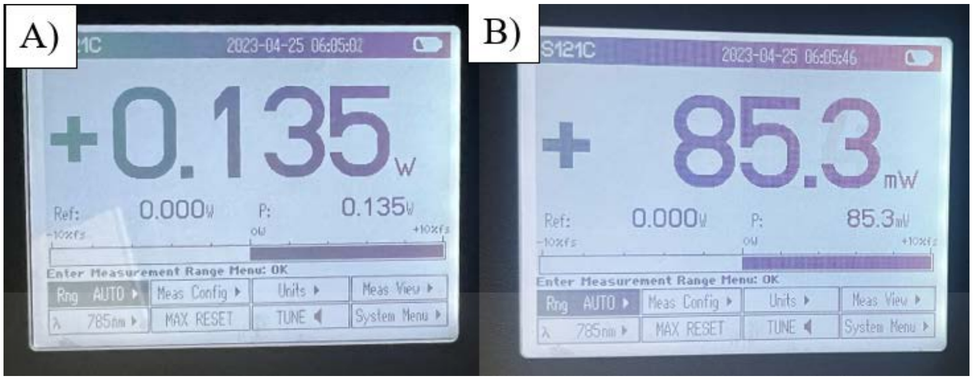


Figure 4-20. A) Laser collimation input power of 135 mW , B) Laser output power of 85.3 mW at the objective lens.

4.3.1.5 Raman Sampling for the Raman System.

We have used a configuration of 180 degree for Raman collection as a reflection from the slide holder controlled by XYZ stage to find the most optimized height ensuring the laser is focused at the sample.

Optimal distance for a focusing distance.

We have installed a XYZ stage to control location and position of the aluminum stage to find the most optimal height of the objective lens target stage which will start from 0 mm (home) to 6 mm closer to an objective lens. Using both images quality of a laser light with the graph plotting to see the maximum peak value from all of the results to find the most optimal distance.

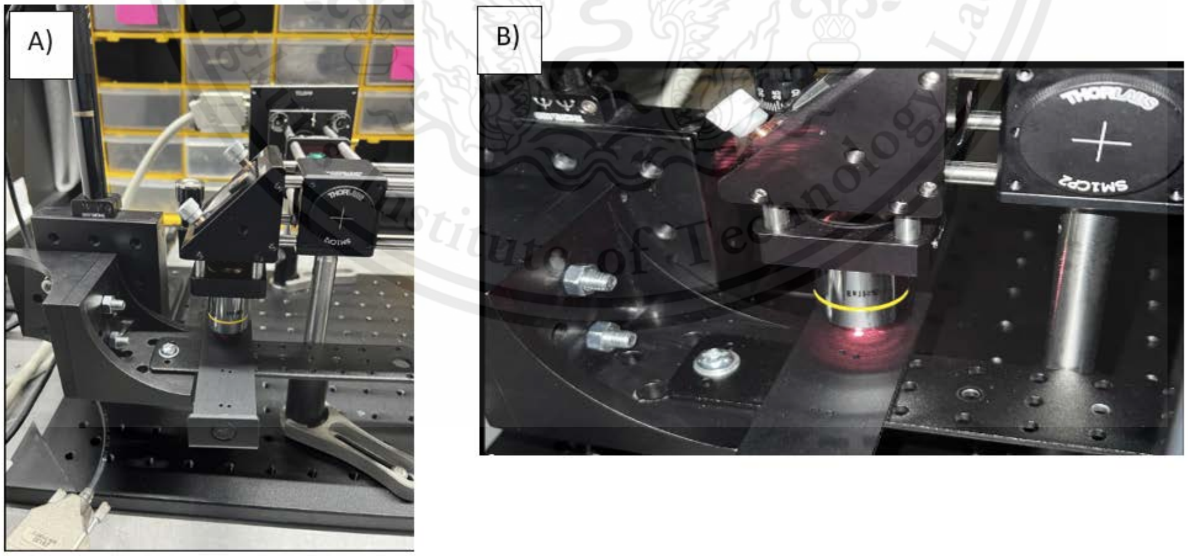


Figure 4-21. The setup used to find the focusing distance A) installation of XYZ stage with a raman system, B) Laser testing in general with XYZ stage.

The result is that as the stage moves closer the image and maximum peak value is increasing to a point where the laser is focused but after that the value will decrease a bit as it will start to be off focusing.

The result for original position in z-axis (0 mm).

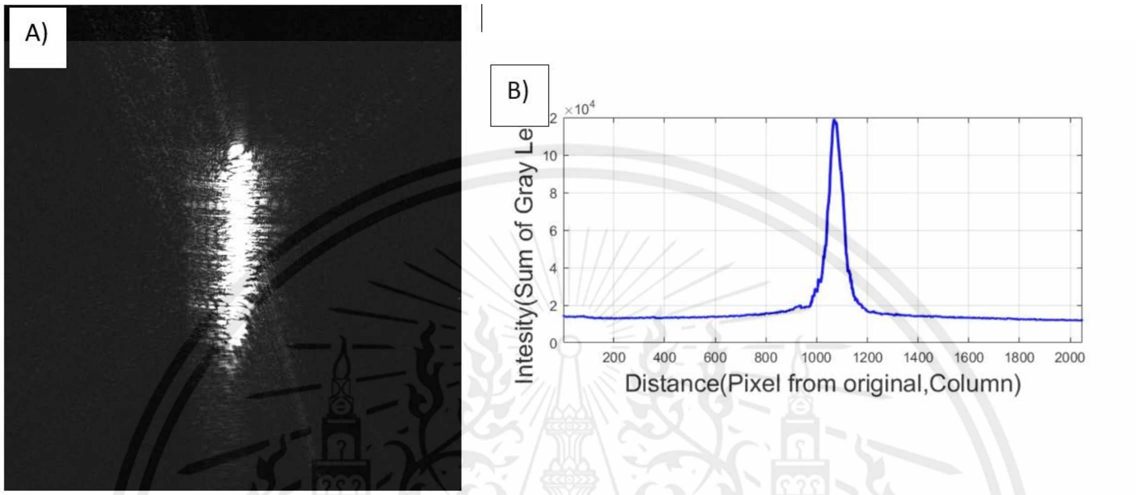


Figure 4-22. The setup used to find the focusing distance A) Image capturing from camera, B) Graph plotting using Matlab finding maximum peak value.

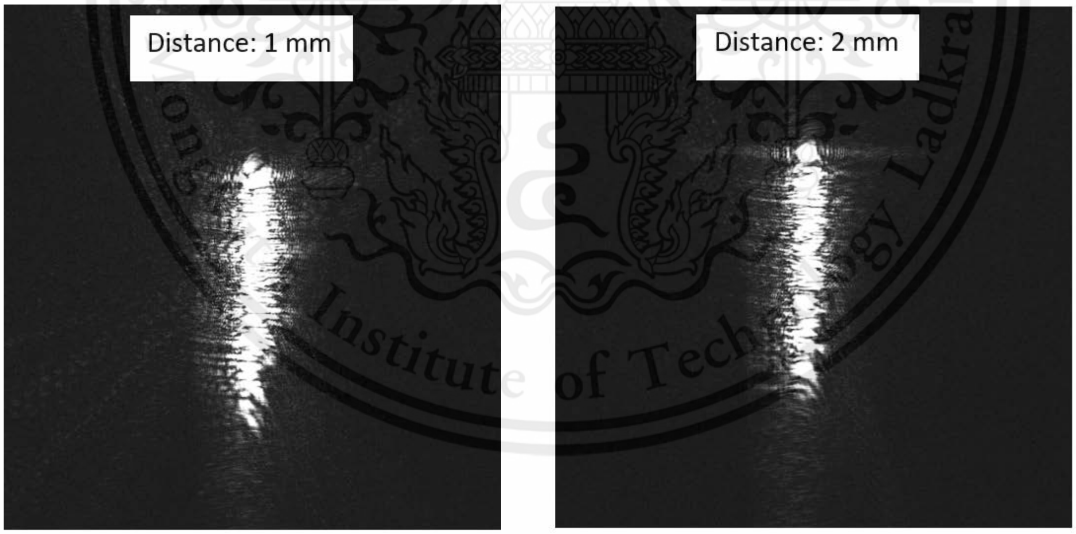


Figure 4-23. Image for distance 1 mm and 2 mm using only aluminum plate stage.

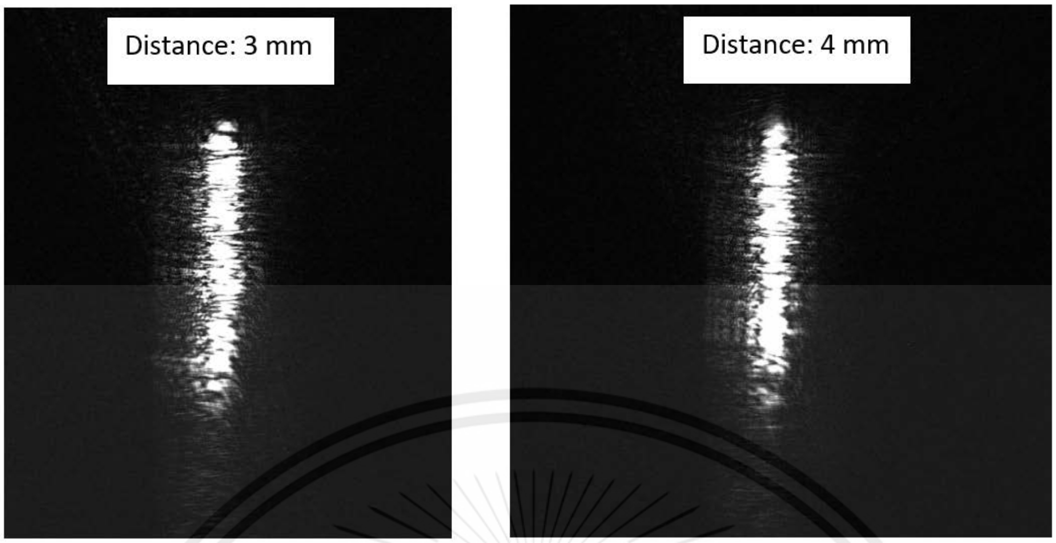


Figure 4-24. Image for distance 3 mm and 4 mm using only aluminum plate stage.

