PORTABLE BILIRUBIN DETECTION DEVICE USING MACHINE LEARNING

A Dissertation for

Course code and Course Title: ELE-625 Project

Credits: 16

Submitted in partial fulfillment of Masters Degree
M.Sc. in Electronics

by

XAVIER COELHO

Seat no: 22P0360004
ABC ID: 885700306290
PR No: 201905899

Under the Supervision of

PROF. JIVAN PARAB

School of Physical and Applied Sciences
Electronics

Goa University

Date:

Examined by :

Seal of the School

DECLARATION BY STUDENT

I hereby declare that the data presented in this Dissertation report entitled, **Portable Bilirubin Detection Device Using Machine Learning** is based on the results of investigations carried out by me in the M.Sc Electronics at the School of Physical and Applied Sciences, Goa University under the Supervision of Prof. Jivan Parab and the same has not been submitted elsewhere for the award of a degree or diploma by me. Further, I understand that Goa University or its authorities will be not responsible for the correctness of observations / experimental or other findings given the dissertation.

I hereby authorize the University authorities to upload this dissertation on the dissertation repository or anywhere else as the UGC regulations demand and make it available to any one as needed.

Xavier Coelho

Seat no: 22P0360004

Date:

Place: Goa University

COMPLETION CERTIFICATE

This is to certify that the dissertation report **Portable Bilirubin Detection Device Using Machine Learning** is a bonafide work carried out by Mr Xavier Coelho under my supervision in partial fulfillment of the requirements for the award of the degree of Masters in the Discipline Electronics at the School of Physical and Applied Sciences, Goa University.

Prof. Jivan Parab

Date:

Signature of Dean of the School

Date:

Place: Goa University

School Stamp

CONTENTS

PREFACE

In this topic we will be aiming for the development of a small portable device that can be used remotely to test for detecting small rise in blood bilirubin and potentially prevent the situation from escalating to a serious matter or even fatality.

This report is divided into four chapters;

Chapter 1 is about the background of the area of work. Understanding the functions and working of the liver and how bilirubin is formed and what it can do to human body specially neonates. Some of the common methods of testing for jaundice have been discussed.

Chapter 2 will be for literature review of some of the works done by other researchers and experts that are related directly or indirectly to the topic on which the project is being done.

Chapter 3 will be about the methodology of this study. The steps which will be needed to be taken, and in what particular order to go on and reach the end goal of the study and also looking at the various Machine Learning Algorithms used.

Chapter 4, the results that have been obtained out of the study will be analysed and also various outcomes will be compsred to analyse and determine which outcome is the best for the case.

ACKNOWLEDGEMENTS

I would like to express my sincerest thanks to everyone who has contributed to this project. First and foremost I want to thank my project guide Prof. Jivan. S. Parab for guiding me and supporting me during the course of this project.

I express my deep gratitude towards all the teaching faculty Prof Rajendra S. Gad, Prof. Marlon Sequeira, Prof. Narayan Vetrekar and Prof. Aniketh Gaonkar for their support and guidance throughout the completion of the project. Also thanks to all the non teaching faculty Mr Vishant Malik, Mr. M. Lanjewar. Their knowledge and suggestions have been crucial in helping me plan, carry out, and execute the project.

I would like to express my gratitude to my family and close friends for their unfailing support and inspiration during this project.

TABLES AND FIGURES

ABBREVIATIONS USED

Entity	Abbreviation
Machine Learning	ML
K-Nearest Neighbour	KNN
Partial Least Square	PLS
Mean Absolute Error	MAE
Root Mean Square Error	RMSE
R-Squared	R2
Bilirubin	BR
Analytical Research	AR
Convolution Neural Network	CNN

ABSTRACT

The goal of the "Portable Bilirubin Detection Device Using Machine Learning" project is to create a novel and user-friendly bilirubin level monitoring system. If left undiagnosed or untreated, elevated bilirubin levels can cause serious side effects such jaundice and kernicterus.

Based on the tests performed and observations made we are able to safely say that the spectras observed during the testing of the samples, the absorption was increasing with increase in bilirubin concentration.

Several machine learning algorithms were used to train and predict the values of concentrations of the samples and among the ones that were used, out of all the ML algoritms that were tested, naming;KNN, Decision Tree, Random Forest and PLS, KNN outperformed all the other models with an R2 score = 0.9907 and MAE = 0.0275 and RMSE = 0.0686.

Furthermore we have proposed a setup that is portable, affordable and non invasive for testing of bilirubin levels using a Triad Spectral sensor.

C H A P T E R 1

INTRODUCTION

1.1 Background

Blood is a vital bodily fluid that circulates through the arteries and veins, delivering oxygen, nutrients, hormones, and other essential substances to tissues and organs while removing waste products. Composed of plasma, red blood cells, white blood cells, and platelets, blood plays crucial roles in maintaining homeostasis, fighting infections, clotting to prevent excessive bleeding, and transporting gases like oxygen and carbon dioxide. Plasma, the liquid component, carries cells and proteins throughout the body. Red blood cells contain hemoglobin, which binds oxygen for transport. White blood cells are key components of the immune system, defending the body against pathogens. Platelets are involved in blood clotting, preventing excessive bleeding when injury occurs. Blood composition and function are tightly regulated to ensure the body's overall health and survival.[a]

Blood serves a variety of purposes, such as:

delivering nutrients and oxygen to the tissues and lungs; creating blood clots to stop excessive bleeding; conveying immune system components and cells; delivering waste products to the liver and kidneys, which clean and filter the blood; controlling body temperature.[b]

Liver:

Situated in the upper right section of the abdomen, above the stomach and below the diaphragm, lies the liver, an essential organ. The largest internal organ in the human body, it carries out a wide range of vital tasks for immunity, digestion, metabolism, and detoxification. Processing food nutrients into forms the body may use for growth, repair, and energy is one of its main functions. In order to assist maintain constant energy levels, the liver stores glycogen, a type of glucose, which it releases into the bloodstream when blood sugar levels fall. Furthermore, the liver produces clotting factors and albumin, two proteins that are essential for blood clotting and fluid homeostasis.[c]

The liver's detoxification process, which involves removing toxins, medications, and dangerous materials from the bloodstream, is another essential job of the liver. These chemicals are broken down by the liver's primary cells, called hepatocytes, into less toxic forms that the body can eliminate. In order to preserve overall metabolic balance, the liver is also essential for the breakdown and removal of excess hormones, cholesterol, and waste products.[c]

In addition, the liver secretes bile, a digestive juice that aids in the small intestine's absorption of fat-soluble vitamins and the emulsification of lipids. To help with the breakdown and absorption of dietary fats, bile is kept in reserve in the gallbladder and released into the small intestine when required.

In addition, the liver is involved in the production of triglycerides and cholesterol as well as the transformation of ammonia into urea, which is eliminated in the urine.

In addition to its metabolic duties, the liver is essential to the immune system, which aids in protecting the body from illnesses and infections. Specialized immune cells found in the liver called kupffer cells aid in the elimination of bacteria, viruses, and other pathogens from the bloodstream while also enhancing immunological function in general.

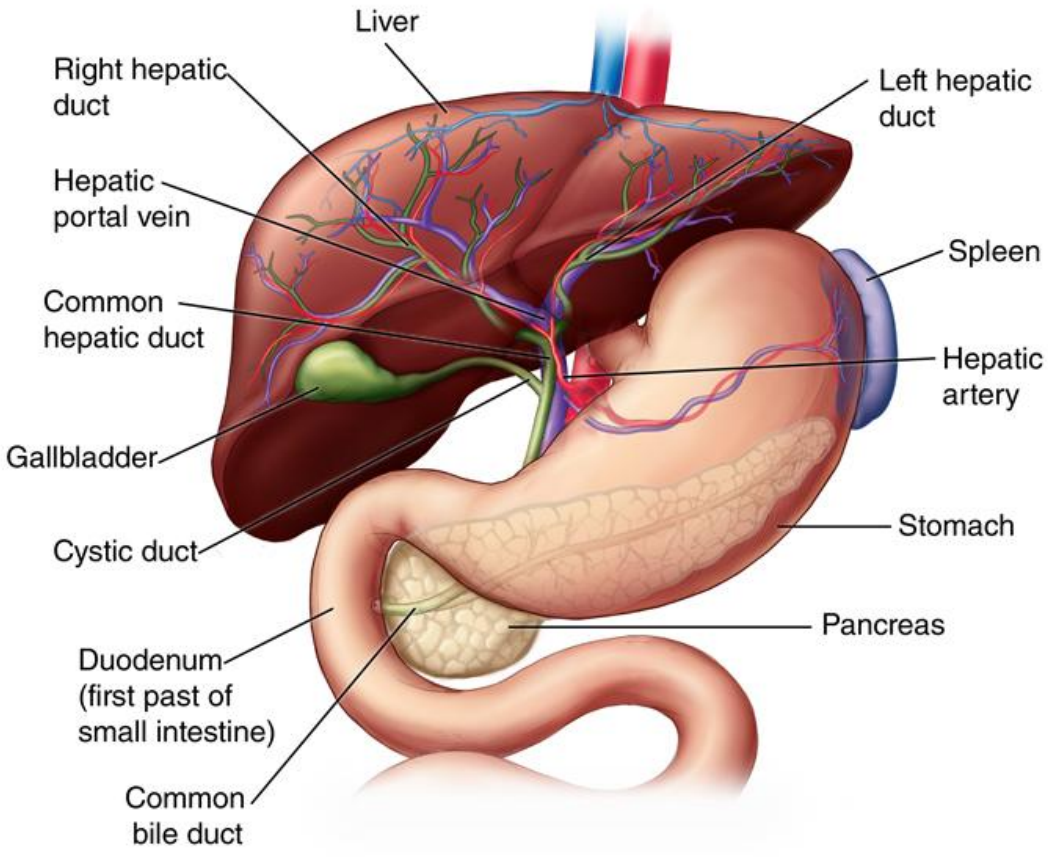


Fig.1.1: liver Anatomy

After being released into the bloodstream, bilirubin is taken to the liver and changes in a number of ways. Hepatocytes, or the cells that make up the liver, are essential to this process. They undergo a process known as conjugation in which they take up the bilirubin from the blood and combine it with glucuronic acid. Because it is soluble in water, this conjugated bilirubin is easier for the body to eliminate. Bilirubin is released into bile, a digestive fluid that the liver produces, once it has been conjugated. It then makes its way to the intestines, where it facilitates the breakdown of lipids. Bilirubin eventually leaves the body through excretions. Maintaining appropriate blood levels of bilirubin is essential for preventing jaundice and other related disorders, as well as for the liver to effectively filter and process bilirubin.

Unhealthy Liver:

The liver is an incredibly resilient organ that can tolerate a wide range of insults and injuries before beginning to fail if it is overworked or continuously harmed. Many conditions, such as chronic hepatitis B or C, fatty liver disease, autoimmune disorders affecting the liver, genetic disorders like hemochromatosis or Wilson's disease, prolonged exposure to specific toxins or medications, and autoimmune disorders affecting the liver can all contribute to liver failure. A representation of various liver disease in given in figure 1.2.

At first, the damage may be offset by the liver's ability to regenerate healthy cells, but eventually, the damage can accumulate and compromise liver function. This may show itself as a progressive deterioration of liver function, accompanied by symptoms including lethargy, jaundice, abdominal or leg

swelling from fluid retention, bleeding easily, disorientation, and ultimately hepatic encephalopathy, a dangerous consequence marked by impaired cognitive function.[d]

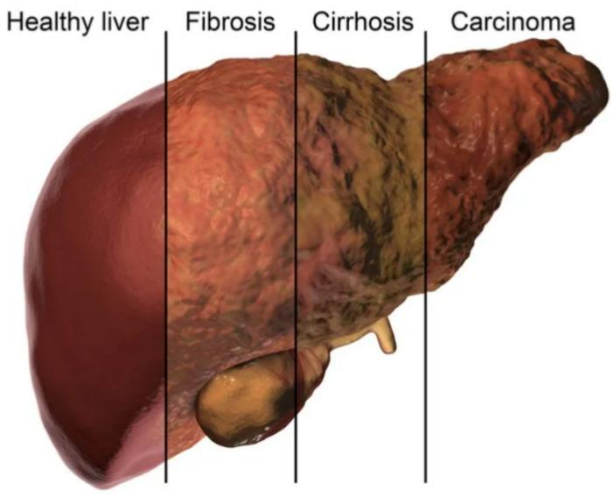


Fig.1.2:different types of liver disease

If liver failure is not treated, it can worsen quickly and become fatal, requiring immediate medical attention, including liver transplantation, to preserve a patient's life. Preventing liver failure and maintaining liver health mostly involves regular medical monitoring, lifestyle modifications, and timely treatment of underlying liver disorders.

Hepatitis, cirrhosis, and fatty liver disease are examples of chronic liver illnesses that can progressively deteriorate liver function over time, decreasing the liver's capacity to carry out vital functions. Liver damage is mostly caused by viral infections, autoimmune diseases, excessive alcohol consumption, and genetics. Furthermore, liver dysfunction can be made

worse by lifestyle choices such an unhealthy diet, obesity, and exposure to pollutants. The ability of the liver to digest nutrients, eliminate toxic chemicals, control blood sugar, and generate vital proteins is compromised as liver function declines, which has a series of negative implications on general health. Progressive liver damage can result in major consequences, such as liver failure, which calls for medical intervention such as transplantation, if it is not addressed promptly and lifestyle changes are not made. Consequently, maintaining liver health with a healthy diet, frequent exercise, avoiding alcohol abuse, and managing underlying medical conditions is paramount in maintaining its efficiency and ensuring optimal physiological function.

Bilirubin:

The body produces bilirubin, a yellow-orange pigment, as red blood cells degrade. It is created when heme and globin, the proteins that carry oxygen in red blood cells, separate from haemoglobin. After additional metabolism, heme becomes biliverdin, which subsequently becomes bilirubin. After travelling via the bloodstream to the liver, bilirubin is converted, which makes it more easily excreted and water soluble. After being released into bile, conjugated bilirubin is eventually excreted from the body through faeces. Blood bilirubin levels are a crucial sign of liver and gallbladder health. Increased bilirubin levels can be a sign of liver disease or blockage of the bile supply, which can be brought on by gallstones, cirrhosis, or hepatitis. A substantial elevation in blood bilirubin levels causes jaundice, or yellowing of the skin and eyes. Blood tests used to measure bilirubin levels aid in the diagnosis and monitoring of a variety of liver and gallbladder problems,

directing the use of the most appropriate medical interventions and therapies.[e]

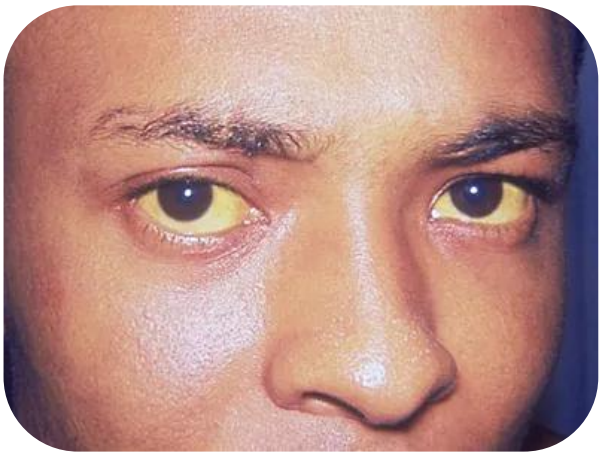


Fig.1.3: sclera turning yellow

Furthermore, because excessive levels of unconjugated bilirubin in infants can cause jaundice and even neurological issues if left untreated, bilirubin is important in evaluating the health of newborns. Thus, maintaining general health requires a grasp of bilirubin metabolism and its importance in clinical settings.

