Plastic solid waste identification system based on near infrared spectroscopy in combination with support vector machine

Sensor & actuator case study- Jan 2023

Group members:

Amin Nazari- 12204809

Mohamed Elhalwagy- 22208525

Mohamed Mady- 22207560

Naveen Kumar Mooram- 12201340

TECHNISCHE HOCHSCHULE DEGGENDORF- CAMPUS CHAM

Table of Contents

List of Figures

Abstract

The objective of this project was to create a method for identifying and classifying various types of solid waste made of plastic utilizing near-infrared spectroscopy (NIR) and support vector machine (SVM) classification. A huge amount of plastic components is gathered, and 900–100 nm NIR spectroscopy data was gathered. With multiple kernel functions, including linear, polynomial, sigmoid, and radial basis function (RBF) kernels, SVM models were trained and tested using the data. The outcomes demonstrate the system's capability to distinguish between various types of plastic with a high level of precision and accuracy. This study highlights the potential of combining NIR spectroscopic data with SVM for plastic trash identification and sorting, which could be a financially advantageous and practical method for managing and recycling plastic garbage. However, the conveyor belt, which is essential for the automation of the process, is not yet developed or implemented. Further research is needed to explore the possibilities of incorporating such a device in future, to make the system a cost-effective and efficient solution for plastic waste management and recycling.

Introduction

Since plastic garbage takes hundreds of years to decompose and has an impact on a variety of ecosystems, wildlife, and human health, it is a serious environmental issue. Recycling plastic garbage is essential for protecting the environment and natural resources. Accurately identifying the various forms of plastic is one of the biggest obstacles to recycling plastic waste.

Considering different forms of plastic have distinct qualities and ultimate uses, identifying plastic waste is just an important stage in the recycling process. Accurate identification enables effective sorting and processing of plastic waste, which eventually produces recycled goods of a higher caliber.

The many forms of plastic can be distinguished using a variety of techniques, such as visual inspection, mechanical testing, and chemical analysis. These approaches, however, can sometimes be time-consuming, expensive, and even harmful. Near-infrared spectroscopy is one method for classifying various plastics (NIRS).

NIRS is a non-destructive technique for determining a material's chemical make-up by observing how much infrared light is absorbed. By examining the distinctive absorption spectra of each type of plastic, this technique can be used to distinguish between various types of plastic. It has been demonstrated that NIRS is a quick, precise, and non-destructive method for classifying various kinds of plastic.

A machine learning approach called Support Vector Machine (SVM) can be used for classification tasks like determining the various varieties of plastic. A dataset containing labeled samples can be used to train the supervised learning algorithm SVM, which can then be used to categorize new, untainted examples. To classify and identify plastic garbage, SVM and NIRS can be used. Different types of plastic garbage, including mixed plastic samples, can be categorized using this method.

In this research, we suggest creating a system for identifying plastic solid waste based on NIRS in conjunction with SVM classification. The system will be evaluated on a dataset of mixed plastic samples after being trained on a dataset of NIR spectra from various types of plastic. The system's objective is to increase the recycling industry's ability to identify plastic waste accurately and practically.

From online sources, we collected three distinct datasets of plastic sample data, and for each dataset, we trained an SVM model. Using this method, we can assess the model's performance on each dataset and spot any potential changes in the consistency or quality of the data. We can choose which model to employ for the final plastic trash identification system after assessing the models' performance, or we can combine the models using ensemble methods to increase the system's performance and robustness.

As stated previously, we received three distinct datasets of plastic samples from internet sources. Considering the potential variances in data quality and consistency, we combined the three datasets. To increase the effectiveness and resilience of the model, we adopted a similarity-based dataset mixing strategy, which includes blending various datasets based on their similarity. We enhanced the number of samples and improved model performance by integrating the datasets. To increase the performance of the model when the new dataset is tiny or has little diversity, we executed an SVM model on the mixed dataset after combining the datasets. This strategy can also be used to strengthen the model's robustness by minimizing the effects of differences in the consistency and quality of the data.

Several performance indicators, including precision, recall, F1-score, and confusion matrix, will be used to assess the proposed system's correctness. The outcomes will be contrasted with those of other cutting-edge techniques for identifying plastic garbage.

Overall, the suggested system for classifying plastic solid waste based on NIRS and SVM has the potential to be a quick, precise, and non-destructive way to detect various kinds of plastic garbage. This strategy can be used to categorize various kinds of plastic trash, including samples made up of mixed plastics, and enhance the precision and usefulness of plastic waste identification in the recycling sector. In the end, this will produce recycled goods of higher quality and lessen the negative effects of plastic waste on the environment.

Literature review

Identification of plastic solid waste is a crucial challenge in the recycling and waste management sectors. Utilizing support vector machine (SVM) classification along with near infrared spectroscopy (NIRS) is one method for classifying various types of plastic.

NIRS is a non-destructive technique for analyzing a material's chemical make-up by gauging its infrared radiation absorption. SVM is a machine learning technique that can be employed for classification jobs like detecting multiple plastic types.

In [1] Zh and its group represent a SVM approach for classifying plastic solid waste including polypropylene (PP), polystyrene (PS), polyethylene (PE), poly (methyl methacrylate) (PMMA), acrylonitrile butadiene styrene (ABS) and polyethylene terephthalate (PET). 186 samples were pressed into uniform round plates (diameter of 100 mm, thickness of 3 mm). All samples were chosen to be transparent or in a bright color, due to the limitations of NIR spectroscopy in identifying dark-colored samples. Also, to simulate real-world conditions, the samples were not cleaned.

To acquire NIR spectra, a NIR optical fiber spectrometer (QUEST 512, Ocean Optics Inc., USA) with an InGaAs detector was selected as the acquisition unit. The spectrometer was chosen because it operates in the most suitable NIR region of 1000-1700 nm. The spectra were recorded with a resolution of 2 nm, and an accumulation of 4 scans was used. They have applied principal component analysis (PCA) to reduce their spectral data dimension and it’s been figured out first the spectral data can be fully represented using just the first 7 principal components as shown in Fig.1. Finally, using a Radial Basis Function (RBF) kernel they have reached 97.5% of accuracy in identifying plastic types.

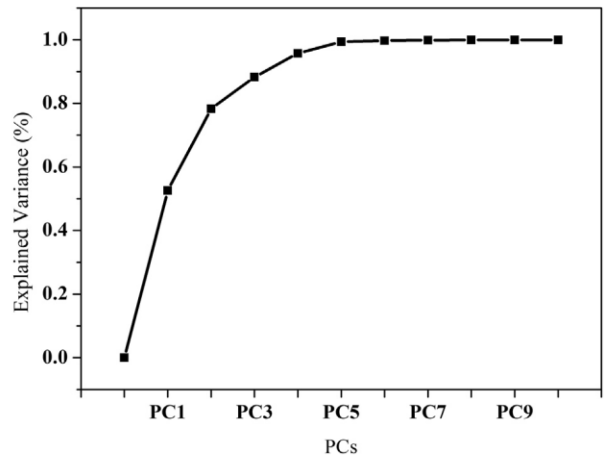

Figure 1-Explained variance plot of the PCA model based on spectra of training set. [1]

Masoumi and Safavi in [2] introduces a “Two-filter” method for classifying plastic types. Relative reflectance is used for this purpose where reflectance in 1724 nm and 1656nm are just considered. In the relative reflectance method, the proportion between reflectances is used instead of using the absolute value of the reflectances. It has been shown that this method performs highly satisfactory as its not sensitive to the sample thickness. Furthermore, different types of plastics make their different spectral graphs in the 900-1700 nm range in which the plastic samples of the same type follow almost the same Graph as it is shown on Fig.2. The reflectance spectra in Fig. 2 are plotted with an offset to prevent overlap and improve visibility.

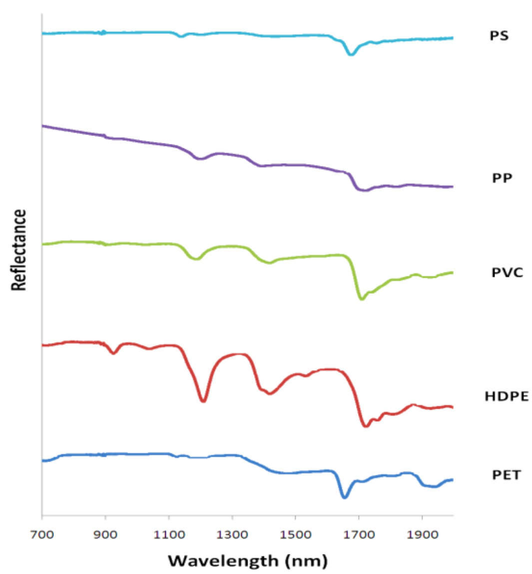

Figure 2-. The reflectance spectra of five resins[2]

Overall, the research contributes to the combination of NIRS and SVM as a promising method for identifying solid waste made of plastic. To increase the precision and utility of this approach, further research is required.

Objective

The goal of this project is to create and train a support vector machine (SVM) model to accurately classify different types of solid plastic waste using near infrared spectroscopy (NIRS) data. This will be accomplished by utilizing existing NIRS datasets containing spectra of various types of plastic waste. The SVM model will be trained and optimized using these datasets to improve its accuracy in identifying different types of plastic.

