**Purpose of the Project:**

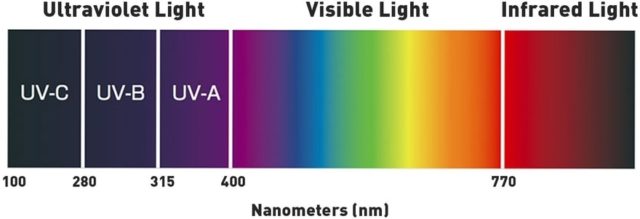
The project aims to develop a **Handheld Arduino-Based Near-Infrared (NIR) Spectrometer** for **Non-Destructive Quality Evaluation of Food**. The spectrometer operates by measuring the **Visible to Near-Infrared Reflectance** of food samples (410 nm – 940 nm), utilizing the reflectance properties of light to determine compositional traits such as sugar, moisture, and other quality parameters.

**What is Light Spectroscopy?**

Spectroscopy is a technique that uses electromagnetic radiation to analyse the structure and properties of matter.

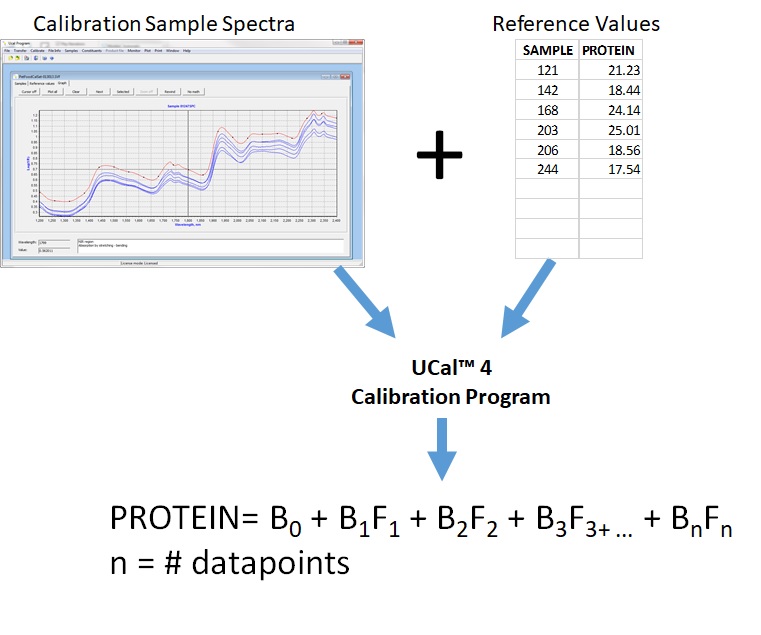
**How does it Work?**

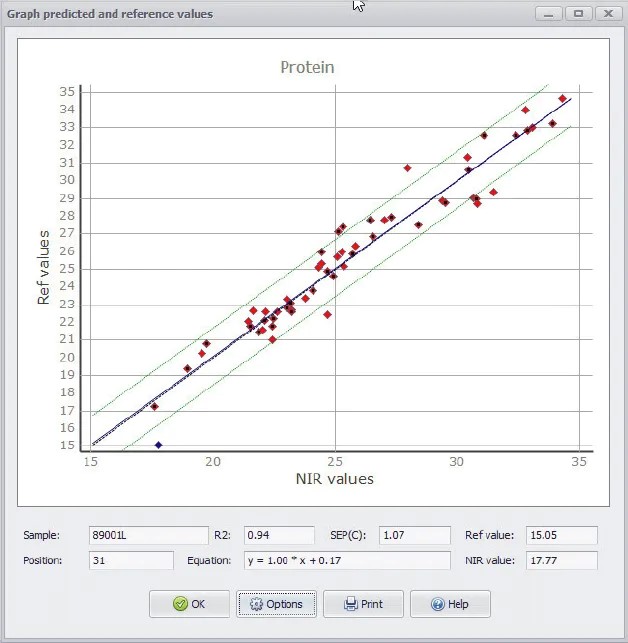
When light is presented to samples high in chemical compounds containing these bonds, some of energy is absorbed by the sample in these specific wavelengths, and thus the reflected light has less intensity in these regions.  The differences in the reflected signal (spectrum) can be correlated to chemical concentration differences, and this forms the basis of an calibration.  Once this calibration is established, it can be used to predict the chemical concentration of unknown samples.



**How is Calibration Made?**

* Gather samples that represent future unknown samples, considering factors like constituent range, origin, and seasonal variation.
* Ensure sufficient sample variety for full-spectrum instruments (50-100+ samples).
* Selected samples are sent for reference analysis using approved methods
* The reference data is combined with the raw sample spectra. These data sets are regressed against each other, commonly using Partial Least Squares (PLS) regression, Back Propagation Neural Networks (BPNN), Multiple Linear Regression (MLR) etc.
* The output of the regression is a linear equation that can predict the constituents of interest in future unknown samples.
* Apply calibration to unknown samples for analysis.
* Use outlier indicators to ensure sample similarity for accuracy.





**Why and where do we need to do this?**

NIR has become a popular and widespread analytical technique for the analysis of food, agricultural, pharmaceutical and chemical products.  NIR analysers have the following benefits

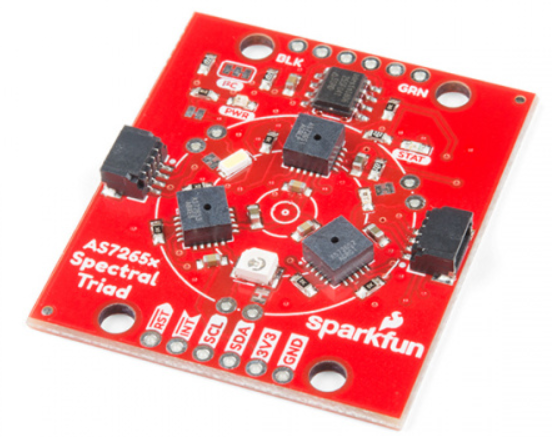
* Easy to use - normal operation consists of loading a sample cell and starting the instrument.
* Little sample preparation - most samples can be analysed as-is or with simple grinding or particle size reduction.
* No hazardous chemical waste - no chemicals are used at all.
* Fast analysis - typical analysis times are 10 seconds - 2 minutes.
* Simultaneous results for multiple parameters - multiple constituents are predicted with one sample analysis.
* Cost effective - one analyst can typically analyse several hundred samples in a day with no consumable costs.

**MOTIVATION**

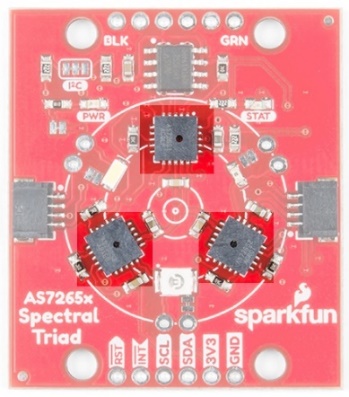
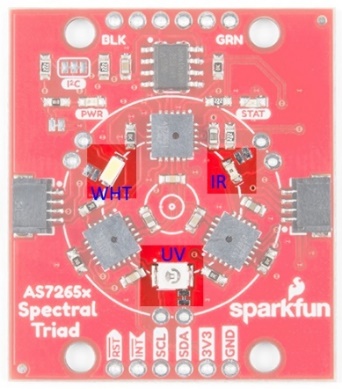
The motivation behind this project stems from the growing need for **non-destructive, real-time food quality evaluation**. Traditional methods of assessing food quality, like chemical tests, are often destructive, time-consuming, and expensive. By developing a **handheld Arduino-based NIR spectrometer**, I aim to create a cost-effective and portable solution that leverages **visible to near-infrared reflectance** to evaluate essential food parameters such as sugar content, moisture, and adulteration. This technology can have significant applications in agriculture and the food industry, helping improve **food safety, quality control**, and **nutrient monitoring**. Ultimately, this project can contribute to advancements in **precision agriculture**, **supply chain management**, and even **consumer-level food testing** by making **hyperspectral imaging accessible** in a compact, affordable form.