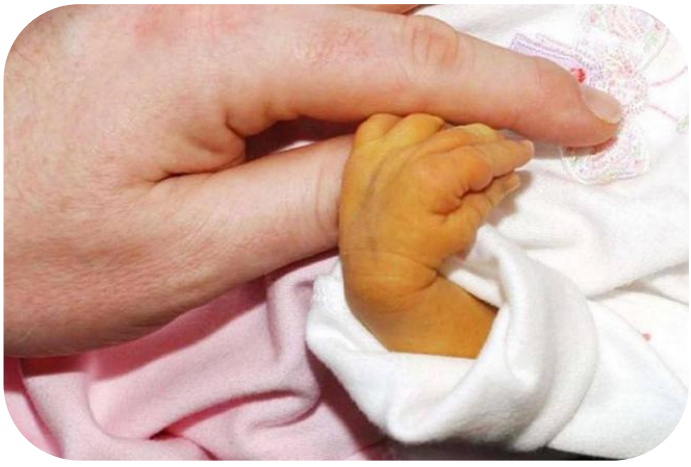


Fig.1.4: skin yellowing due to jaundice

When bilirubin levels in the bloodstream rise beyond normal levels, a condition known as hyperbilirubinemia occurs, which can have significant implications for health, particularly in newborns. Bilirubin is a yellow pigment produced by the breakdown of red blood cells, and its accumulation can lead to jaundice, characterized by yellowing of the skin and eyes. In infants, the immature liver may struggle to efficiently process and eliminate bilirubin, resulting in elevated levels. If left untreated, high bilirubin levels can pose risks such as kernicterus.[f]

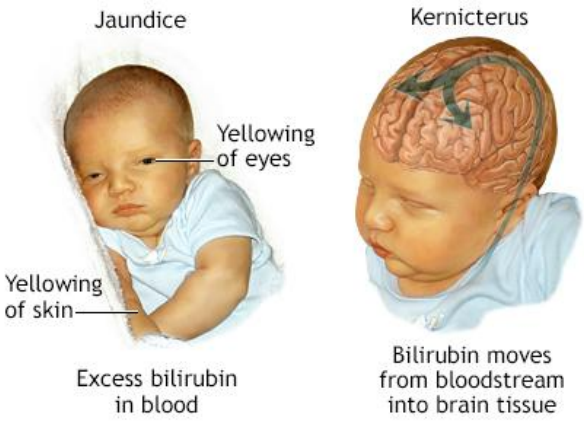


Fig.1.5: kernicterus

Bilirubin typically attaches to the serum protein albumin and travels to the liver, where enzymes change it into a form that is soluble in water and excreted from the blood into bile. However, in infants whose livers have not yet produced enough enzymes to do this, bilirubin builds up in the blood, surpasses albumin's capacity to store it, and seeps into the skin, causing yellowing of the skin, as well as into the brain, where it causes irreversible brain damage, potentially causing neurological damage, hearing loss, or even death.

Total bilirubin levels in healthy individuals normally fall between 0.2 to 1.9 mg/dL (milligrams per deciliter) of blood. Both conjugated (direct) and unconjugated (indirect) bilirubin are included in this value. It's crucial to remember that reference ranges can change slightly between laboratories and medical professionals. Furthermore, a number of variables, including age, sex, ethnicity, and underlying medical disorders, can cause variations in bilirubin levels. Consequently, the clinical context and individual characteristics should be taken into account when interpreting bilirubin levels.

Liver disease results in approximately 57,000 fatalities in the United States and approximately 2 million deaths worldwide each year.

Jaundice in newborns is highly prevalent. About 60% of babies, or 3 out of 5, experience jaundice.

Jaundice can occasionally go undiagnosed until later on, particularly if the underlying reason manifests slowly or if the symptoms are not severe. Sometimes people don't notice that their skin or eyes are getting yellower, especially if it happens gradually over time. Therefore, an additional aim of our study is to develop a non-invasive and user-friendly device that can detect subtle changes in bilirubin levels, aiding in the early identification and management of jaundice in individuals who may not readily recognize its symptoms.

The following are some simple methods to test for jaundice:

- Observation of Skin and Eyes:

  Skin: Check for yellow discoloration, particularly on the face, chest, and palms.

  Eyes: Look for yellowing in the sclera (the white part of the eyes). This is often more noticeable than skin yellowing.

- Blanch Test:

  Press on the skin, particularly the chest or the tip of the nose, for a few seconds.

  When you release the pressure, observe the color of the blanched area. If it appears yellow instead of white, it may indicate jaundice.

- Urine Examination:

  Observe the color of urine. Dark yellow or brownish urine can be a sign of jaundice, indicating high bilirubin levels being excreted by the kidneys.

- Stool Examination:

  Pale or clay-colored stools can be a sign of jaundice, as they may indicate a lack of bilirubin being excreted into the intestines.

- Detection of Itchiness:

  Chronic itching without a rash can be associated with jaundice due to bile salt deposition in the skin.

- Visual Comparison:

  Compare the skin and eyes of the person suspected of having jaundice with those of another person with a known healthy bilirubin level. This can sometimes help in identifying subtle yellowing.

- Total Serum Bilirubin Test:

  Measures the total amount of bilirubin in the blood, including both direct (conjugated) and indirect (unconjugated) bilirubin.

  Elevated levels indicate jaundice.

- Direct (Conjugated) Bilirubin Test:

  Measures the level of bilirubin that has been processed by the liver and is water-soluble.

  High levels suggest liver or bile duct issues.

- Indirect (Unconjugated) Bilirubin Test:

  Measures the level of bilirubin before it reaches the liver.

  Elevated levels may indicate increased breakdown of red blood cells or a problem with bilirubin transport to the liver.

- Liver Function Tests (LFTs):

  Include tests for enzymes like Alanine Aminotransferase (ALT) and Aspartate Aminotransferase (AST), which can indicate liver damage.

  Also measure Alkaline Phosphatase (ALP) and Gamma-Glutamyl Transferase (GGT), which can indicate bile duct obstruction or liver disease.

- Complete Blood Count (CBC):

  Can help identify conditions like hemolytic anemia, which may lead to jaundice due to increased red blood cell breakdown.

- Prothrombin Time (PT) and International Normalized Ratio (INR):

  Measure the blood's clotting ability, which can be impaired in liver disease.

- Albumin Test:

  Measures the level of albumin, a protein made by the liver. Low levels can indicate chronic liver disease.

  [g]

There are drawbacks to the current methods of diagnosing jaundice, which include physical examinations for yellowing of the skin and eyes and blood tests measuring bilirubin levels. Some are costly, invasive and slow to give an outcome. They might not constantly offer a thorough evaluation of liver function or the underlying reason for jaundice. Furthermore, these techniques might not be appropriate for the early diagnosis or surveillance of specific

liver disorders. They are also less available in some areas or in hospital settings because they need specific tools and skilled workers.

Fig.1.6: blood sampling for testing of bilirubin levels

Fig.1.7: bilirubinometer for testing bilirubin

Thus, prompt detection and management of hyperbilirubinemia are crucial to prevent complications and ensure the well-being of affected individuals, especially newborns. Treatment options may include photo-therapy, which helps break down bilirubin in the skin, or in severe cases, exchange transfusions to remove excess bilirubin from the bloodstream. Vigilant monitoring and timely intervention are essential to mitigate the risks associated with elevated bilirubin levels and safeguard the health of those affected.

1.2 Aim and Objectives

The primary aim of this project is to design and develop a novel device capable of accurately determining bilirubin concentration in biological samples using ML techniques. This device aims to offer a non-invasive and efficient method for bilirubin assessment, contributing to enhanced diagnostic capabilities in healthcare settings.

It needs to be determined if based on the concentration of the bilirubin

Objectives:

- Data Collection and Spectral Analysis:
  1. Collect spectral data of bilirubin samples using appropriate spectroscopic techniques.
  2. Analyze the collected spectra to identify relevant features indicative of bilirubin concentration.
- Model Training and Optimization:

  Train the ML models using the collected spectral data, ensuring robust performance across different sample variations.

  Optimize model parameters to improve accuracy, sensitivity, and specificity in bilirubin concentration prediction.
- Device Design and Integration:
  1. Design a compact and portable device capable of acquiring spectral data from bilirubin samples.
  2. Integrate the ML algorithms into the device's software for real-time analysis and prediction.

- Validation and Calibration:

  Validate the performance of the developed device against known bilirubin concentrations using standard reference methods.

  Calibrate the device to ensure consistent and reliable results across different environmental conditions and sample types.

- Evaluation and Comparison:

  Evaluate the accuracy and efficiency of the developed device in comparison to existing methods for bilirubin concentration determination.

  Assess the practical usability and cost-effectiveness of the device in clinical and research settings.

- Documentation and Dissemination:

  Document the entire development process, including methodologies, results, and insights gained.

  Disseminate findings through research publications, conference presentations, and potential patents to contribute to the scientific community and healthcare industry.

By accomplishing these objectives, this project aims to deliver an innovative solution for bilirubin concentration determination, with the potential to improve diagnostic accuracy, streamline healthcare processes, and enhance patient care outcomes.

Many spectra have been seen based on other people's earlier research, and they all generally indicate that bilirubin absorption typically occurs in the 300–550 nm range. One example of the spectra observed during literature review is given below in figure 1.8.

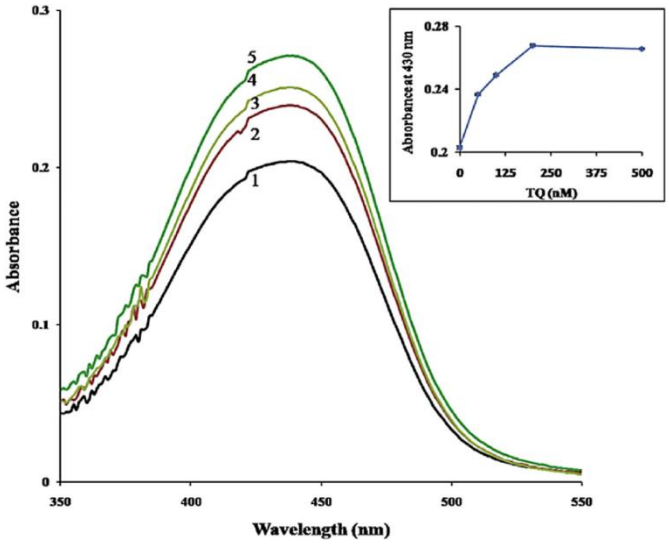


Fig.1.8: observed spectra of bilirubin

1.3 Hypothesis

The present study aims to investigate the effectiveness of Machine Learning (ML) algorithms in learning discriminative features from input data, particularly spectral data obtained from samples, and predicting bilirubin concentration levels. It is hypothesized that leveraging ML techniques can significantly improve the efficiency and accuracy of systems tasked with determining bilirubin levels in human blood samples. By harnessing the computational power and pattern recognition capabilities of ML algorithms, the system is anticipated to exhibit enhanced predictive performance compared to traditional methods. Moreover, these ML based systems are expected to demonstrate robustness to noise, allowing for accurate bilirubin concentration estimation even in the presence of environmental variability or measurement errors. The successful validation of this hypothesis would not only advance the field of medical diagnostics but also pave the way for the development of reliable and efficient tools for monitoring bilirubin-related disorders in clinical settings.

1.4 Scope

The goal of the current project is to explore the possibilities of ML techniques for the creation of a tiny, portable tool that measures blood bilirubin levels.

First a sample of bilirubin is to be prepared and tested for its spectral data. The obtained Data is going to be sent to ML algorithms for training and testing followed by evaluation of the performance of the ML algorithms.

Once the sufficient observations are made, one can proceed with further testing that testing a portable sensor for the collection spectra from the sample followed by testing of data acquisition from human blood.

If carried out properly, this can have a significant positive impact on liver disease prevention and early identification for the general population. Detecting slight increases in bilirubin levels in the blood can help identify liver disease early and stop it from progressing to a severe stage, thus reducing the number of liver failure-related fatalities that occur each year.

C H A P T E R 2

LITERATURE REVIEW

In this literature review, we aim to explore the research and developments in the field of bilirubin detection using spectroscopy and ML. We will examine the principles and methodologies of spectroscopic techniques for bilirubin measurement. Additionally, we will review the application of various ML algorithms, such as support vector machines, neural networks, and random forests, in analyzing spectroscopic data for bilirubin detection.

By synthesizing the existing literature and identifying gaps and opportunities for future research, this review seeks to contribute to the advancement of non-invasive and accurate methods for bilirubin detection, ultimately improving clinical outcomes and patient care in various healthcare settings.

Following are some reviews of the studies done by researchers related to the topic:

- In a study done by Yijia Yuan et al (2022), Jiun Lee & Sangho Kim which evaluated the performance of various ML models for determining bilirubin levels from tissue phantom images in both classification and regression tasks. In the classification task, the mean accuracies of the models were as follows: Decision Tree (DT) at 0.672, K-nearest neighbour (KNN) at 0.737, Random Forest (RF) at 0.774, Support Vector Machine (SVM) at 0.827, and LightGBM at 0.848. SVM had the highest mean accuracy and the highest AUC score of 0.93, indicating its superior ability to distinguish between normal and abnormal bilirubin levels. LightGBM also performed well, with a mean accuracy close to that of SVM. In the regression task, the models LightGBM, Support Vector Regression (SVR), and Random Forest (RF) were the most effective in predicting exact bilirubin concentrations. The inclusion of additional features significantly reduced the

mean square error (MSE) values across all models, enhancing their predictive accuracy. Overall, the study concluded that the Support Vector Machine (SVM) model was the most effective, excelling in both classification and regression tasks.[1]

- Marimuthu Muthuvel in his study (2022) concluded that various supervised learning models, including Decision Tree, Random Forest, Support Vector Machine (SVM), and LightGBM, were used to predict jaundice with differing accuracy levels. The models were evaluated based on factors such as gender, age, total bilirubin, direct bilirubin, total protein levels, albumin, SGPT, and SGOT. Logistic Regression achieved an accuracy of 79% in predicting jaundice, while a Neural Network model using Principal Component Analysis explained 75.8% of the data variation. Factor Analysis identified direct bilirubin as the most significant factor influencing jaundice occurrence. Among the models, SVM had the highest performance in distinguishing between normal and abnormal bilirubin levels, with an AUC score of 0.93, and was also effective in predicting bilirubin concentrations accurately. Overall, the study demonstrated the success of ML models in predicting jaundice, with SVM being the most accurate and successful model tested.[2]

- Ahmad Yaseen et al (2024) published a study that was focused on the development of a system for non-invasive detection of neonatal jaundice using ML algorithms. The study aimed to diagnose jaundice in infants based on a dataset of 767 infant images captured using a computer device and a USB webcam. Four ML algorithms, namely Support Vector Machine (SVM), K-

nearest neighbor (KNN), Random Forest (RF), and Extreme Gradient Boost (XGBoost), were evaluated for their performance in classifying infant images as normal or jaundiced. Regarding the accuracy of the methods, the study found that the XGBoost algorithm outperformed the other algorithms with an accuracy of 99.63%. The RF algorithm followed closely with an accuracy of 98.99%, while KNN achieved an accuracy of 98.25%. SVM had the lowest performance among the algorithms, with an accuracy of 96.22%. Based on these results, the XGBoost algorithm was chosen as the classifier for the proposed system due to its high accuracy in diagnosing neonatal jaundice.[3]

- Yunus Karamavuş and Mehmed Özkan in 2019 published their article which focuses on the use of reflectance spectroscopy and regression tools to predict jaundice levels in newborns non-invasively. The study proposes the use of multiple polynomial regression (MPR), artificial neural network (ANN), and support vector regression (SVR) algorithms to predict jaundice levels based on transcutaneous bilirubin (TcB) measurements obtained from 314 babies. The study compares the performance of these methods with a commercially available TcB device, Draeger JM-103. In terms the accuracy of the methods, the study found that all three methods (MPR, ANN, and SVR) accurately predicted the jaundice level with correlation values between 0.932 and 0.943. The ANN method, in particular, showed promising results by converging to more accurate invasive serum bilirubin measurements. The study suggests that the proposed algorithms improve the accuracy of transcutaneous bilirubinometers and may increase their clinical usage.[4]

- Daisaku Morimoto et al (2024) conducted a study that aimed to develop ML models to predict total serum bilirubin by correcting errors in transcutaneous bilirubin measurements using neonatal biomarkers. This retrospective study included 683 infants born at ≥36 weeks gestation and ≥2,000 grams, without a history of phototherapy, between January 2020 and December 2022. Robust linear regression, gradient boosting tree, and neural network models were utilized. The neural network, designed with three layers, showed a root mean square error of 1.03 mg/dL and a mean absolute error of 0.80 mg/dL, both significantly lower than those of transcutaneous bilirubin. The 95% limit of agreement between the neural network's estimates and total serum bilirubin was −2.01 to 2.01 mg/dL. The ML approach reduced the total serum bilirubin estimation error by 25%, potentially decreasing unnecessary blood draws by up to 78%. This integration of ML with transcutaneous bilirubin measurements can enhance accuracy, reduce infant discomfort, and simplify procedures, providing a smart alternative to traditional blood draws for determining phototherapy thresholds.[5]

- Mirhadi Mussavi et al conducted a study that aimed to compare three different methods for measuring neonatal bilirubin concentrations to determine the most appropriate method for specific clinical situations. In this prospective study conducted in 2011, 428 full-term neonates from emergency departments and neonatal wards in Kerman city were evaluated using "Capillary", "Cutaneous" (JM103), and "Laboratory" methods. The correlation coefficients were high, with 0.91 for "David Icterometer" vs. "JM103", 0.96 for "David Icterometer" vs. "Capillary", and 0.85 for "JM103" vs. "Capillary". The David Icterometer measured bilirubin concentrations 2.36 mg/dl higher on average compared to the

JM103 method. The Capillary method showed bilirubin levels 0.91 mg/dl lower than the venous concentrations, while the JM103 measured levels 0.57 mg/dl higher than the Capillary method. The study concluded that due to the minimal differences (less than 1 mg/dl) between the JM103 and Capillary methods, either could be used as alternatives to the standard laboratory method for measuring neonatal bilirubin concentrations.[6]

