



UNIVERSITY OF
CAMBRIDGE

PRINCIPLE OF SENSING COURSEWORK

UNIVERSITY OF CAMBRIDGE

DEPARTMENT OF CHEMICAL ENGINEERING AND BIOTECHNOLOGY

SENSOR CDT

Low-cost Spectrophotometer for Juice Classification

Junjian Chi (CID: jc2592)

Contents

1 Abstract	2
2 Introduction	2
3 Physical Principle of Method	3
3.1 Beer-Lambert Absorption Law	3
3.2 AS7343 Multi-Spectral Sensor	4
4 Description and Justification of the Device Design	4
4.1 3D-Printed Mechanical Model	4
4.2 Embedded Electronic System and Sensor Choice	5
4.3 Data Collection and Pre-processing	6
4.4 Bill of Materials and Costs	9
5 Evaluation	10
5.1 Evaluation of the Device's Performance	10
5.2 Quantitative and Qualitative Assessment of the Device's Uncertainties	11
5.3 Improvements in the Next Iteration of the Device	13
6 Conclusion	13
7 Reference	14

1 Abstract

Juice concentration estimation often uses commercial spectrophotometers that measure the absorption of light transmitting through the liquid sample. Due to the expensive cost of such instruments, this motivates the design a low-cost version that could be reproduced from inexpensive components in other laboratories. In this work, we present a 3D-printed reproducible spectrophotometer system integrating an Arduino MCU with AS7343 spectral sensor.

The system captures 12-channel spectral measurements for three types of juice, with three different brands for each type. These measurements are pre-processed and used to train machine learning models for juice type and concentration classification. Based on the performance on an independent test set, the best two models achieve a test accuracy of 100% for juice type classification, and 82.35% for concentration classification. This system provides a cost-effective and extensible alternative with customized machine learning models to commercial spectrophotometers, enabling broader adoption in resource-limited environments.

2 Introduction

A spectrophotometer is an instrument that quantitatively measures the absorption or transmission of light as a function of wavelength when it passes through a sample [1]. Commercial spectrophotometers are typically expensive, limiting the accessibility for open-source use. This report presents the design and evaluation of a low-cost spectrophotometer as shown in Fig. 1. It could classify juice types and their concentrations in water. In section III, fundamental principles of method are explained and discussed. Section IV describes the mechanical and electronic design of the device, as well as the signal processing approach. Moreover, the experimental setup and evaluation are presented in section V and VI respectively. Conclusions are in section VII.

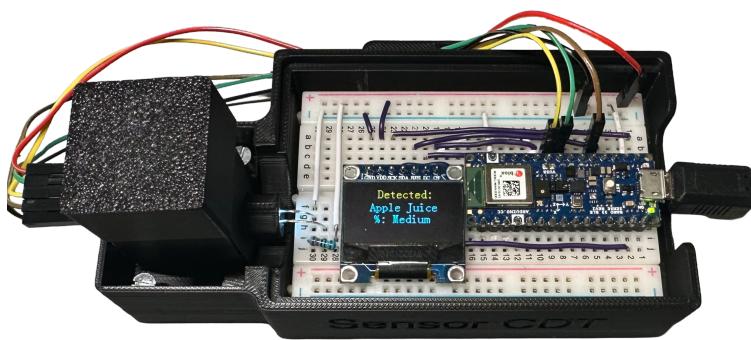


Figure 1: Low-cost 3D printed spectral sensor system

3 Physical Principle of Method

This proposed method uses a multi-spectral sensor to analyze juice samples for category classification and concentration estimation. A white LED (Light-Emitting Diode), covering the visible spectral range, illuminates a cuvette containing the sample. Light at different wavelengths is absorbed according to the molecular composition of the solution. By measuring transmitted light intensity across multiple wavelengths, the concentration and types of solution can be estimated.

3.1 Beer-Lambert Absorption Law

Based on the summarized method described above, it is necessary to dive deeper into the quantitative relationship between transmitted light intensity and sample concentration.

As shown in Fig. 2, an incident light beam I_0 enters the cuvette with the solution and propagates through the medium over an optical path of length l .