The project will begin with the selection and acquisition of appropriate NIRS datasets, which will be used to train the SVM model to learn the distinct properties of each plastic type, such as polyethylene, polypropylene, and polystyrene. Once the datasets have been collected and prepared, the SVM model will be trained on them. To improve its performance and accuracy in identifying different types of plastic waste, the model will be fine-tuned and optimized.

The overall objective is to increase recycling rates, reduce the environmental impact of plastic waste, and achieve high accuracy in plastic identification. The project will also concentrate on developing the NIR system, which will generate NIR data for plastic using an NIR Photodiode, NIR LEDs (800 nm-1700 nm), a microcontroller, and other electrical components. We created an electrical circuit for the NIR system.

Scope

The scope of this project is to create and train an SVM model to accurately distinguish and identify different types of solid plastic waste using near-infrared (NIR) spectroscopy data. This will be accomplished by collecting NIR spectra from various plastic samples using a lab prototype circuit for an NIR sensor. The breadboard prototype circuit will be used to precisely detect the infrared radiation absorption of the plastic samples.

The following step is to manually collect a set of plastic sample data that will be used to train the SVM model. The dataset will concentrate on the five most common types of plastic waste, which account for more than 75% of consumer waste plastic. After the dataset has been prepared, the

SVM model will be trained using data from the NIR sensor. In our case, we gathered pre-existing NIR data.

The model's performance and accuracy in identifying different types of plastic waste will be fine-tuned and optimized. The project will be more concerned with developing and optimizing the SVM model than with the NIR system, data collected using the NIR sensor, and circuit.

Requirements

NIR sensor: A cheap, readily available NIR sensor that can be used to gauge how much infrared light plastic samples absorb is widely available on the local market.

Plastic samples: To train the SVM model and test the system, a variety of plastic samples, including PE, PET, PP, PS, and PVC, will be employed.

System for reflect spectroscopy and NIR illumination: a system that consists of an infrared light source, optics to control the light's illumination and collection, and a sensor to measure the radiation's absorption.

SVM model: To classify and detect new plastic samples, a Support Vector Machine model will be trained on the dataset.

Electronic components and a breadboard: A breadboard will be used to create the NIR sensor's circuit, making experimentation and customization simple.

Software: Software will be required for the SVM model implementation, data preprocessing, and result visualization.

Laboratory or workspace: The prototype can be set up and tested in a laboratory or workspace.

Sensor Comparison

In this study, NIR sensors from several producers, including Excelitas, Hamamatsu, and API, were compared. The comparison was primarily concerned with the sensor's price, packaging, and wavelength range. It was discovered that Excelitas sensors span a wide range of wavelengths, from 800-1700 nm. The wavelength coverage of Hamamatsu sensors, on the other hand, was typically between 900 and 1700 nm. API sensors cover a wide range of wavelengths, often between 900 and 1700 nm. Excelitas sensors were discovered to have the smallest packages, which makes them perfect for portable applications. It was discovered that API sensors had the largest package, slightly larger than that of Hamamatsu sensors. Finally, Excelitas sensors were discovered to be the most expensive, followed by API and Hamamatsu sensors in terms of price. The exact application and the trade-off between the wavelength range, package size, and price will ultimately determine which sensor is best.

Table 1- Photodiode sensors comparison

Manufacturer	Series	Type	Range	Packages	Part number	Price(euro)
Excelitas	C30617	InGaAs	800-1700	TO-18 with ball glass lens	C30617BH	55
Excelitas	C30618GH	InGaAs	800-1700	TO-18 with flat glass lens	C30618GH	78
Hamamatsu	G11193	InGaAs	900-1700	Surface mount	G11193-10R	-
API	-	InGaAs	900-1700	Surface mount	0090-3111-185	7.5

Our chosen sensor for this case is from Advanced Photonix (API) with a wavelength range of 900-1700nm, type of indium gallium arsenide (InGaAs), and priced at 7.5 euros.

The sensor we've chosen for this situation is an Advanced Photonix (API) InGaAs sensor with a 900–1700 nm wavelength range.

Spectrometer in Industrial Applications

A versatile and powerful instrument for a variety of industrial applications is the Specim FX17 sensor. It delivers a high degree of sensitivity and precision for a variety of measurements with a spectral range of 900-1700 nm. It also has a high spatial resolution of 640 pixels, which makes it perfect for applications that demand accurate and detailed images.

The[14] Specim FX17 sensor has a fast image speed, with a frame rate of 527 FPS for the GigE version and 670 FPS for the CameraLink version. This is one of its main features. This makes it the perfect option for real-time imaging-required applications including recycling, waste sorting, and food and feed quality management.

Moisture measurement is another application where the Specim FX17 sensor excels. It is an effective tool for determining the amount of moisture in a variety of materials, including food goods and industrial materials, thanks to its high-resolution imaging capabilities and wide spectral range.

The Specim FX17 sensor is useful for industrial applications as well as security and danger detection. It is an effective tool for spotting potential threats and keeping an eye out for suspicious activities due to its high-speed photography and broad spectral range. A purchase of 36000 euros will get you a Specim FX17 sensor.

Overall, the Specim FX17 sensor is a powerful and versatile tool for a wide range of industrial applications, offering high-performance imaging capabilities and a wide spectral range for a range of measurements, at a price of 36000 euros.

Figure 3- Specim FX17 spectrometer[14]

Bill of Materials

Table 2- Bill of materials

S. No	Components Name	Part-Number	Quantity
1.	Arduino_UNO_R3	-	1
2.	capacitor 100nf	-	6
3.	capacitor 47nf	-	1
4.	capacitor 100pf	-	2
5.	capacitor 22mf	-	3
6.	capacitor 18pf	-	2

7.	capacitor 10pf	-	2
8.	Led-1050nm	-	1
9	Led-1200 nm	-	1
10	Led-1300nm	-	1
11	Led-1450 nm	-	1
12	Led-1550nm	-	1
13	Led-1650 nm	-	1
14	Led-860nm	-	1
15	Led-940nm	-	1
16	InGaas-sensor	0090-3111-185	1
17	Amplifier	MCP6281	2
18	Analog to Digital Converter	AD7124-4BRUZ-RL7	1
19	Voltage Reference IC Adjustable	AS431AZTR-E1	2
20	Voltage Reference IC Fixed	ZXR E330S-7	2
21	LED Driver IC	MAX6971 NG+	1
22	Ferrite Bead Axial	-	1
23	Tactile Switch	-	1
24	TTL Oscillator	ECS-100A-080	2
25	Circuit pins	-	2

Description of Components

Arduino UNO R3

A microcontroller board called Arduino UNO R3 is based on the ATmega328P microcontroller. It has a 16 MHz quartz crystal, 6 analogue inputs, 14 digital input/output pins, a USB port, a power jack, an ICSP header, and a reset button. The Arduino Software (IDE), a free, open-source tool that enables users to write, upload, and run code on the board, is used to program the device.

Due to its simplicity of use, extensive library of tutorials, and active user base, Arduino boards are frequently used in a wide range of projects, including robotics, automation, and Internet of Things (IoT) applications. The Arduino UNO R3 can be used in our situation as a microcontroller to operate and communicate with other components.

LEDs

When a current is carried through semiconductor LEDs (Light Emitting Diodes), they release light. They are extensively utilized for a wide range of purposes, such as lighting, displays, and signaling. The semiconductor material used in an LED determines the wavelength of the light it emits, which can range from infrared to ultraviolet. Our research uses LEDs with wavelengths between 860 nm and 1650 nm.

Near-infrared (NIR) applications typically employ LEDs with wavelengths between 860 nm and 1650 nm. Although this wavelength range is invisible to the human eye, it is still useful for a number of tasks like sensing, communication, and imaging.

InGaAs photodiode

For detecting the reflectance, the InGaAs (Indium Gallium Arsenide) photodiode (part number 0090-3111-185) [7] is being used in our project to gather information on various forms of solid plastic trash. The sensor is sensitive to light in the near-infrared (NIR) region, which has a wavelength range of 900–1700 nm. InGaAs photodiodes are also known for their high-speed response, low noise and high quantum efficiency, making them a reliable choice for high-performance optical systems. The plastic waste's NIR signals are captured and turned into electrical impulses by the device. The responsivity of 0090-3111-185 InGaAs photodiode is shown in Figure 4.

This sensor can be put directly onto the surface of the circuit board thanks to its surface mount packaging. The sensor is essential to the system because it collects the NIR signals given off by the waste plastic and gives the SVM algorithm the information it needs to correctly detect and classify the various types of plastic.