- Based on a study done by research students at Mount Sinai Hospital in New York, it has been determined that the normal range for unconjugated bilirubin levels in a healthy adults body is from 0.3 to 1.9 mg/dl.[7]

- Kwang-Sun Lee and Lawrence M. Gartner discusses the spectrophotometric characteristics of bilirubin under various physical conditions, such as agitation, oxidation, and changes in pH. The study aims to understand the different states of bilirubin in solution and their spectral properties. Various experiments were conducted to observe the spectral changes in bilirubin solutions under different conditions, including the presence of antioxidants, exposure to light, and alterations in pH. In terms of the conclusion regarding the spectra of bilirubin, the study found that the spectral curves of bilirubin may represent a composite resulting from four physical states of bilirubin: bilirubin truly in solution, bilirubin in fine colloidal dispersion, bilirubin flocculant giving a shoulder at 480-560 nm, and oxidation products of bilirubin with spectral peaks lower than 440 nm. The study highlights the importance of understanding the physical states of bilirubin and their spectral characteristics, which can vary based on factors such as concentration, pH, and exposure to light or agitation.[8]

- Steven L. Jacques and Scott A. Prahl In 1998 Oregon Graduate Institute determined that at 460 nm, the extinction coefficient of bilirubin is ε = 53,846 [cm-1M-1].[9]

- Roland Stocker et al in 1987 said in their research that bilirubin contains an extended system of conjugated double bonds and a reactive hydrogen atom and thus could possess antioxidant properties.[10]

- Antony McDonagh (1990) in his research concluded that bilirubin, traditionally considered a toxic compound at high concentrations, possesses antioxidant properties. The research found that bilirubin effectively scavenges peroxyl radicals, particularly under low oxygen concentrations, and can protect vitamin A and linoleic acid from oxidative destruction in the intestinal tract. The study also demonstrated that bilirubin's antioxidant activity surpassed that of alpha-tocopherol in certain conditions. These findings suggest a potential physiological role for bilirubin as a chain-breaking antioxidant, particularly in lipid oxidation processes.[11]

- According to an article published by Anna Samoc In 2003 a critical review of the values of linear refractive indices of important solvents, such as chloroform, toluene, benzene, and carbon disulfide, was conducted due to their frequent use in studies of nonlinear optical effects in the near-infrared region. The study aimed to evaluate refractive indices of the solvents at longer wavelengths, either from measurements or from the interpolation of the experimental data provided a

proper dispersion equation was available. The research derived precise constants for the Cauchy's dispersion equations of linear refractive indices at 20 °C from available literature data for these solvents in the UV-visible-near-infrared range. The study also highlighted discrepancies in the temperature derivative of refractive index data for these solvents and emphasized the need for accurate refractive index data for the study of nonlinear optical effects. Additionally, the research discussed the use of semiempirical quantum chemical calculations to estimate the values of refractive index and dispersion, but found that the theoretical values did not align with the experimental data. Overall, the research aimed to provide accurate values of coefficients of dispersion equations for these solvents, essential for various optical and optoelectronic applications.[12]

- In a paper from 1882 the Lambert H. Ormsby concluded saying that while it remains a part of the pharmacopeia and can be useful in certain medical contexts, there is a call for caution in its administration due to its history of fatalities, even under careful supervision and by experienced practitioners. The author emphasizes the need for vigilance and careful consideration when using chloroform, suggesting that its risks should not be overlooked despite instances where it has been administered without immediate adverse effects.[13]

- Putcha Venkateswarlu in 1951, did a study which explores the infrared spectrum of methyl chloroform in the 1.6μ-20μ region, determining vibrational assignments and fundamental frequencies. The research reveals PQR structures in most bands, with detailed analysis of symmetric and degenerate vibrations. The study also discusses overtone and combination bands, highlighting the presence

of various combinations and overtones. The author acknowledges Dr. G. Herzberg for guidance and interest in the research.[14]

- You Zung. Hsieh and Michael D. Morris (1988) published an article that discusses the gas-phase chemistry of naked transition metal ions, focusing on the reactivity of Mn+ ions with organic substrates. The study explores the unique reactivity of Mn+ ions compared to other transition metal ions like Fe+, Co+, and Ni+ when interacting with organic molecules. The specific reaction studied is the Mn+-induced demethanation of 4-octyne, which involves an unprecedented 1,6-elimination mode across the CC triple bond. In terms of the conclusions related to the accuracy of the methods, the study provides valuable insights into the reactivity of Mn+ ions with organic substrates. The results obtained from the gas-phase ion chemistry of Mn+ highlight the unique behavior of manganese compared to other transition metals. The study discusses the implications of the unexpected and rich gas-phase ion chemistry of Mn+ in relation to theoretical models. The accuracy of the methods used in the study is supported by the detailed analysis of the reaction mechanisms and the comparison with theoretical concepts, providing a deeper understanding of the reactivity of Mn+ ions.[15]

- The study by Noriyoshi Suzuki and Masayuki Toyoda from 1966 focuses on the infrared absorption spectra of bilirubin and calcium bilirubinate for gallstone analysis using infrared spectroscopy. It corrects previous misinterpretations of the spectra, especially regarding carboxyl groups. The study on the infrared absorption spectra of bilirubin and calcium bilirubinate aimed to establish standard spectra for these compounds to aid in gallstone analysis using infrared

spectroscopy. The research corrected previous misinterpretations of the spectra, particularly regarding carboxyl groups, and emphasized the importance of meticulous analysis and interpretation of standard spectra in bile pigment studies and gallstone component identification. The findings contribute to a better understanding of gallstone composition and analysis techniques.[16]

- The study by Jergen Jacobsen and Rolf Brodersen delves into the kinetic and spectroscopic aspects of the binding mechanism of bilirubin and xanthobilirubic acid to human serum albumin. It explores the relaxational conformational changes post-bilirubin binding, the competitive binding of xanthobilirubinate, and the absence of late conformational changes after xanthobilirubinate binding. The study concludes that there are distinct differences in the binding behaviors of these compounds. The research highlights the competitive binding of xanthobilirubinate with bilirubin, the absence of late conformational changes after xanthobilirubinate binding, and the implications of these findings for understanding the interactions of these compounds with albumin. The study provides valuable insights into the binding mechanisms and conformational changes associated with bilirubin and xanthobilirubinate, contributing to the understanding of their complex interactions with albumin.[17]

- Doron Kaplan and Gill Navon in their in 1983, performed a study which focused on the NMR spectroscopy of bilirubin and its derivatives. It reviews literature NMR data of bilirubin and related compounds, discussing their structure, solution conformation, dynamic properties, and various spectroscopic parameters. The study highlights the advancements in NMR techniques, such as high field

Fourier-transform spectrometers, which have expanded the range of solvents for NMR studies of bilirubin and its derivatives. The research emphasizes the importance of NMR in providing insights into the basic structure, conformation, tautomerism, hydrogen bonding, acidity constants, and other dynamic properties of these compounds in solution. the overall content suggests that NMR spectroscopy plays a crucial role in elucidating the structural and dynamic properties of bilirubin and its derivatives in solution. The study underscores the significance of NMR techniques in providing detailed information on the conformational changes, intermolecular interactions, and dynamic behaviors of these compounds, contributing to a better understanding of their chemical properties and behavior in various solvents.[18]

- The study done by Xia Wang and Jayanta Roy Chowdhury discusses the NMR spectroscopy of bilirubin and its derivatives, focusing on their structural and dynamic properties in solution. It reviews literature NMR data, highlighting advancements in NMR techniques that have expanded the range of solvents for studying these compounds. The discussion covers the structure, solution conformation, dynamic properties, and spectroscopic parameters of bilirubin and related compounds. In conclusion, the study emphasizes the importance of NMR spectroscopy in providing insights into the basic structure, conformation, tautomerism, hydrogen bonding, acidity constants, and dynamic properties of bilirubin and its derivatives in solution. NMR techniques play a crucial role in elucidating the structural and dynamic behaviors of these compounds, contributing to a better understanding of their chemical properties and interactions in various solvent environments.[19]

- The study done by Jr. E. Clinton Texter et al in 2016, focuses on the NMR spectroscopy of bilirubin and its derivatives, highlighting the structural and dynamic properties of these compounds in solution. The study emphasizes the significance of NMR techniques in elucidating the conformational changes, intermolecular interactions, and dynamic behaviors of bilirubin and its derivatives. NMR spectroscopy plays a crucial role in providing detailed insights into the chemical properties and behavior of these compounds in various solvent environments.

- According to a survey conducted by Guofang Ding et al in 2001 China, it was found that Jaundice in most infants was detected at 2-3 days after birth. The bilirubin level usually reached a peak level of 204±54.69 μmol/L at 5 days after birth and then fell. Among the 875 infants, the serum bilirubin levels in 34.4%of neonates were higher than 220.5μmol/L. The mean serum bilirubin level of the infants during the first week after birth varied with geography ($P < 0.001$) and season ($P<0.001$). The serum bilirubin level was significantly associated with gestation age ($P<0.01$), delivery method ($P < 0.01$), weight loss ($P<0.001$), and PCV elevation ($P<0.001$) during the first three days after birth.[20]

- According to March of Dimes, a non profit organization committed to ending preventable maternal health risks and death, ending preventable preterm birth and infant death and closing the health equity gap for all families, Newborn jaundice is very common—about 3 in 5 babies (60 percent) have jaundice. Jaundice usually happens a few days after birth. Most of the time, it’s mild, doesn’t hurt

your baby and goes away without treatment. But if a baby has severe jaundice and doesn't get quick treatment, it can lead to brain damage.[21]

- Based on a research paper published by Bolajoko O Olusanya, Michael Kaplan, Thor W R Hansen in 2018, approximately 24 million out of 134 million live born babies in 2010 developed clinically significant jaundice. It further estimates that 481,000 late-preterm and term neonates developed extreme hyperbilirubinaemia, with 114,000 deaths and over 63,000 survivors experiencing moderate or severe long-term neurological impairments. Additionally, in 2016, neonatal jaundice accounted for 1309.3 deaths per 100,000 livebirths globally in the early-neonatal period (0-6 days), ranking seventh among all causes of neonatal deaths.[22]

- Edward J. Milton, Nigel P. Fox, Michael E. Schaepman in their research concluded that The focus of field spectroscopy on spectral reflectance has hindered its impact on quantitative Earth observation. Reflectance data, while valuable, lacks reproducibility without detailed metadata. To enhance accuracy, a new approach in field spectroscopy should prioritize quantitative measurements of directional radiance and irradiance. Combining different methods is crucial for achieving high-precision measurements of Bidirectional Reflectance Distribution Function (BRDF) in the field. Additionally, field spectroscopy serves a vital role in education and training, with technological advancements making instruments more accessible for diverse measurements.[23]

- In 1976 I.E. McCarthy and E. Weigold published a paper which provided a detailed exploration of the symmetric (e, 2e) reaction in various systems, aiming to investigate electron momentum distributions and angular correlations to understand atomic and molecular structures. Key conclusions include the importance of a symmetric experimental setup for maximizing momentum transfer, the potential neglect of binding effects in high-velocity free electron collisions, and the utility of (e, 2e) spectroscopy as a tool for atomic structure determination using momentum space wave functions. Additionally, the study emphasizes the comparison of experimental and theoretical results, including configuration interaction calculations, and discusses the interpretation of single-particle separation energy and the role of quadruple core excitation in ions like argon, krypton, and xenon. Overall, the paper underscores the value of (e, 2e) spectroscopy in revealing electron distributions and correlations in atoms and molecules, offering valuable insights into their structural properties and interactions.[24]

- Stuart B. in 2005 discussed various aspects of Infrared (IR) Spectroscopy, including microscope techniques, FTIR spectrometer operation, sample preparation considerations, applications in studying biological molecules and industrial analysis, and advanced spectral analysis techniques like difference spectra and deconvolution. A comprehensive overview of IR spectroscopy, highlighting its significance in various fields such as pharmaceuticals, food analysis, and biological research is also provided. The research emphasizes the importance of proper sample preparation, instrument operation, and advanced

spectral analysis techniques for accurate and meaningful results in IR spectroscopy applications.[25]

- J. T. Kindt and C. A. Schmuttenmaer (1997) provided a study which utilized Instantaneous Normal Mode (INM) theory to compute the absolute far-infrared absorption spectra of chloroform and compared it with spectra obtained through conventional formulas. The agreement between INM results and Molecular Dynamics (MD) analysis was satisfactory. However, discrepancies with experimental data, particularly in the range of 4–100 cm^-1, were attributed to the simplicity of inter-molecular potentials rather than the computational method itself. Although the INM analysis effectively replicated the shape and intensity of chloroform's absorption band, including the high-frequency tail, the overall absorbance was consistently underestimated, indicating the need for further refinement of the inter-molecular potential for improved accuracy.[26]

- Amaj A. Laskar et al (2016) conducted a study which investigated the binding of TQ to bilirubin through absorbance measurements. TQ, at concentrations ranging from 0 to 500 nM, was incubated with bilirubin (0.2 μM) for 30 minutes at 25°C. Absorption spectra were recorded in the wavelength range of 350–550 nM. An increase in absorbance at 430 nm was observed with increasing TQ concentration, indicating hyperchromicity. Absorbance reached saturation at 200 nm TQ concentration, beyond which no further increase was observed. This suggests that TQ binds to bilirubin, as evidenced by the rise in absorbance at 430 nm with increasing TQ concentrations. Bilirubin is normal to be found in blood in some low concentrations but needs to be administered if its concentration goes too high.

Bilirubin shows absorption in the range of 200 to 550 nm with usually showing peaks at 430-450 nm. Chloroform is a suitable solvent for bilirubin but needs to be handled carefully in a well ventilated room and also avoiding skin contact.[27]

- Joon-Soo Hahm et al (1992) published a paper on a study which focused on the ionization and self-association of unconjugated bilirubin (UCB) using rapid solvent partition from chloroform. The research minimized crystal effects and degradation by solvent partitioning UCB into buffered aqueous NaCl. Results indicated UCB's pKa values around 8.1 and 8.4, with H2B0 as the dominant UCB species. The study also highlighted the challenges in interpreting ionization and solubility values from crystal dissolution studies.[28]

- Amareshwar Kumar Rai et al in the paper titled Spectroscopic studies and normal coordinate analysis of bilirubin (2001), discussed the vibrational assignments observed in the IR and Raman spectra of bilirubin, comparing calculated and experimental frequencies. It highlights the presence of various bands and their assignments, such as NH stretching, OH stretching, C-O stretching, C-C stretching, and bending modes. The electronic absorption spectrum of bilirubin in CHCl3 solution is also discussed, emphasizing the absorption peak at 454.2 nm. Additionally, the study touches on the photoacoustic spectrum of bilirubin powder, noting interesting features and possible excited states.[29]

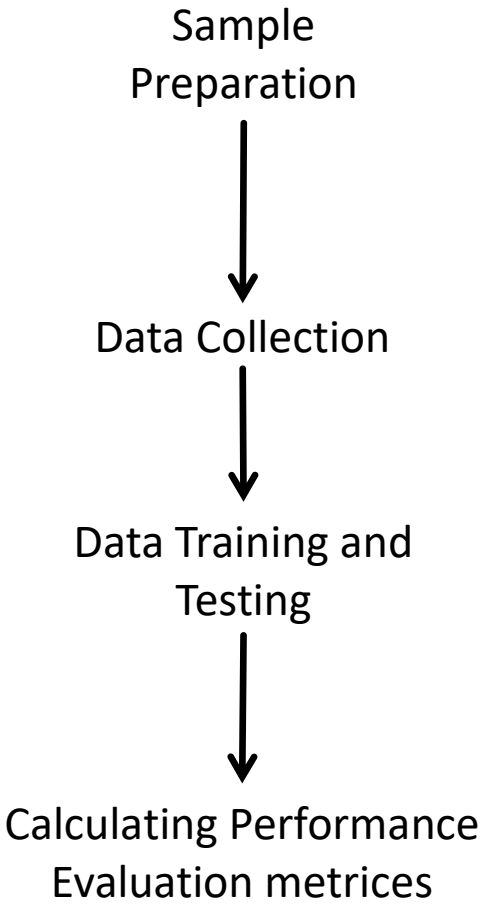
- G. Blauer and G. Wagniere in 1974, presented a paper which discusses the chiroptic properties of pigment chromophores, the conformation of bilirubin and biliverdin in complexes with serum albumin, and the spectroscopic features

observed in the study. It also touches on the computational methods used to analyze the conformation of the pigments and their interactions with albumin. The document emphasizes the specific binding between the pigments and the protein, the influence of pH on chirality, and the challenges in interpreting experimental data.[30]

C H A P T E R 3

METHODOLOGY

In order to collect and process the spectroscopic data before real word trials, we need to test the setup to detect the levels of bilirubin, our methodology employed a series of steps.



3.1 Sample Preparation

For the purpose of testing, Bilirubin powder of AR grade was used.