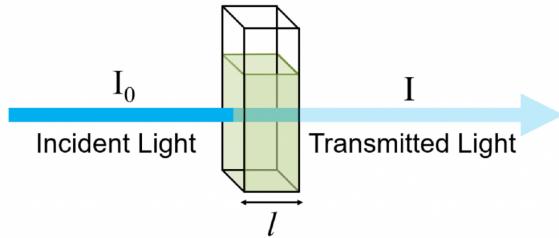


Figure 2: Transmission of light through a sample solution in a cuvette

During propagation, incident light at I_0 is partially absorbed, resulting in a reduced transmitted intensity I . To quantitatively describe this attenuation, absorbance A is defined:

$$A(\lambda) = -\log_{10}\left(\frac{I(\lambda)}{I_0(\lambda)}\right) \quad (3.1)$$

Equation 3.2 is the Beer–Lambert law, where ε , the molar absorption coefficient, is a measure of how strong an absorber the sample is at a specific wavelength. The concentration c is simply the moles $L^{-1}(M)$ of the sample dissolved in the solution, and the optical path length l is the width of the cuvette [2].

$$A(\lambda) = \varepsilon(\lambda) c l \quad (3.2)$$

As the Beer-Lambert law states that there is a linear relationship between the concentration and the absorbance of the solution, it enables the concentration of a solution to be calculated by measuring its absorbance.

3.2 AS7343 Multi-Spectral Sensor

In this system, transmitted light is measured using the AS7343 multi-spectral sensor as shown in Fig. 3. The spectral response obtained is defined by individual channels covering approximately 380 nm to 1000 nm with 11 channels centered in the visible spectrum (VIS), plus one near-infrared (NIR) and a clear channel [3]. They are converted into 16 bit integer to represent relative intensity measurements.

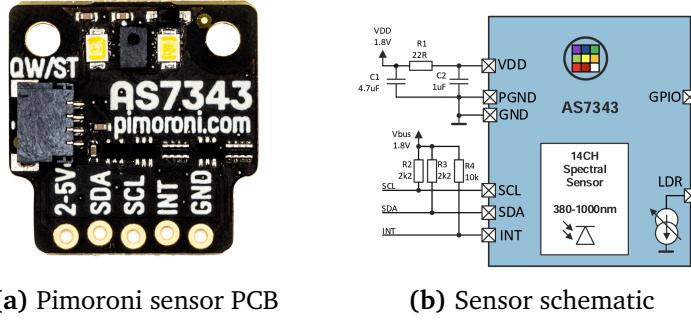


Figure 3: AS7343 spectral sensor

Blank sample like water is used as a reference measurement. Equation 4.2 can be applied to obtain the discrete absorption spectrum of the sample. As a result, changes in the overall magnitude of absorbance could be used to estimate the concentration, while differences in absorbance spectrum across wavelengths can be used to classify different juice types.

4 Description and Justification of the Device Design

4.1 3D-Printed Mechanical Model

The mechanical structure of this system is designed to ensure measurement stability and repeatability for a low-cost prototype. The overall structure ensures that light propagates linearly along a fixed optical path length within the sample and minimizes the impact of random placement errors on the measurement results.

As shown in Fig. 4 (a), the system consists of two main 3D-printed components: a cuvette holder and an open case. The cuvette holder features a top opening to allow insertion of the cuvette and is equipped with a removable lid, which is used to block ambient light and prevent liquid leakage during measurements. On the other hand, the bottom case mechanically holds the cuvette holder with four M3 screw bolts and a regular medium-sized breadboard through a snap-fit interface. Electronic components such as MCU (Micro-Controller Unit) or other custom modules can be arranged on the breadboard, providing flexible circuit configuration while facilitating heat dissipation and convenient testing.

In terms of geometric details, a circular hole with a diameter of 6 mm is designed on the right hand side of the cuvette holder to accommodate a 5 mm white

LED. On the opposite side of the cuvette holder, the AS7343 is mounted vertically using two M2 screw bolts, ensuring stability and repeatable assembly. The vertical position of the LED is designed such that the emission center is aligned with the photosensitive area of the AS7343 sensor, establishing a stable optical path. The height of the cuvette holder is approximately two-thirds of the height of the cuvette, which minimizes the risk of liquid damage to surrounding electronic components.