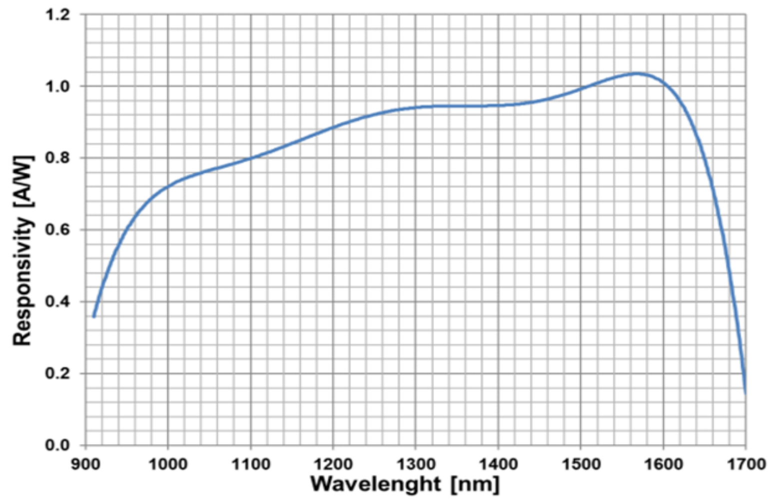

Figure 4- Responsivity of 0090-3111-185 InGaAs photodiode

Amplifier

In our project, an amplifier works in tandem with an InGaAs sensor. The amplifier's function is to boost the electrical impulses that the sensor produces. This is crucial because the signals produced by the sensor might not have enough strength or amplitude to be analyzed effectively by the SVM algorithm. By ensuring that the signals generated by the sensor are of sufficient amplitude to be analyzed by the SVM algorithm, the amplifier contributes to the improvement of the overall accuracy and performance of the plastic waste identification system.

The system's electrical layout uses an InGaAs sensor and amplifier to gather and process the NIR signals that the plastic garbage emits. The NIR signals are picked up by the sensor, which then transforms them into electrical impulses that are amplified by the amplifier to become stronger. The SVM algorithm is then used to identify and classify various forms of plastic garbage using these amplified signals as input data. The near-infrared (NIR) light intensity collected by the InGaAs sensor is represented numerically by the signals it produces. The amplifier is essential to this electrical design because it makes sure that the signals produced by the sensor have enough amplitude for the SVM algorithm to process them accurately.

Analog-to-Digital Converter

In our project, It is important to keep in mind that the InGaAs sensor produces analogue signals, which must be converted into digital form in order to be processed by a digital algorithm like SVM. This is where ADC comes into play. By guaranteeing that the analogue signals produced by the sensor and amplifier can be processed precisely by the SVM algorithm, the ADC plays a crucial function in the system. The analogue signals produced by the sensor and amplifier are sampled and quantized by the ADC to provide digital values that the SVM algorithm can comprehend.

Voltage Reference IC Adjustable

To supply a consistent and precise reference voltage for the ADC in our project, an adjustable voltage reference IC is probably being used. The analogue signals produced by the InGaAs sensor and amplifier are converted into digital signals by the ADC using this reference voltage as a point of reference. The output voltage can be modified using the adjustable voltage reference IC to meet circuit requirements or to account for changes in temperature or other environmental conditions. This guarantees that the ADC has a stable and reliable reference voltage to deal with, which enhances the system's overall accuracy and performance.

Voltage Reference IC Fixed

In the case of our project, it's likely that a fixed voltage reference IC is being used to give the ADC a steady and precise reference voltage. The analogue signals produced by the InGaAs sensor and amplifier are converted into digital signals by the ADC using this reference voltage as a point of reference. A fixed voltage that is stable under a variety of circumstances, including temperature, supply voltage, and time, is provided by the fixed voltage reference IC. This guarantees that the ADC has a stable and reliable reference voltage to deal with, which enhances the system's overall accuracy and performance.

A fixed voltage reference IC provides a fixed output voltage that cannot be adjusted, whereas an adjustable voltage reference IC allows the output voltage to be adjusted. This is the main distinction between the use of a fixed voltage reference IC and an adjustable voltage reference IC in our project.

LED driver IC

Our plastic waste identification system relies on the LED driver IC, which controls the current flowing through the 8 LEDs with various wavelengths and supplies the necessary voltage to turn them on. By illuminating the plastic parts with different wavelengths of light, it is able to capture the distinct NIR spectra of each plastic. These 8 LEDs are used to illuminate the plastic pieces being studied.

The LED driver IC is essential to the system because it supplies the voltage and current required to illuminate the 8 LEDs with various wavelengths. In addition to guaranteeing that LEDs are used within their acceptable working parameters, it also offers stable, reliable lighting conditions for accurate NIR spectra collecting.

The system's overall efficiency is additionally increased by the LED driver IC, which controls the current going through the LEDs and supplies the required voltage to make them glow. As a result, the LEDs' lifespan is increased, and power consumption is decreased.

In conclusion, the LED driver IC plays a crucial role in the system by controlling the current flowing through the 8 LEDs with various wavelengths, providing the required voltage to light them up, and ensuring that the LEDs are operated within their safe operating range for precise NIR spectra collection, which in turn helps to improve the overall accuracy and performance of the plastic waste identification system.

Ferrite Bead Axial

Ferrite beads, sometimes referred to as EMI (electromagnetic interference) filters, are passive parts used in electronic circuits to reduce unwanted electromagnetic interference (EMI). In electronic circuits, ferrite beads are used to reduce high-frequency noise. In order to efficiently prevent high-frequency noise from entering the circuit, they function by offering a high resistance to the noise.

The axial type of ferrite beads is one of many varieties available. Axial ferrite beads are cylindrical ferrite beads with leads or pins attached to each end. The ferrite bead is connected to the circuit via these leads or pins.

It is likely that the ferrite bead axial in our project is being utilized to block undesired high-frequency noise in the circuit. By lowering interference and ensuring that the circuit operates as intended, this could improve the circuit's performance.

Tactile switch

Pushing down on a tactile switch, commonly referred to as a momentary switch or push-button switch, activates it. The switch returns to its original position as soon as the button is released, opening the circuit. Control panels, keyboards, and other devices that call for a straightforward on/off or fleeting switching action frequently use tactile switches.

It is likely that in the case of our project, the tactile[8] switch is being used as an input device to regulate how the system operates. The switch can be used, for instance, to start or stop data collecting, turn the system on or off, or alter the system's settings. [NM1]

TTL Oscillator

A TTL oscillator is a type of electronic oscillator that creates a periodic square wave signal using transistor-transistor logic (TTL) technology. A logic gate, such as a NAND gate or a NOT gate, and passive parts like resistors and capacitors make up the oscillator circuit in most cases. The square wave signal is produced by the logic gate, and the frequency of the signal is controlled by the passive components.

The TTL oscillator is probably being used in our project to produce a periodic square wave signal that is used as a clock signal or timing signal for the system. The oscillator's square wave output could be used to regulate the timing of data collection, signal processing, or other system operations.

Circuit pins

Circuit pins are tiny, cylindrical connections that are used to join electronic parts to a circuit board. They are often referred to as headers or connector pins. They are frequently composed of metal, and a tiny hole in the center enables for insertion into a PCB or breadboard. Different sized circuit pins are available for both through-hole and surface-mount components.

Circuit pins are probably being utilized in our project to connect the various electronic components to the breadboard. The circuit pins act as the intermediary connectors between the two component types because we have a mixture of surface-mount and through-hole components. We can quickly and safely attach the surface-mount components on the breadboard and join them to the circuit by utilizing circuit pins.

Circuit Design

Our near-infrared spectroscopy and support vector machine-based plastic waste identification system's circuit design will incorporate important parts like an NIR sensor, LED driver IC, LEDs, amplifier, microcontroller, voltage reference IC, Ferrite bead, tactile switch, and TTL oscillator to gather data, regulate operation, and increase the system's accuracy. Additionally, passive parts like resistors and capacitors will be utilized to regulate the voltage in the circuit and filter out noise. To connect all these parts together, a breadboard will be used to hold them all.

KiCad

Powerful open-source software called KiCad is frequently used to create printed circuit boards and electronic circuit schematics (PCBs). It is a user-friendly program that offers a full range of tools for PCB designing, 3D visualization, and schematic capture. In this project, we used KiCad to construct the electrical circuit for our NIR spectroscopy and SVM-based plastic trash recognition system. Using the software, we were able to quickly and accurately draw a circuit diagram that served as the basis for the PCB layout design. This made it easier to make sure the circuit worked as intended and that the various parts were linked correctly. It also made it simple for us to spot any possible problems and make any necessary revisions prior to fabricating the final PCB.

LED Circuit Design

The 8 LEDs' current is controlled and regulated by the LED driver IC, enabling them to emit light with the desired wavelength. The plastic components being examined are lit using LEDs as a light source. The system may be readily controlled by the user thanks to the LED reset switch. The circuit is powered by a 3.3V power source, and capacitors and resistors are employed to regulate the circuit's voltage and filter out external noise. This circuit design is utilized to regulate the 8 LEDs' light with various wavelengths, which is then used to record the distinctive NIR spectra of each plastic.