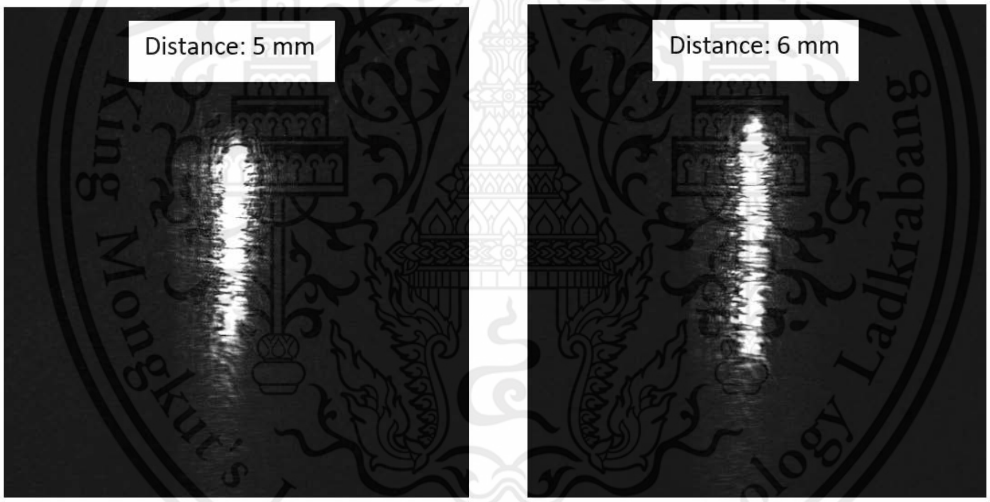


Figure 4-25. Image for distances 5 mm and 6 mm using only aluminum plate stage.

Based on all of the captured images it is a hard decision to find the most suitable distance using only image observation, so we will use a graph plotting from Matlab to see the maximum value.

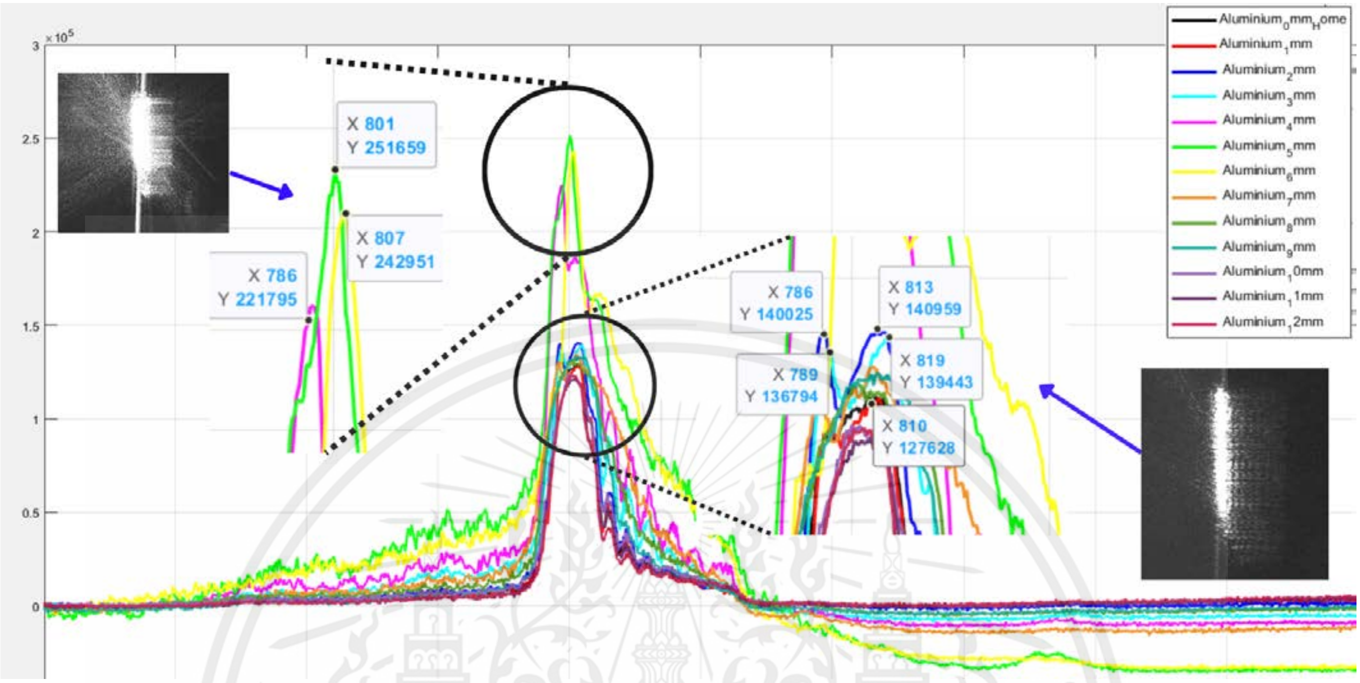


Figure 4-26. Diagram for comparing a 2 difference system signal.

The results show that the 5 mm shows a better resolution than the other result and also has a decent intensity value as the highest value. 5 mm is the standard height when we are dealing with the current Raman system as it is the most optimal height for now.

4.3.2 Spectrometer

Spectrometers are the other parts of Raman spectroscopy as the Raman scattering light being collected also needs to be diffracted at a certain angle. Using Reflective grating we have to calculate for the angle of incident light that can diffracted a 785 nm wavelength upward range as with too less or too large angle the diffraction light will be at off angle and 785 nm wavelength will not be detected by camera.

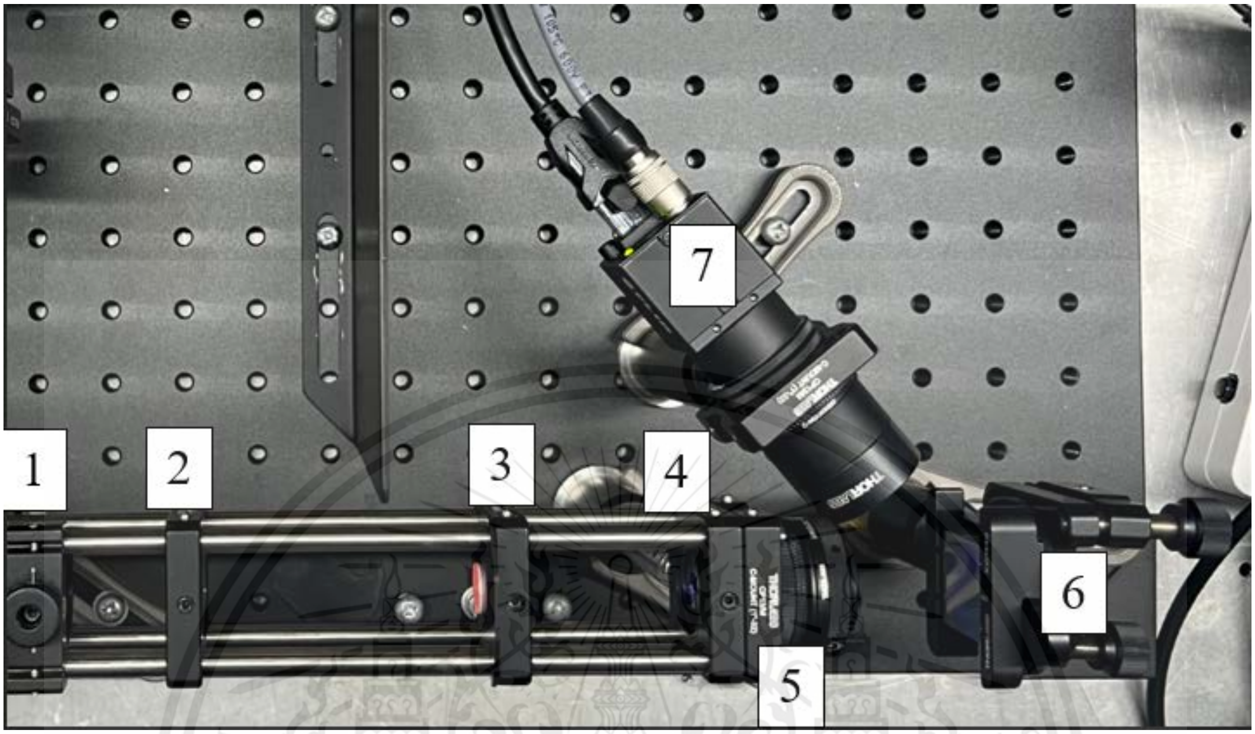


Figure 4-27. The parts of the system called a spectrometer parts 1) 805 nm Long pass filter (FELH0800) 2) focusing lens (AC254-080-B) 3) slit of 50 µm Wide, 3 mm Long 4) Collimating lens (AC254-050-B) 5) Aperture (SM1D12C) 6) Grating of 1200 lines/mm (GR25-1208) 7) CMOS Detector (Camera BFLY-PGE-31S4M-C)

Setup for grating experiment.