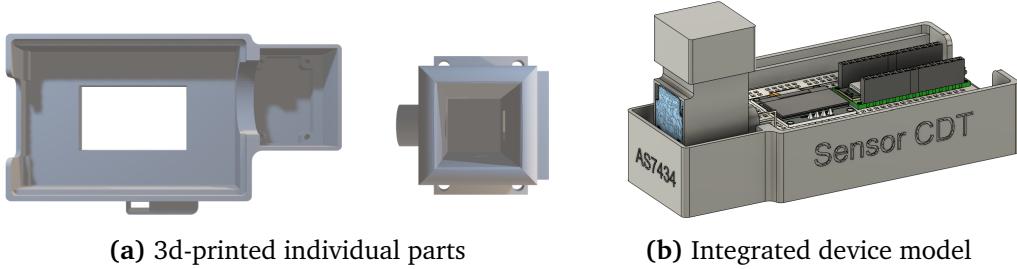


Figure 4: Device mechanical design

The overall system as shown in Fig. 4 (b) is manufactured using PLA material via 3D printing, with a nozzle diameter of 0.4 mm. This manufacturing method offers advantages such as low cost and rapid iteration. However, due to the rough layer textures on the 3D-printed material surface, this design avoids using optical components, which require high reflective surface. Instead, it achieves a linear optical path and light shielding through closed structure. This mechanical model achieves a reasonable trade-off between optical performance and manufacturing cost, making it suitable for a low-cost spectrophotometer.

4.2 Embedded Electronic System and Sensor Choice

The proposed embedded system is organized using a three-layer architecture including a hardware driver layer, an MCU application layer and a PC-side (Personal Computer) processing layer.

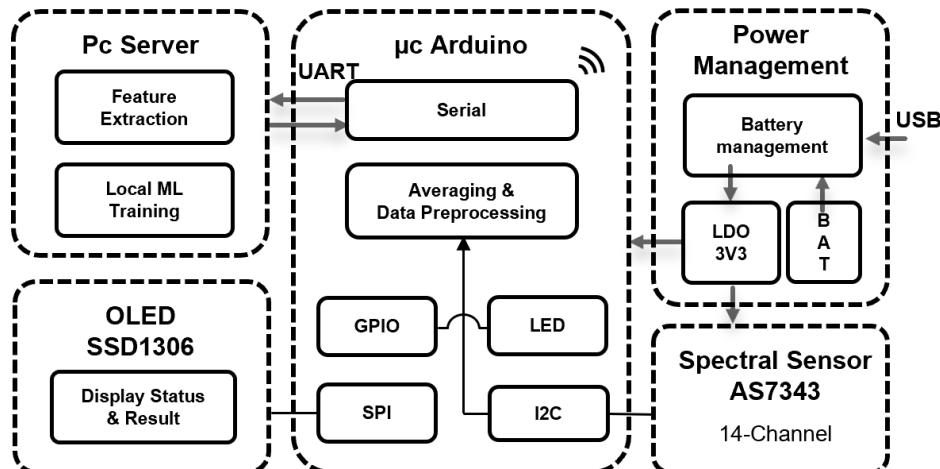


Figure 5: Block diagram of the embedded electronic system

As shown in Fig. 5, the lowest hardware layer handles direct communication between MCU and peripheral devices: 1) AS7343 is interfaced with the MCU via I2C protocol. 2) OLED SSD1306 is interfaced with the MCU via SPI interface. 3) White LED is interfaced with the MCU via GPIO.

Prior to measurement, three acquisition precision modes (Low, Medium and High) are configured by adjusting the AS7343 integration-related registers (ATIME and ASTEP) to control the integration time and balance measurement speed against signal quality. For a given spectral channel k , the relative raw sensor output can be expressed as

$$D_k \propto G \cdot T_{\text{int}} \cdot P(\lambda_k) \quad (4.1)$$

where G is the sensor gain, T_{int} is the integration time, and $P(\lambda_k)$ is the transmitted optical power.

Increasing T_{int} allows a larger amount of photocharge to be accumulated within a single measurement cycle [3], thereby improving the 16-bit ADC dynamic range and enhancing measurement resolution. Accordingly, the high-precision acquisition mode employs a long integration setting (ATIME = 0, ASTEP = 65532) to maximise resolution, at the expense of increased acquisition time.

Above the hardware interface, the embedded application layer programmed on the MCU controls the system logic. The firmware has two public modular operating modes: a data acquisition mode and an inference mode.