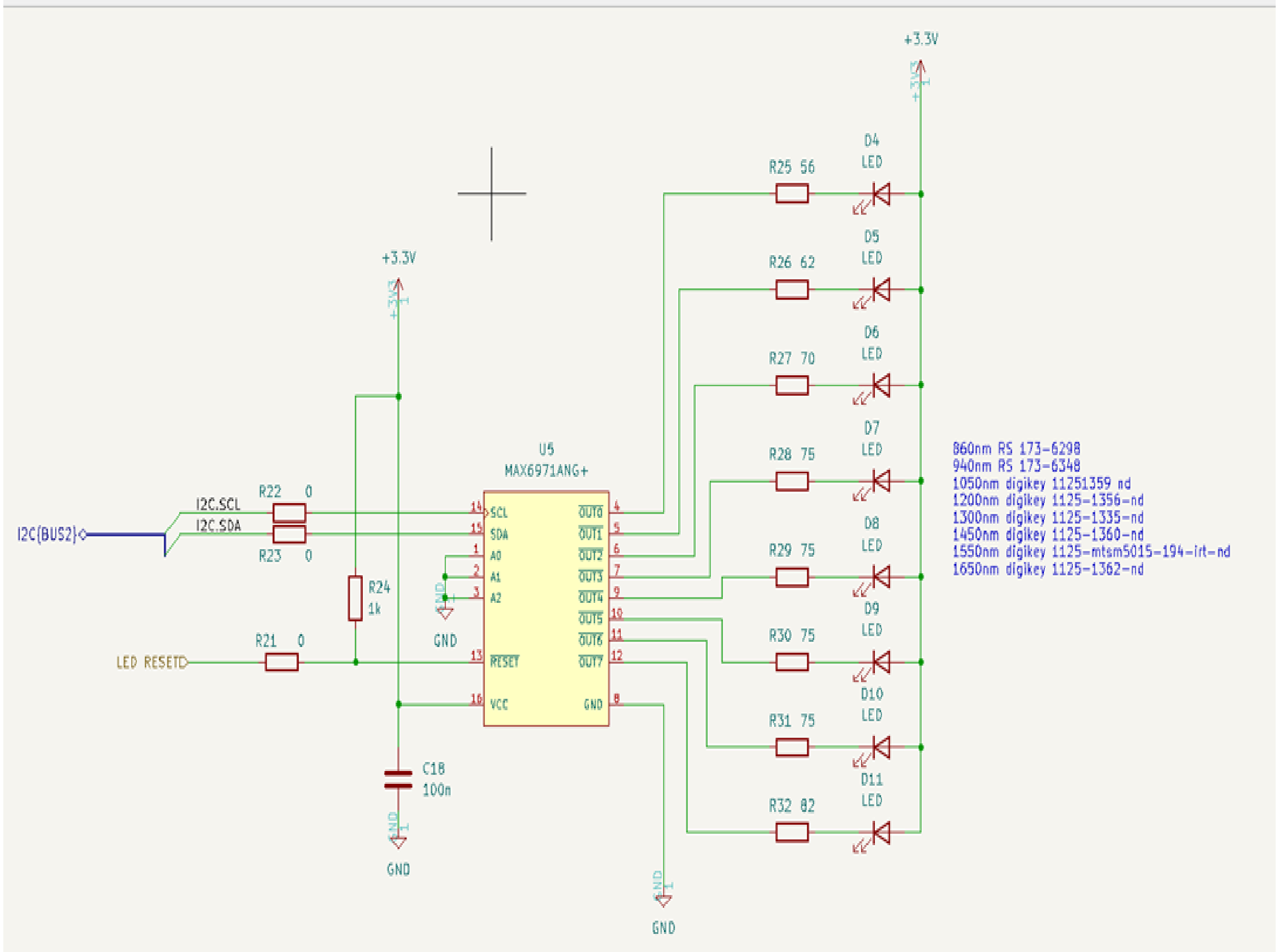

Figure 5- LED Circuit Design

Microcontroller Circuit

The microcontroller, an Arduino UNO, has connections to reset buttons, LED debug, ADC debug, and LEDs in its circuit design. The user can quickly reset the microcontroller and the circuit using the reset buttons. The circuit's health and performance are tracked using the LED debug and ADC debug, which provides helpful data for troubleshooting and debugging. The LEDs are employed to inform the user visually of the circuit's condition. The microprocessor serves as the system's brain, managing circuit operation and processing data gathered from the NIR sensor. Overall, the system control and monitoring interface provided by this circuit design is user-friendly, and it makes debugging and troubleshooting simple.

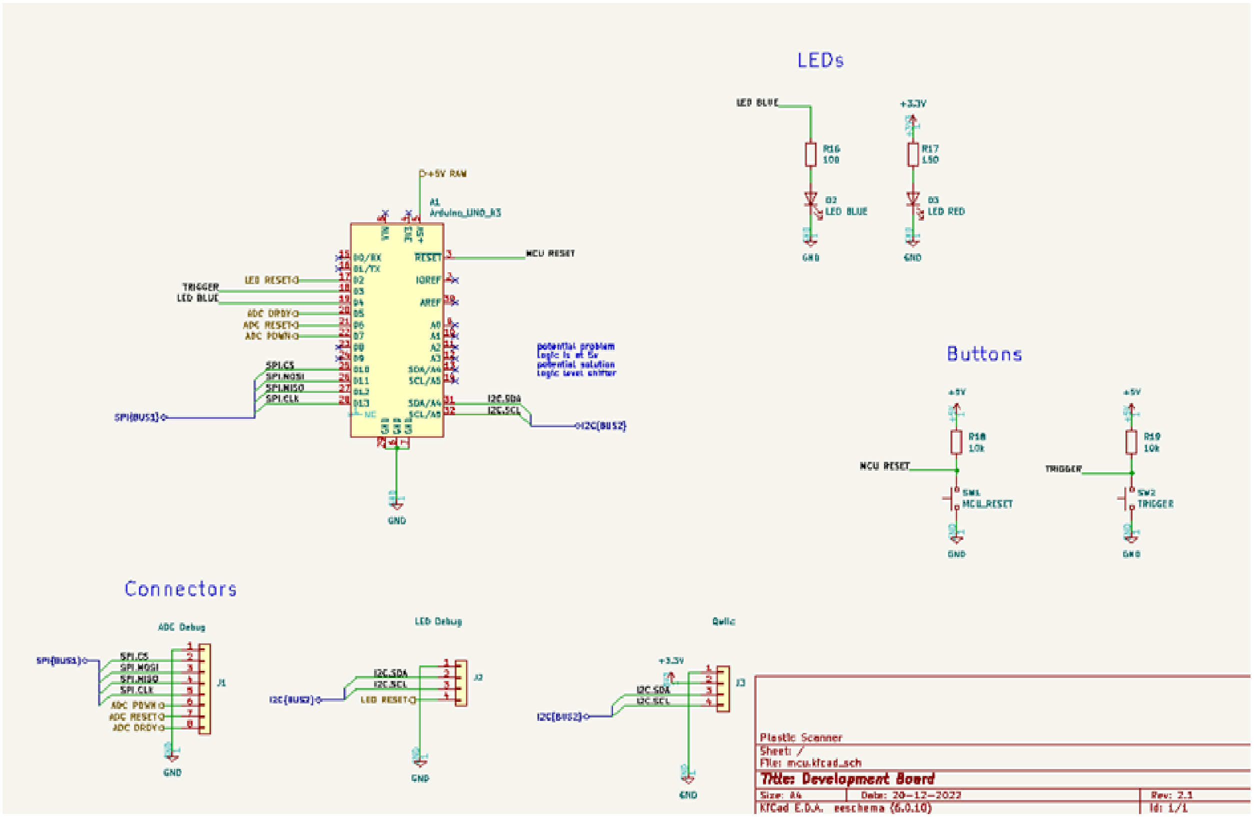

Figure 6- Microcontroller circuit design

ADC Circuit

A 5V power supply is included in the ADC circuit diagram along with a variety of resistors and capacitors. Together, these parts transform the analogue signals from the NIR sensor into digital signals that the microcontroller can read and handle. The circuit's voltage is stabilized and noise is filtered using resistors and capacitors, resulting in accurate and dependable data conversion. The circuit receives power from the 5V power supply, enabling optimal operation. This circuit, which is linked to the microcontroller and NIR sensor, is essential for collecting the distinctive NIR spectra of each plastic and identifying and classifying the many types of plastic garbage.