**What Is NIR AS7265x?**

1. The AS7265x chipset consists of three sensor devices AS72651 with master capability, AS72652 and AS72653
2. The multispectral sensors can be used for spectral identification in a range from Visible to Near Infrared.



1. The Triad contains a 5700k white LED, a 405nm UV LED, and a 875nm IR LED mounted alongside the sensors. These LEDs were chosen to illuminate the target. If the On-board LEDs are not satisfactory, external LEDs can also be used.

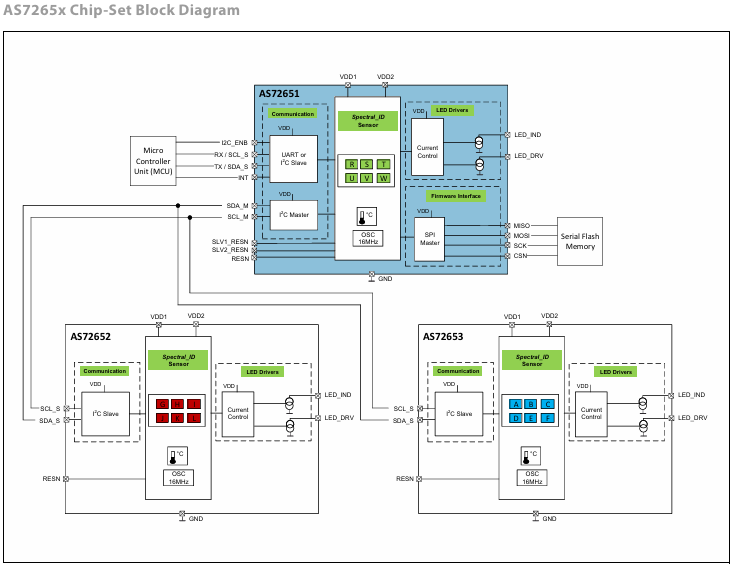
Integrated LEDs

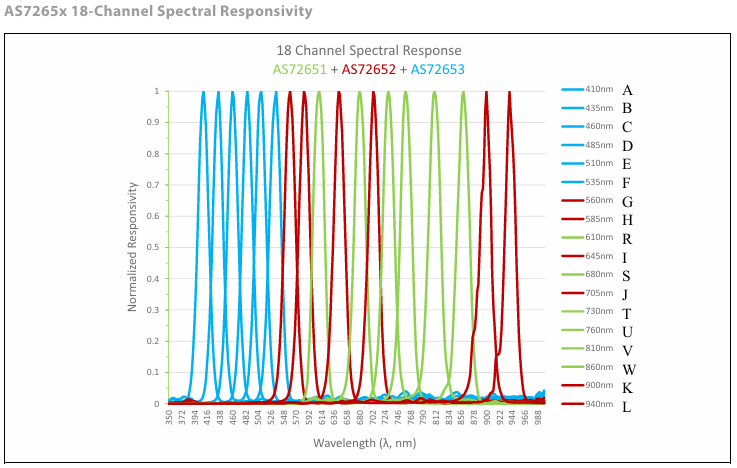
The Three Sensors

1. Every of the three sensor devices has 6 in dependent on-device optical filters whose spectral response is defined in a range from 410nm to 940nm with FWHM of 20nm
2. The components AS72651, AS72652 and AS72653 are pre-calibrated with a specific light source.

* This means that the sensors are exposed to a controlled known light source during manufacturing. The sensors are then adjusted or calibrated to ensure that the measurement they produce are accurate

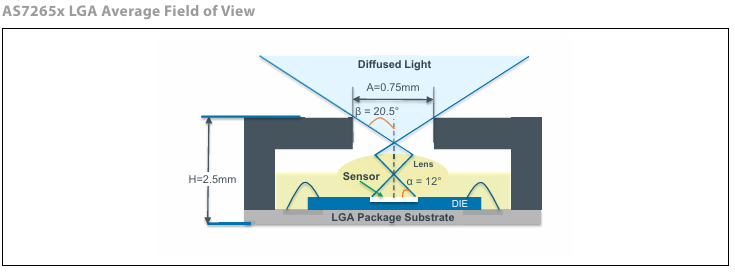
1. The AS7265x gives 18 Channel Spectral Responsivity (410nm-940nm)
2. The output of each channel is the measured light intensity in that particular spectral band (Expressed in terms of counts or raw ADC values corresponding to the amount of light detected.

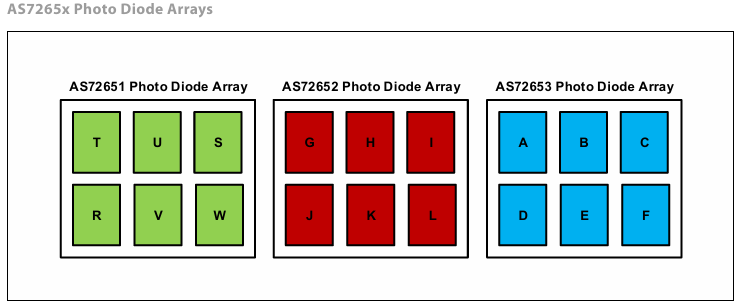




1. Optical Characteristics

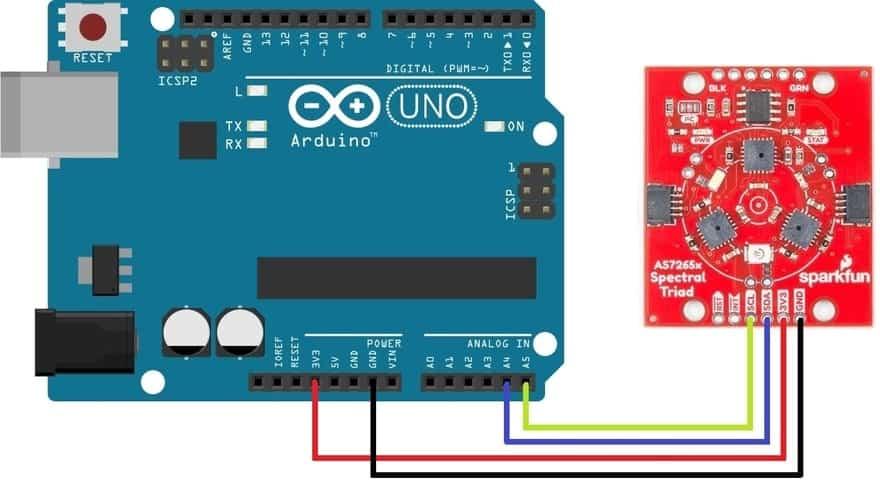
All optical characteristics are optimized for diffused light. When using a point light source or collimated light on the sensor, the sensor opening must be covered by Lambertian diffuser with achromatic characteristics.