In the data acquisition mode, the system performs repeated spectral measurements, averaging five consecutive readings to reduce measurement noise. Then, the average sensor readings are transmitted into the PC serial port via UART protocol. In the inference mode, spectral data from an unknown sample are acquired and transmitted to the PC. Trained models are deployed to predict sample type and concentration. The computed results are then sent back to the MCU and presented to the user via the OLED display.

4.3 Data Collection and Pre-processing

The highest layer consists of the machine learning training on the PC for the collected data and the local inference of the trained model back to the Arduino display.

As shown in Fig. 6 (a), a pipette with a volume range from $10 \mu\text{L}$ to $1000 \mu\text{L}$ is used to dilute pure juice samples with water, producing three concentration levels for machine learning classification labels. The low, medium, and high concentration levels correspond to 10%, 50%, and 100% juice concentration respectively.

As shown in Fig. 6 (b), we include three juices (orange, apple and grape) with three different brands into the collected dataset, considering the generalization capability of the juice type classification model.

4.3 Data Collection and Pre-processing

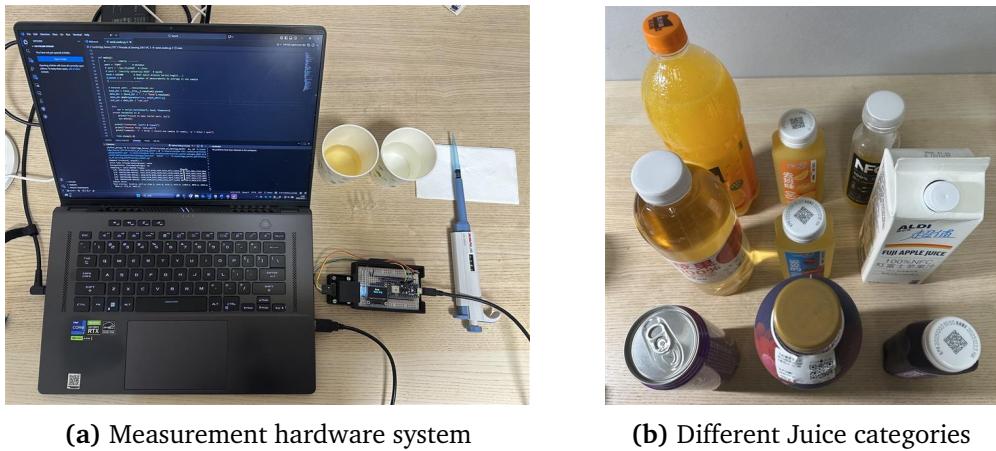


Figure 6: Experimental setup for spectral data acquisition

After the raw spectral data and corresponding labels are collected and written into the .csv file, pre-processing techniques are applied to them before being fed into the classification models. The L1 normalization method is applied to mitigate overall intensity scale variations:

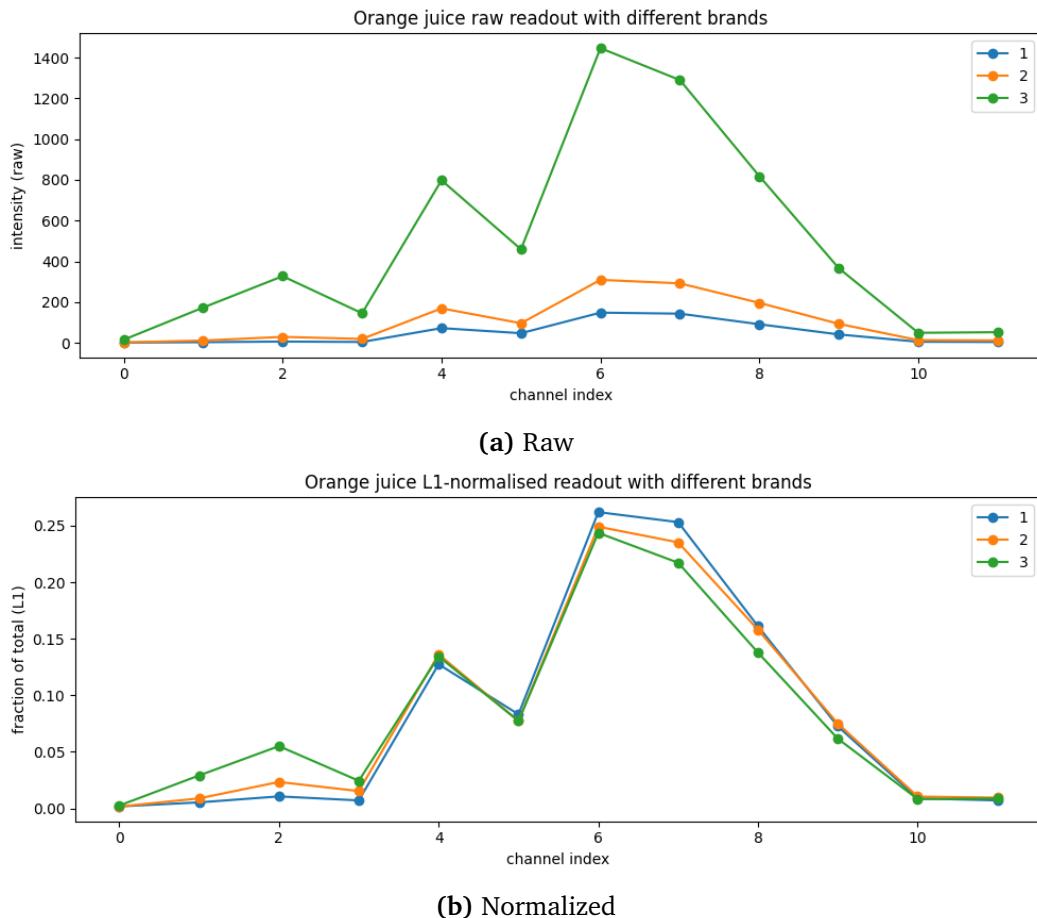


Figure 7: Orange juice spectra with different brands

$$\hat{s}_k = \frac{s_k}{\sum_{i=1}^{12} |s_i|} \quad (4.2)$$

where s_k is the measured intensity from the k -the channel.