Fig.3.1: bilirubin powder

Since bilirubin doesn't dissolve in water, we need to use a solvent like chloroform or benzene. For this project we employed chloroform as it is far less toxic compared to benzene.

Fig.3.2: chloroform

Samples for a sample of 0.1mg of bilirubin per decilitre an equivalent sample of 0.00001g/10ml was prepared. Similarly samples from 0.1 - 2.5mg/dL had their equivalents of g/10ml prepared.

Ten millilitre samples were prepared, and each was then further divided into two portions, each containing 5ml. Four spectra of each 5 ml sample were obtained. In total 200 spectras were collected from various combinations of bilirubin. The combinations of bilirubin concentrations and their equivalent 10ml samples are provided in table 3.1.

Table3.1: mg/dL concentrations and their equivalent g/10ml

Concentration (mg/dl)	**Concentration (g/10ml)**
0.1	0.00001
0.2	0.00002
0.3	0.00003
0.4	0.00004
0.5	0.00005
0.6	0.00006
0.7	0.00007
0.8	0.00008
0.9	0.00009

1.0	0.00010
1.1	0.00011
1.2	0.00012
1.3	0.00013
1.4	0.00014
1.5	0.00015
1.6	0.00016
1.7	0.00017
1.8	0.00018
1.9	0.00019
2.0	0.00020
2.1	0.00021
2.2	0.00022
2.3	0.00023
2.4	0.00024
2.5	0.00025

each spectra was then divided into two parts of 5ml each. Each 5ml sample was poured in 10mm quartz cuvettes.

Fig.3.3: quartz cuvette used

3.2 Data Collection

The equipment used for the spectra measurement was Jasco V-770.

Fig.3.4: spectral data collection setup

High-performance UV-Vis-NIR spectrophotometer JASCO V-770 is renowned for its accuracy and adaptability in measuring transmittance, absorbance, and reflectance over a broad spectral range, spanning from ultraviolet (UV) to near-infrared (NIR) wavelengths. It has sophisticated features like fast data capture, high sensitivity, and compatibility with different sample kinds and measurement modes. The tool is widely used for both qualitative and quantitative analysis, as well as the study of molecular structures and interactions, in research, pharmaceutical, chemical, and materials science applications. Along with it, an integrating sphere was employed to obtain the absorption spectra of the sample.

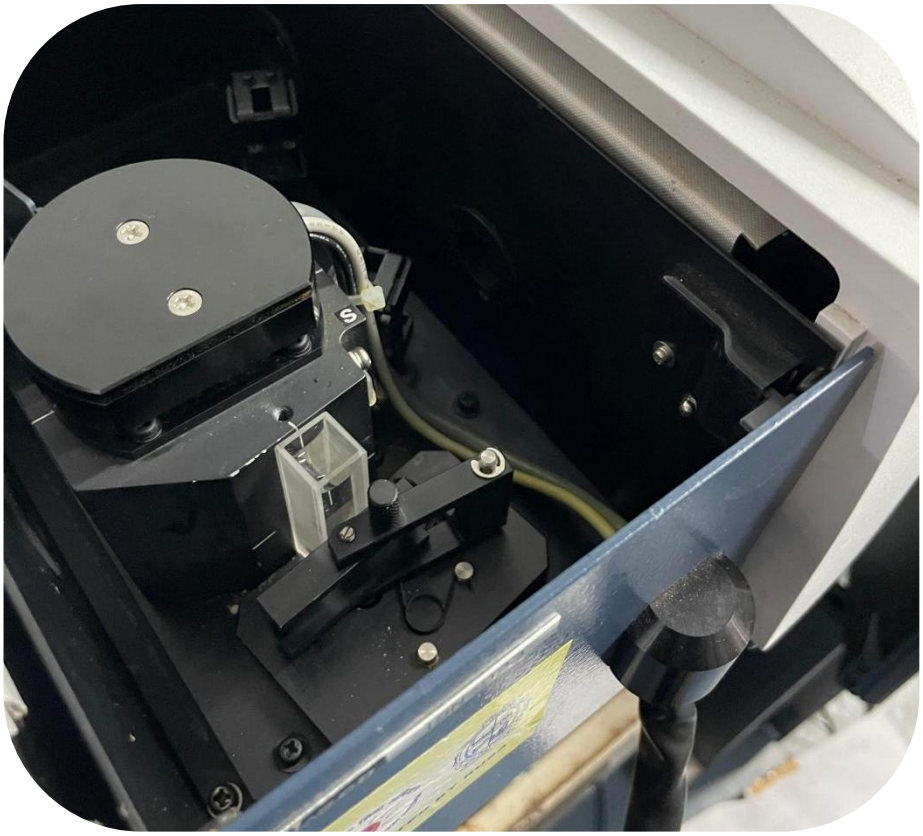


Fig.3.5: cuvette placed in the integrating sphere

Additionally, a spectrolon was positioned behind the cuvette to block any additional light that might be reflected from the spectrophotometer's surface.

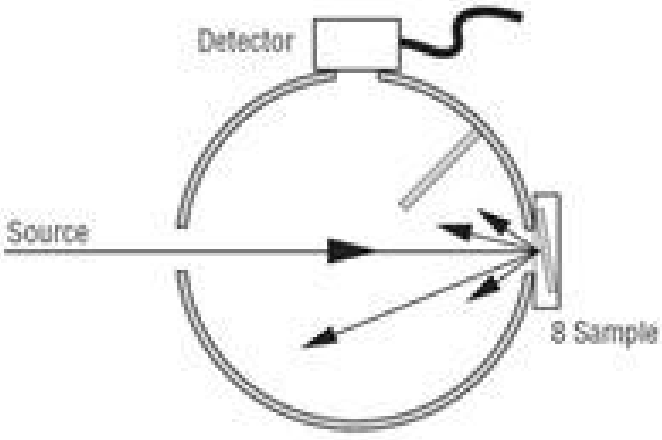


Fig.3.6: integrating sphere ray diagram

"Spectra Manager" was the software used to control the spectrophotometer, and the 200nM to 800nM wavelength range was selected for capturing.

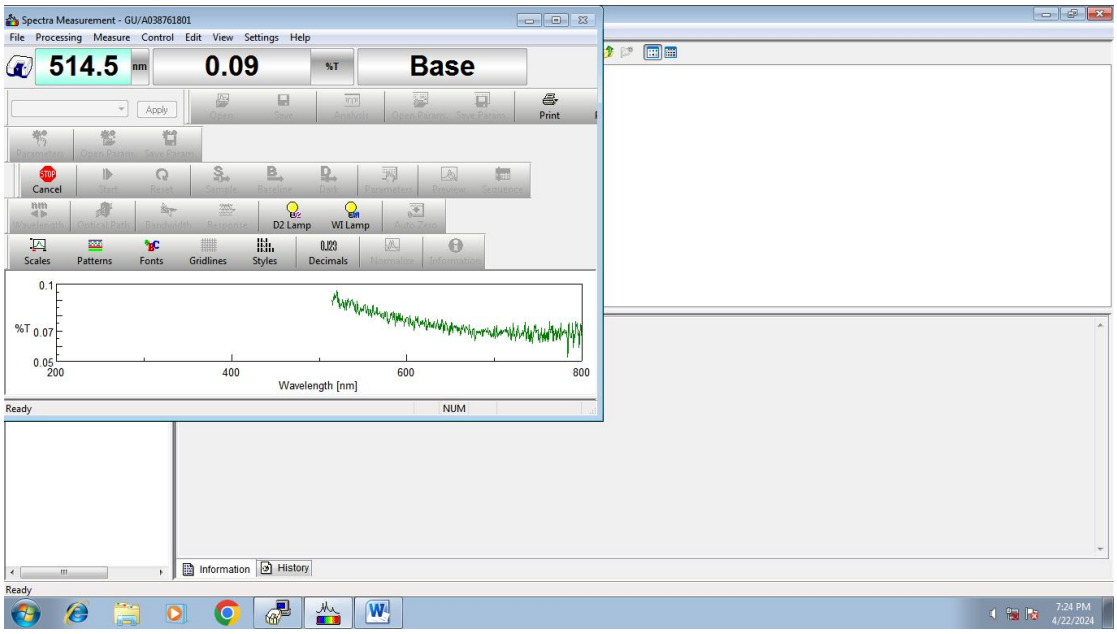


Fig.3.7: spectra manager used to observe and perform data collection

Before the spectral data of the samples was captured, a baseline spectra was taken of the solvent that is chloroform in the cuvette. Each of the 10ml samples was divided into two parts of 5ml each and for each each 5ml spectra was taken four times, for all the samples from 0.1 to 2.5ml/dL.

Upon capturing the spectra, a particular pattern was observed in the range of 250-500 nm wavelength. This wavelength looked familiar to that of the wavelengths observed during literature review.

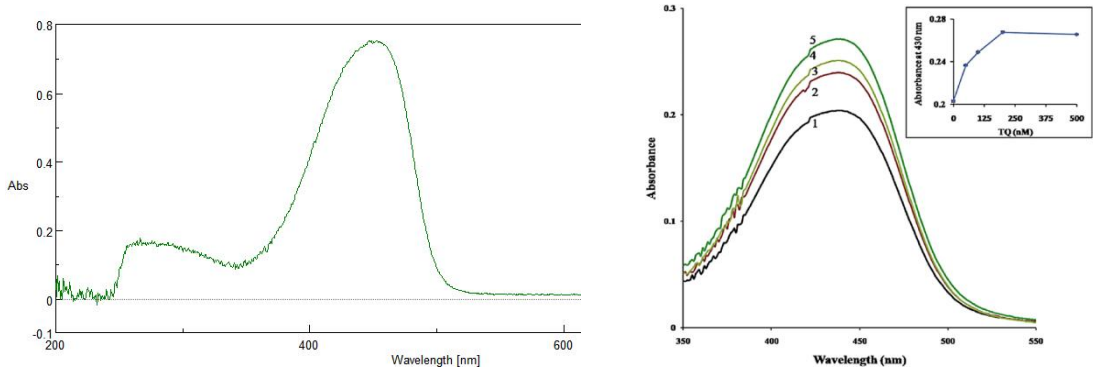


Fig 3.8: comparison of obtained an researched spectra

It was also observed that as the concentration of bilirubin was increased, the absorption was also increasing.

Following is an example comparing two spectras of 1.2mg/dL and 1.4mg/dL

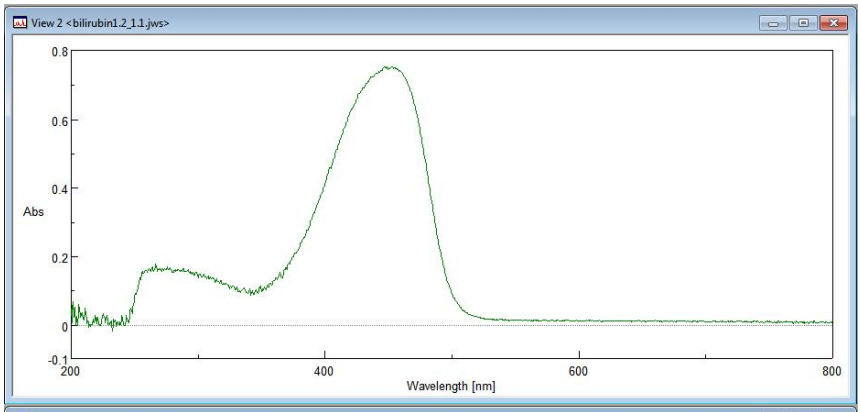


Fig.3.9: Bilirubin 1.2mg/dL

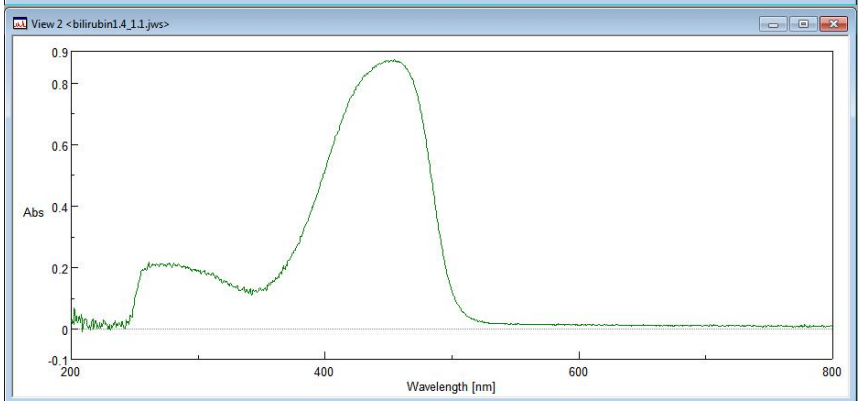


Fig.3.10: Bilirubin 1.4mg/dL

The software was used to save the data in .csv format and was later compiled altogether.

The compiled data was then used to make a dataset.

3.3 Data Training and Testing

Once the data was compiled into one singular dataset, it was then divided into a training dataset and a testing dataset. Code for reading and splitting dataset is provided in Appendix on page 77.

The training dataset was used to train a few ML algorithms.

The data was divided in an 80-20 split. That is, 80% data was used for training and the remaining 20% was used for testing.

The data was trained on four ML Algorithms;

- K-Nearest Neighbour
- Decision Trees
- Random Forest
- Partial Least Square

3.3.1 K-Nearest Neighbour:

Popular and simple, the K-Nearest Neighbour (KNN) technique is used in ML for both regression and classification tasks. The idea behind it is simple: the algorithm guesses the class or value of a query instance by looking at the average value or majority class of its K nearest neighbour in the feature space. First, all training instances that are accessible are saved, together with the classes or values that correspond to them. The algorithm determines the distance, usually in Euclidean distance, between each training sample and the query instance before making any predictions. On the basis of these distances, it then chooses the K closest neighbour. In classification tasks, the majority class of the query instance's K nearest neighbour determines the predicted class; in regression tasks, the predicted value is the mean of the values of the query instance's K-Nearest Neighbour. Because KNN uses the training instances for prediction directly during the training phase rather than creating an explicit model, it is referred to as a lazy learning algorithm. KNN is computationally efficient, but because it depends on all training instances, it might predict things more slowly for huge datasets. Nevertheless, it is a useful tool in many ML applications because to its simplicity and efficacy. A brief flowchart of KNN is provided in figure 3.11, and the code used for training and testing the datasets using KNN is given in appendix on page 78.

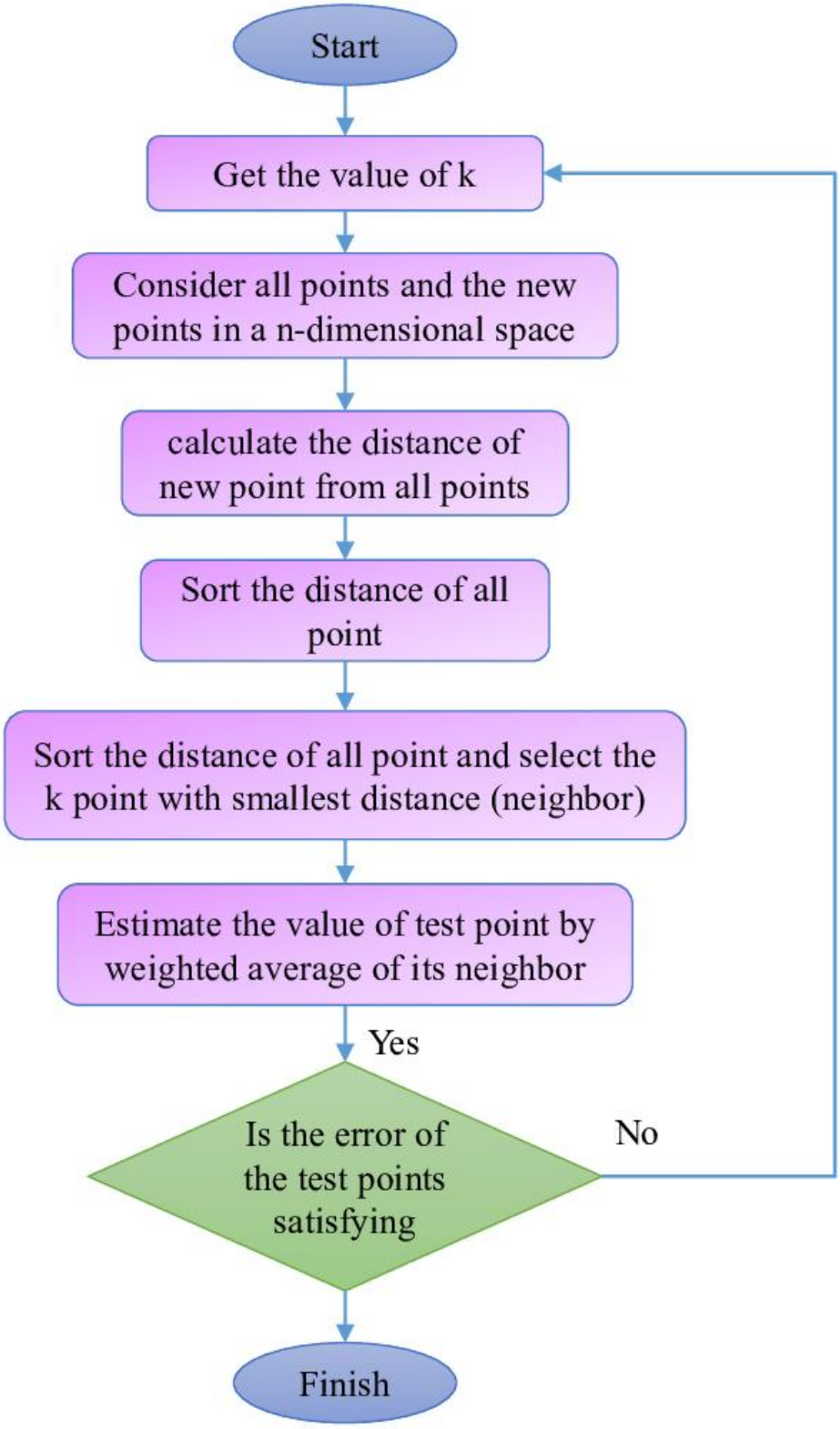


Fig.3.11: flowchart for KNN

3.3.2 Decision Trees:

A basic ML algorithm for classification and regression problems is the decision tree. They work by dividing the feature space into a flowchart-like hierarchical structure. Based on the value of a feature, a choice is taken at each node of the tree, resulting in subsequent branches that represent various outcomes. Until a stopping requirement is met, such as a maximum tree depth or a minimum amount of samples in a node, this process repeats recursively. The secret to using decision trees is figuring out which characteristic, by maximizing information gain or minimizing impurity measures like entropy or Gini impurity, is best for splitting the data at each node. Once built, decision trees provide clearly visualized and interpretable models, which makes them useful tools for comprehending decision-making processes across a range of areas. Decision trees, however, are susceptible to overfitting, especially if noisy data is used or if the tree is allowed to grow too deeply. Pruning, restricting tree depth, and utilizing ensemble techniques like Gradient Boosting or Random Forests might assist lessen these problems and enhance the generalization performance of the model. Decision trees continue to be widely used and valued due to their interpretability, simplicity, and capacity to handle both category and numerical data, despite certain restrictions. General flowchart of decision trees is provided further in figure 3.12 and the code used to read the datasets into training and testing for decision tree is provided in appendix on page 79.