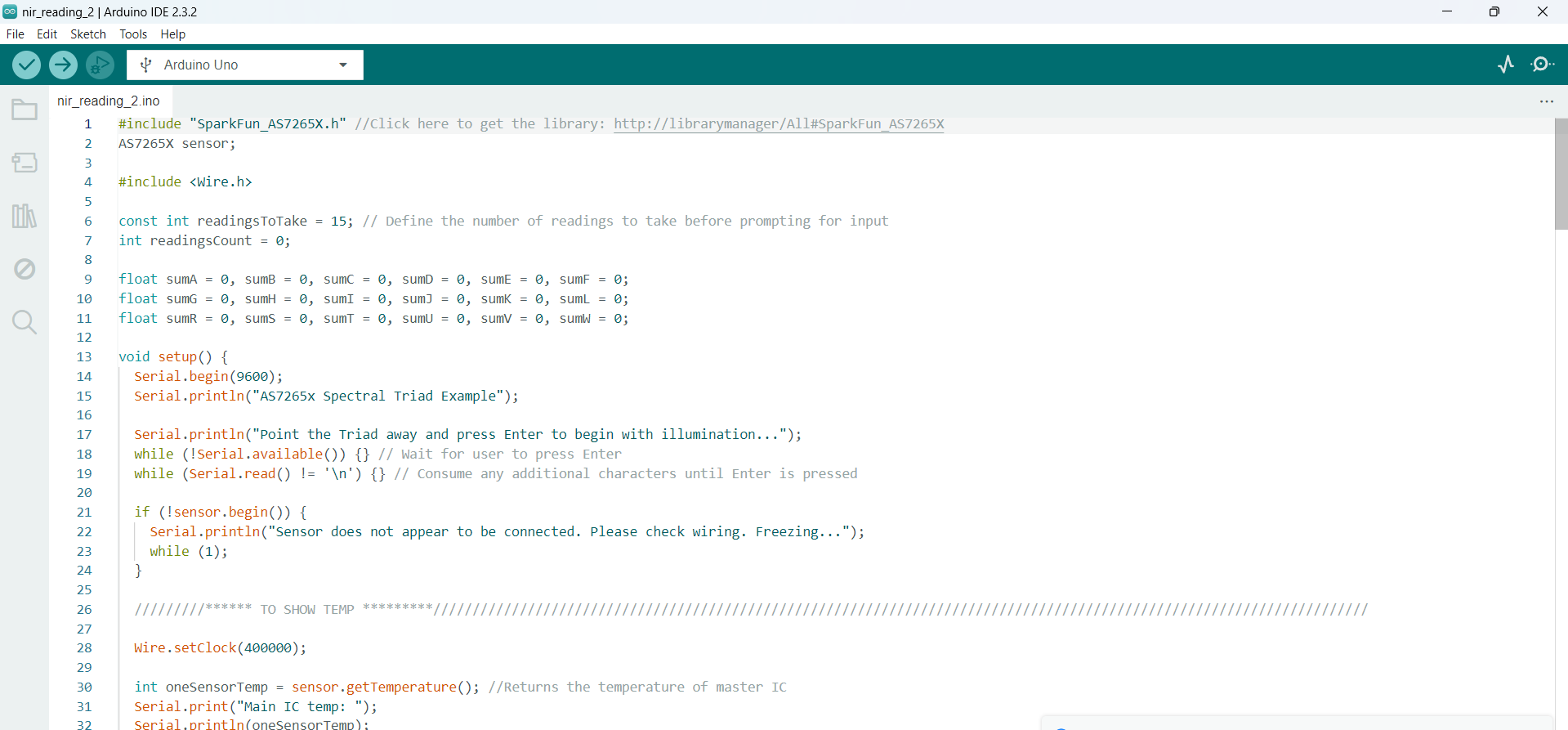




**Summary of Things to Avoid:**

1. **Do not exceed the maximum voltage, current, or temperature ratings** listed in the datasheet.
2. **Do not operate the sensor outside the recommended Electrical Characteristics**, even if temporarily.
3. **Avoid prolonged operation at the maximum rated conditions**, as it can degrade the sensor's performance and reliability.
4. **Do not expose the sensor to high-energy UV light**, especially in outdoor environments where it faces direct sunlight.
5. **Avoid exposure to extreme temperatures** (both high and low), which can damage the sensor's internal components.
6. **Ensure proper power supply** to avoid over-voltage, reverse voltage, or excessive current that could burn out components.
7. **Shield the sensor from harsh environmental conditions** like humidity, dust, and vibrations to maintain its longevity.
8. The accuracy of the AS7265x sensor’s optical filters depends on the angle of light entering the sensor. To ensure accurate spectral readings, the sensor’s field of view is limited to ±20.5°.