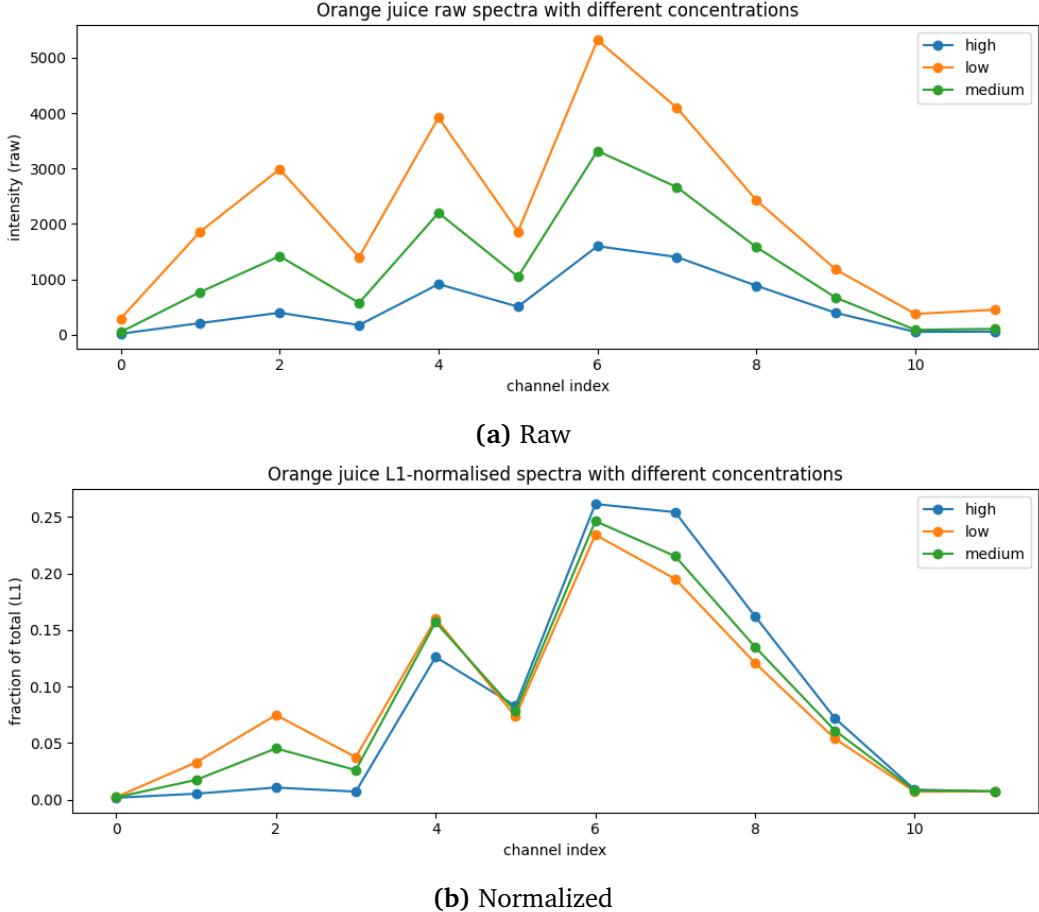


Figure 8: Orange juice spectra with different concentrations

As shown in Fig. 7 and 8, the raw spectra exhibit large amplitude variations, making it difficult to compare spectral shapes directly. After L1 normalization, the spectral profile of orange juice becomes largely invariant to brand and concentration level. In addition, as shown in Fig. 9, the spectral shapes of the three different juices are completely different, which indicates that the model can easily classify the spectra of these different juices. As a result, the L1-normalizaton method is used for the input pre-processing for the juice type classification model while keeping the juice type encoded as integers using a label encoder.

For the concentration classification tasks, from Fig. 8, it shows that the measured intensities of orange juice are different for juices with different concentrations. Hence, we simply take the mean of the raw spectra as one of the inputs. However, as the overall intensities of different juices vary, the juice type is also used as an

input for the concentration classification task. Hence, this concentration model will be able to estimate the concentration level for the specific juice type.

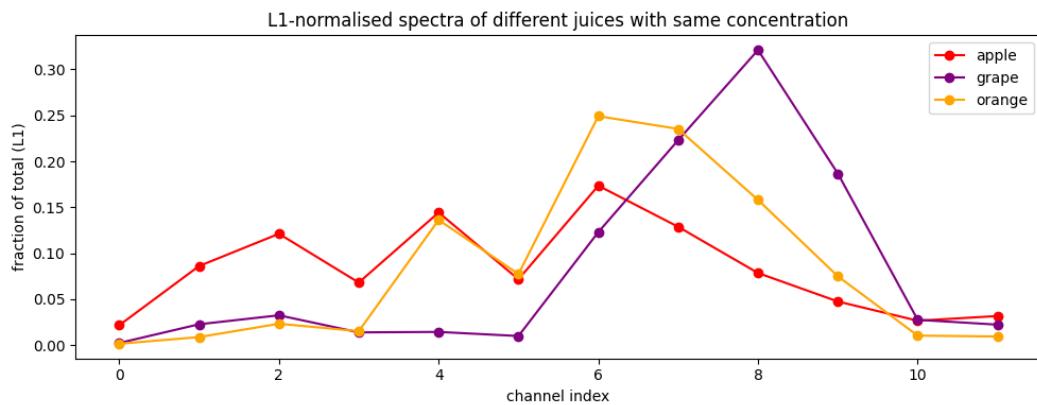


Figure 9: L1-normalized spectra of different juices

4.4 Bill of Materials and Costs

The total cost of the prototype is £67.32, which is substantially lower than that of commercial laboratory spectrophotometers.

No	Item	Description	Price / unit (£)	Qty	Line Total (£)
1	OLED	0.96" SSD1306 OLED display module	4.00	1	4.00
2	Arduino MCU	Arduino Nano 33 BLE Sense Rev2	37.81	1	37.81
3	AS7343	Pimoroni AS7343 multi-spectral sensor	18.90	1	18.90
4	PLA Material	PLA filament, 1.75 mm (£0.02 per g)	0.02	50 g	1.00
5	White LED	White 5 mm LED	0.27	1	0.27
6	Resistor	220 Ω resistor	0.04	1	0.04
7	Wire	Jumper wires (£0.07 per wire)	0.07	5	0.35
8	Breadboard	Solderless breadboard	4.95	1	4.95
Total (£)					67.32

Table 1: Bill of Materials and Costs

5 Evaluation

5.1 Evaluation of the Device's Performance

After collecting 81 samples (3 types of juices * 3 brands * 3 concentration levels * 3 cuvettes), 80% of the samples are used for training and model selection, and the remaining 20% functioned as independent test set. K -fold cross-validation ($K = 5$) is used on the training set for model selection due to the small size of the dataset. This method prevents data leakage from the test set during model selection. The average accuracy and macro-averaged F1-score across the validation folds are used as the primary criteria for comparison.

