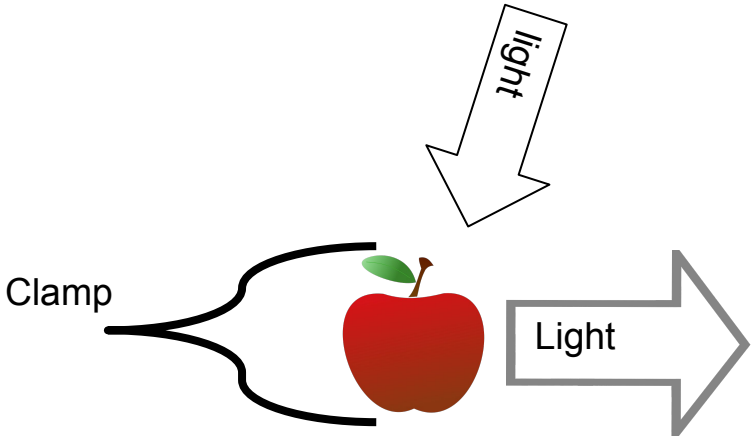
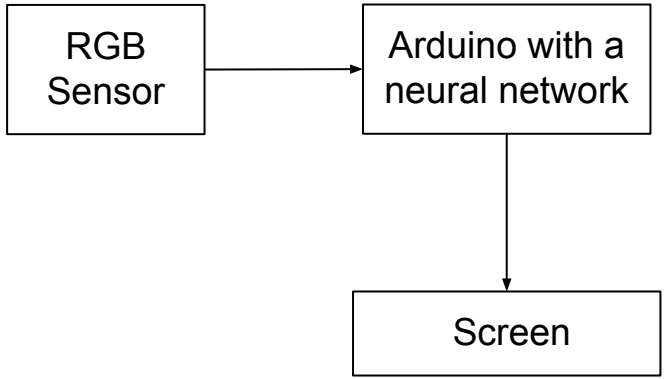
Identifying Fruits Using Data Science & Machine Learning