**TIMELINE:**

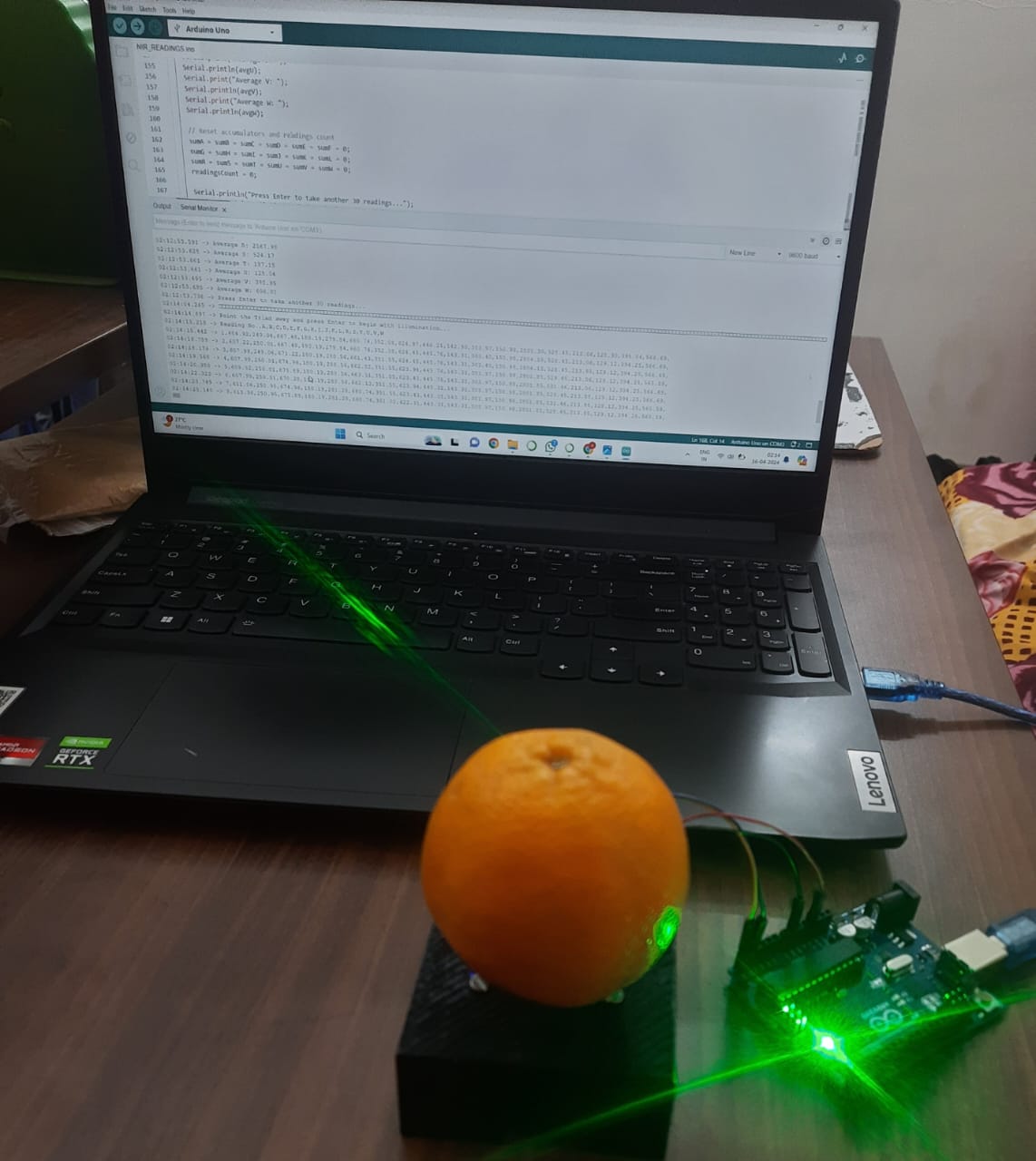
MID EVALUATION

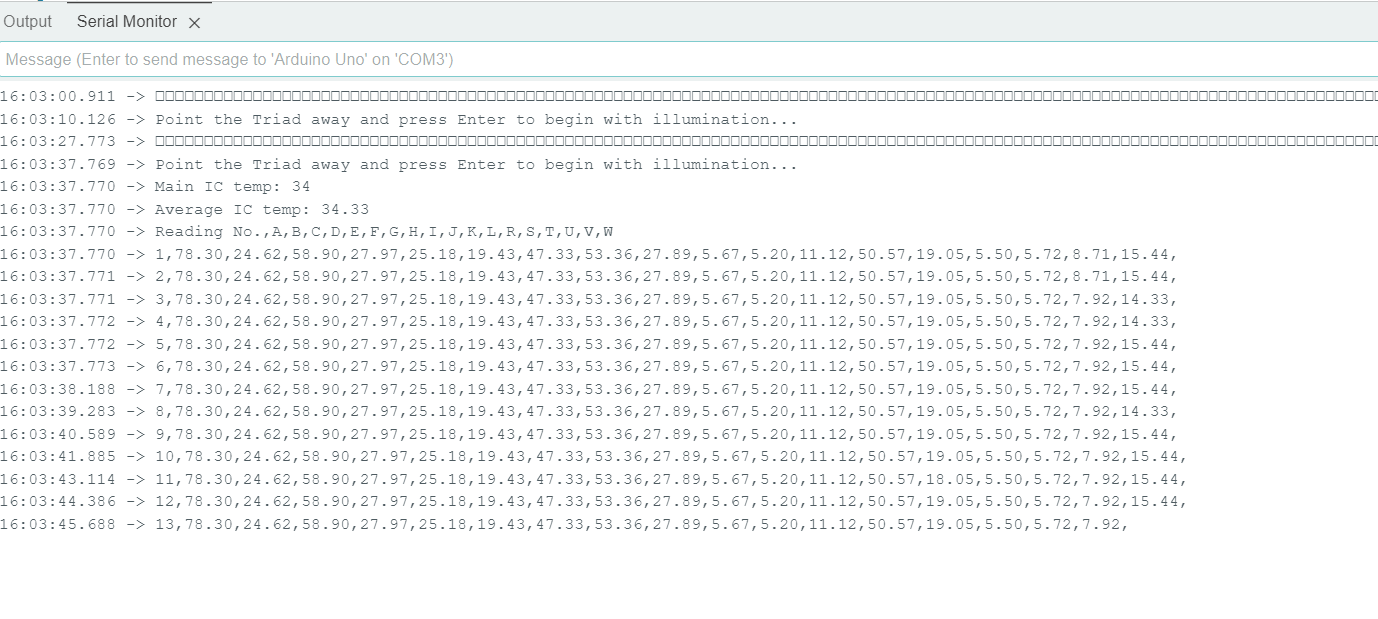
Glucose Sampling:

* Prepared 40 glucose concentration samples ranging from 0 mg/ml to 600 mg/ml.
* Incremented concentration by 15 mg/ml for each sample.
* Created stock solution of 600 mg/ml by dissolving 42 mg of dextrose in 70 ml distilled water.
* Utilized micropipette for sample accuracy.
* Each sample volume: 3 ml.
* The AS7265x chipset readings were calibrated with the already known glucose concentrations.

Sugar Content Estimation in Oranges and Tomatoes:

* The AS7265x chipset was used to capture the reflectance of oranges/tomatoes.
* The sensor was activated using an Arduino-based program, and the LED emitted bright light to capture reflectance periodically.
* The reflectance signals were displayed as outputs on the computer.
* A Brix refractometer measured the sugar content in the oranges/tomatoes, providing Brix values.
* These Brix values were used as reference data for training and testing various machine learning algorithms.





Machine Learning Algorithms used:

Back Propagation Neural Networks (BPNN):

**Neural Network Architecture:**

* **Input Layer**: 18 neurons representing the spectrometer output.
* **Output Layer**: 1 neuron representing sugar concentration/Brix.
* **Hidden Layer**: The number of nodes is determined by the formula
* is the number of input neurons
* is the number of output neurons
* is the number of training samples
* **Training and Optimization:**
* **Loss Function**: Mean Absolute Error (MAE).
* **Optimizer**: Adam.
* **Learning Rate**: 0.1.
* **Epochs**: Determined based on suitable convergence during training.

**Data Preparation and Splitting:**

* **Data Split**: 80% for training and 20% for testing.
* **Cross-Validation**: Applied to further divide data for robust evaluation.

**Activation Functions:**

* **Activation Functions**: Experimented with different functions like ReLU and Linear to introduce non-linearity and enable learning of complex patterns.

**Performance Evaluation:**

* **Error Analysis**: Mean Absolute Percentage Error (MAPE) was used to assess the model's performance.

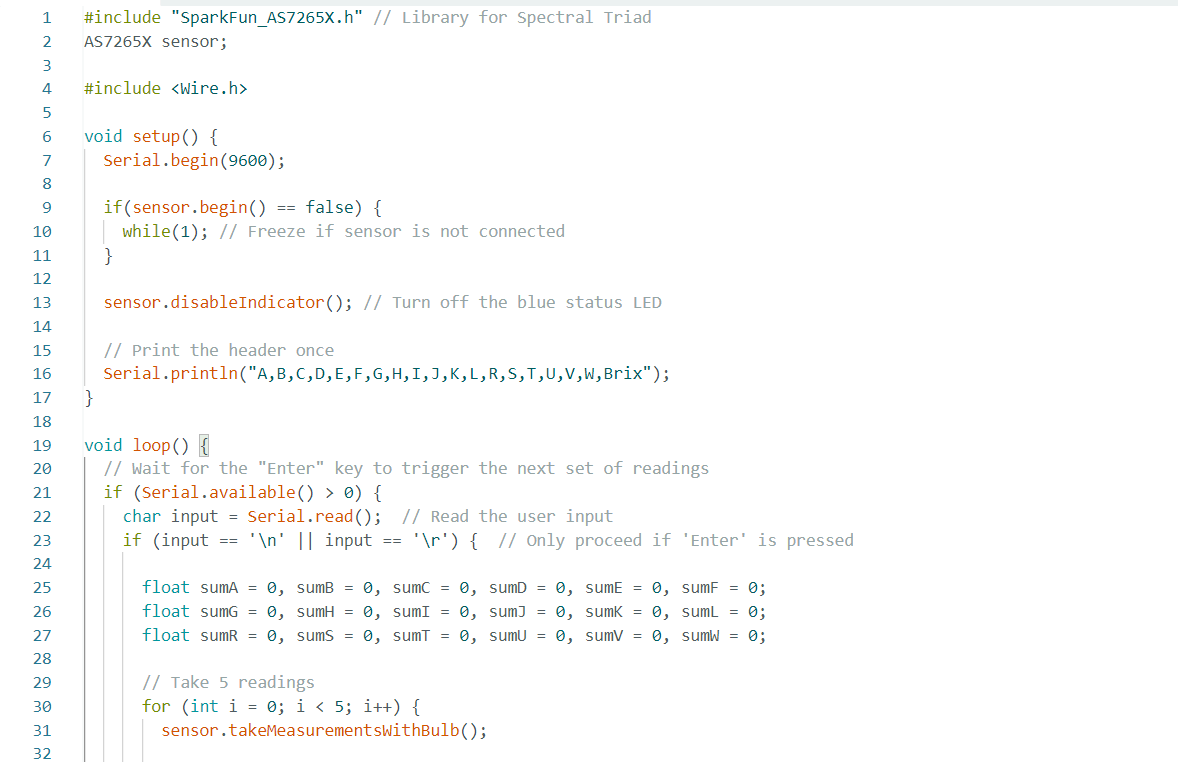
Multiple Linear Regression (MLR):

1. **Model Architecture**:
   * **Input**: Spectrometer data as features.
   * **Output**: Predicted sugar concentrations (Brix).
   * **Coefficients**: Estimated to minimize the error between predicted and actual values, capturing linear relationships.
2. **Training and Optimization**:
   * **Loss Function**: Mean Squared Error (MSE).
   * **Optimization**: Ordinary Least Squares (OLS) method to fit the regression model.
   * **Data Split**: 80% for training, 20% for testing, to ensure a robust model evaluation.
3. **Data Preparation and Splitting**:
   * **Preparation**: Input data from spectrometer, target variable as sugar concentrations.
   * **Data Split**: Training/testing data split for model validation, with the majority used for training.
4. **Model Evaluation**:
   * **Evaluation Metrics**: R-squared for variance explanation, and train/test scores to ensure generalization.
   * **Cross-Validation**: Applied for robust performance checking.
5. **Performance Evaluation**:
   * **Error Analysis**: Mean Absolute Percentage Error (MAPE) to assess the model’s predictive accuracy in estimating sugar concentration (Brix).