Classical machine learning classifiers are employed for both juice type identification and concentration classification, including Logistic Regression (LR), linear Support Vector Machine (SVM-Lin), radial basis function Support Vector Machine (SVM-RBF), Random Forest (RF), and Gradient Boosting (GB).

This test plan evaluates the device across multiple juice types, brands, and concentration levels to assess generalization. Table 2 and Table 3 summarize the classification performance on the unseen test dataset:

Table 2: Test-set performance of juice classification models

Model	Accuracy (%)	Macro-F1	Weighted-F1
Logistic Regression	76.47	0.77	0.77
SVM (Linear)	82.35	0.83	0.83
SVM (RBF)	94.12	0.94	0.94
Random Forest	100.00	1.00	1.00
Gradient Boosting	100.00	1.00	1.00

Table 3: Test-set performance of concentration classification models

Model	Accuracy (%)	Macro-F1	Weighted-F1
Logistic Regression	35.29	0.32	0.32
SVM (Linear)	41.18	0.39	0.37
SVM (RBF)	41.18	0.30	0.30
Random Forest	82.35	0.81	0.81
Gradient Boosting	47.06	0.42	0.40

The confusion matrices corresponding to the best-performing model are shown in Fig. 10 (a) and Fig. 10 (b) for juice classification and concentration classification.

5.2 Quantitative and Qualitative Assessment of the Device's Uncertainties

These matrices provide a detailed view of per-class prediction behavior and indicate the major error patterns.

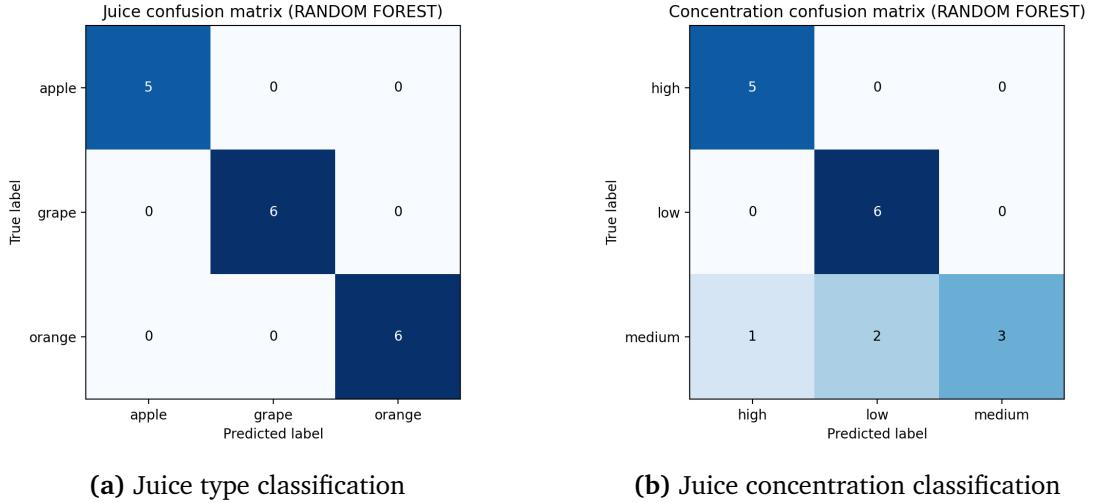


Figure 10: Confusion matrices of RF model

As discussed in the last section, different juices exhibit specific absorption patterns across the 12 spectral channels. Hence, the classification accuracy observed for juice identification is perfect due to the strong spectral difference between different juice types. Simple non-linear classifiers, such as Random Forest and Gradient Boosting, are able to distinguish these differences effectively. In addition, juice classification relies primarily on the shape of spectrum rather than subtle intensity variations, making it independent of sensor noise and measurement error.

In contrast, concentration classification is a more challenging task as different concentration levels of the same juice exhibit highly similar spectral shapes. When the input features are reduced to the mean spectral value, it makes the information that models can learn becomes more limited. As a result, linear models struggle to separate concentration levels, while models such as Random Forest achieve better performance by capturing non-linear decision boundaries.