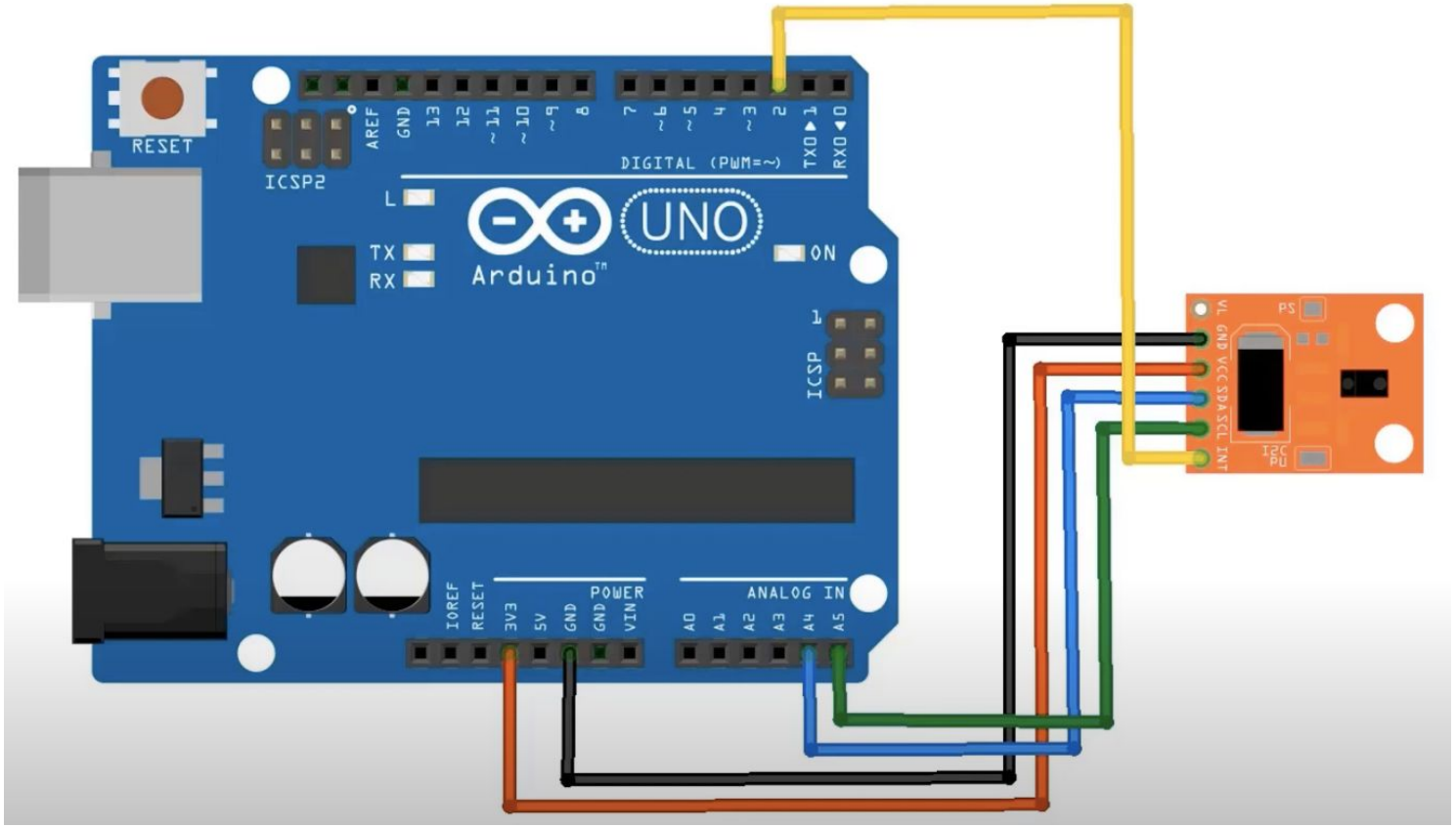
Ryan Z. Nie and Vera Degtiareva

Abstract

We propose an object recognition sorting device that will use machine learning to recognize fruits. We will use an APDS9960 RGB sensor to determine the color of a fruit, process the RGB data through a neural network and KNN and output the probability that the fruit is of a certain type according to each algorithm. Neural network training and other machine learning algorithms will be done on Python and deployed onto the Arduino.


light
Clamp
Light


RGB Sensor
Arduino with a neural network
Screen


RESET
ICSP2
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM=~)
L
TX
RX
UNO
Arduino™
ON
ICSP
IOREF
RESET
3V3
5V
GND
GND
VIN
POWER
A0
A1
A2
A3
A4
A5
ANALOG IN
VL
GND
VCC
SDA
SCL
INT
PS
I2C PU

List of materials

- APDS9960 sensor
- Arduino Mega
- Clamp
- Lamp
- Breadboard/wires
- Apples, oranges, lemons

Update April 11, 2022

- Added components to apparatus
- Collected some test RGB data to play around with
  - Figured out how to collect and preprocess data on Python
- Implemented few algorithms (KNN and DNN)
  - K-Nearest Neighbors achieved an average of 85 percent accuracy
  - Neural Network achieved a best of 95 percent accuracy

Plans for next week?

- Update apparatus to ensure good data collection
- Collecting actual data (fruits RGB)
- Cross validation on ML algorithms

Update April 20, 2022

- Collected data for orange, apple, lemon
- Preprocessed data on Python
- Created DNN architecture and cross validated to find optimal hyperparameters (~97% accuracy)
- KNN is working on the Arduino Mega - Arduino had memory problems

Plans for next week?

- Converting code to Arduino using TinyML and testing out models
- Writing code to print soft classification probabilities