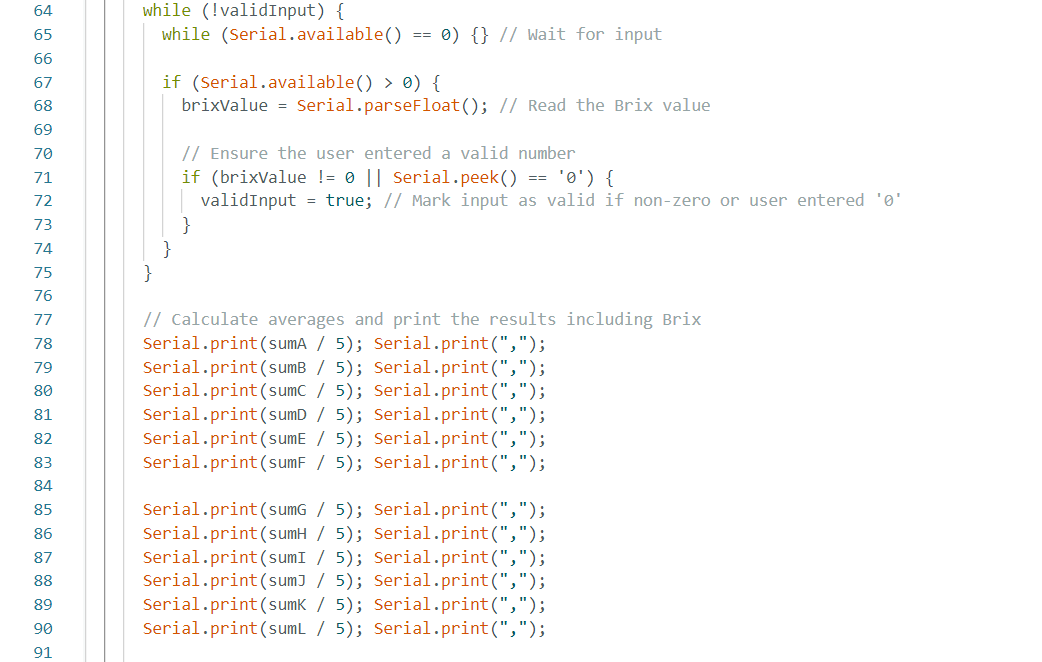
Partial Least Squares Regression (PLSR)**:**

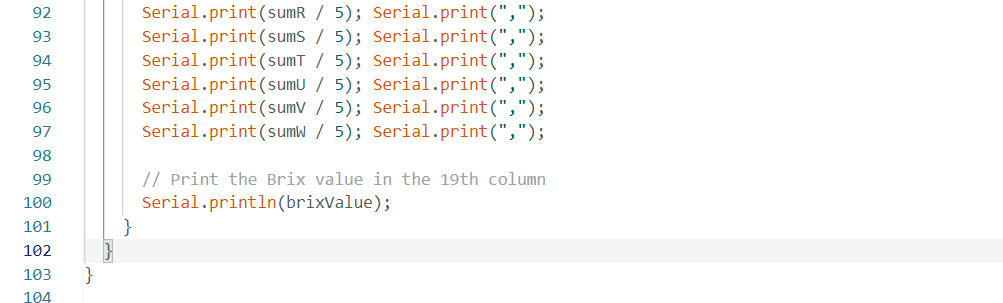
1. **Introduction to PLSR**:
   * **Purpose**: Ideal for high-dimensional or multicollinear data.
   * **Latent Variables**: Determines and uses latent variables to reduce dimensionality while balancing model complexity and performance.
2. **Maximizing Covariance**:
   * **Objective**: Maximizes covariance between predictors and target variables.
   * **Performance Goals**: Minimizes Mean Absolute Error (MAE) and maximizes Residual Prediction Deviation (RPD).
3. **Implementation Tools**:
   * **Tools**: Implemented using Python's scikit-learn library or R's pls package.
   * **Functions**: These tools facilitate model fitting, specification of latent variables, training, and performance assessment.
4. **Model Evaluation**:
   * **Train-Test Split**: 80% for training and 20% for testing.
   * **Performance Metric**: Evaluated using Mean Absolute Percentage Error (MAPE) for accuracy.
5. **Enhancing Predictive Modelling Efficiency**:
   * **PLSR Advantage**: Captures data variability efficiently while minimizing prediction errors.
   * **Outcome**: Improves predictive accuracy and efficiency of the model.