1) Laser diode source: 785 nm (L785P090).
1) Collimation lens: Achromatics doublet (AC254-050-B).
2) Light gratings (GR25-1208: 1200 Lines/mm).
3) Reflection surface

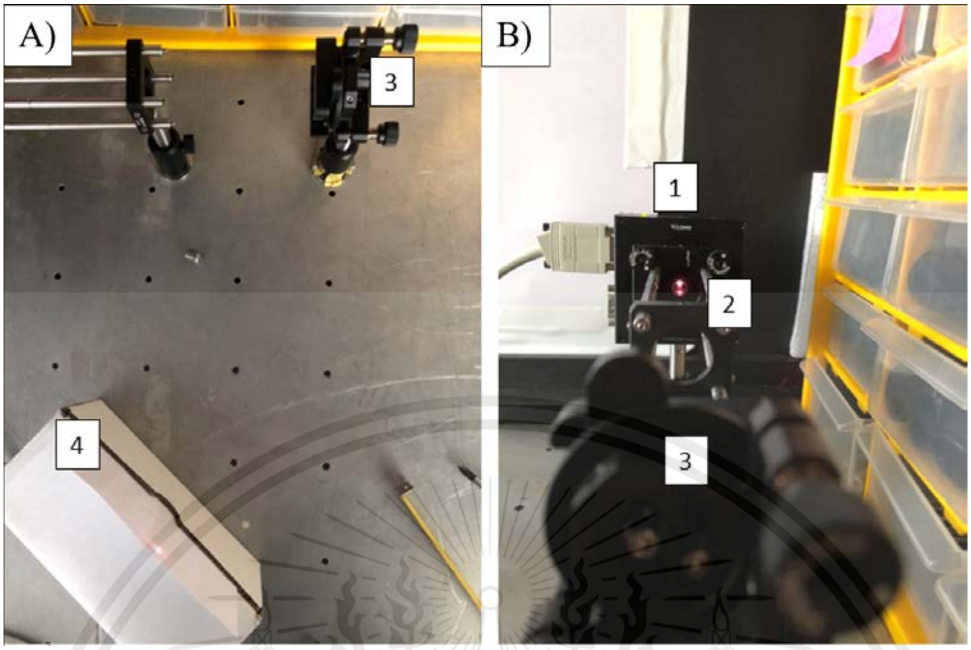


Figure 4-28. Setup for grating experiment: A)Set up from the top , B)Set up from the side. 1) Laser diode source: 785 nm (L785P090), 2) Collimation lens: Achromatics doublet (AC254-050-B), 3) Light gratings (GR25-1208: 1200 Lines/mm), 4) Reflection surface.

For this experiment we will try to calculate the wavelength of our laser using grating using a diffraction equation.

Diffraction equation

- $d\sin(\theta) = n\boldsymbol{\lambda}$

  d = line spacing (distance between slits)
  θ = diffraction angle
  $\boldsymbol{\lambda}$ = wavelength of incident light
  n = diffraction order

- $\tan(\theta) = \frac{distance\ from\ zero\ order\ to\ first\ order\ diffraction}{distance\ from\ grating\ to\ reflected\ surface}$

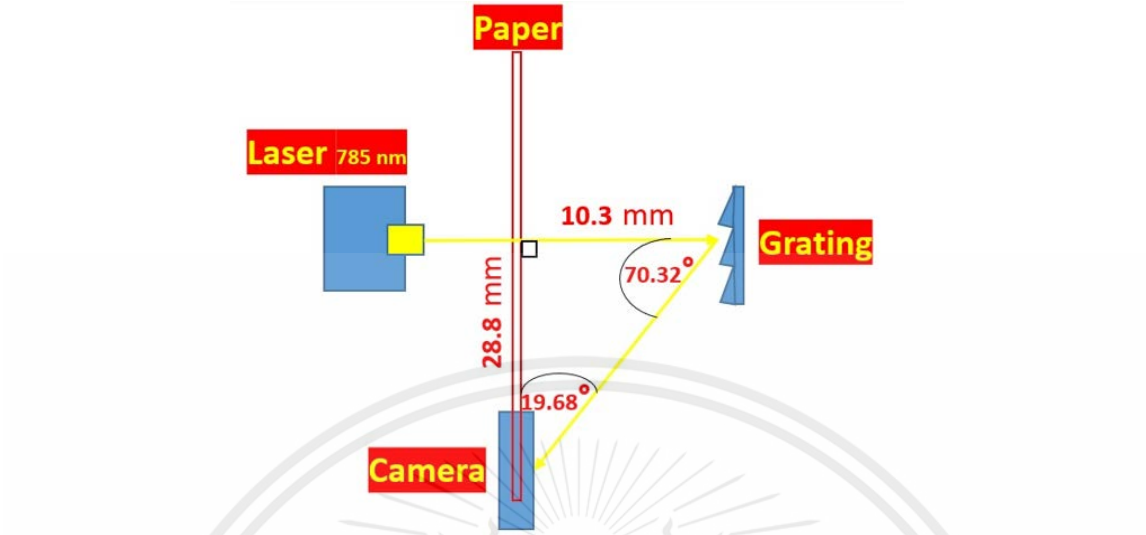


Figure 4-29. Diagram from the set up with angle and distance required for calculation.

From the light diffraction experiment

$$d \sin(\theta) = n\boldsymbol{\lambda}$$

$$d = 1200 \text{ lines/mm} = 1200/10^{-3} = 1/1.2\text{x}10^{6} = 8.33\text{x}10^{-7} \text{ m}$$

$$\theta = tan^{-1}\left(\frac{distance\ from\ zero\ order\ to\ first\ order\ diffraction}{distance\ from\ grating\ to\ reflected\ surface}\right) = tan^{-1}\frac{28.8\ mm}{10.3\ mm} = 70.32$$

$$d \sin(\theta) = n\boldsymbol{\lambda}$$

n = 1 as first order of light diffraction.

$$\boldsymbol{\lambda} = \frac{sin(70.32)\ x\ 8.33x10^{-7}}{1} = 784.34 \text{ nm [error : } \frac{785-784.32}{785} = 0.087\% \text{]}$$

From the new setup the wavelength that is being calculated is much closer to the real laser wavelength with only an error of 0.087%. With this we can use this equation to determine the angle that wavelength will be shown. Then we can use a detector to receive a Raman signal that chooses the right range of wavelengths that will be significant for our signal dectection.

Grating equation

$$n\lambda = d(\sin\theta r + \sin\theta i)$$

d = line spacing (distance between slits).

θr = first order diffraction angle.

θi = Incident angle.

$\boldsymbol{\lambda}$ = wavelength of the first order diffraction.

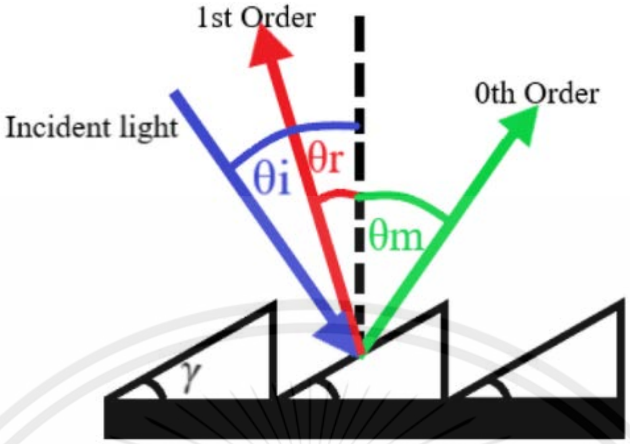


Figure 4-30. Grating diffraction order diagram.

n = diffraction order.

$n\lambda = d(\sin\theta r + \sin\theta i) = \frac{n\lambda}{d} - \sin\theta i$

$\theta r = \arcsin(\frac{n\lambda}{d} - \sin\theta i)$

d = line spacing (distance between slits): 1200 lines/mm = $\frac{1x10^3}{1200}$ = 833.33 nm

$n\lambda = d(\sin\theta r + \sin\theta i) = \frac{n\lambda}{d} - \sin\theta i$

$\theta r = \arcsin(\frac{n\lambda}{d} - \sin\theta i)$

At 30 degree $\theta r = \arcsin(\frac{1x785\ nm}{833.33\ nm} - \sin(30)) = 26.23$ degrees.

At 35 degree $\theta r = \arcsin(\frac{1x785\ nm}{833.33\ nm} - \sin(35)) = 21.62$ degrees.

At 40 degree $\theta r = \arcsin(\frac{1x785\ nm}{833.33\ nm} - \sin(40)) = 17.41$ degrees.

At 45 degree $\theta r = \arcsin(\frac{1x785\ nm}{833.33\ nm} - \sin(45)) = 13.58$ degrees.

For this calculation we will try to calculate the angle that is best for 785 nm wavelength upward to set up our grating and detector angle to ensure that the diffraction light will be collected at maximum value. We will use a setup of indicated angle to be around 30-45 degree as the calculation will show that the 785 nm diffraction will be near the surface normal of the grating that can help us to set up the detector and grating easier.

4.4 Results

4.4.1 Results from the first system design

During the first design the system on both sides of Raman parts and spectrometer parts were still not properly installed and Aligned. Resulting in images and signals that are not shown or cannot classify as a Raman spectrum.