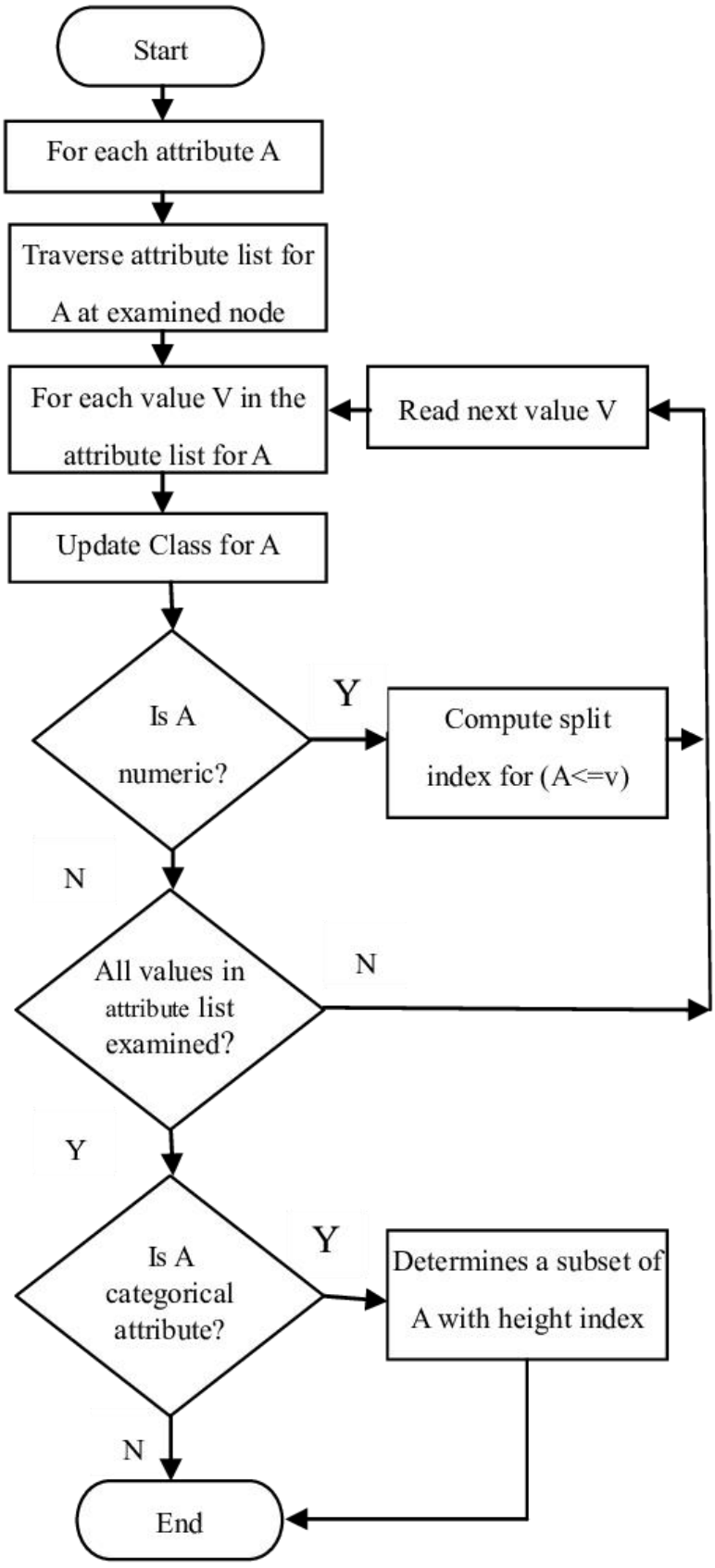


Fig.3.12: flowchart for decision tree

3.3.3 Random Forests:

One popular ML technique for both classification and regression applications is Random Forests, which is strong and adaptable. During training, it builds a large number of decision trees. To determine the optimal split at each node, each tree is trained on a random part of the data and chooses a random selection of characteristics. Each member of the tree ensemble makes a prediction during prediction, and the final result is decided by a voting mechanism for classification or by averaging for regression. By improving generalization and decreasing overfitting, this method strengthens Random Forests' resistance to outliers and noisy data. It also offers insights into the significance of features, which helps with feature selection and result interpretation. A flowchart to demonstrate the basic working of random forests has been given in figure 3.13 and the code used to implement random forests for the dataset has been provided in appendix on page 80.

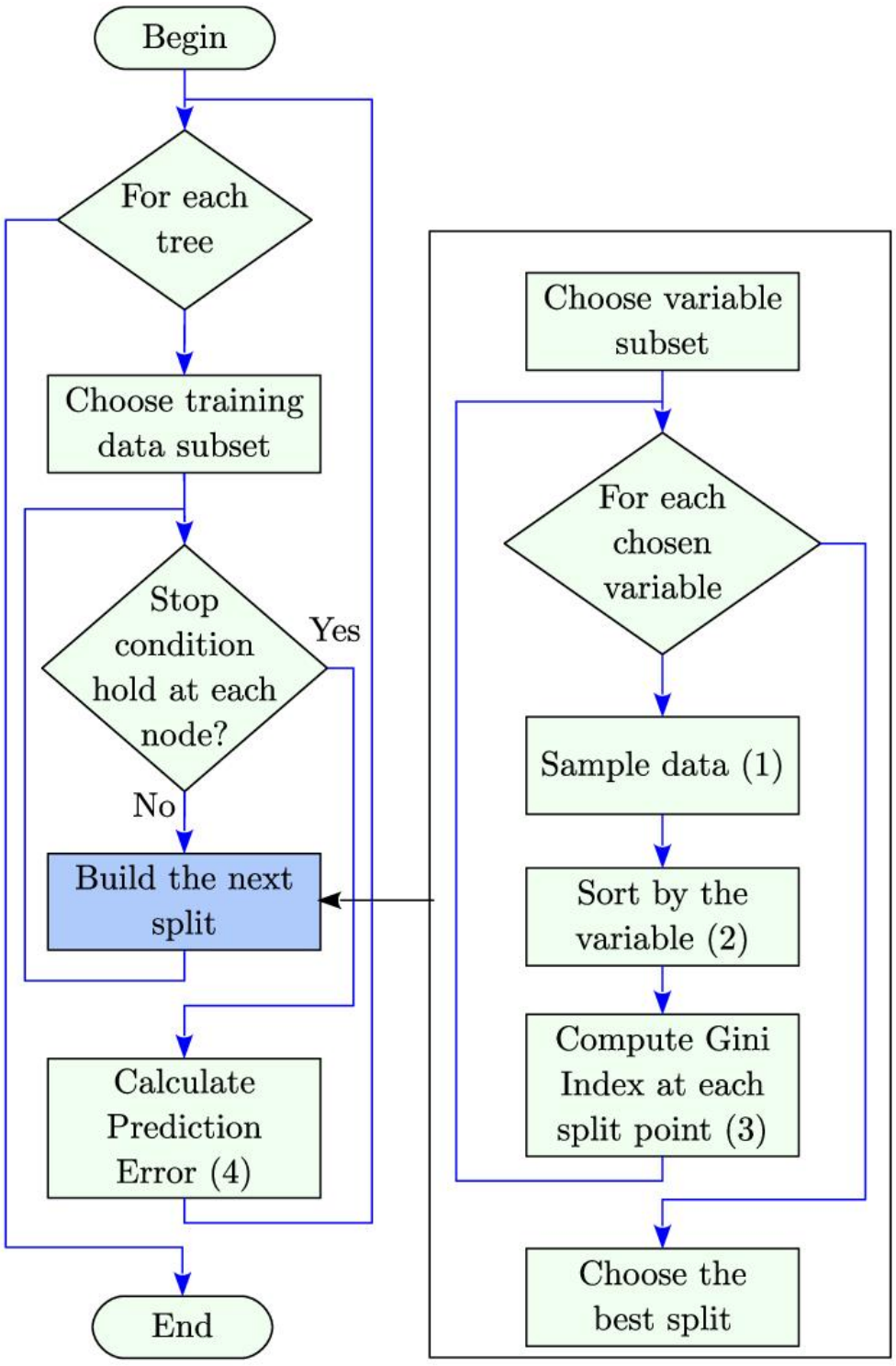


Fig.3.13: flowchart for random forests

3.3.4 Partial Least Square:

A statistical technique that's frequently used in data modelling and regression analysis is called partial least squares, or PLS. It works especially well with datasets that include a large number of predictors, or independent variables, or when there are problems with multicollinearity, or the presence of correlations between the predictors. PLS works by finding a set of latent variables, known as components, that capture the maximum variance in both the predictors and the response variable simultaneously. Unlike traditional regression methods, PLS creates these components iteratively, by maximizing the covariance between the predictors and the response variable. This iterative process enables PLS to handle complex datasets efficiently, making it a valuable tool in fields such as chemometrics, bioinformatics, and marketing research. A general idea of partial least square can be observed in the figure 3.14 and the code to implement PLS can be found in appendix on page 81.

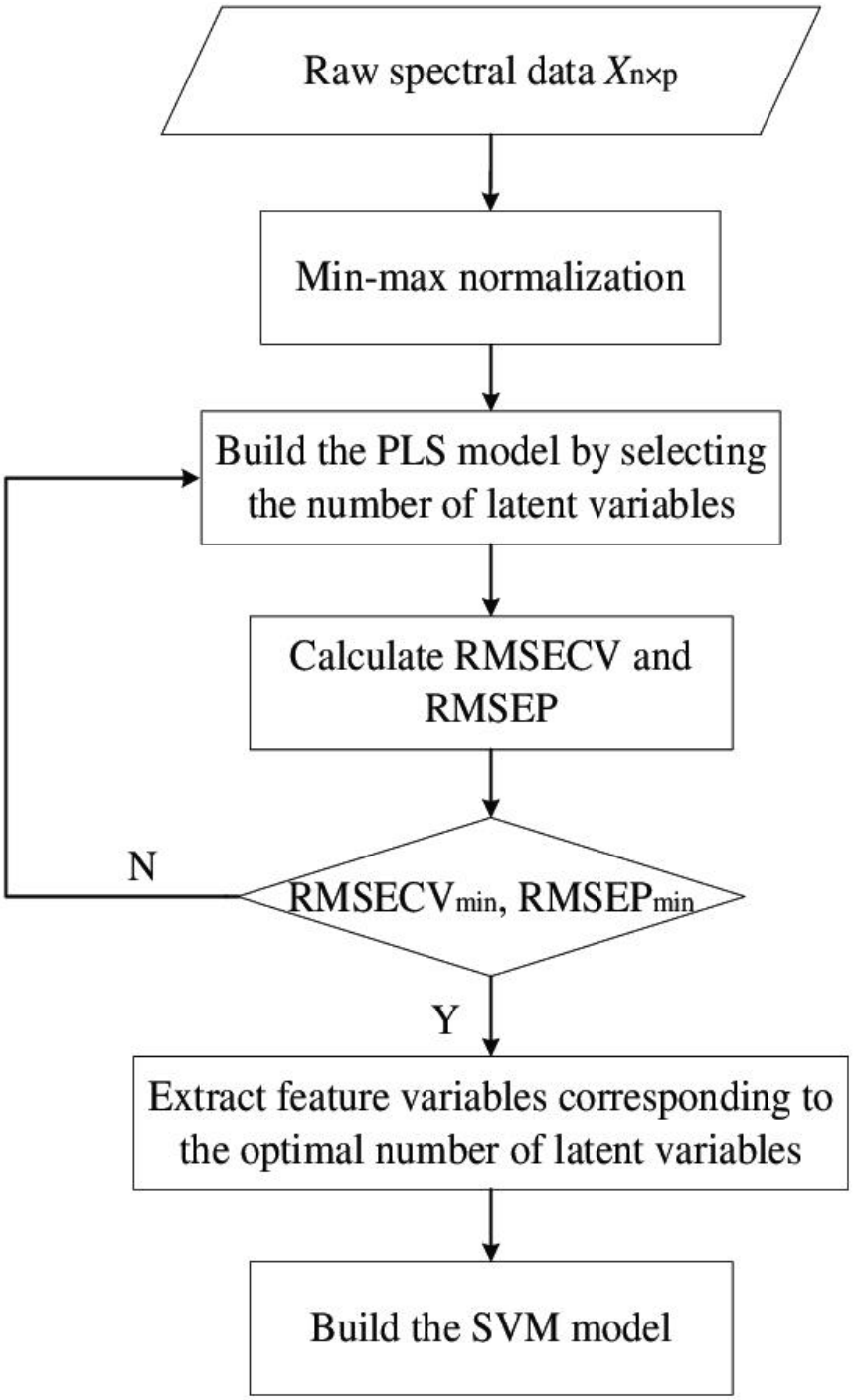


Fig.3.14: flowchart for partial least square

3.4 Calculating Performance Evaluation Metrices

To evaluate the prediction accuracy errors and performance of the algorithms after training and testing, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R-Squared (R2) are calculated along with the plotted graphs.

3.4.1 Mean Absolute Error:

$$\frac{1}{N}\sum_{i=1}^{N} \left|\text{actual value - predicted value}\right|$$

a measurement of the typical amount of errors in a set of forecasts, without consideration for direction.

3.4.2 Root Mean Square Error:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N} (\text{actual value - predicted value})^2}$$

measures the average difference between values predicted by a model and the actual values.

3.4.3 R-Squared:

$$1 - \frac{\sum (\text{actual value - predicted value})^2}{\sum (\text{actual value - mean of dependant variable})^2}$$

a statistical metric used to assess how well a regression model fits data.

C H A P T E R 4

RESULTS

For this study the aim was to develop a portable device to detect small rises in bilirubin concentrations. Samples of various concentrations were taken and a total of 200 sample data was collected and compiled. The data was then split and trained into four ML algorithms. This was followed by testing the accuracy of the ML algorithms by calculating the evaluation matrices.

The obtained dataset was compiled and plotted to visualize the data.