**Arduino Code Used:**

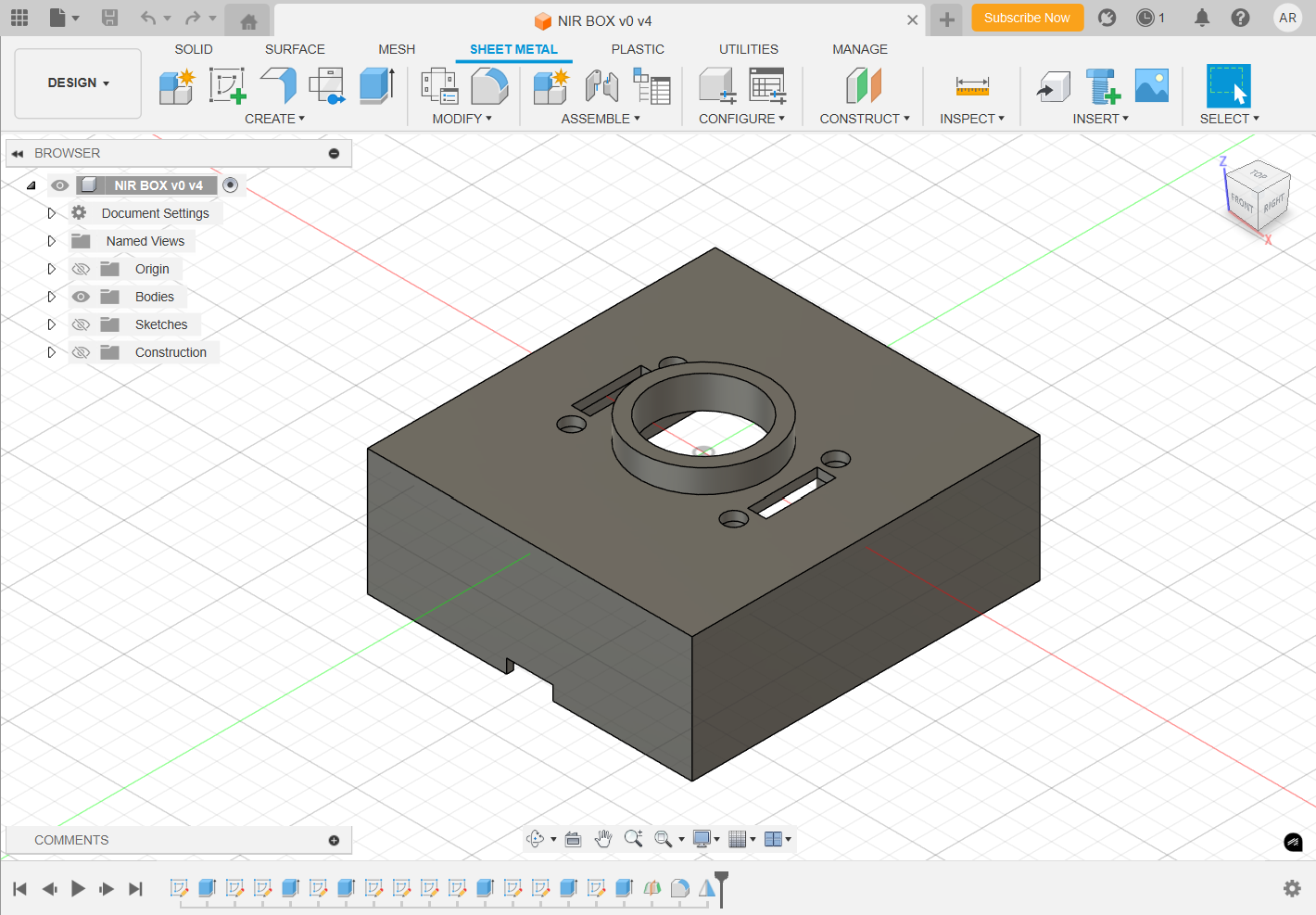
****

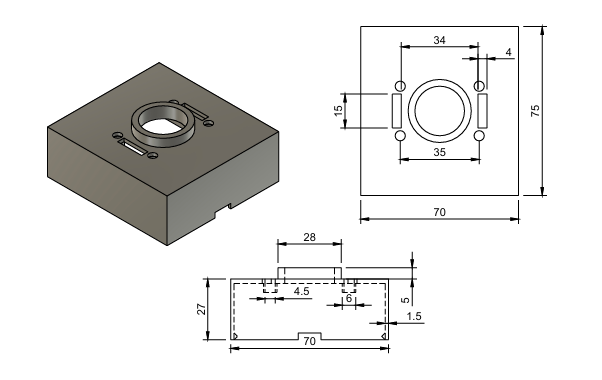
****





**NIR BOX 3D DIAGRAM:**

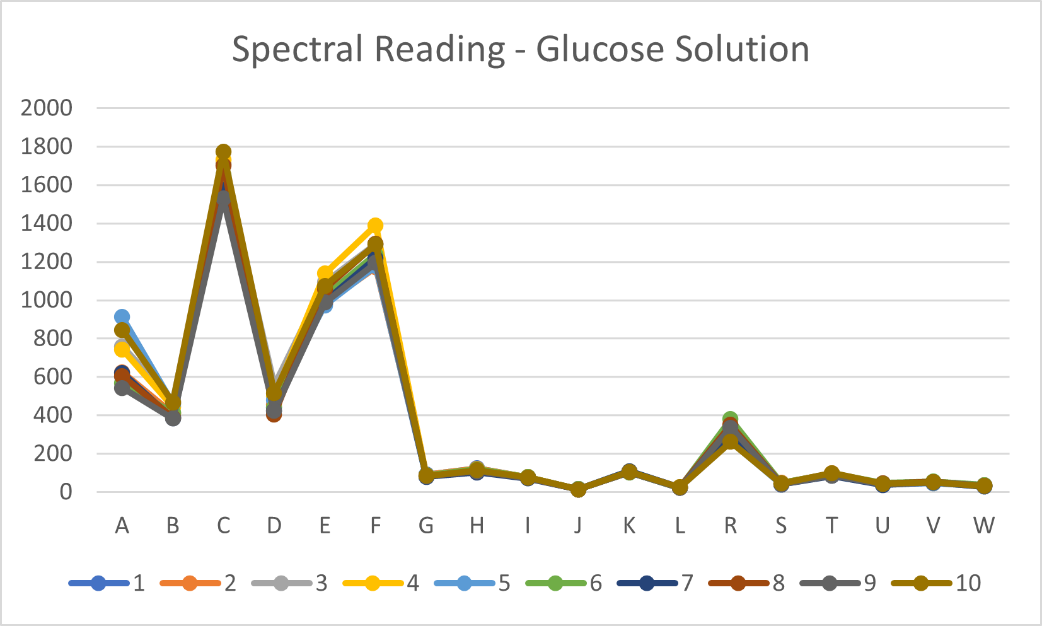


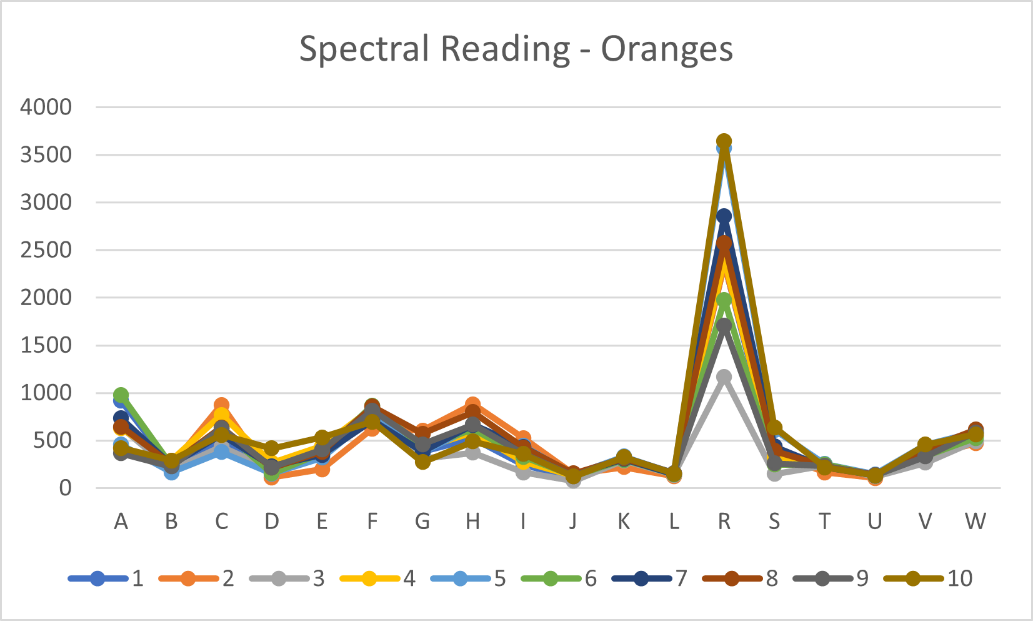


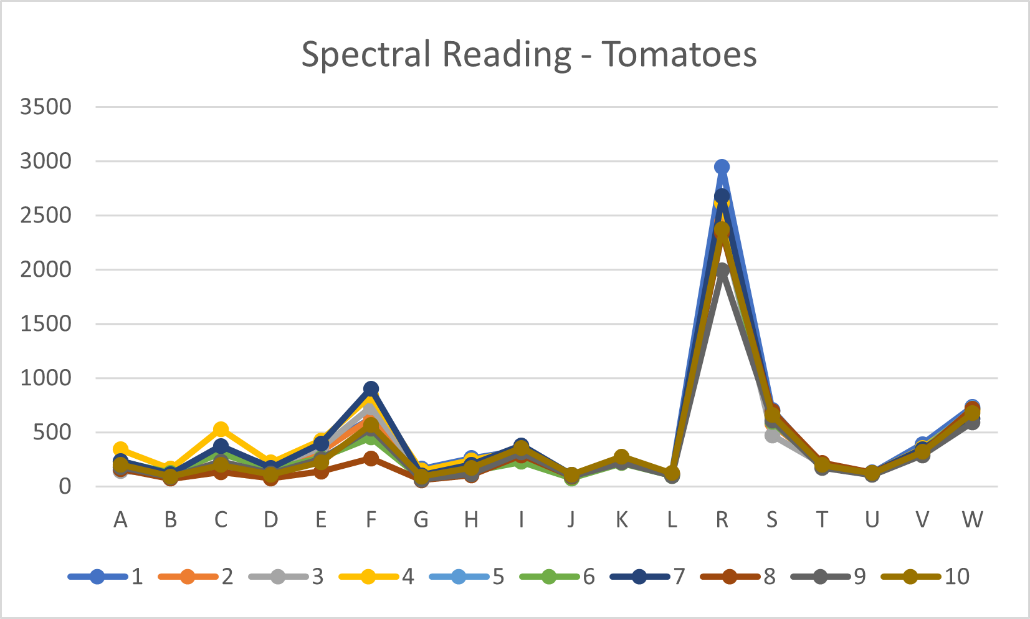
Note: All dimensions are in mm.