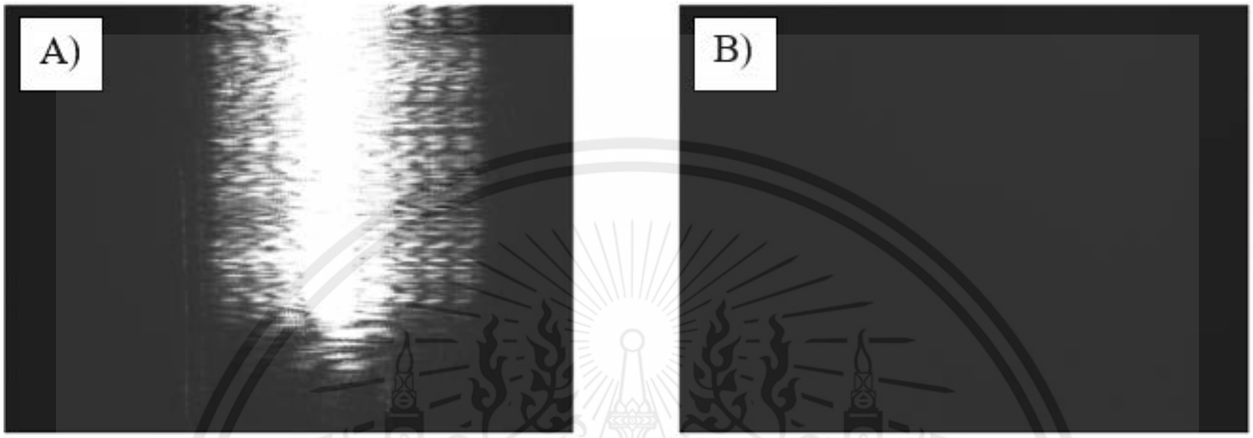


Figure 4-31. A) Glucose Powder reflection is captured while passing to the spectrometer parts with no long pass filter of 830 nm (BLP01-830R-25). B) Glucose Powder reflection is captured while passing to the spectrometer parts with a long pass filter of 830 nm (BLP01-830R-25).

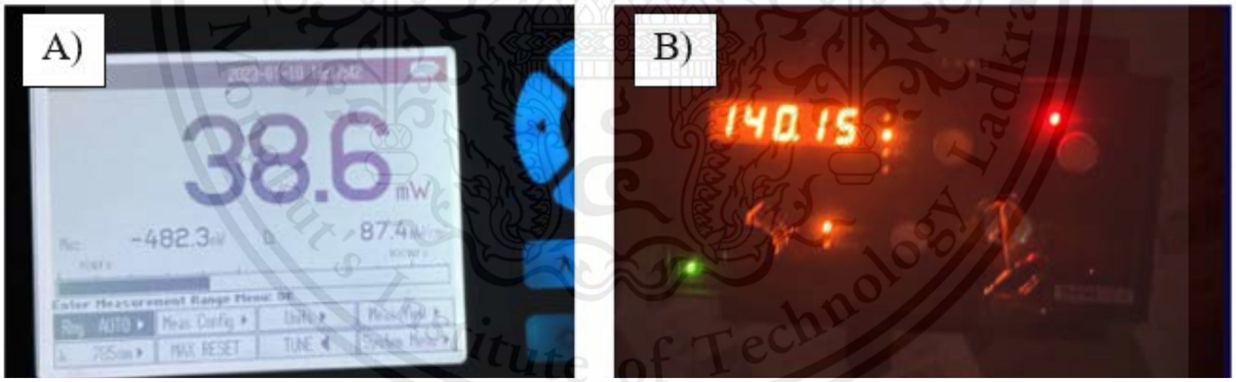


Figure 4-32. A) Power meter measurement at objective given out power of 38.6 mW, B.) Display of laser diode controller that given the current of 140.15 mA equivalent to around 80 mW.

Since the alignment of the system is not properly done the whole system is still not at the absolute best as it will be reflected from the result of image and laser power reduction. It can be seen that the raw image of the signal from the objective is not having enough clarity as shown in Figure 4-31. As in terms of power there is a power reduction from 80 mW to 38.6 mW at the objective lens that could result in lowering the Raman signal to occur.

4.4.2 Results from the second system design

During the second design the system before using the beam expander we have further done the system alignment in Raman parts that result in better quality of raw image but still the system does not show a signal of Raman spectrum as we are installing the long pass filter of 830 nm (BLP01-830R-25) in Figure 4-32 b). With this image we had assumed that the Raman signal is very weak so we had improved the system to be in second design with the addition of a beam expander.

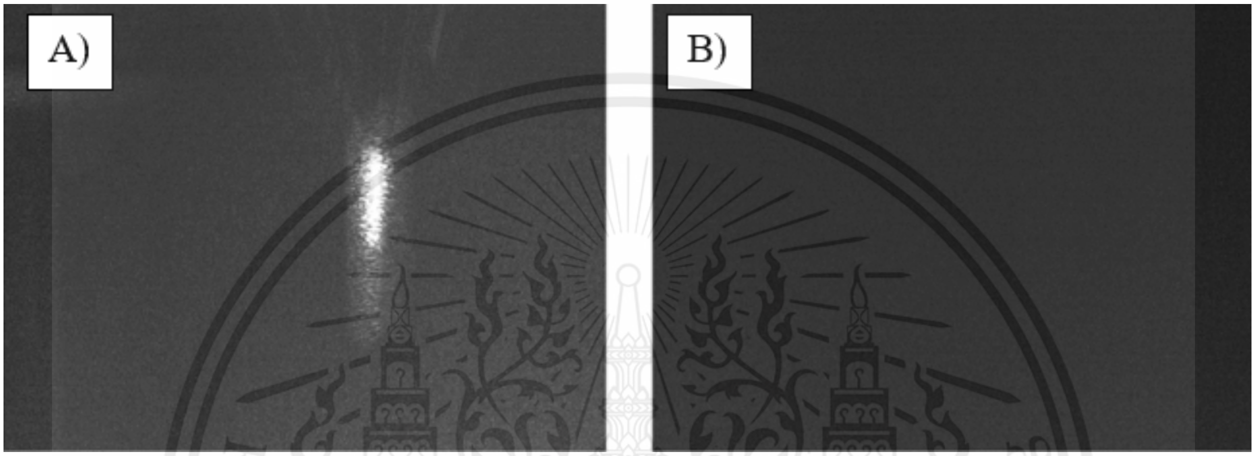


Figure 4-32. Improved system of first design A) Glucose Powder is captured while passing to the spectrometer parts with no long pass filter of 830 nm (BLP01-830R-25). B) Glucose Powder reflection is captured while passing to the spectrometer parts with a long pass filter of 830 (BLP01-830R-25).

In this second design we try to use Keplerian beam expander to improve the Raman signal and expect that a signal will appear. But still with the addition of Keplerian beam expander the signal of Raman spectrum is not yet to be detected as we are installing the long pass filter of 830 nm (BLP01-830R-25) in Figure 4-33 b). But with the beam shape and the improvement in the beam Clarification in Figure 4-33 a). With some improvement we further made a Gailiean beam expander type system to compare between the 2 beam expander types and expect a better result in our third design.

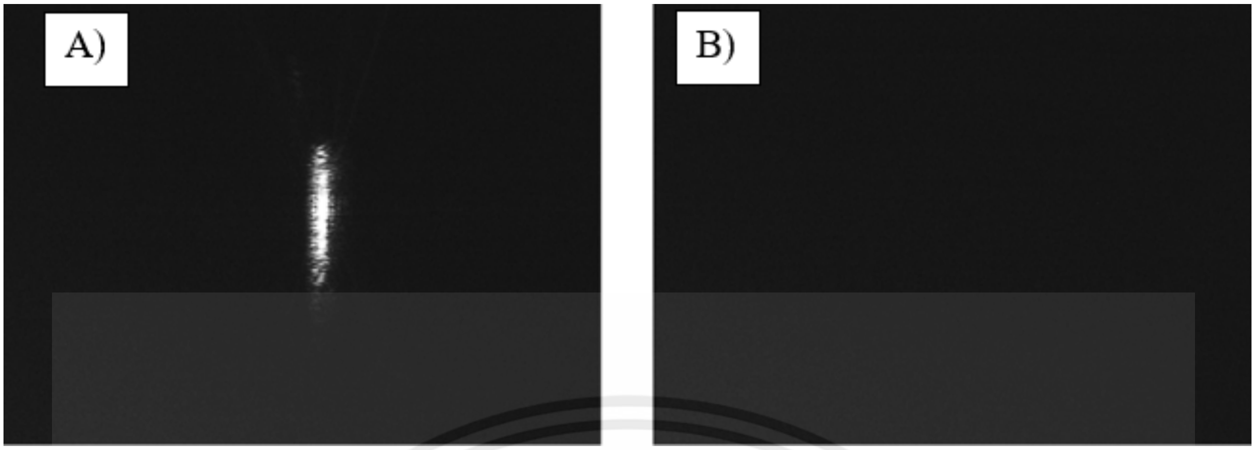


Figure 4-33. System of second design with Keplerian beam expander A) Glucose Powder reflection is captured while passing to the spectrometer parts with no long pass filter of 830 nm (BLP01-830R-25). B) Glucose Powder reflection is captured while passing to the spectrometer parts with a long pass filter of 830 (BLP01-830R-25).

4.4.3 Results from the third system design

From the second design that we had used a Kelpherain beam expander to improve the laser beam and power in this third design we have made a change in beam expander type to be Gailiean beam expander to compare the beam profile and power at the objective lens.