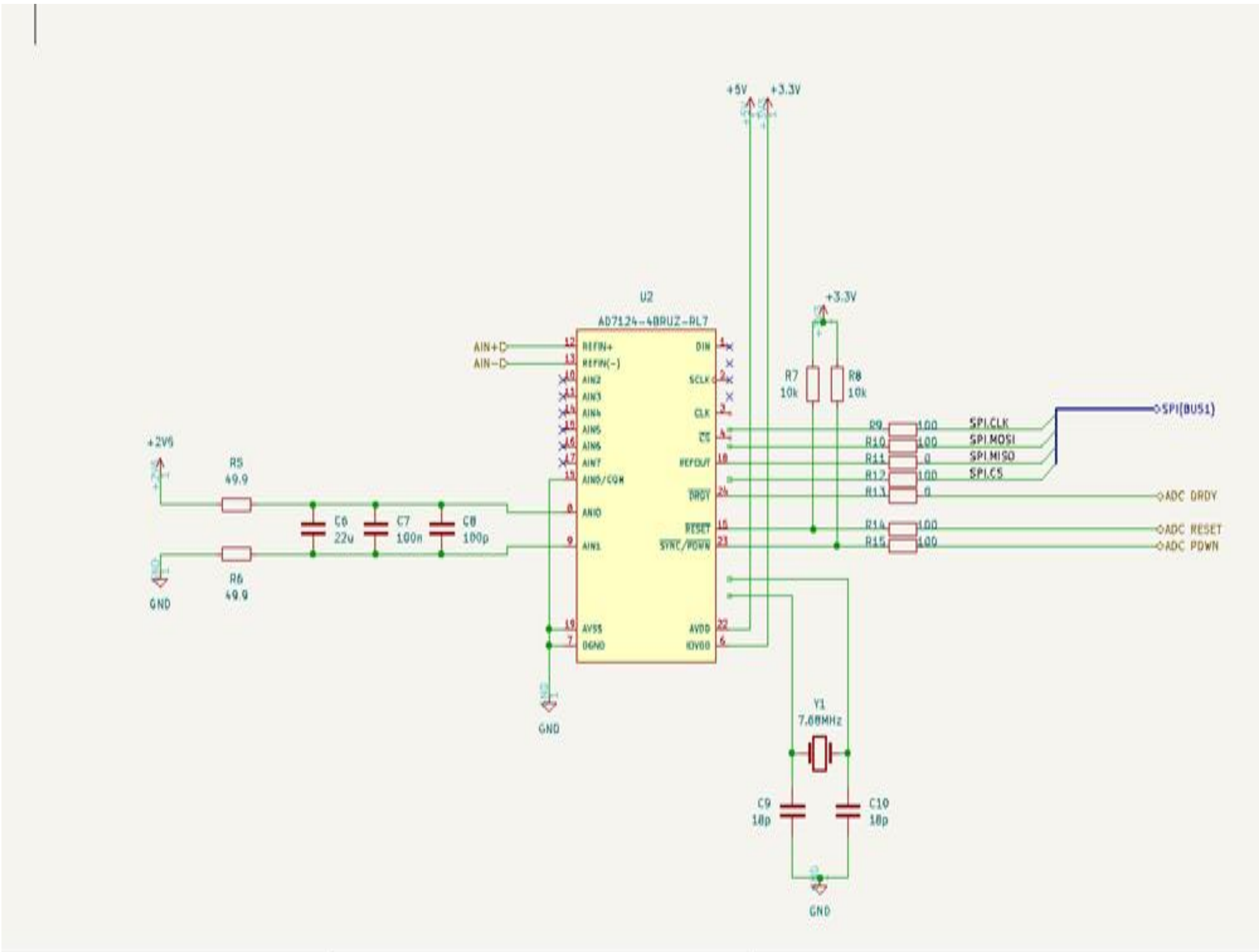

Figure 7- ADC circuit design

Sensor circuit

The sensor circuit design shows an amplifier in addition to numerous capacitors and resistors. The NIR sensor's weak signals are amplified by the amplifier, enabling more precise and dependable data collecting. Resistors and capacitors are employed in the circuit to stabilize the voltage and filter out noise, which enhances the sensor's overall accuracy and performance. Each plastic's NIR spectrum is recorded by the sensor, and this information is then sent through the amplifier and other circuit elements to create a powerful, distinct signal that the microcontroller can easily read and analyze. The ADC circuit, which transforms analogue signals into digital signals that can be examined by the SVM algorithm to precisely identify and classify the various types of plastic garbage, is designed to work in conjunction with the microcontroller.

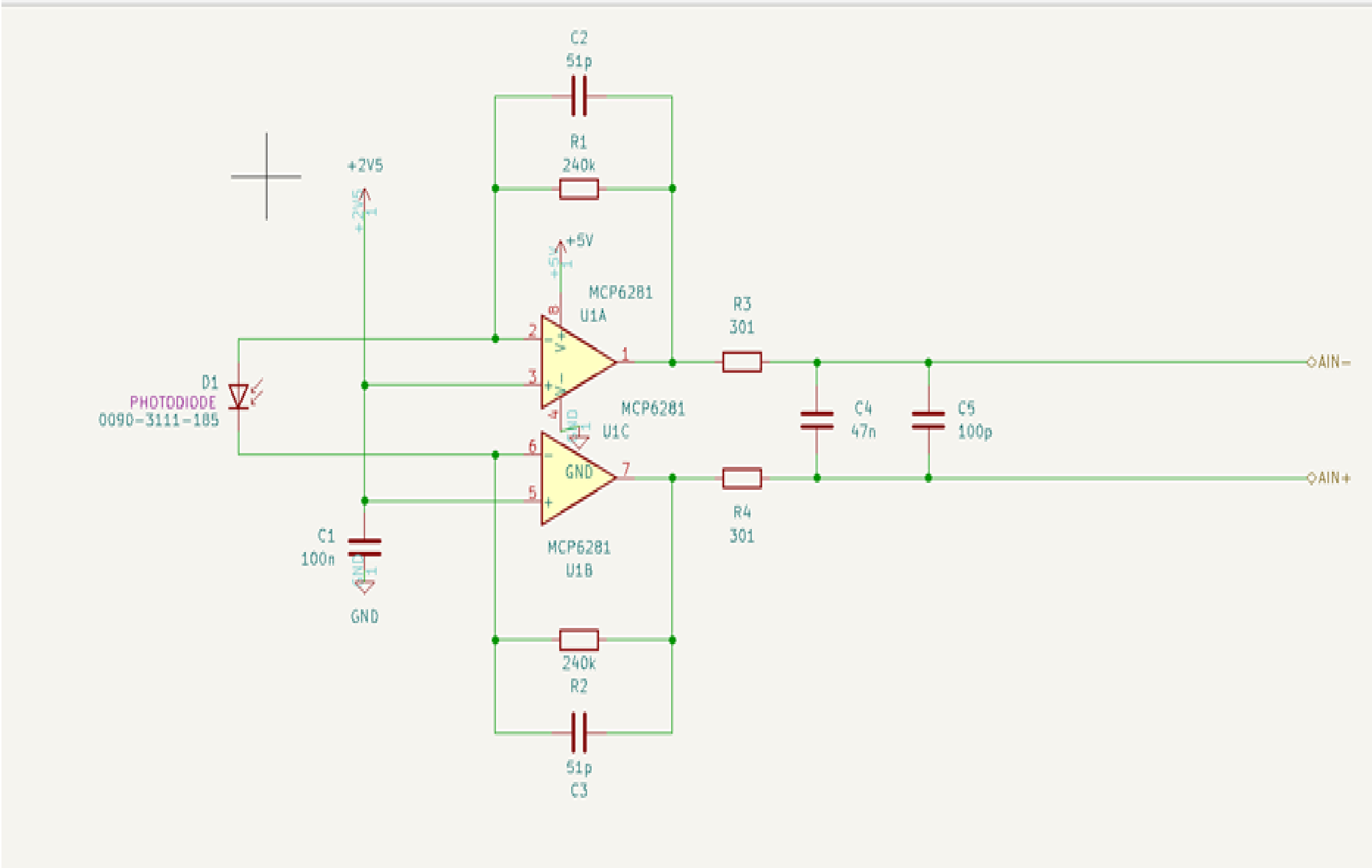

Figure 8- Photodiode electrical design

Overall Circuit

All three circuits (the ADC circuit, the sensor circuit, and the LED circuit) are combined in the circuit diagram. The output of the sensor circuit is connected to the input of the ADC circuit, and the output of the ADC circuit is connected to the microcontroller. This is how all of these distinct circuits are connected to one another. To regulate the LEDs' illumination, the microcontroller and power supply are connected to the LED circuit.

All the components are connected on a single breadboard by combining all the circuit layouts, making it simple to test, troubleshoot, and debug the system. Additionally, this makes it simpler to comprehend how the system as a whole works and how each component is connected. The whole plastic waste identification system's network of connections and cooperative operation is shown in this diagram.