**RESULTS SO FAR:**

**The spectral reading graphs for Glucose, Oranges and Tomatoes.**





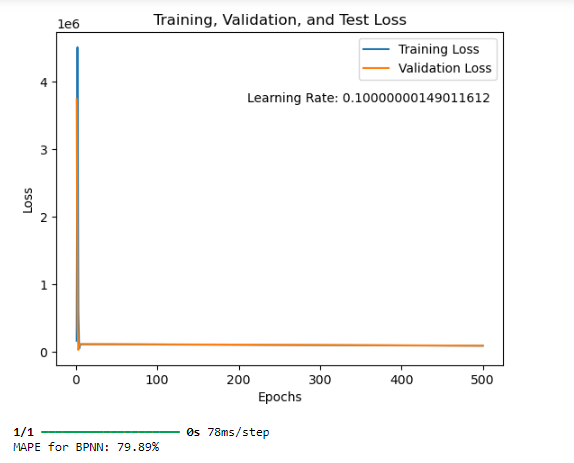


A Few Important Deductions:

* **Inherent Spectral Patterns:**  
  Each item (glucose, oranges, and tomatoes) has its own unique spectral graph, which represents how that specific substance interacts with light across 18 different wavelengths.
* **Consistent Pattern Across 10 Readings:**  
  For each item, 10 separate readings have been plotted, and all show a consistent pattern within each item, indicating reproducibility and reliable reflection behaviours for each specific wavelength.
* **Oranges and Tomatoes Show Similar Reflection Behaviour:**  
  The spectral graphs for oranges & tomatoes are very similar ye unique in shape, particularly in the positioning of their main peak at Channel R.
* **Application for Food Adulteration Detection:**  
  This data is useful as it confirms that the sensor can potentially detect **adulteration** in food items by identifying deviations in their spectral patterns.
* **Adulteration in Milk Products (700-940 nm):** NIR spectroscopy can detect adulterants like water, starch, and synthetic chemicals in milk, identifying diluted or mixed products.
* **Detection of Toxic Dyes in Spices (410-700 nm):** The visible range helps identify toxic dyes, such as lead chromate, in spices by comparing absorption patterns of pure and adulterated samples.
* **Adulteration in Edible Oils (700-940 nm):** NIR spectroscopy detects low-quality oils mixed with high-quality ones by analysing their distinct fatty acid absorption at various wavelengths.
* **Watered-Down Fruit Juices (700-940 nm):** NIR measurements can identify dilution in fruit juices by comparing water and sugar content to natural juice compositions.
* **Adulteration in Honey (410-940 nm):** Both visible and NIR spectra can reveal sugar syrups or artificial sweeteners in honey through specific absorption pattern analysis.
* **Detection of Harmful Chemicals in Food (410-700 nm):** The visible range detects harmful chemicals, like synthetic waxes or pesticides on fruits, by analysing their reflective properties compared to untreated produce.

**5) Data trained on different Models -**

1.Input: Used 40 NIR sensor readings.

2. Output: Predicted glucose concentrations (0-600 mg/ml).

**1)BPNN**

1. Model: Single hidden layer BPNN with 16 ReLU neurons.

2. Results: Training MAE:254.49 Testing MAE:294.57

3. Evaluation: Assessed error using MAPE in predicting glucose from NIR data which is 79.8%

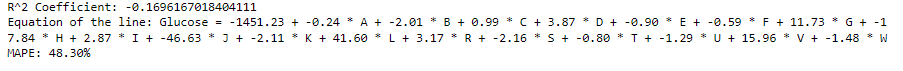
**2. Multiple Linear Regression:**

1. Model: Linear Regression Model

2. Results: Train Score:0.9295, Test Score:-0.1696 R-squared: -0.16961

3. Evaluation: Assessed error using MAPE in predicting glucose from NIR data which is 48%

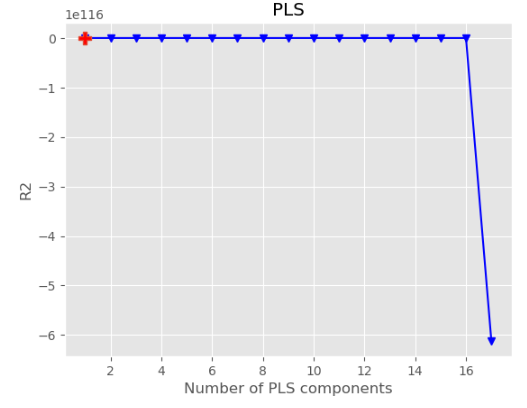
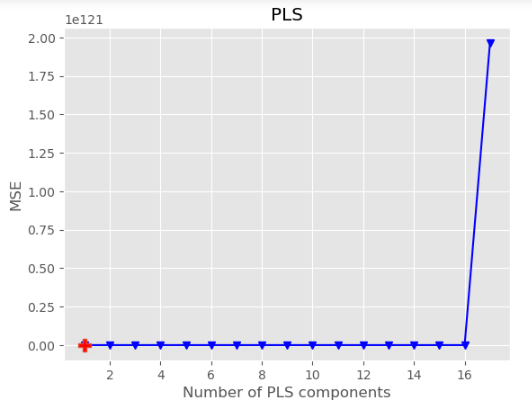




**3. Partial Linear Square Model:**

1. Linear PLSR with 1 latent variable

2. Results: MSE: 34976.73 R-squared: -0.0896 RPD: 0.9580

3. Evaluation: Assessed error using MAPE in predicting glucose from NIR data which is 88.1%



To view the ML code - <https://drive.google.com/file/d/1fjG4QTQF35NIIEKgCsiN74KBBtXZCn2I/view?usp=sharing>

**9) Final Dataset trained on ML -**

1. Input: Used 30 NIR sensor readings.

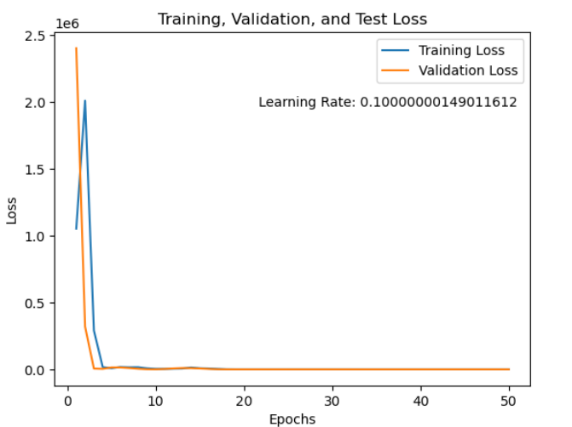
2. Output: 30 BRIX values collected from the refractometer of the random 30 orange sample