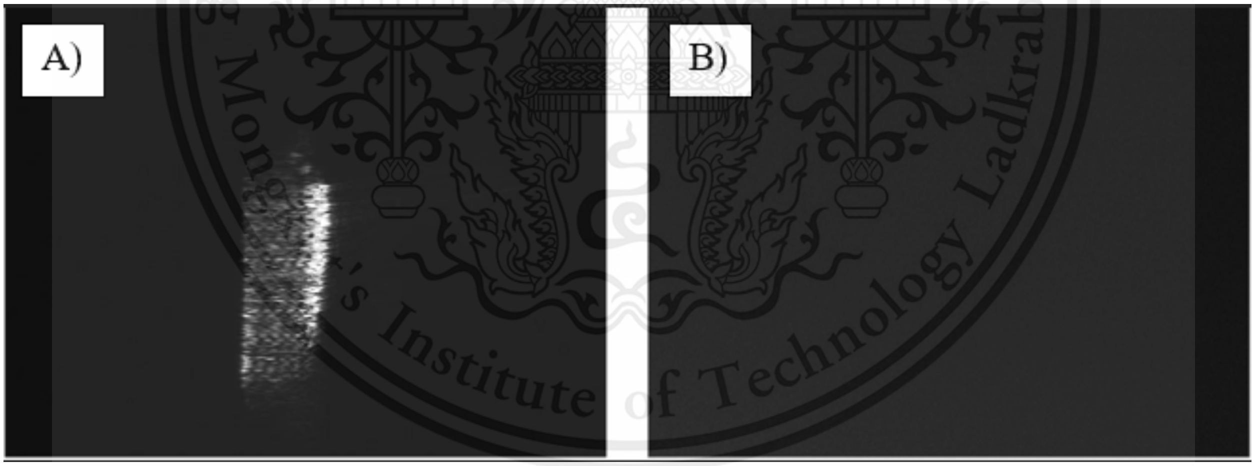


Figure 4-34. Galilean beam expander system of third design A) Glucose Powder is captured while passing to the spectrometer parts with no long pass filter of 830 nm (BLP01-830R-25) . B) Glucose Powder reflection is captured while passing to the spectrometer parts with a long pass filter of 830 (BLP01-830R-25).

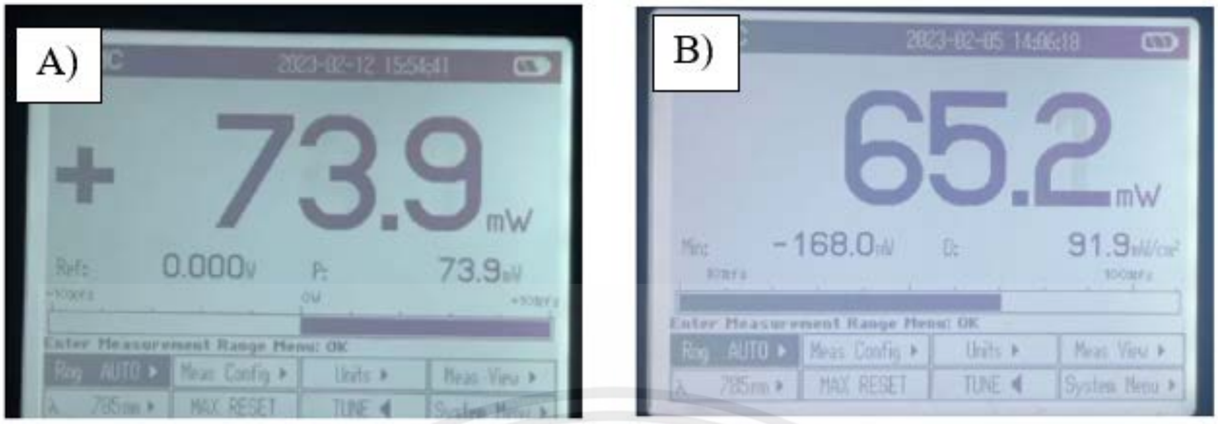
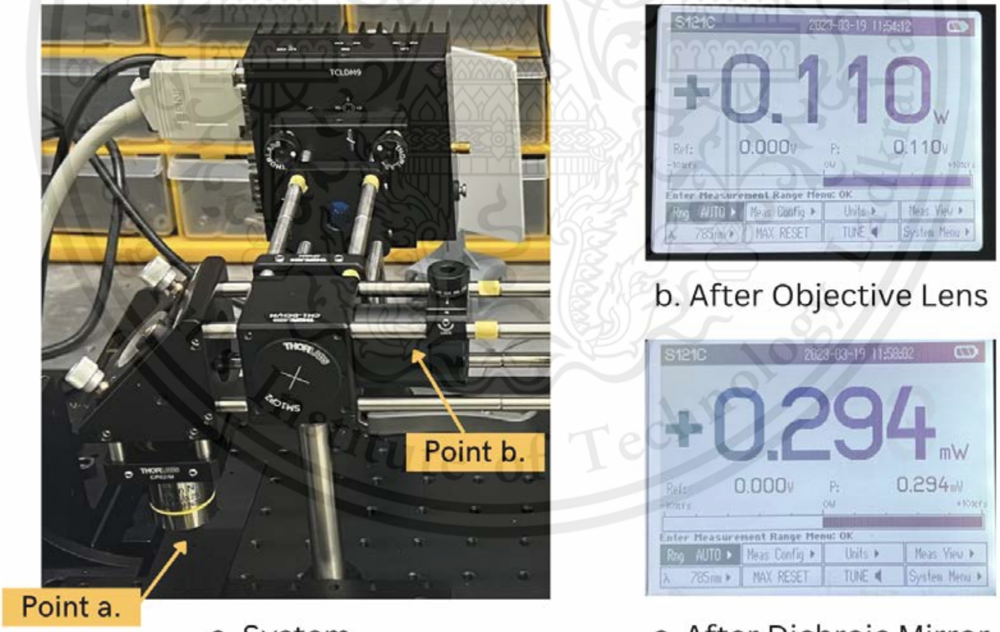


Figure 4-35. Power meter measurement A) Power at an objective lens of Galilean beam expander of 73.9 mW. B) Power at an objective lens of Keplerian beam expander of 65.2 mW.

In this third design there is a change in beam profile with this change we have using this second system for a decent time. With the expectation that the image of this third system might show a sign of Raman spectrum, also with the improvement of laser power at an objective lens from 65.2 mW in Keplerian design to 73.9 mW in Galilean design.

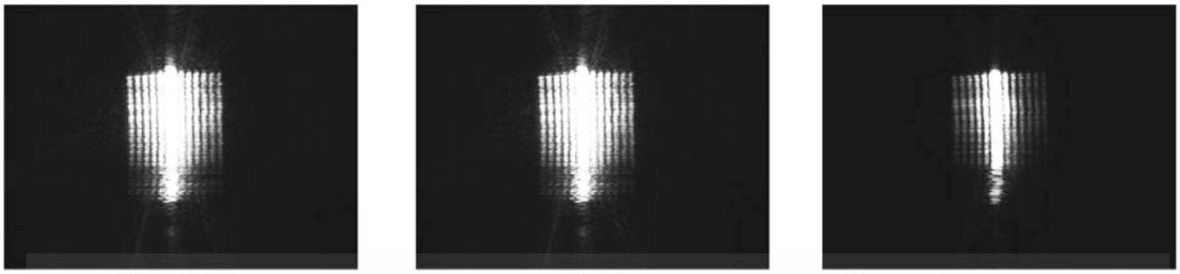


b. After Objective Lens

a. System

c. After Dichroic Mirror Transmission

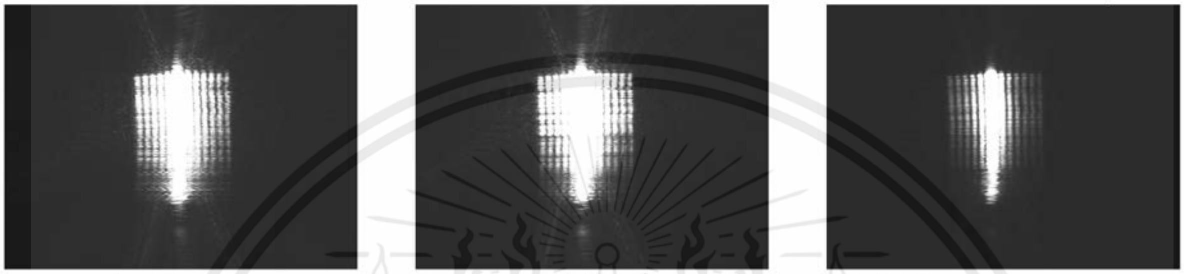
Figure 4-36. Further improvement in alignment and change of laser diode (L785H1) a.) The third design system of Raman parts. b.) Power after an objective lens of 110 mW from a 200 mW power input c.) Power after the dichroic mirror that transmits from the slide holder.



a. Aluminium

c.SiO2

e. Glucose_reduce_exposure



b. Glucose

d. ICG

f. SiO2_reduce_exposure

Figure 4-37. Further improvement in alignment and change of laser diode (L785H1) with a result of laser with no long pass filter of 830 nm (BLP01-830R-25) a.) Result of using Aluminum plate as sample. b.) Result of using glucose powder as sample. c) Result of using SiO2 as a sample. d) Result of using ICG as sample. e) Result of using glucose powder as a sample with a reduction of exposure time. f) Result of using SiO2 as a sample with a reduction of exposure time.