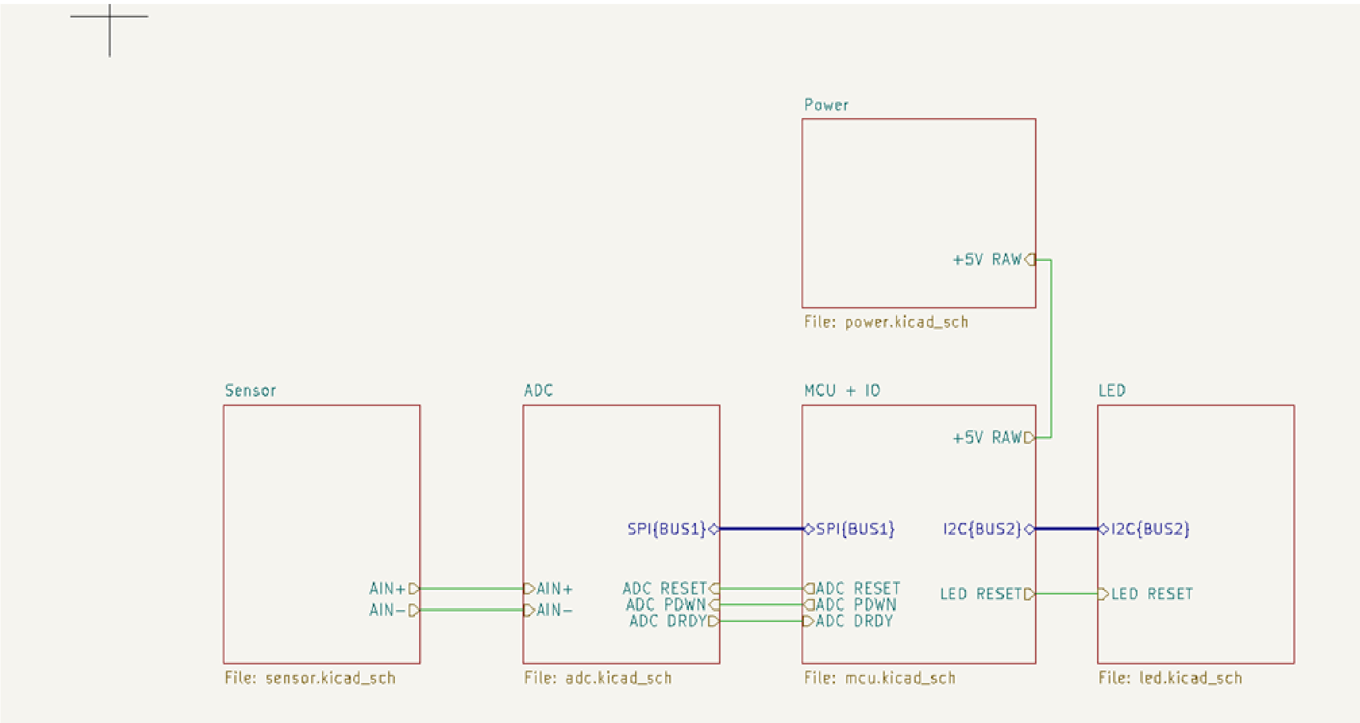

Figure 9- Overall electrical schematic

Plastic samples

As of 2015, we have cumulatively created 7.8 billion tons of plastic, which enumerate that every human being has at least 1000kg of plastic, per person, alive today. The amount of plastic cumulated around the world is expected to still grow in the coming years since global plastic production rates are still growing yearly.

Figure 10- Plastic waste harming nature

For implementing the required identification model, samples made of PET, HDPE, PVC, PP, and PS were collected. That dataset was collected from plastic samples obtained from a domestic waste disposal site. A total of 256 samples of plastic bottles and containers were collected with the help of Technology Campus Cham students and staff. Some of the samples were clear, some were white, and some were colored. These samples generally included various bottles and containers of different sizes. The samples were collected for testing and consisted of four different types of plastic: PETE, HDPE, LDPE, and PP.

We did Manual sorting for the collected plastics and resulted in Specifically, 147 samples of PETE, 18 samples of HDPE, 6 samples of LDPE, and 85 samples of PP were surveyed and collected as part of the dataset. Reflectance of infrared light determines the type of plastic, it is noninvasive but requires a database of known samples.

Table 3- Plastic collection

Type 1 PETE	Type 2 HDPE	Type 4 LDPE	Type 5 PP
147	18	6	85

Flattening plastic samples is a good idea to consider It presses the bottles so they do not roll or shake while being scanned by the NIR sensor. Flattening the containers reduce rolling and shaking movement on the scanning surface. Unflatten plastics can lead to preventing the sensor from getting the right reading. Not all samples were cleaned prior to testing, and in some cases were notably dirty or greased on their surfaces which could also lead to false results.

Figure 11- Collected plastic samples

Datasets

Through our search for literature we find a paper [4] from Armin Straller from Augsburg Hochschule in which he had implemented a low cost spectrometer for classifying different plastic types called the Reremeter.

Figure 12- Reremeter developed by Armin Straller [4]

This low-cost spectrometer encompasses 8 variant NIR illuminator in different wavelengths between 850 to 1650. The Reremeter is capable of differentiate 4 types of plastic including polyethylene terephthalate (PET), high density polyethylene (HDPE), polypropylene (PP) and polystyrene (PS). Straller use the relative reflectance idea for classifying the plastic types. We find this implementation inspiring for starting to build our own spectrometer. Also Armin has published his data regarding this spectrometer on GitHub which is included the data regarding electrical and mechanical design and also outcome data from scanning 70 plastic samples.

Following his project, we found Jerry de vos from Delf university, who continued Straller's idea for his master thesis. Also he had shared the project steps in GitHub and the project website [3]. There we find another dataset consist of 60 samples.

We didn't see this amount of data enough for training our model, specially as they had used same structure and same idea. So we continued our search where we found a group of researchers in Eindhoven university which are working on the idea of implementing an integrated NIR scanner [5]. Ms.Fang kindly shared their group dataset with us in respond of our email. Their dataset encompasses 80 plastic samples which are scanned by commercial spectrometer in the range of 900-1700nm in 565 variant wavelengths. The scanning step is not fix. each sample is been scanned 5 times and average value has been used as a final value.

Table 4- Summaration of datasets

Provider	NO of samples	Wavelengths(nm)	Plastic Types
Armin Straller	70	White,850,940,1200,1300,1450, 1550, 1650	1,2,5,6
Jerry de Vos	60	940,1050,1200,1300,1450,1550 ,1650,1720	1,2,5,6
Fang Ou	80	900-1700(565 wavelengths)	1,2,4,5,6

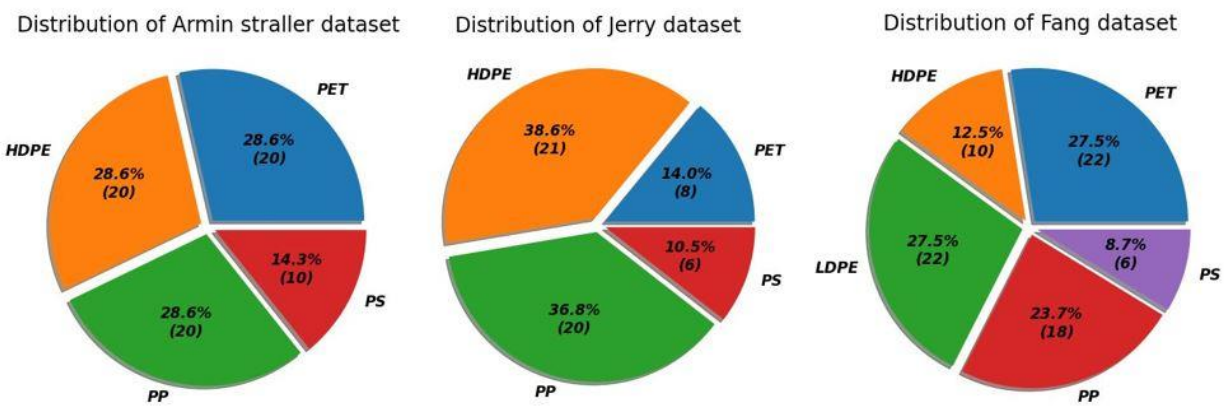

Figure 13- Plastic samples distribution in datasets. from left: 1. Straller 2. Jerry De Voss 3.Fang

Machine Learning Modelling

We have tried a lot of machine learning models to find the best model that can be used to classify plastic types, we also tried all these models on the various datasets we have. The models were trained to tell apart the individual types of plastic using all the measurement samples. The trained model is stored in a Jupyter notebook file and can be compiled as a python script. The models can be fit anytime with newer or larger datasets fed to the machine learning models and predict the outputs and also calculate the accuracy of each model. Within that Jupyter notebook Predictions can then be also visualized using various Data visualization techniques.

We have been switching between various models on top of them Logistic regression, K Nearest Neighbors Classifier, Naive Bayes Classifier, Decision Tree Classifier, Random Forest Classifier, and Support Vector Machine Classifier.

We focused more on Support Vector Machine models as it's the one that was supposed to be used during this project and also because it's the most appropriate approach for this task.

Support Vector Machine

Support Vector Machine (SVM) is one of the most robust and accurate methods in all machine-learning algorithms. SVM Classifiers offer good accuracy and perform faster prediction compared to Naïve Bayes algorithm They also use less memory because they use a subset of training points in the decision phase. SVM works well with a clear margin of separation and with high dimensional space. It primarily includes Support Vector Classification (SVC) and Support Vector Regression (SVR). The SVC is based on the concept of decision boundaries. The basic idea behind the SVM algorithm is to find the best boundary (or "hyperplane") that separates the data points of different classes in a high-dimensional feature space. The boundary is chosen so that it maximally separates the classes while keeping the margin of error as small as possible. A decision boundary separates a set of instances having different class values between two groups. SVM can also be used to handle non-linearly separable data by transforming the input data into a higher-dimensional space where a linear boundary can be applied. So the SVC supports both binary and multi-class classifications.