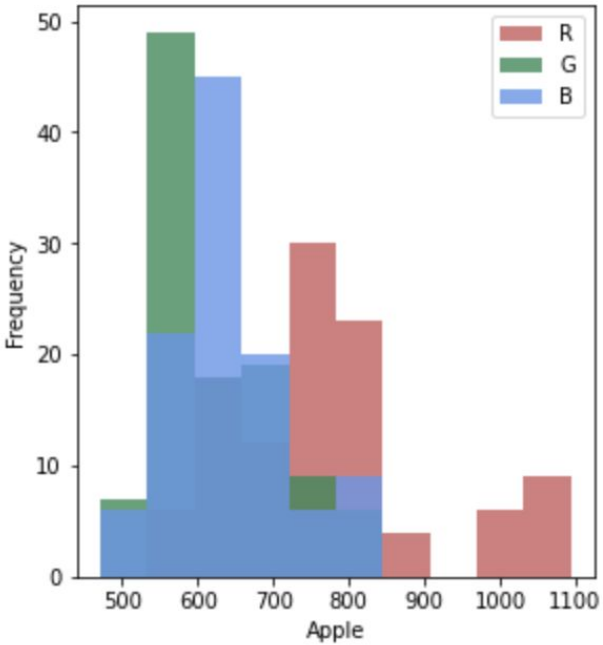
Update April 25, 2022

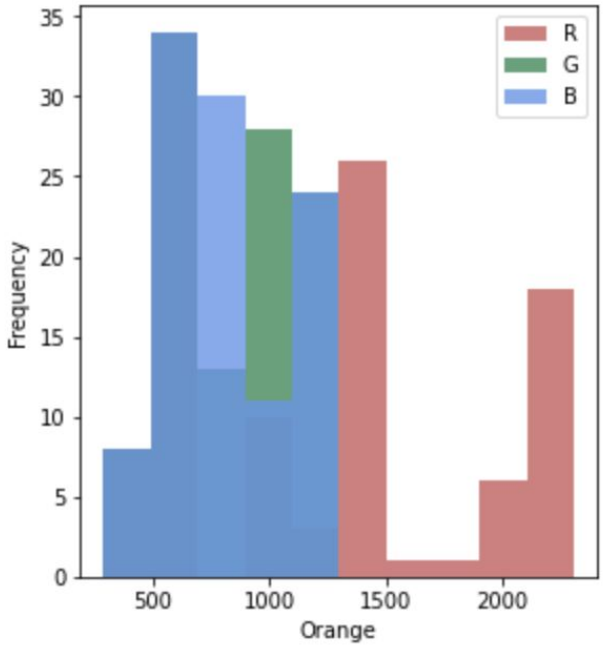
- Collected more data (train on more variety?)
- Generated graphs to describe data
- Created confusion matrices for KNN and neural network

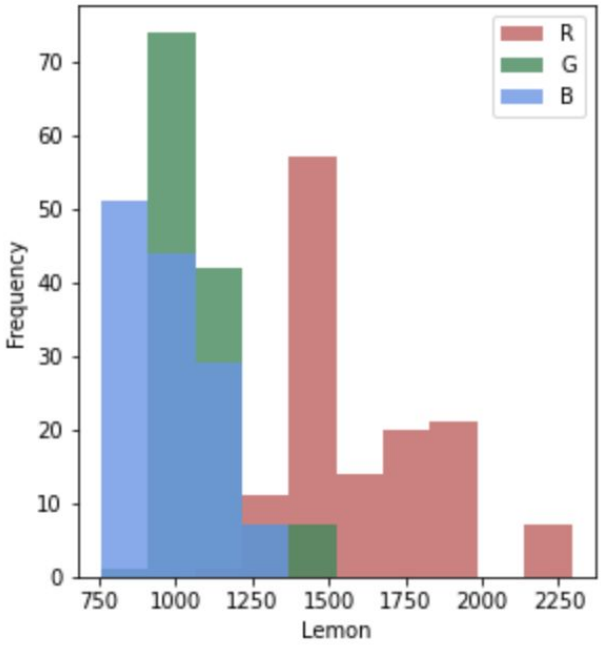
Plans for next week?