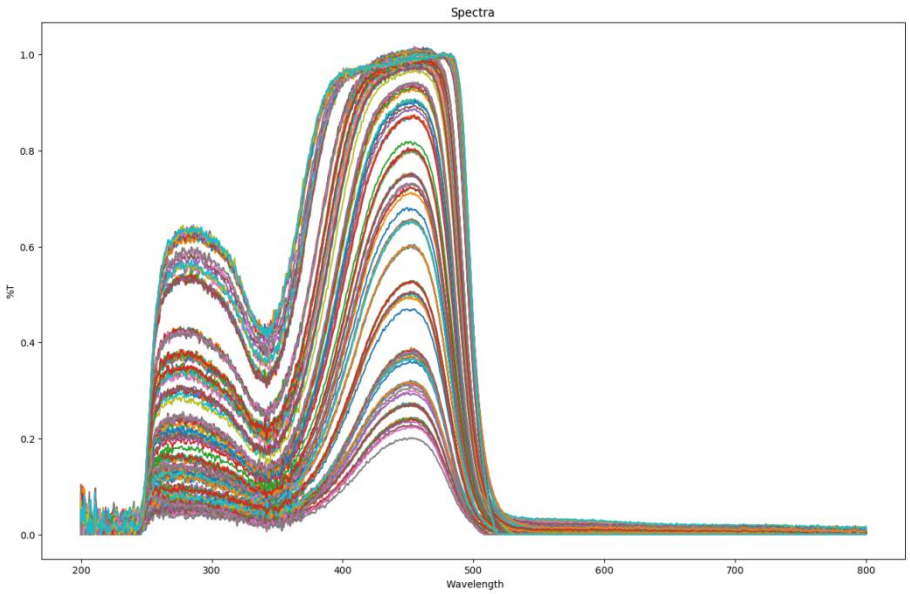


Fig.4.1: spectral plot of the dataset

A solid mix to use in the final result can be found by comparing different ML algorithms and data splits.

Once the data was divided into an 80-20 split, it was given to the four ML algorithms for training and testing. Once the testing was done a plot for comparison of predicted and actual values was done and the evaluation matrices MAE; RMSE; R2, were also calculated using the code provided in appendix on page [88].

MAE :

$$\frac{1}{N}\sum_{i=1}^{N} |\text{actual value - predicted value}|$$

a measurement of the typical amount of errors in a set of forecasts, without consideration for direction.

RMSE:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N} (\text{actual value - predicted value})^2}$$

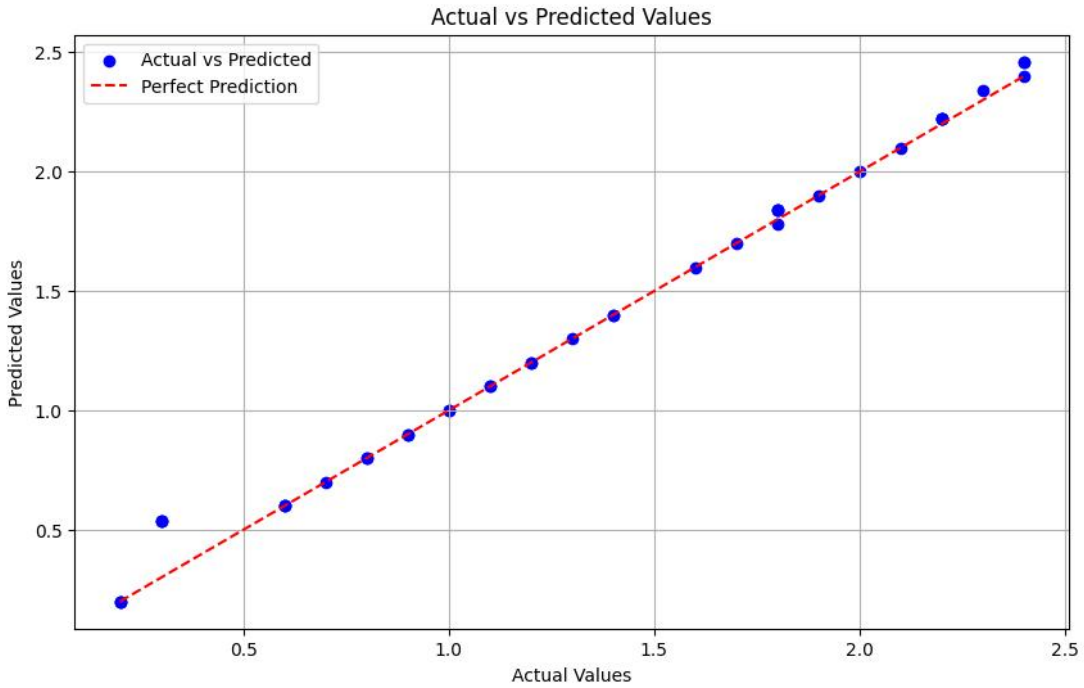
measures the average difference between values predicted by a model and the actual values.

R2:

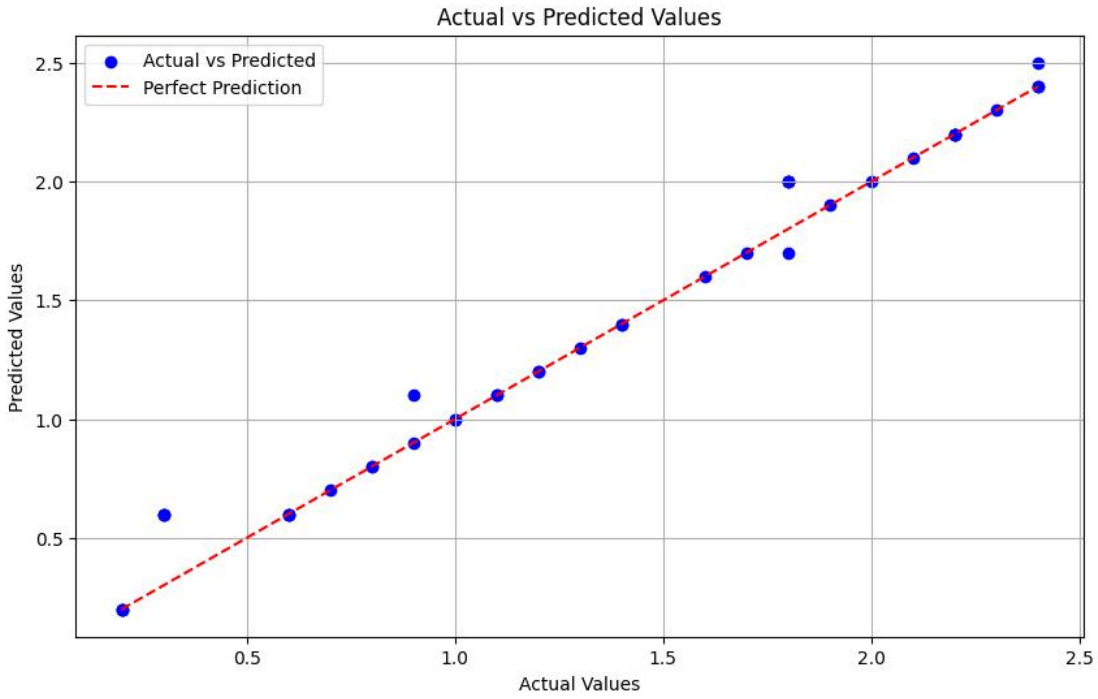
$$1 - \frac{\sum (\text{actual value - predicted value})^2}{\sum (\text{actual value - mean of dependant variable})^2}$$

a statistical metric used to assess how well a regression model fits data.

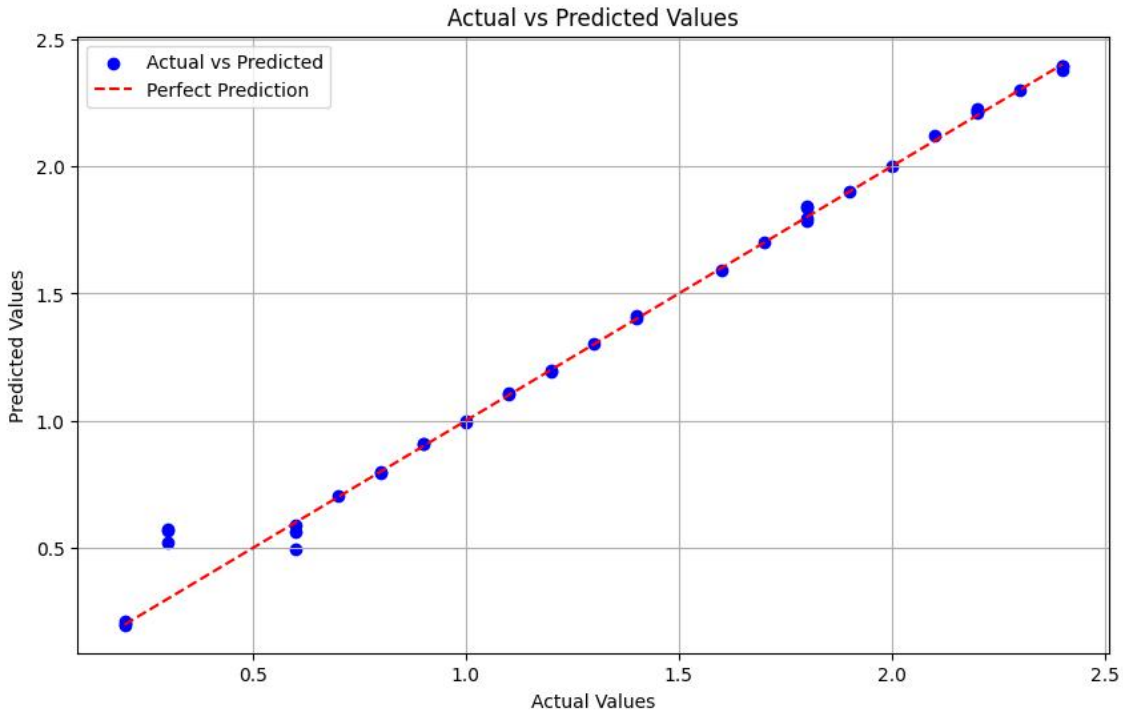
For the 80-20 data split the plots and evaluation matrices were calculated and as per the calculations KNN gave a MAE=0.0275, RMSE=0.0686 and a R2=0.9907.



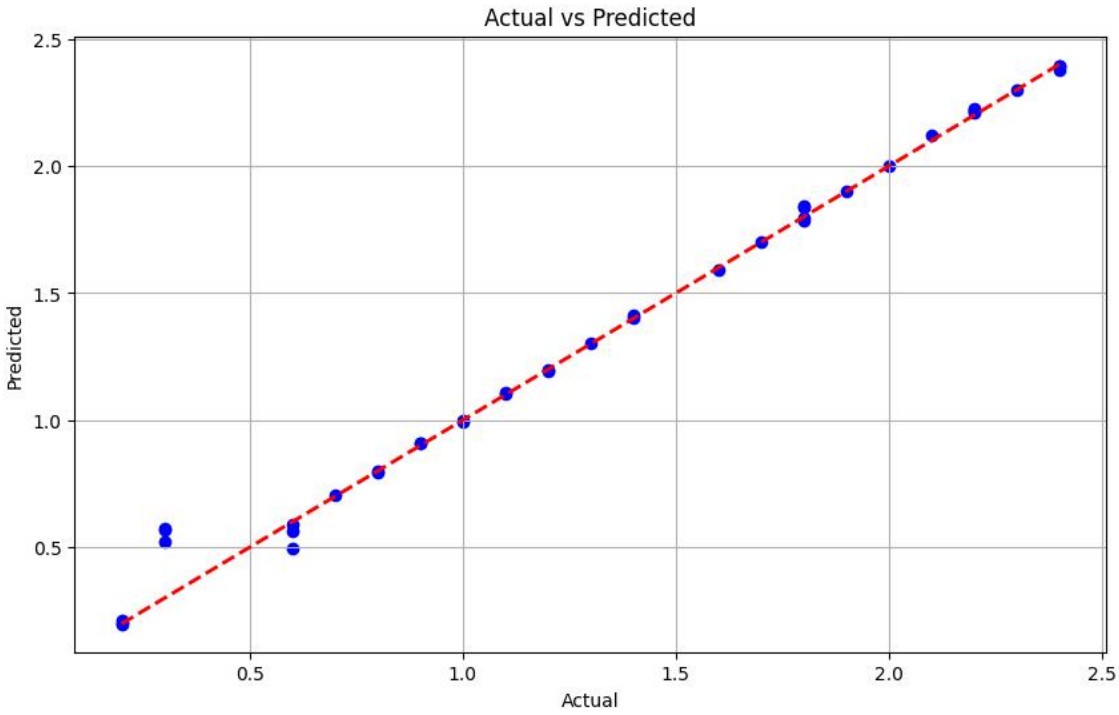
Decision Trees also performed in a similar fashion with MAE=0.0475, RMSE=0.1061 and R2=0.9777.



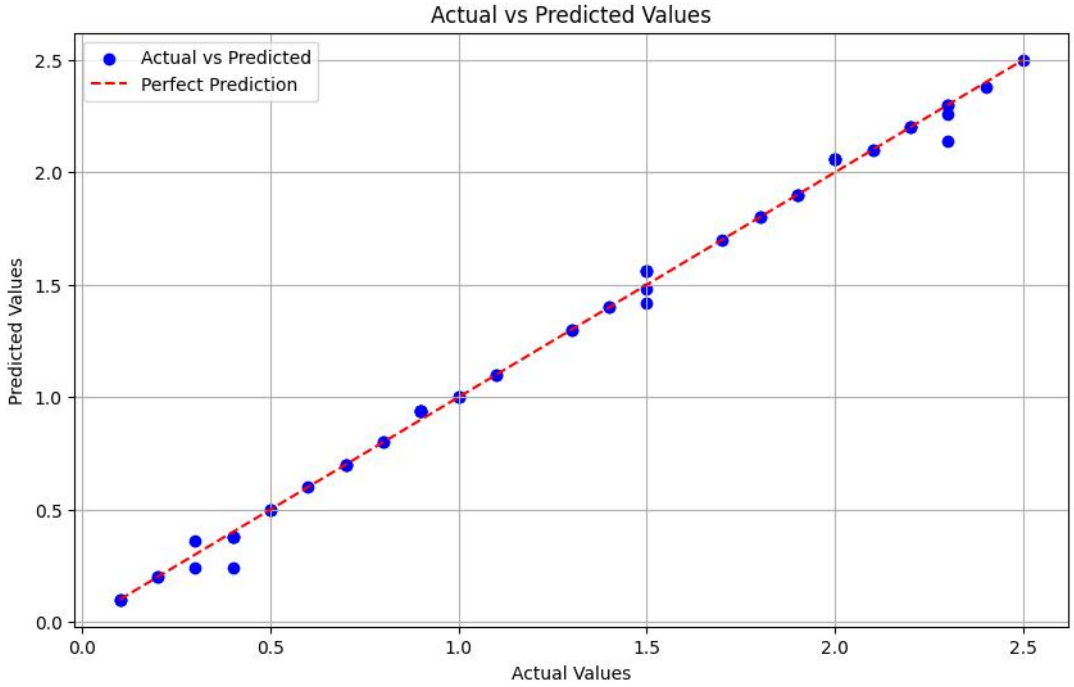
Random Forest was not far off with an MAE=0.0314, RMSE=0.0735 and R2=0.9893.



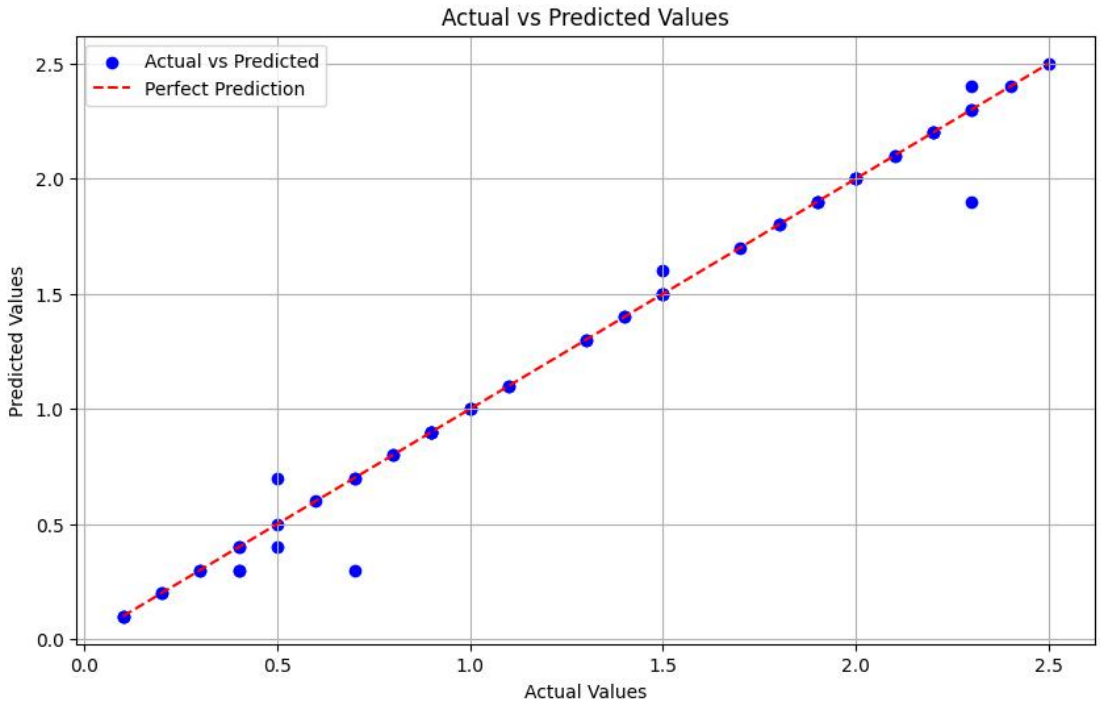
Partial Least Square gave a performance quite similar to the first the ML algorithms with MAE=0.031, RMSE=0.0735 and R2=0.9893.