It is crucial to note that the dataset size used in this work is relatively small, with only 17 samples on the test set. While the final results demonstrate promising performance, the limited sample size may lead to optimistic estimates of generalization capability.

5.2 Quantitative and Qualitative Assessment of the Device's Uncertainties

Uncertainties of the current device could be separated into quantitative and qualitative ones. Starting from quantitative uncertainties, it could be calculated by using the root-sum-square method of systematic and random uncertainties.

5.2 Quantitative and Qualitative Assessment of the Device's Uncertainties

First relative systematic uncertainty is the LED drift due to temperature variation. For a 5 mm white LED, the temperature coefficient of luminous intensity is

$$TC_{IV} \approx -0.5\%/\text{K}.$$

Therefore, for a temperature rise of ΔT , the relative systematic uncertainty due to the LED intensity drift is

$$u_{\text{LED,rel}} = |TC_{IV}| \cdot \Delta T.$$

Assuming a typical temperature rise of $\Delta T = 10\text{ K}$:

$$u_{\text{LED,rel}} = 0.5\%/\text{K} \times 10\text{ K} = 5.0\%.$$

According to the AS7343 datasheet, the ADC (Analogue to Digital Converter) noise is specified as approximately 0.05% of full scale. Assuming the systematic contributions are independent, we combine them using the optimistic (RSS) method:

$$\begin{aligned} u_{\text{sys,rel}} &= \sqrt{u_{\text{LED,rel}}^2 + u_{\text{sens,rel}}^2} \\ &= \sqrt{(5.0\%)^2 + (0.05\%)^2} \approx 5.0\%. \end{aligned}$$

To estimate the random uncertainty introduced by cuvette cleaning after each measurement, $n = 5$ repeated measurements under identical conditions are acquired. As the sample mean and the standard error of the mean (SEM) are

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (5.1)$$

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}. \quad (5.2)$$

Substituting the collected channel values, the cuvette-handling uncertainty is

$$u_{\text{A,rel}} = \frac{s_{\bar{x}}}{\bar{x}} \approx 2.14\%.$$

The overall relative uncertainty is obtained by combining the systematic and Type A components using the root-sum-square method:

$$\begin{aligned} u_{\text{overall,rel}} &= \sqrt{u_{\text{sys,rel}}^2 + u_{\text{A,rel}}^2} \\ &= \sqrt{(5.0\%)^2 + (2.14\%)^2} \approx 5.44\%. \end{aligned}$$

Therefore, the overall relative uncertainty of the proposed measurement system is estimated to be approximately 5.44%.

Other quantitative uncertainties also occur in this device, including minor mechanical misalignment between the LED, cuvette, and spectral sensor, which may introduce small range of error in the detected intensity.

5.3 Improvements in the Next Iteration of the Device

Potential future iterations could be divided into hardware improvements and software improvements. From the hardware perspective, we are currently using separate multiple electronic components for the prototype validation, this could be improved by integrating sensor and MCU into on PCB (Printed Circuit Board). This could reduce the noise brought by the breadboard-based setup.

Regarding to the experiment design, future work will focus on collecting larger and more diverse datasets to further validate the robustness of the proposed system and to reduce variance in performance estimation. Furthermore, model inference can be also deployed on the edge side. Instead of running the computation on PC, trained model can be saved on MCU and locally estimate the results, reducing the data transmission time.

6 Conclusion

This work presents a low-cost and reproducible spectrophotometer system with a 3D-printed mechanical structure. Customized machine learning models are trained to distinguish different juice types and achieve reasonable performance in concentration classification. A systematic evaluation indicates the uncertainties and limitations of the current design. Overall, the system provides a cost-effective and extensible alternative to commercial spectrophotometers, particularly suitable for educational and resource-limited applications.

7 Reference

[1] ‘Spectrophotometer - an overview — ScienceDirect Topics’. Accessed: Dec. 14, 2025. [Online].

[2] ‘Beer-Lambert Law — Transmittance Absorbance’. Accessed: Dec. 14, 2025. [Online].

[3] ‘ams AS7343 Spectral Sensor Ambient light, Color, Spectral Proximity Sensors - ams-osram - ams’, ams-osram. Accessed: Dec. 14, 2025. [Online].