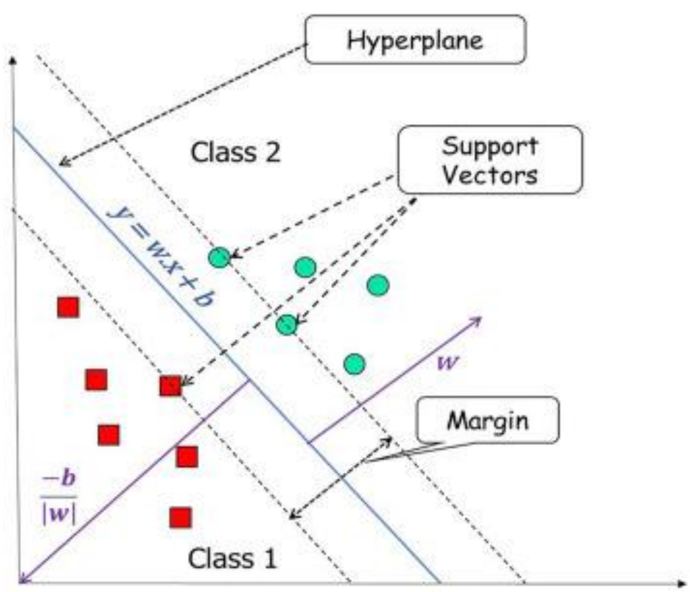

Figure 14- SVM algorithm Illustration

Hyperparameter tuning is an essential part of controlling the behavior of a machine learning model. If we don't correctly tune our hyperparameters, our estimated model parameters produce suboptimal results, as they don't minimize the loss function. This means our model makes more errors. Hyperparameter tuning must be used to avoid over-fitting. Because overfitting may fail to fit additional data, and this may affect the accuracy of predicting future observations. And careful tuning of the regularization parameter, C, and in the case of non-linear SVMs, careful choice of kernel and tuning of the kernel parameters. And in our case we have to tune the SVM parameters to reach the best accuracy, we have done that using:

Kernel: a set of mathematical functions used in SVM providing the window to manipulate the data. And the Kernel function is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

C parameter: tells the SVM optimization how much you want to avoid misclassifying each training example and adds penalty to each misclassified point. If the C value is small, then essentially, the penalty for misclassified points is also small, thus resulting in a larger margin based.

Gamma parameter: It defines how far the influence of a single training example reaches. So, it can be seen as the inverse of the radius of influence of samples selected by the model as support vectors.

Confusion Matrix: The Confusion Matrix is a very useful machine learning method which allows you to measure Recall, Precision, Accuracy, and AUC-ROC curve. As we see above all the predictions are on the diagonal line which means there’s no single plastic type predicted wrong.

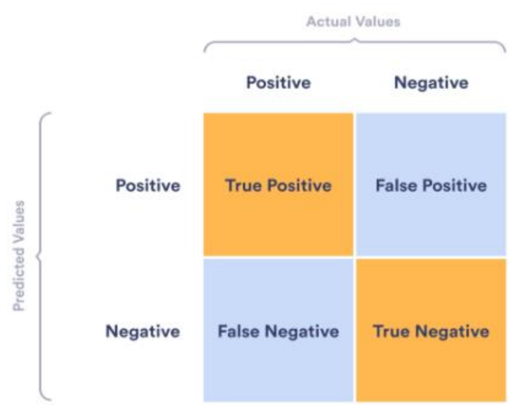

Figure 15-confusion matrix syntax

Data Preprocessing

Data structure preparation

As It has been discussed in literature search, for identifying plastic types picking up 6-8 variant wavelength reflectance in the 900-1700nm is enough. Considering this point and for mixing the datasets in the next steps, the common wavelengths are parsed from each dataset which are 940, 1200,1300,1450,1550,1650nm.

Moreover, for better performance, relative reflectance concept has been employed to add new attributes to each dataset. As a result, 5 new columns are added to the datasets. Final attributes are as below:

Table 5- Attributes and their description

Attribute	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11
Description	940nm reflectance	1200nm reflectance	1300nm reflectance	1450nm reflectance	1550nm reflectance	1650nm reflectance	(w2-w1)/1200-940	(w3-w2)/1300-1200	(w4-w3)/1450-1300	(w5-w4)/1550-1450	(w6-w5)/1650-1550

MinMaxScaler

Each of the providers used different device for reading reflectance data. As a result, their values are in different ranges. For avoiding the errors due to this issue we use MinMaxScaler(). Using the MinMaxScaler(), the reflectance values normalize to values between 0 and 1 for each dataset.

Results

In this section the provided datasets will be studied. Firstly they will be studied separately and then mixing the datasets for providing more reliable model will be discussed. For each dataset multiple SVM models with linear, rbf and polynomial kernels developed using variant C factors and Gamma factors. Defining Lists as below:

c_list=[1,3,10,30,100,300,1000,3000,10000,30000,100000,300000] gamma_list=[0.1,0.3,1,3,10,30,100,300]

models are developed for each possible of C factor and gamma factor combination. Normally for each kernel there are more than one combination which leads to best result. Best result for each kernel type are shown in the table. Also for each dataset, confusion matrix for best result is provided.

1.Straller

Table 6- Table 7- SVM models results based on Straller’s dataset

Model	Accuracy	F1 score
Linear Kernel	100 %	1.0
Rbf Kernel	100 %	1.0
Polynomial Kernel	100 %	1.0

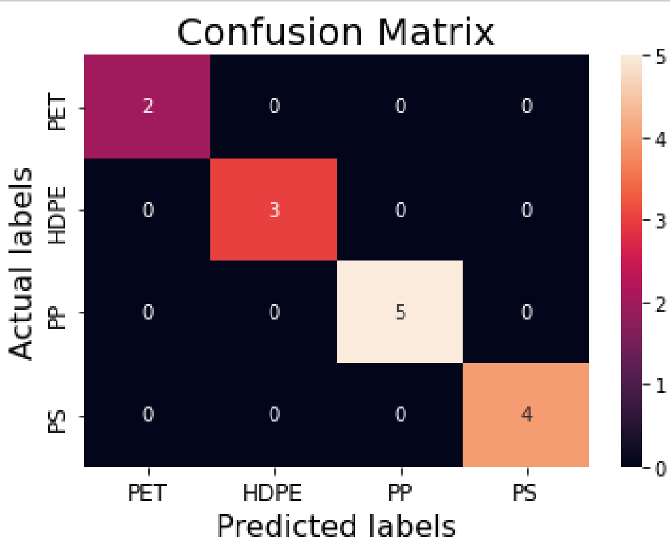

Figure 16- confusion matrix for final model based on Armin Straller’s dataset

2. Fang

Table 8- SVM models results based on Fang's dataset

Model	Accuracy	F1 score
Linear Kernel	100 %	1.0
Rbf Kernel	100 %	1.0
Polynomial Kernel	100 %	1.0

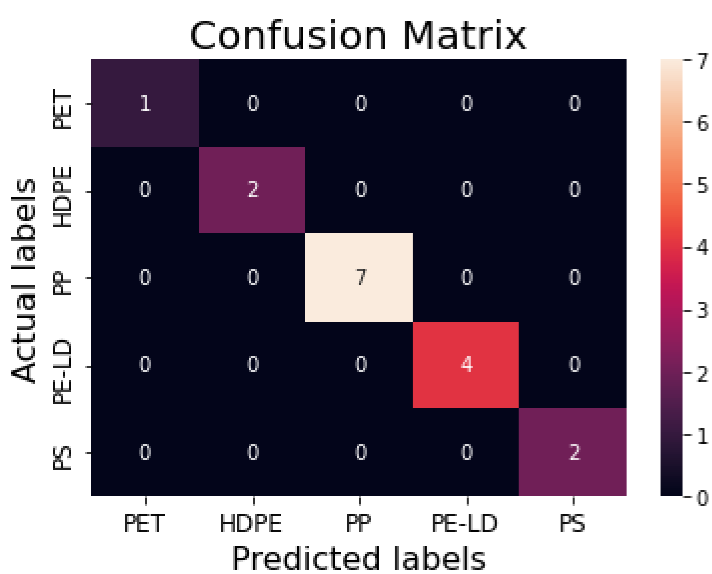

Figure 17- confusion matrix for final model based on Fang's dataset

3.Jerry De Voss

Table 9-SVM models results based on Jerry's dataset

Model	Accuracy	F1 score
Linear Kernel	58.33%	0.49
Rbf Kernel	66.7%	0.58
Polynomial Kernel	58.33%	0.49

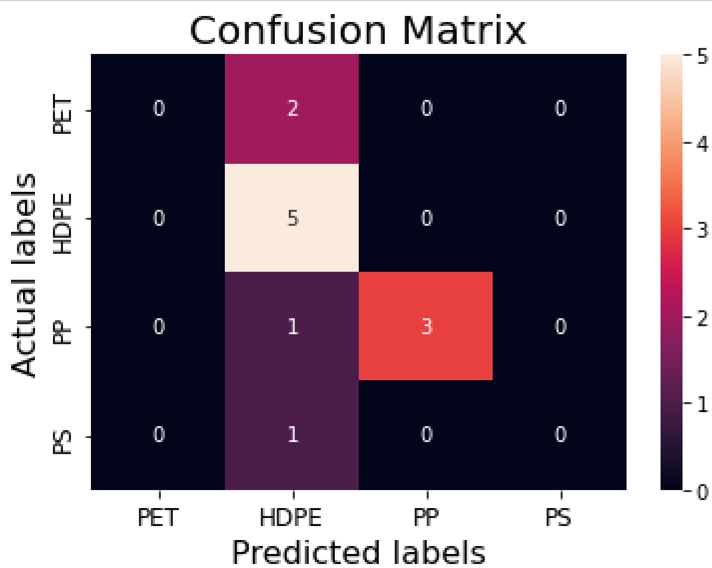

Figure 18-confusion matrix for final model based on Jerry's dataset

4.Mixing dataset

As it is shown, using Straller's and Fang's dataset SVM model provides 100% accuracy whereas this value for Jerry's dataset couldn't raise to more than 66.7%. As a result, for mixing the datasets to train a better model, Straller's and Fang's are mixed. Developing an SVM model using these two datasets will lead to around 97% accuracy where the model is trained using 120 samples as train dataset and it was evaluated using 30 samples as a test dataset. Also this model can recognize plastics of type PET, HDPE, PE-LD, PP and PS. Furthermore, Jerry's dataset is also added to check the results. Not surprisingly, model final performance decreased to around 81% adding Jerry’s dataset.