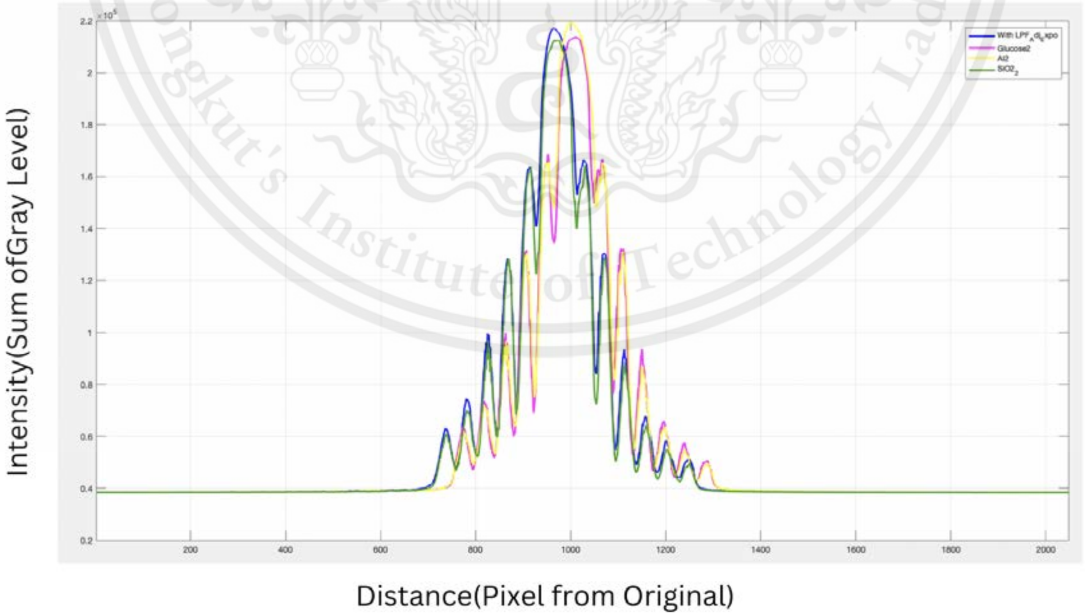


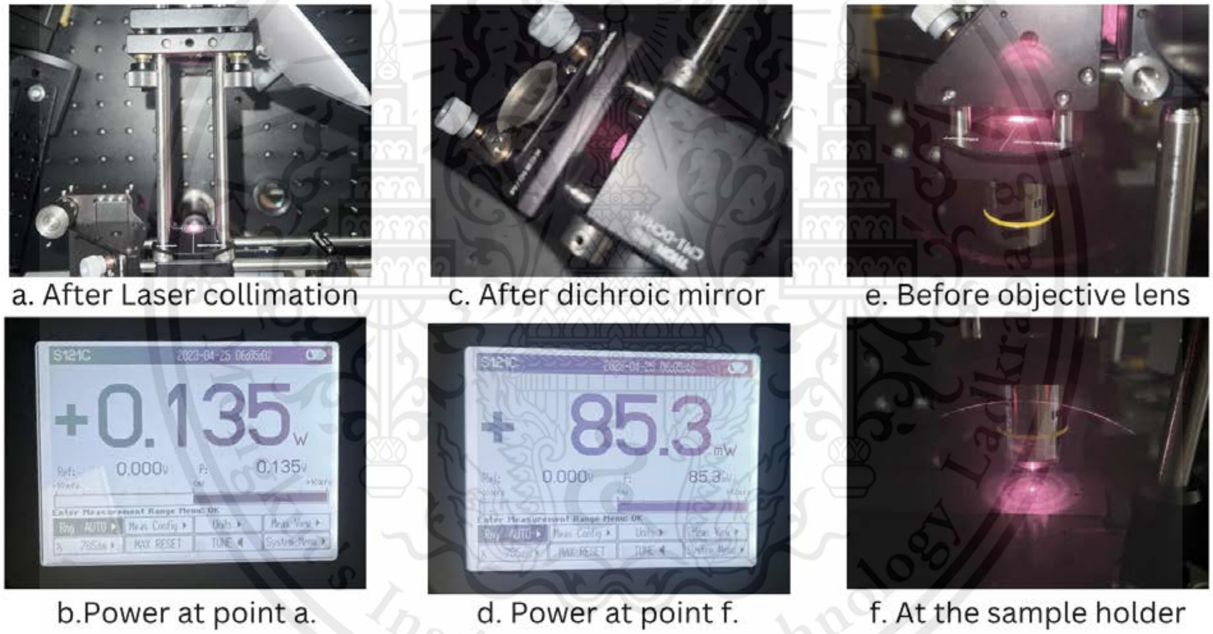
Figure 4-38. MATLAB coding for comparing the signal between each sample of ICG, Glucose, Aluminium and SiO2.

From the comparison using graph plotting in MATLAB code. We found that there is a shift in signal intensity and location between each sample but in terms of Raman spectrum or signal it is not significant enough to detect a sample fingerprint of each sample.

4.4.4 Results from the fourth system design

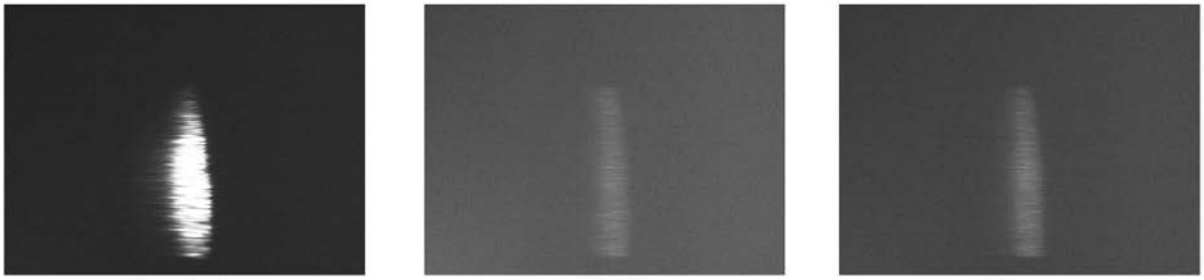
In this fourth system design it is the latest version of the Raman system as we will be using the result of this design to further conclude the final result.

With this fourth system design we have changed the collimating setup to use only one achromatic doublet (AC080-020-B) with the focal length of 20 mm and changing of long pass filter to 805 nm (FELH0800) from 830 nm (BLP01-830R-25). With this design and better system alignment using an IR card and IR target.



a. After Laser collimation
c. After dichroic mirror
e. Before objective lens
b.Power at point a.
d. Power at point f.
f. At the sample holder

Figure 4-39. The fourth system is designed using a collimation lens (AC080-020-B). a) Laser beam after laser. b) Power of 135 mW from 200 mW after the lens collimator c) Laser beam after dichroic mirror. d) Power of 85.3 at the objective lens. e) Laser beam profile that enters the objective lens. f)Laser beam at the objective lens on sample holder.



a. Aluminium

c.SiO2

e. Glucose

b. SERs Ethanol

d. SERs Glucose

f. SERs

Figure 4-40. Fourth system design image collection with a long pass filter of 805 nm (FELH0800) to reduce the range of wavelength blockage . a.) Result of using Aluminum plate as sample. b.) Result of using SERs Ethanol as a sample. c) Result of using SiO2 as a sample. d) Result of using SERs glucose as sample. e) Result of using glucose powder as a sample with a reduction of exposure time. f) Result of using SERs as a sample.

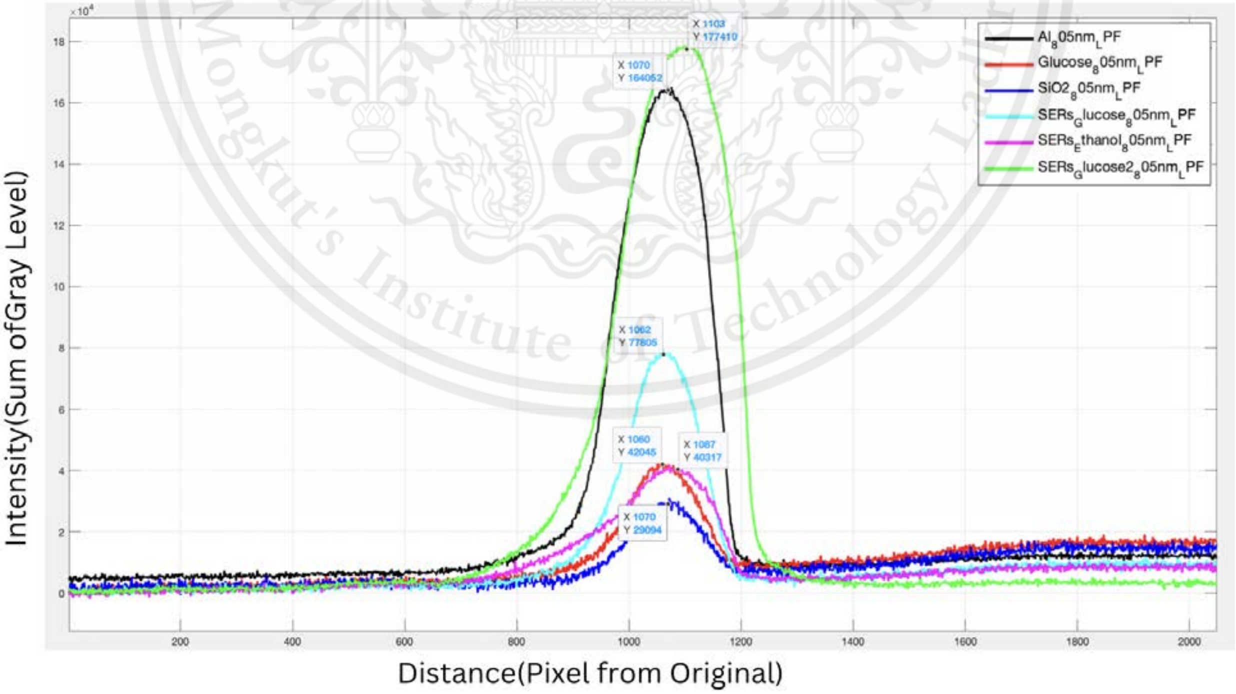


Figure 4-41. MATLAB coding for comparing the signal between each sample of Aluminum plate , Glucose, SiO2, SERs glucose, SERs Ethanol and SERs background.

With the comparison of different types of samples there is some peak shift between each sample and there is a significant difference between the SERs sample and non SERs sample as expected that SERs solution is improving the signal increasing the intensity as shown in Figure 4-38. With this result we further did some image processing and data analysis as a baseline subtraction.

4.4.5 Experimental on baseline subtraction

The baseline subtraction is one of the methods of improving the signal quality of the Raman scattering signal. For the experiment with powder substrate we use aluminum as the sample slide. From the figure 4-xx the glucose and SiO2 have some difference peak trend without the long pass filter. The baseline subtraction with the system that include 805 nm long pass filter(FELH0800) have a sharper peak at the slighting different point of the distance of pixel from origin. However, we can not detect the three significant peaks of the glucose substrate at any experiment.