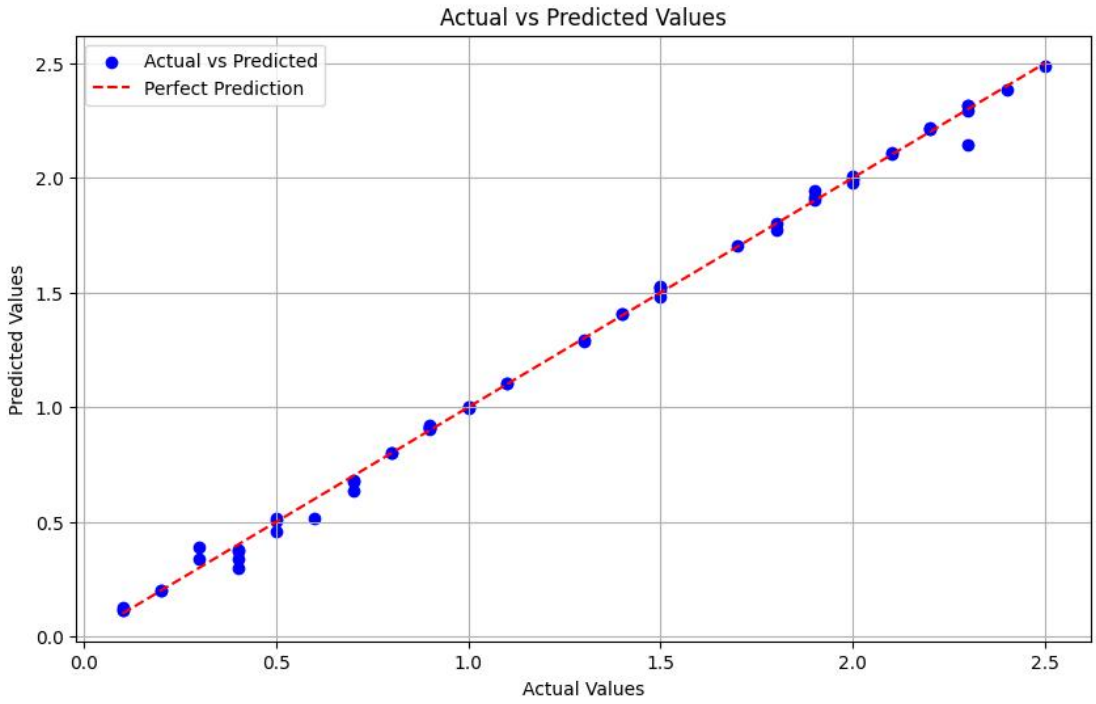
Similarly the plots and evaluation matrices of other combinations was also found and in the case of the 70-30 split, KNN gave a performance with MAE=0.0240, RMSE=0.0562 and R2=0.9926.



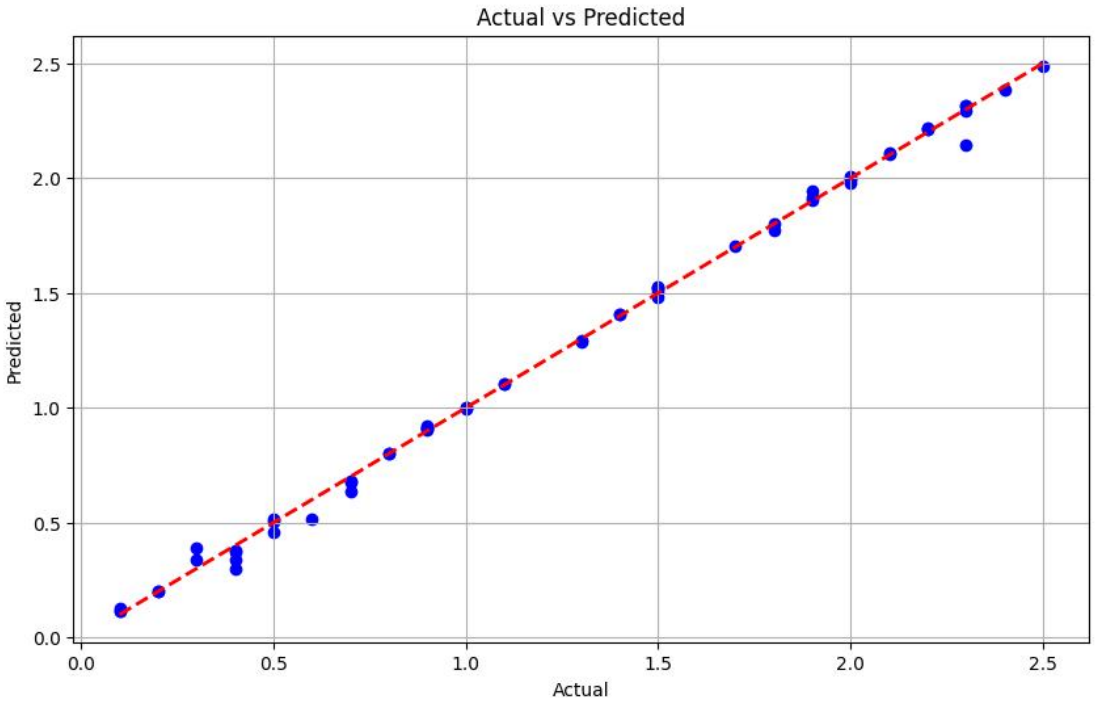
For the same 70-30 split Decision Tree gave the values of MAE=0.0117, RMSE=0.0563 and a R2=0.9926.



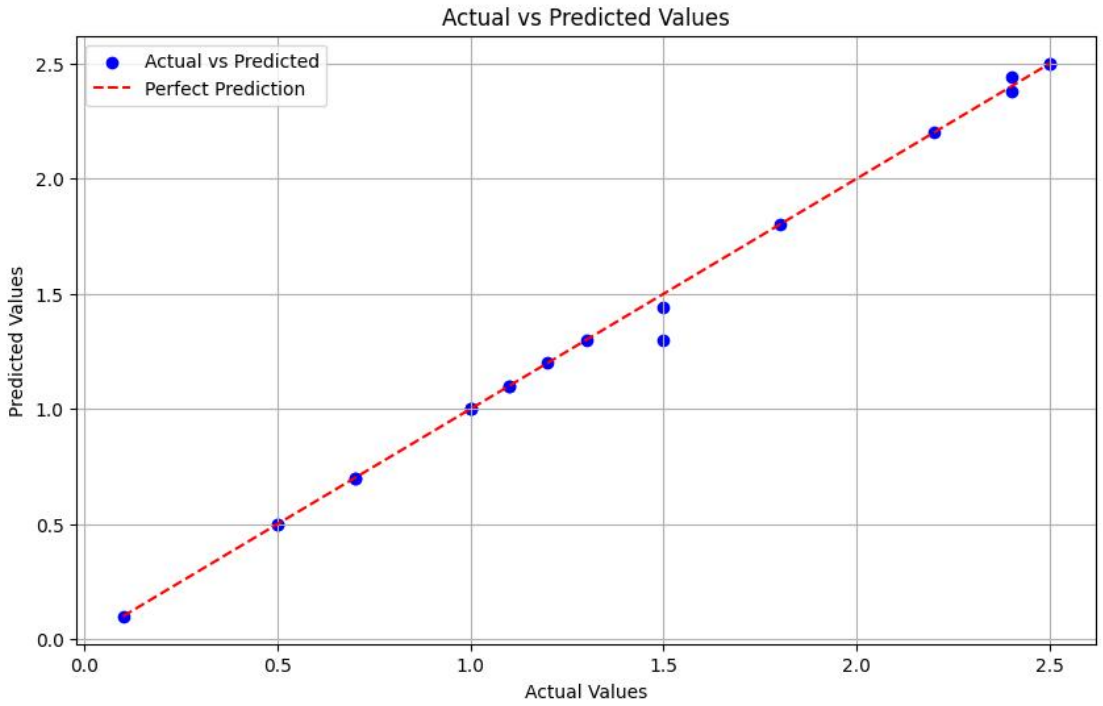
Random Forest for 70-30 split gave the evaluation values of MAE=0.0277, RMSE=0.0513 and R2=0.9938.



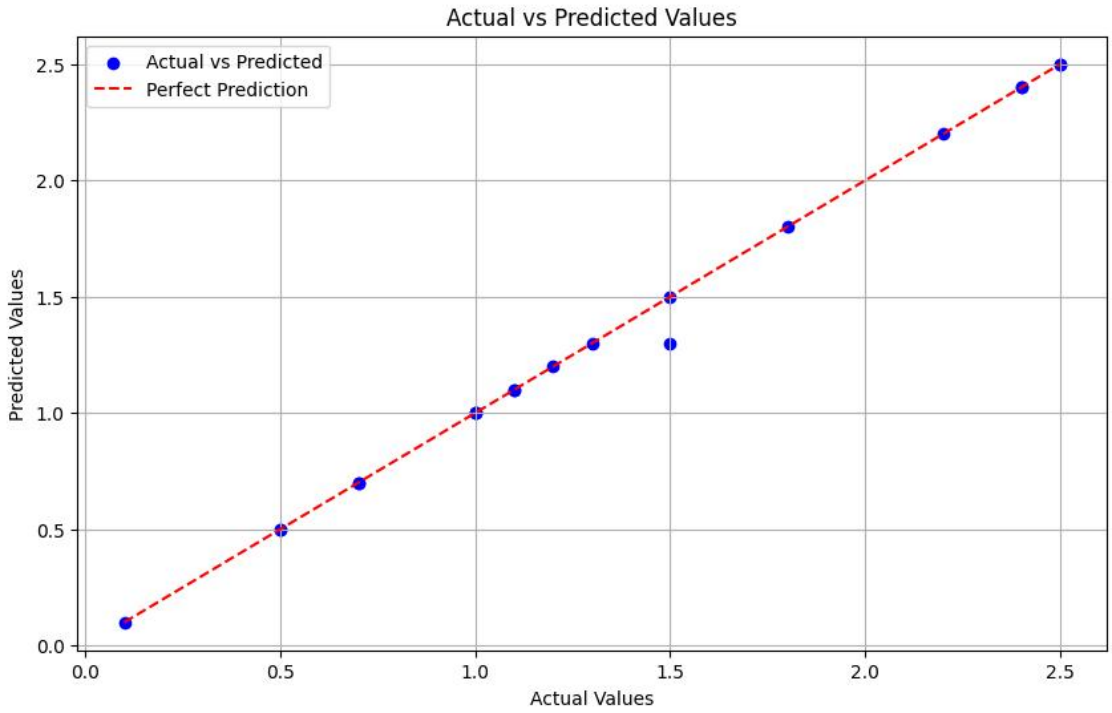
Partial Least Square for 70-30 split gave the evaluation values of MAE=0.0277, RMSE=0.0513 and R2=0.9938.



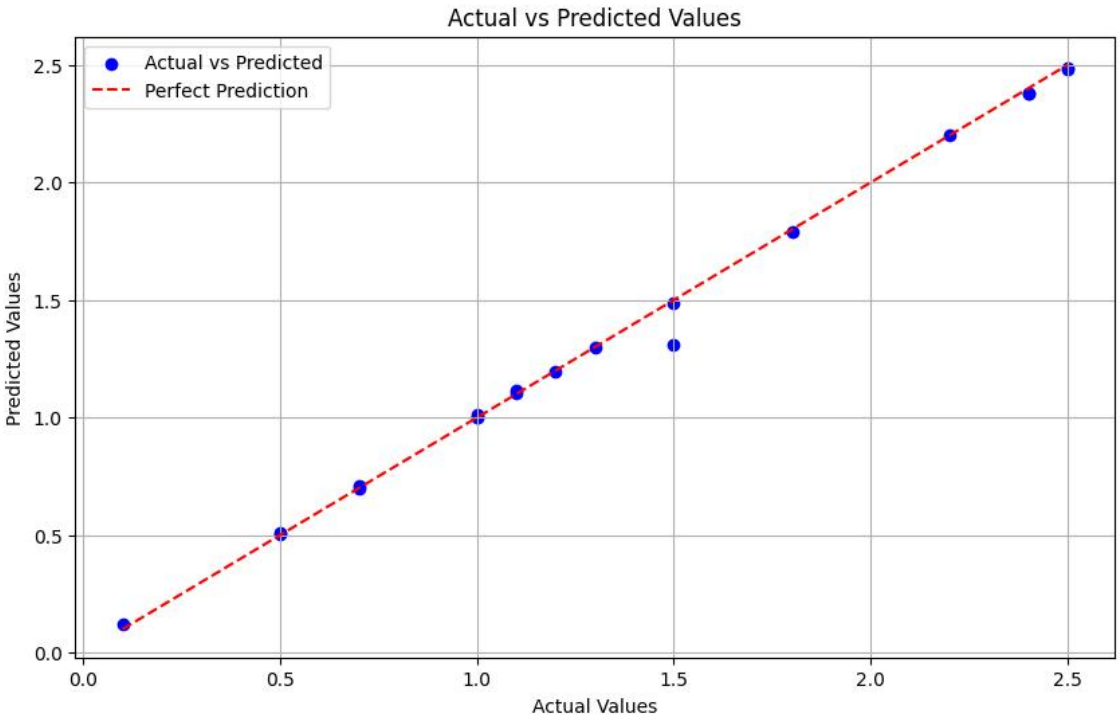
Similarly for the Split 90-10 KNN gave an evaluated performance with MAE=0.0100, RMSE=0.0290 and R2=0.9982.



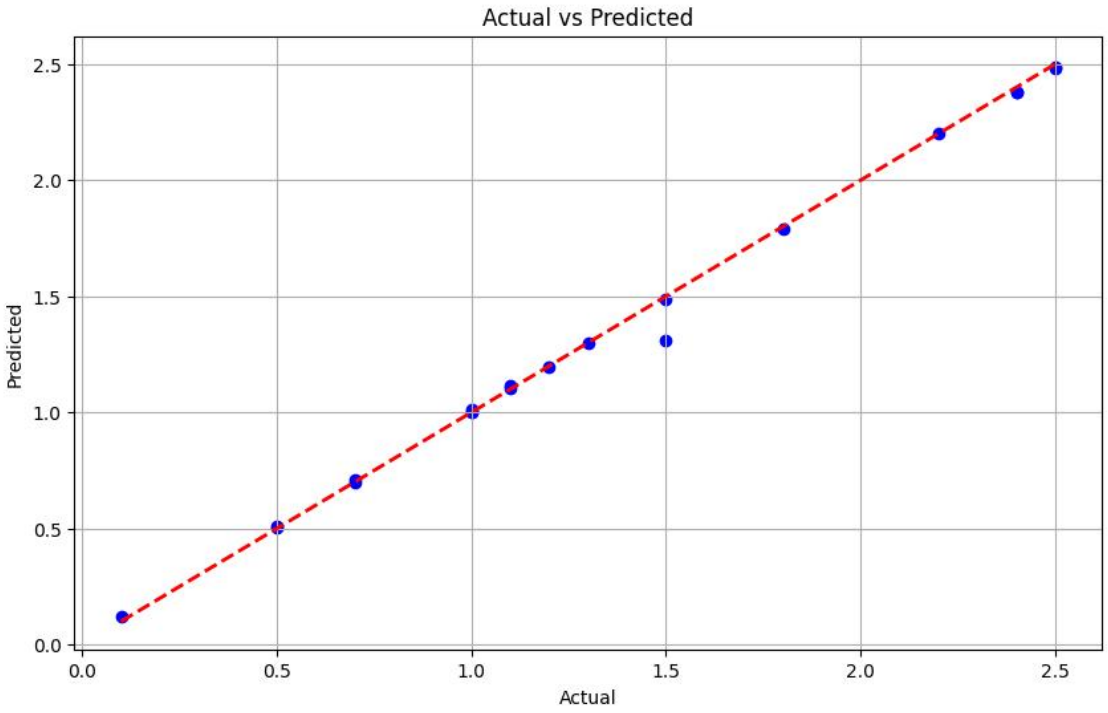
For decision tree the 90-10 data split performed with MAE=0.0100, RMSE=0.0316 and R2=0.9979.



Random forest with the 90-10 data split gave the values for MAE=0.0148, RMSE=0.0249 and R2=0.9986.



And similarly Partial Least Square for the 90-10 data split gave the evaluation matrices as MAE=0.0148, RMSE=0.0249 and R2=0.9987.



Analysis:

It is very noticeable that as the concentration of bilirubin was going up in the solvent, the absorption was also increasing along with it. And this is a crucial part which can be used for predicting the values of bilirubin concentrations.