Table 10- SVM models result based on the mixed datasets

	Kernel type		
Datasets	Linear	rbf	Polynomial
Straller+fang	96.7%	96.7%	96.7%
Straller+fang+Jerry03	81.0%	81.0%	81.0%

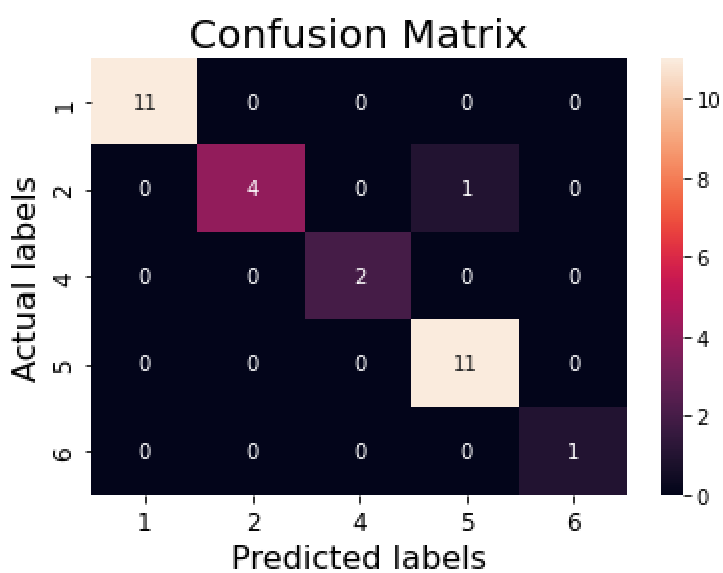

Figure 19-confusion matrix for final model based on Straller+Fang dataset

Risk assessment

Finding an available NIR Sensor with low cost in the domestic market: Starting for implementing the low cost NIR spectroscopy syste, we found them expensive and, in some cases, not capable in domestic market. As it discussed before in sensor selection section, we shortlisted to 4 InGaAs sensors and successfully provide a one suitable sensor in our desire spectra in 900-1700nm.

Electronic design error: Mistakes and errors in design step are hard to avoid and if something is not considered in this step it would lead to redesigning. Hopefullywe find some electrical designs for the same purpose that has made satisfactory results. So we modified their electrical design based on available items in market which ensures our design to perform correctly. We also decided to implement the breadboard type first to test the electrical configuration before applying it on the printed circuit board (PCB).

Collecting enough plastic samples and making a balanced collection: As more samples you can feed to your model in training procedure, the better will be your model. So we aimed to collect lots of plastic samples from our normal daily life environment. Our final outcome is satisfactory in the terms of quantity but its not a balanced collection due to limited sources. For further steps, it would be better to focus on other sources for gathering plastic samples to make the sample collection more balanced.

Identifying solid plastics with different colors (especially dark colors): As it has been discussed in literature review, using NIR spectroscopy dark plastics could not be classified whereas other colors affect the reflectance but it’s still possible to detect their type.

Justifying object distance from the sensor: Different distance from spectroscopy system can affect the result greatly. So, finding a solution to keep this distance unique is important to consider. In our system, we are considering keeping this distance unique manually by putting the plastic on our scanner manually. Although, in a case of doing this procedure automatically this point should be considered.

Fix mechanical topology: For getting consistent result for same plastic samples, not only the distance between plastic and scanner should be constant but also other distances and configuration should be same during the scans over time. For this, we find the configuration proposed by Armin Straller in Reremeter a good approach.

Intensive Timeline: This project firstly aimed to developing a SVM model using available datasets in internet, then we find a low-cost method to implement our own spectrometer which made our timeline too full. Finding datasets from sources on the web and studying the data and building a SVM model based on them was already lots to do in this case study which been summed up with designing a electrical design based on available items on domestic market.

Timeline

Below the timeline for our project is provided. All the steps went almost on its time except for electrical implementation. Unfortunately, due to intensive timeline we couldnt finish this part on this project although we made the electrical design, and we bought the electrical items.

WBS	Task	Start	End	Days	% Done	Work Days
1	**Basic Design**	Wed 10/12/22	Tue 11/15/22	35	100%	25
1.1	Literature search	Wed 10/12/22	Tue 11/15/22	35	100%	25
1.2	Electronic design	Wed 10/19/22	Tue 11/15/22	28	100%	20
1.3	Item Specification	Wed 10/19/22	Tue 11/15/22	28	100%	20
2	**Data**	Tue 11/08/22	Sun 12/04/22	27	100%	
2.1	search	Tue 11/08/22	Mon 11/21/22	14	100%	10
2.2	mix	Mon 11/21/22	Sun 12/04/22	14	100%	10
3	**Procurement**	Tue 11/29/22	Sun 12/18/22	20	100%	14
3.1	Electronics	Tue 11/29/22	Sun 12/18/22	20	100%	14
3.2	Plastic samples	Wed 10/12/22	Mon 1/09/23	90	100%	64
4	**Installation& commissioning**	Sun 12/18/22	Sat 1/14/23	28	25%	20
4.1	Breadboard Installation	Sun 12/18/22	Sat 12/24/22	7	50%	5
4.2	programming & commissioning	Sun 1/08/23	Sat 1/14/23	7	0%	5
5	**Model Developement**	Sun 12/04/22	Mon 1/02/23	30	100%	21
5.1	Developing SVM model	Sun 12/04/22	Mon 1/02/23	30	100%	21
6	**Finalising**	Sat 1/14/23	Fri 1/20/23	7	100%	5
6.1	Report	Sat 1/14/23	Fri 1/20/23	7	100%	5

Figure 20- Project Timeline (For fitting the timeline in page, first 5 weeks are hided)

Conclusion

Through this project its been shown that using SVM approach a ML model can be developed for classifying plastic types by using NIRS data. For applying this, its not necessary to take all the spectrum reflectance data and extracting 7 reasonable wavelengths is enough. Although for providing better result relative reflectance concept has been applied in which along side our normal reflectance readout, their proportional value would be considered. We have seen for 2 out of 3 different available datasets we could reach 100% accuracy on test dataset, where for the other dataset the results couldn’t be more than 75% using different C, gamma values and kernels. Also for providing better and more reliable model, we considered mixing datasets. Not surprisingly, the best result is for mixing Armin Straller’s dataset with Fang’s with near 97% accuracy. Adding the data from Jerry de Vos’s dataset decreases the model performance to 85% accuracy. For further steps, developing models based on other machine learning methods like decision tree or logistic regression can be considered.

Although, at the first it was assumed that implementing a NIR spectrometer for classifying plastic types is expensive, through this project it has been shown that some low-cost solutions are available for set-up. This method which uses 6 to 8 individual reflectors in 900-1700nm wavelength, has been executed before with satisfactory results and we consider it for our lab prototype implementation. An InGaAS photodiode has been used for reading reflectance as it provides a high responsivity in range of 900-1700nm. Through this project, We developed a design for first version lab prototype implementation. After finalizing this prototype, next steps could be providing a PCB based spectrometer based on lab prototype and finally performing a automatic sample scan.

References

[1] Zhu, Shichao, Honghui Chen, Mengmeng Wang, Xuemei Guo, Yu Lei, and Gang Jin. "Plastic solid waste identification system based on near infrared spectroscopy in combination with support vector machine." Advanced Industrial and Engineering Polymer Research 2, no. 2 (2019): 77-81.

[2] Masoumi, Hamed, Seyed Mohsen Safavi, and Zahra Khani. "Identification and classification of plastic resins using near infrared reflectance." Int. J. Mech. Ind. Eng 6 (2012): 213-220.

[3] de Vos, Jerry , Diehl, J.C. , van Engelen, J.M.L., “Plastic Identification Anywhere: Development of open-source tools to simplify plastic sorting”, Delft University of Technology, master thesis, 2021-02-25.

[4] Straller, Armin, and Bernhard Gessler. "Identification of Plastic Types Using Discrete Near Infrared Reflectance Spectroscopy."

[5] Ou, Fang, Anne van Klinken, Petar Ševo, Maurangelo Petruzzella, Chenhui Li, Don MJ van Elst, Kaylee D. Hakkel, Francesco Pagliano, Rene PJ van Veldhoven, and Andrea Fiore. "Handheld NIR Spectral Sensor Module Based on a Fully Integrated Detector Array." Sensors 22, no. 18 (2022): 7027.

[6] https://docs.plasticscanner.com/boards/DB2.1_build

[7] https://media.digikey.com/pdf/Data%20Sheets/Phoenix%20Contact%20PDFs/0090-3111-185.pdf

[8] https://omronfs.omron.com/en_US/ecb/products/pdf/en-b3f.pdf

[9] https://ww1.microchip.com/downloads/en/DeviceDoc/20001811F.pdf

[10] https://www.analog.com/media/en/technical-documentation/data-sheets/AD7124-4.pdf

[11] https://www.diodes.com/assets/Datasheets/AS431.pdf

[12] https://www.diodes.com/assets/Datasheets/ZXRE330.pdf

[13] https://ecsxtal.com/store/pdf/ecs_100.pdf

[14] https://www.specim.com/