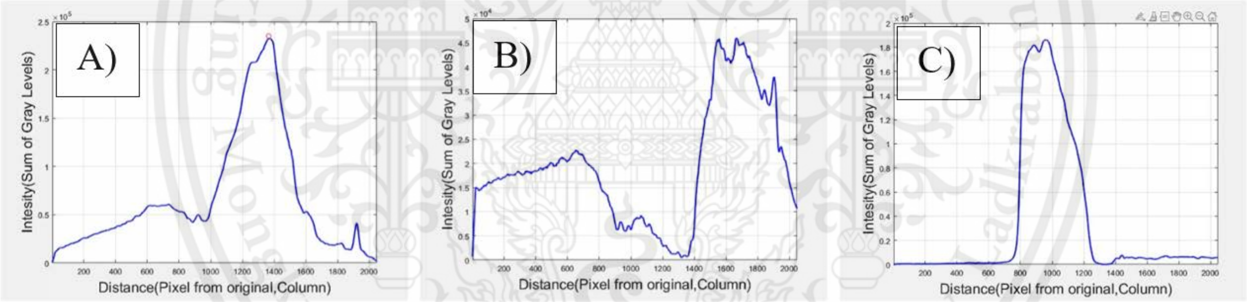


Figure 4-42. The baseline subtraction without long pass filter A) Aluminum substrate(Baseline), B) Glucose substrate, C) SiO2 substrate.

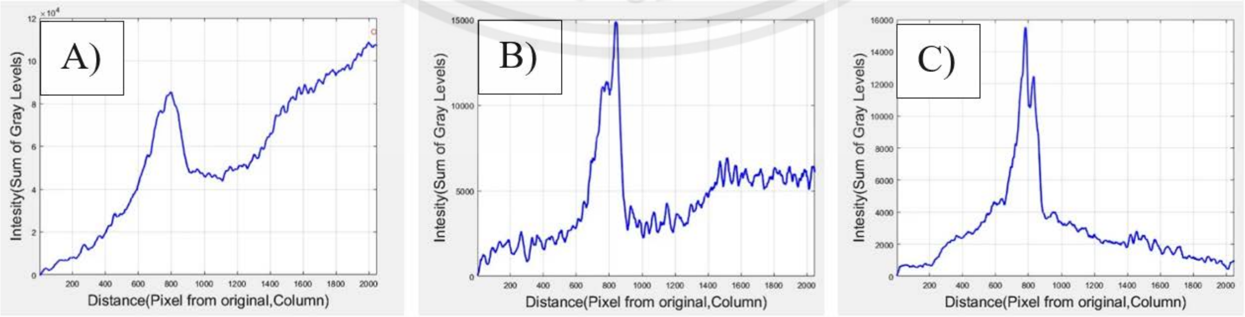


Figure 4-43. The baseline subtraction with 805 nm long pass filter (FELH0800) A) Aluminum substrate(Baseline), B) Distil water, C) Glucose substrate.

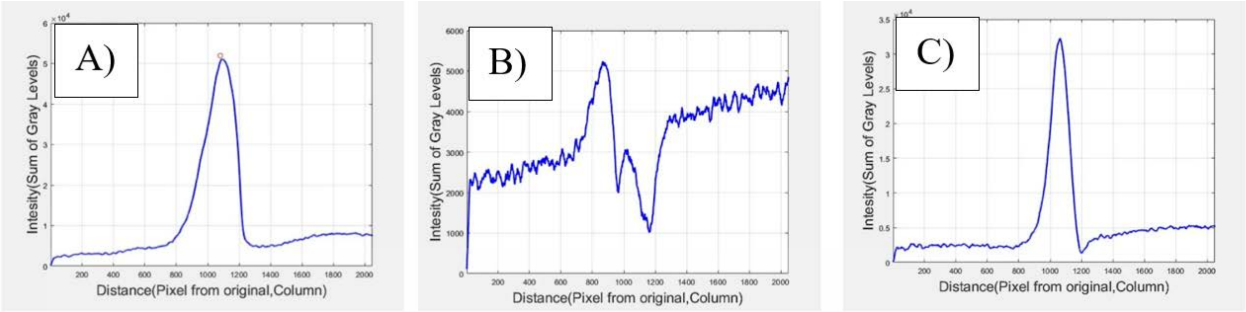


Figure 4-44. The baseline subtraction with 805 nm long pass filter(FELH0800) A) SERs substrate(Baseline), B) Ethanol on SERs substrate, C) Glucose on SERs substrate.

4.5 Discussion

The intention of this project was to build a low-cost Raman spectrometer that can measure glucose non-invasively using Surface-Enhanced Raman Spectroscopy (SERS) substrate. However, we faced limitations in achieving this goal due to the need for high sensitivity components in the in vivo or the 785 nm Raman system, which cost a lot. As a result, we shifted our focus to using SERs substrate samples, but even then, we encountered limitations.

Major issue we faced was the wavelength of the 785 nm laser, which required a high quantum efficiency on the 830-910 nm camera detector. This limitation made it challenging to obtain accurate and reliable results. Furthermore, we changed the fiber laser to a diode laser to lower the cost, but this caused a decrease in the excitation of electrons, resulting in less power for the laser beam. Another crucial factor in improving the system is better control of the environment. We observed that the results varied based on the surrounding environment, and a more controlled environment would ensure more accurate and consistent results.

The alignment of the Raman system also posed a challenge, and we concluded that it could be improved by removing unnecessary components and redesigning it to use a 532nm laser wavelength base to be compatible with SERs substrate and other particles not used in vivo. The cost of the system can also be reduced by using a cheaper wavelength.

4.6 Summary

In conclusion, building a low-cost Raman spectrometer for non-invasive glucose measurement using SERs substrate samples is a promising concept. However, we faced limitations in achieving accurate and reliable results due to the need for high sensitivity components and challenges with alignment. To improve the system, we recommend using a 532nm laser wavelength base, removing unnecessary components, and controlling the environment to ensure more accurate and consistent results.

CHAPTER 5

CONCLUSION

5.1 Introduction

In this chapter, we first summarize the work described in this report (Section 4.3). Then we draw a number of conclusions about key parts of the work undertaken in Section 5.2, and finally, in 5.4 we discuss future work and how to improve the research method and experiments.

5.2 Summary

This is a summary of each chapter intro and summary

Chapter 1 introduced the problem of the method of invasive diabetes detection and potential to use Raman spectroscopy to be the non-invasive method of detection.

Chapter 2 reviewed the state-of-the-art in blood particle analysis, method of non-invasive detection, previous studies on this subject. SERs substrate enhancement based on urine was introduced and the wavelength of IR laser described. The potential for artificial intelligence for analysis.

Chapter 3 describes the design of the Raman spectrometer. The method of calibration and instrumentation have been described. The method of sample collection has been included. Using the picture summation method.

Chapter 4 described the results of each designed system and limitations we have encountered.

Chapter 5 described the considerations for demonstrating and improving this research.

5.3 Conclusions

In conclusion, our senior project aimed to build a Raman spectrometer and measure glucose on SERs substrate while calibrating the results using a benchmark experiment. Despite our best efforts, we were unable to obtain significant results due to alignment issues and limitations of our equipment. However, we learned valuable lessons throughout the process, including the importance of careful planning, attention to detail, and perseverance in the face of setbacks.

We also gained a deeper understanding of Raman spectroscopy, its potential applications in medical diagnostics, and the challenges of constructing a functional instrument on a limited budget. Our experience underscores the need for continued innovation and research in this field to improve the accuracy and reliability of non-invasive glucose monitoring methods.

Although we did not achieve our original goal, we believe that our project has laid a solid foundation for future students to build upon and further refine. We hope that our efforts will inspire others to pursue similar projects and contribute to the advancement of Raman spectroscopy and medical diagnostics as a whole.

5.4 Suggestion

1. Consider using a different wavelength laser, such as 532nm, for better compatibility with SERs substrate and other particles not in vivo.
2. Explore options for more sensitive camera detectors, ideally with a quantum efficiency range of 830-910 nm for use with the 785 nm laser.
3. Review the design of the system to remove any unnecessary components and minimize distance between components to improve alignment.
4. Consider implementing better environmental control measures to reduce potential sources of error.

REFERENCES

[1] Ogurtsova K., Rocha Fernandes, J.D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N.H., Cavan, D., Shaw, J.E., Makaroff, L.E. *IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040*, Diabetes Research and Clinical Practice, 2017.

[2] Watson, Stephanie. *Diabetes: Symptoms, Causes, Treatment, Prevention, and More.* Healthline. February 27, 2020.

[3] Aggidis, A. G. A., Newman, J. D., and Aggidis, G. A. *Investigating pipeline and state of the art blood glucose biosensors to formulate next steps*, Biosensors and Bioelectronics, vol. 74, pp. 243–262, 2015.

[4] Lisa J. Jacob, Hans-Peter Deigner, *Nanoparticles and Nanosized Structures in Diagnostics and Therapy,* Precision Medicine, Pages 229-252.

[5] Emmanuel, N., Nair, R. B., Abraham, B., and Yoosaf, K. *Fabricating a low-cost Raman spectrometer to introduce students to spectroscopy basics and applied instrument design,* Journal of Chemical Education, vol. 98, no. 6, pp. 2109–2116, 2021.

[6] Eduardo, H., Montoya, R., Aurelio Arbildo, L., Oscar, R., and Baltuano, E., *A Homemade Cost Effective Raman Spectrometer with High Performance,* in Journal of Laboratory Chemical Education, 2015, 3(4): 67-75.

[7] Kang, J. W., Park, Y. S., Chang, H., Lee, W., Singh, S. P., Choi, W., Galindo, L. H., Dasari, R. R., Nam, S. H., Park, J., and So, P. T. *Direct observation of glucose fingerprint using in vivo Raman spectroscopy*, Science Advances, vol. 6, no. 4, 2020.