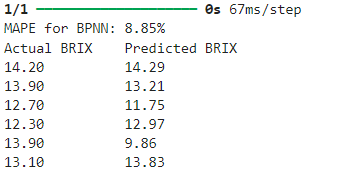
**1)BPNN**

1. Model: Single hidden layer BPNN with 24 ReLU neurons and 0.2 Dropout

2. Results: Training MAE:1.68 Testing MAE: 1.19

3. Evaluation: Assessed error using MAPE in predicting BRIX from NIR data which is 8.85%



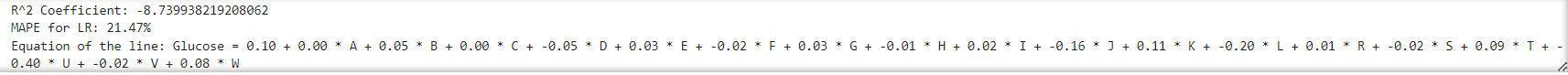


**2. Multiple Linear Regression:**

1. Model: Linear Regression Model

2. Results: Train Score: 0.91261, Test Score: -8.7399 R-squared: -8.37399

3. Evaluation: Assessed error using MAPE in predicting BRIX from NIR data which is 21.42%

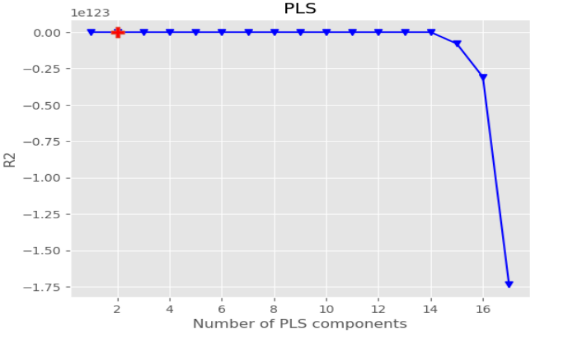
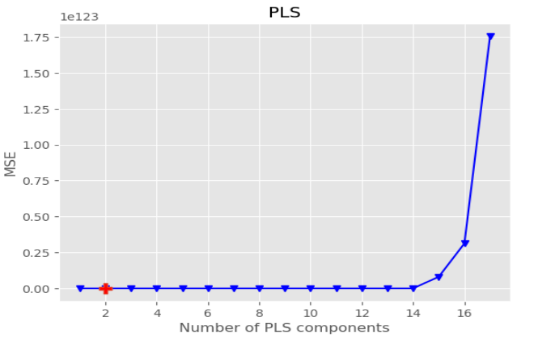


**3. Partial Linear Square Model:**

1. Linear PLSR with 2 latent variables

2. Results: MSE: 1.326 R-squared: -0.3116 RPD: 0.8732

3. Evaluation: Assessed error using MAPE in predicting BRIX from NIR data which is 6.21%

****

**FUTURE PROSPECTS:**

* **Adulteration Detection in Solids and Liquids:** Future research could focus on identifying inorganic and organic contaminants in food items, ensuring food safety through non-invasive methods.
* **Detection of Sugar in Clear Solutions and Beverages:** Expanding sugar detection capabilities to carbonated drinks and juices can improve quality control and adulteration detection
* **Sugar Content in Fruits and Vegetables for Condition Monitoring:** Monitoring sugar levels in produce can aid in determining ripeness and freshness, enhancing supply chain management
* **Nitrogen and Water Content in Leaves:** Adapting the sensor for leaf analysis can help monitor nitrogen and water levels, optimizing fertilizer use and irrigation in agriculture.
* **Pigments and Colorants (Visible Range ~410-700 nm):** The sensor can distinguish materials based on their colour by detecting natural and artificial pigments in the visible spectrum.
* **Haemoglobin and Blood Oxygenation (Visible Range ~500-600 nm):** Haemoglobin absorption peaks allow detection of blood oxygenation levels or the presence of blood.
* **Chlorophyll in Plants (Visible Range ~410-700 nm):** Chlorophyll absorbs blue and red light while reflecting green, making plants appear green, and serves as an indicator of plant health and photosynthetic activity.
* **Water Content (Near-Infrared Range ~700-940 nm):** Water absorbs light strongly in the NIR range, especially around 900 nm, which helps detect water content in various biological materials.
* **Protein Content in Lentils and Other Pulses:** The sensor could measure protein content in pulses, aiding in the nutritional assessment of these staple foods.
* **Organic Compounds (Visible to Near-Infrared Range ~410-940 nm):** Distinctive absorption features of organic compounds in this range enable measurement of nutritional content and adulteration detection.
* **Food Spoilage Detection (Visible to Near-Infrared Range ~410-940 nm):** The sensor can detect spoilage in food products by identifying changes in chemical composition and absorption patterns as they degrade over time.

Resources

1. [AS7265x\_Datasheet](https://cdn.sparkfun.com/assets/c/2/9/0/a/AS7265x_Datasheet.pdf)
2. [Spectral Triad (AS7265x) Hookup Guide](https://learn.sparkfun.com/tutorials/spectral-triad-as7265x-hookup-guide)
3. [Handheld arduino-based near infrared spectrometer for non-destructive quality evaluation of siamese oranges](https://iopscience.iop.org/article/10.1088/1755-1315/653/1/012119/pdf)
4. [A Cost‐Effective and Portable Optical Sensor System to Estimate leaf Nitrogen and Water Contents in Crops](https://www.researchgate.net/publication/339761781_A_Cost-Effective_and_Portable_Optical_Sensor_System_to_Estimate_Leaf_Nitrogen_and_Water_Contents_in_Crops)
5. [Detection of adulterated cane sugar in granulated coconut sugar](https://www.researchgate.net/publication/367382565_Design_and_performance_test_of_portable_spectrometer_using_AS7265x_multispectral_sensor_for_detection_of_adulterated_cane_sugar_in_granulated_coconut_sugar)
6. [Low-cost IoT-based multichannel spectral acquisition systems for roasted Coffee beans](https://www.sciencedirect.com/science/article/pii/S088915752400512X?pes=vor)
7. [Glucose Monitoring Using Optical Sensor for diabetes applications](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9427517)
8. [Rapid Identification of Apple Maturity Based on Multispectral Sensor Combined with Spectral Shape Features](https://www.mdpi.com/2311-7524/8/5/361)
9. [Classification of Organic and Conventional Vegetables Using Machine Learning: A Case Study of Brinjal, Chili and Tomato](https://www.mdpi.com/2304-8158/12/6/1168)
10. [Machine-Learning-Based Spectroscopic Technique for Non-Destructive Estimation of Shelf Life and Quality of Fresh Fruits Packaged under Modified Atmospheres](https://www.mdpi.com/2071-1050/15/17/12871)
11. [Fish Farming Water Monitoring with Spectrometer Sensor AS7265x](https://youtu.be/7J0p1xn5vMk?si=JPrblSSQnNQmxNBQ)
12. [AS7265x I2C C library for Raspberry Pi / Linux](https://github.com/jdesbonnet/as7265x)
13. [SparkFun\_AS7265x\_Arduino\_Library](https://github.com/sparkfun/SparkFun_AS7265x_Arduino_Library)
14. [Qwiic\_Spectral\_Sensor\_AS7265x](https://github.com/sparkfun/Qwiic_Spectral_Sensor_AS7265x)
15. [Interfacing Triad Spectroscopy Sensor AS7265x with Arduino](https://how2electronics.com/interfacing-triad-spectroscopy-sensor-as7265x-with-arduino/)
16. [AS7265x-spectrometer-GitHub](https://github.com/LiamsGitHub/AS7265x-spectrometer)
17. [py\_Spectral\_Triad-Github](https://github.com/hwreverse/pySpectralTriad)
18. [Non-linear Correlation of Absorbance with Respect to Concentration of Sugar](https://core.ac.uk/download/pdf/249334886.pdf)