The following table 4.1 shows a comparative analysis of all the performances of the ML algorithms based MAE, RMSE and R2.

Table4.1: Comparison of all the performance matrices

DATA SPLIT	MLA	MAE	RMSE	R2 SCORE
70 - 30	KNN	0.0240	0.0562	0.9926
	Decision Trees	0.0117	0.0563	0.9926
	Random Forests	0.0277	0.0513	0.9938
	Partial Least Square	0.0277	0.0513	0.9938
80 - 20	KNN	0.0275	0.0686	0.9907
	Decision Trees	0.0475	0.1061	0.9777
	Random Forests	0.0314	0.0735	0.9893
	Partial Least Square	0.0313	0.0735	0.9893
90 - 10	KNN	0.0100	0.0290	0.9982
	Decision Trees	0.0100	0.0316	0.9979
	Random Forests	0.0148	0.0249	0.9986
	Partial Least Square	0.0148	0.0249	0.9987

These results, along with a comparison of the evaluation matrices, allow us to conclude that for the 80-20 split KNN produced the best results with MAE of 0.0275, RMSE of 0.0686 and R2 score of 0.9907.

We can confidently state that the decision tree fared better for the 70-30 split than the other three methods with values for MAE as 0.0117, RMSE of 0.0563 and R2 score of 0.9926.

Ultimately, all of the algorithms performed incredibly well for the 90-10 split, but KNN proved to be the most accurate with MAE of 0.0100, RMSE of 0.0290 and R2 score of 0.9982.

FUTURE OUTLOOK

a model can be suggested where a sensor like a Triad Spectroscopy Sensor can be integrated with an Arduino micro-controller and the data received data from the sensor can be used to predict the bilirubin level.

A powerful remote sensing tool that can collect precise spectrum data at various wavelengths is the Triad spectrum Sensor. It works by identifying and quantifying light reflected from surfaces, which enables the examination of different substances and environmental factors. The Triad Spectral Sensor is commonly utilized in fields including agricultural, environmental monitoring, and military reconnaissance. It has the ability to distinguish between various vegetation kinds, evaluate the quality of water, and identify objects that are concealed. Exact data collecting is made possible by its high-resolution capabilities, which are essential for exact analysis and decision-making. The sensor is a useful instrument for practical applications in a variety of industries as well as scientific research due to its superior technology and versatility.

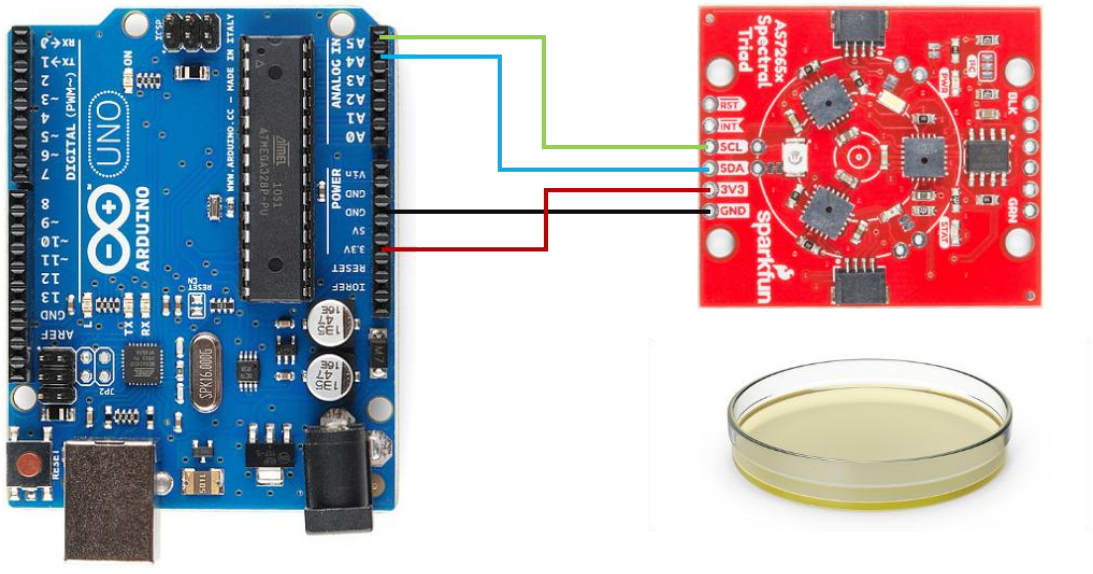


Fig.4.2: setup for using triad spectral sensor for spectroscopy

Triad spectral sensor has a total of 18 distinct channels and each of the tree sensors take care of 6 channels each.

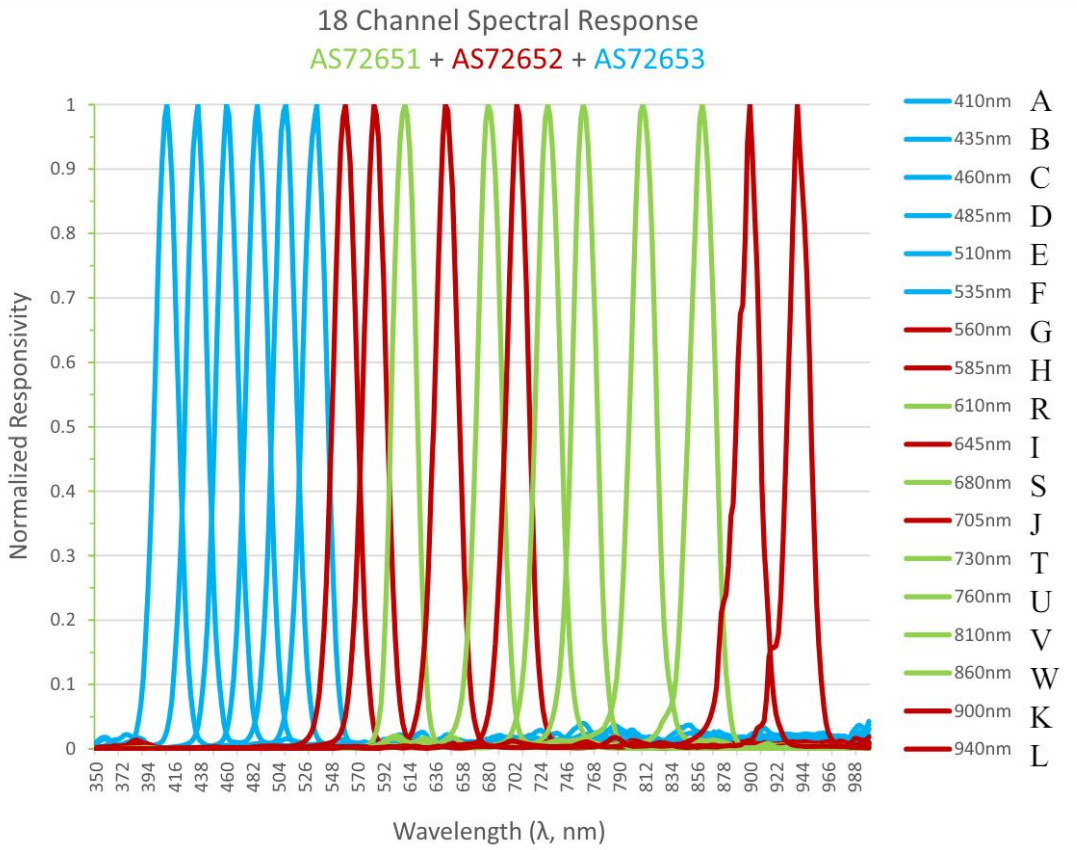


Fig.4.3: all the 18 channels triad spectral sensor can capture

Since the variation in absorption spectra for bilirubin occurs in the range of 350-500, it is possible to prepare a setup with just one setup with sensor AS72653, as it covers the needed range and hence the cost can be further reduced.

A simple code to run the triad with all the 18 channels by integrating it on an Arduino can be found in appendix on page 82.

Preliminary testing on a prepared bilirubin samples should be done before proceeding with human blood testing.

With further development, the proposed set up can be a cost effective and efficient replacement for the current available options for testing bilirubin concentration and also available in rural hospitals.

APPENDIX

Code for Splitting the dataset:

```
data=pd.read_excel('/content/dataset.xlsx')

data.head()

y=data.conc

x=data.drop('conc',axis=1)

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.1)
```

Code for KNN:

```
k = 5  # You can adjust the value of k as needed

model = KNeighborsRegressor(n_neighbors=k)

model.fit(x_train, y_train)

y_pred = model.predict(x_test)

mae=mean_absolute_error(y_test, y_pred)

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

rmse = np.sqrt(mse)

print("Mean Absolute Error (RMSE):", mae)

print("Mean Squared Error (MSE):", mse)

print("Root Mean Squared Error (RMSE):", rmse)

print("R-squared (R2) Score:", r2)

plt.figure(figsize=(10, 6))

plt.scatter(y_test, y_pred, color='blue', label='Actual vs Predicted')

plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)],

color='red', linestyle='--', label='Perfect Prediction')

plt.title('Actual vs Predicted Values')

plt.xlabel('Actual Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()
```

Code for Decision Trees:

```
model = DecisionTreeRegressor(random_state=42)

model.fit(x_train, y_train)

y_pred = model.predict(x_test)

mae=mean_absolute_error(y_test, y_pred)

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

rmse = np.sqrt(mse)

print("Mean Absolute Error (RMSE):", mae)

print("Mean Squared Error (MSE):", mse)

print("Root Mean Squared Error (RMSE):", rmse)

print("R-squared (R2) Score:", r2)

plt.figure(figsize=(10, 6))

plt.scatter(y_test, y_pred, color='blue', label='Actual vs Predicted')

plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)],

color='red', linestyle='--', label='Perfect Prediction')

plt.title('Actual vs Predicted Values')

plt.xlabel('Actual Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()
```

Code for Random Forest:

```
model = RandomForestRegressor(n_estimators=100, random_state=42)

model.fit(x_train, y_train)

y_pred = model.predict(x_test)

mae=mean_absolute_error(y_test, y_pred)

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

rmse = np.sqrt(mse)

print("Mean Absolute Error (RMSE):", mae)

print("Mean Squared Error (MSE):", mse)

print("Root Mean Squared Error (RMSE):", rmse)

print("R-squared (R2) Score:", r2)

plt.figure(figsize=(10, 6))

plt.scatter(y_test, y_pred, color='blue', label='Actual vs Predicted')

plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)],

color='red', linestyle='--', label='Perfect Prediction')

plt.title('Actual vs Predicted Values')

plt.xlabel('Actual Values')

plt.ylabel('Predicted Values')

plt.legend()

plt.grid(True)

plt.show()
```

Code for PLS:

```
pls = PLSRegression(n_components=2)

pls.fit(x_train, y_train)

Y_pred = pls.predict(x_test)

mae=mean_absolute_error(y_test, y_pred)

mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

rmse = np.sqrt(mse)

print("Mean Absolute Error (RMSE):", mae)

print("Mean Squared Error (MSE):", mse)

print("Root Mean Squared Error (RMSE):", rmse)

print("R-squared (R2) Score:", r2)

plt.figure(figsize=(10, 6))

plt.scatter(y_test, y_pred, color='blue')

plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()],

'--', color='red', linewidth=2)

plt.xlabel('Actual')

plt.ylabel('Predicted')

plt.title('Actual vs Predicted')

plt.grid(True)

plt.show()
```

Code for integrating Triad spectral sensor with Arduino:

```
#include "SparkFun_AS7265X.h"

AS7265X sensor;

void setup() {

  Serial.begin(9600);

  Serial.println("AS7265x Spectral Triad Example");

 if(sensor.begin() == false)

  {

    Serial.println("Sensor does not appear to be connected. Please

check the wiring. Freezing...");

    while(1);

  }

  Serial.println("A,B,C,D,E,F,G,H,I,J,K,L,R,S,T,U,V,W");

}

void loop() {

  sensor.takeMeasurements();

  Serial.print(sensor.getCalibratedA());

  Serial.print(",");

  Serial.print(sensor.getCalibratedB());

  Serial.print(",");

  Serial.print(sensor.getCalibratedC());

  Serial.print(",");

  Serial.print(sensor.getCalibratedD());
```

```
Serial.print(",");
Serial.print(sensor.getCalibratedE());
Serial.print(",");
Serial.print(sensor.getCalibratedF());
Serial.print(",");
Serial.print(sensor.getCalibratedG());
Serial.print(",");
Serial.print(sensor.getCalibratedH());
Serial.print(",");
Serial.print(sensor.getCalibratedI());
Serial.print(",");
Serial.print(sensor.getCalibratedJ());
Serial.print(",");
Serial.print(sensor.getCalibratedK());
Serial.print(",");
Serial.print(sensor.getCalibratedL());
Serial.print(",");
Serial.print(sensor.getCalibratedR());
Serial.print(",");
Serial.print(sensor.getCalibratedS());
Serial.print(",");
Serial.print(sensor.getCalibratedT());
```

```
    Serial.print(",");
    Serial.print(sensor.getCalibratedU());
    Serial.print(",");
    Serial.print(sensor.getCalibratedV());
    Serial.print(",");
    Serial.print(sensor.getCalibratedW());
    Serial.print(",");
    Serial.println();
}
```

REFERENCES

[1] https://www.nature.com/articles/s41598-024-56319-4

[2]https://www.researchgate.net/publication/358276841_Jaundice_prediction_using_Machine_Learning_approach

[3] https://www.mdpi.com/2673-7426/4/1/34

[4] https://www.sciencedirect.com/science/article/pii/S1746809419300199

[5] https://karger.com/neo/article-abstract/doi/10.1159/000535970/906695/Machine-Learning-to-Improve-Accuracy-of?redirectedFrom=fulltext

[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3684480/#:~:text=The%20Capillary%20method%20showed%20a,an%20average%200.57%20mg%2Fdl.

[7] https://www.mountsinai.org/health-library/tests/bilirubin-blood-test#:~:text=Bilirubin%20is%20left%20in%20the,%2C%20mucus%20membranes%2C%20or%20eyes.

[8] https://www.nature.com/articles/pr1976138.pdf

[9] https://omlc.org/classroom/ece532/class3/bilirubin.html

[10] https://www.science.org/doi/10.1126/science.3029864

[11] https://www.sciencedirect.com/science/article/abs/pii/S0095510818305724

[12] https://openresearch-repository.anu.edu.au/bitstream/1885/15580/2/01_Samoc_Dispersion_of_refractive_2003.pdf

[13] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2371535/

[14] https://journals.aps.org/pr/abstract/10.1103/PhysRev.81.821

[15] https://pubs.acs.org/doi/10.1021/ja00209a009

[16] https://www.jstage.jst.go.jp/article/tjem1920/88/4/88_4_353/_article/-char/ja/

[17] https://www.sciencedirect.com/science/article/pii/S0021925818324116

[18] https://onlinelibrary.wiley.com/doi/abs/10.1002/ijch.198300026

[19] https://www.sciencedirect.com/science/article/abs/pii/S0957583905001442

[20] https://mednexus.org/doi/full/10.3760/cma.j.issn.0366-6999.2001.04.102

[21] https://www.marchofdimes.org/find-support/topics/planning-baby/newborn-jaundice#:~:text=It's%20caused%20by%20the%20build,and%20goes%20away%20without%20treatment.

[22] https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(18)30139-1/abstract

[23] https://www.sciencedirect.com/science/article/pii/S003442570700363X

[24] https://www.sciencedirect.com/science/article/abs/pii/0370157376900053

[25] https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119373421.ch7

[26] https://pubs.aip.org/aip/jcp/article-abstract/106/11/4389/181964/Far-infrared-absorption-spectra-of-water-ammonia

[27] https://www.sciencedirect.com/science/article/pii/S0300908416301092

[28] https://pubmed.ncbi.nlm.nih.gov/1431594/

[29] https://www.sciencedirect.com/science/article/pii/S1386142501006850

[30] https://pubs.acs.org/doi/pdf/10.1021/ja00840a057

[a] https://www.hematology.org/education/patients/blood-basics#:~:text=Blood%20is%20a%20specialized%20body,to%20the%20lungs%20and%20tissues

[b] https://www.hematology.org/education/patients/blood-basics#:~:text=Blood%20is%20a%20specialized%20body,to%20prevent%20excess%20blood%20loss

[c] https://columbiasurgery.org/liver/liver-and-its-functions#:~:text=The%20liver%20filters%20all%20of,fats%20and%20carry%20away%20waste.

[d] https://www.mayoclinic.org/diseases-conditions/liver-problems/symptoms-causes/syc-20374502

[e] https://my.clevelandclinic.org/health/diagnostics/17845-bilirubin

[f] https://www.mountsinai.org/health-library/diseases-conditions/bilirubin-encephalopathy#:~:text=Bilirubin%20encephalopathy%20is%20a%20rare,irreversible%20damage%20to%20the%20brain.

[g] https://medlineplus.gov/lab-tests/bilirubin-blood-test/#:~:text=A%20health%20care%20professional%20will,takes%20less%20than%20five%20minutes.