[8] Stryer, L. *Biochemistry*, 4th ed., W.H. Freeman: New York, 1995.

[9] Kahn, C. R., Joslin, E. P. *Joslin's diabetes mellitus*, 14th ed.; Lippincott Williams & Willkins: Philadelphia, 2005.

[10] Henry, J. B. *Clinical diagnosis and management by laboratory methods*, 19th ed.; Saunders: Philadelphia, 1996.

[11] Sherwani, S. I., Khan, H. A., Ekhzaimy, A., Masood, A., and Sakharkar, M. K. *Significance of hba1c test in diagnosis and prognosis of diabetic patients,* Biomarker Insights, vol. 11, 2016.

[12] Anas, M. N., Nurun, N. K., and Norali and M. Normahira, A. N. *Non-invasive blood glucose measurement,* 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, pp. 503-507, 2012.

[13] Della Man, C., Caumo, A., and Cobelli, C. *The oral glucose minimal model: Estimation of insulin sensitivity from a meal test,* in IEEE Transactions on Biomedical Engineering, May, vol. 49, no. 5, pp. 419-429, 2002.

[14] Sokol, Y., Shchapov, P., Tomashevskyi, R., Veligorskyi, O., Picking, R., and Chakirov, R. *Data analysis of random blood measurements for abnormal condition detection*, Internet Technologies and Applications (ITA), pp. 204-208, 2017.

[15] Shih W.-C. “Quantitative Biological Raman Spectroscopy” *in Quantitative Biological Raman spectroscopy for non-invasive blood analysis*. essay. vol.12, 2007.

[16] Khalil, O. S. *Clinical Chemistry*, 45, 165-177, 1999.

[17] Khalil, O. S. *Diabetes Technology & Therapeutics*, 6, 660-697, 2004.

[18] Defronzo, R. A., Tobin, J. D., Andres, R., *American Journal of Physiology*, 237, E214-E223, 1979.

[19] Arnold, M. A., Burmeister, J. J., Small, G. W., *Analytical Chemistry*, 70, 1773-1781, 1998.

[20] Olesberg, J. T. , Arnold, M. A., Mermelstein, C., Schmitz, J., Wagner J. *Applied Spectroscopy*, 59, 1480-1484, 2005.

[21] Marbach, R., Koschinsky, T., Gries, F. A., Heise, H. M. *Applied Spectroscopy*, 47, 875-881, 1993.

[22] Robinson, M. R., Eaton, R. P., Haaland, D. M., Koepp, G. W., Thomas, E. V., Stallard, B. R., Robinson P. L. *Clinical Chemistry*, 38, 1618-1622, 1992.

[23] Haaland, D. M., Robinson, M. R., Koepp, G. W., Thomas, E. V., Eaton R. P. Applied Spectroscopy, 46, 1575-1578, 1992.

[24] Small, G. W., Arnold, M. A., Marquardt, L. A. Analytical Chemistry, 65, 3279- 3289, 1993.

[25] Samann, A., Fischbacher, C., Jagemann, K. U., Danzer, K., Schuler, J., Papenkordt, L., Muller, U. A. Experimental and Clinical Endocrinology & Diabetes 2000, 108, 406-413.

[26] Maruo, K., Oota, T., Tsurugi, M., Nakagawa, T., Arimoto, H., Hayakawa, M., Tamura, M., Ozaki, Y., Yamada, Y. Applied Spectroscopy, 60, 1423-1431, 2006.

[27] Olesberg, J. T., Liu, L. Z., Van Zee, V., Arnold, M. A. Analytical Chemistry, 78, 215-223, 2006.

[28] Burmeister, J. J., Arnold, M. A., Small, G. W. Diabetes Technology & Therapeutics, 2, 5-16, 2000.

[29] Maruo, K., Tsurugi, M., Tamura, M., Ozaki, Y. Applied Spectroscopy, 57, 1236- 1244, 2003.

[30] Diessel, E., Willmann, S., Kamphaus, P., Kurte, R., Damm, U., Heise H. M. Applied Spectroscopy, 58, 442-450, 2004.

[31] Kim, Y. J., Hahn, S., Yoon G. Applied Optics, 42, 745-749, 2003.

[32] Heise, H. M., Marbach, R. Cellular and Molecular Biology, 44, 899-912, 1998.

[33] Enejder, A. M. K., Koo, T. W., Oh, J., Hunter, M., Sasic, S., Feld, M. S., Horowitz G. L. Optics Letters, 27, 2002.

[34] Qu, J. N. Y., Wilson, B. C., Suria D. Applied Optics, 38, 5491-5498, 1999.

[35] Rohleder, D., Kiefer, W., Petrich, W. Analyst, 129, 906-911, 2004.

[36] Lambert, J. L., Morookian, J. M., Sirk, S. J., Borchert, M. S. Journal of Raman Spectroscopy, 33, 524-529, 2002.

[37] A. M. K. Enejder, T. G. Scecina, J. Oh, M. Hunter, W. C. Shih, S. Sasic, G. L. Horowitz, M. S. Feld Journal of Biomedical Optics 2005, 10, 031114.

[38] J. Chaiken, W. Finney, P. E. Knudson, R. S. Weinstock, M. Khan, R. J. Bussjager, D. Hagrman , P. Hagrman, Y. W. Zhao, C. M. Peterson, K. Peterson Journal of Biomedical Optics 2005, 10

[39] A. J. Berger, T. W. Koo, I. Itzkan, G. Horowitz, M. S. Feld Applied Optics 1999, 38, 2916-2926.

[40] C. C. Pelletier, J. L. Lambert, M. Borchert Applied Spectroscopy 2005, 59, 1024-1031.

[41] R. Luo, J. Popp, & T. Bocklitz Deep Learning for Raman Spectroscopy: A Review 2022. *Analytica*, *3*(3), 287–301.

[42] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning; MIT Press: Cambridge, MA, USA, 2016; pp. 1–481.

[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2006), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1–12.

[44] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation. arXiv 2015, arXiv:1505.04597.

[45] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning; MIT Press: Cambridge, MA, USA, 2016; pp. 1–481.

[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Bengio, Y. Generative adversarial nets. arXiv 2014, arXiv:1406.2661.

[47] Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, UK.

[48] Timothy, M.D. *Non-linear methods for the analysis of metabolic profiles(n.d)* in Elsevier B.V. 2007, 7, 212-214.

[49] Raman, C.V., Krishnan, K. S. *Nature* 1928, *121,* 501-502.

[50] McCreery, R. L. *Raman spectroscopy for chemical analysis*, John Wiley & Sons: New York, 2000.

[51] Gillies, R., Zonios, G., Anderson, R. R., Kollias, N. *Journal of Investigative Dermatology* 2000, 115*,* 704-707.

[52] Hull, E. L.; Ediger, M. N.; Unione, A. H. T.; Deemer, E. K.; Stroman, M. L.; Baynes, J. *W. OpticsExpress* 2004, *12,* 4496-4510.

[53] Muller, M. G.; Georgakoudi, I.; Zhang, Q. G.; Wu, J.; Feld, M. S. *Applied Optics* 2001, 40,4633-4646.

[54] Zeng, H.; MacAulay, C.; Palcic, B.; McLean, D. I. *SPIE* **1983,** *1882,* 278-290.

[55] Zeng, H.; MacAulay, C.; McLean, D. I.; Palcic, B.; Lui, H. *Photochemistry and Photobiology* **1998,** *68,* 227-236.

[56] Jongen, A. J. L.; Sterenborg, H. J. C. M. *Physics in Medicine and Biology* **1997,** *42,* 1701-1716.

[57] Gornushkin, I. B.; Eagan, P. E.; Novikov, A. B.; Smith, B. W.; Winefordner, J. D. *Applied Spectroscopy* **2003,** *57,* 197-207.

[58] Lieber, C. A.; Mahadevan-Jansen, A. *Applied Spectroscopy* **2003,** *57,* 1363-1367.

[59] Vickers, T. J.; Wambles, R. E.; Mann, C. K. *Applied Spectroscopy* 2001, *55,* 389-393.

[60] Caspers, P. J.; Lucassen, G. W.; Puppels, G. J. *BiophysicalJournal* **2003,** *85,* 572-580.

[61] Gunst, R. F.; Mason, R. L. *Regression analysis and its application : a data-oriented approach;*M. Dekker: New York, 1980.

[62] Wold, S.; Martin, H.; Wold, H. *Lecture Notes in Mathematics;* Springer-Verlag: Heidelberg, 1983.

[63] Granite, S. Optimisation of SERS for Glucose Sensing: Raman spectroscopy *in Edinburgh Instruments* Available: https://www.edinst.com/optimisation-of-sers-for-glucose-sensing/ [Accessed Oct 25, 2022]

[64] contributors, P. L. (n.d.). A homemade cost effective Raman spectrometer with high performance *in Public Lab* Available: https://publiclab.org/notes/emontoya57/12-16-2015/a-homemade-cost-effective-raman-spectrometer-with-high-performance [Accessed Oct 25, 2022]

[65] D. Motta, A. Á. B. Santos, I. Winkler, B. A. S. Machado, D. A. D. I. Pereira, A. M. Cavalcanti, E. O. L. Fonseca, F. Kirchner, and R. Badaró, “Application of convolutional neural networks for classification of Adult Mosquitoes in the field,” *PLOS ONE*, 14-Jun-2019. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0210829. [Accessed April 8, 2022].

[66] I. Sarker, M. Faruque, H. Alqahtani, and A. Kalim, “K-nearest neighbor learning based diabetes mellitus prediction and analysis for eHealth Services,” *ICST Transactions on Scalable Information Systems*, p. 162737, 2018.