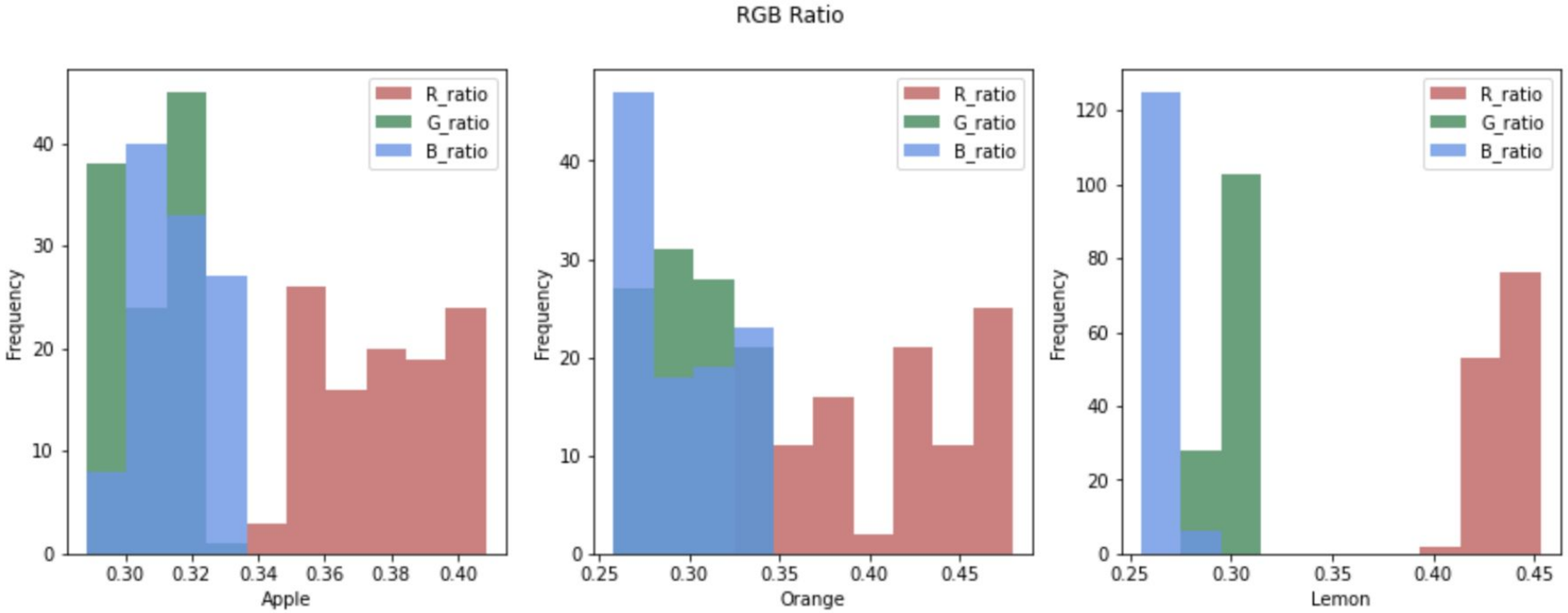
- Figure out how to send data from Arduino to two computer simultaneously
- Collecting more data
- Finishing touches and preparing for open house

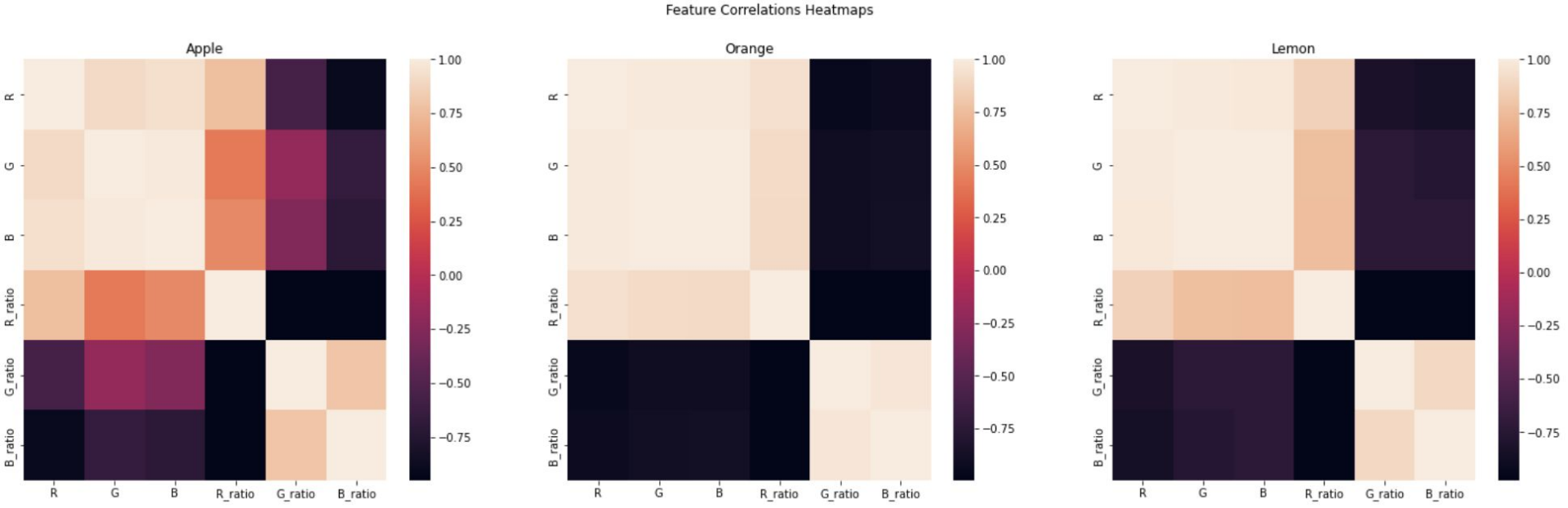
RGB

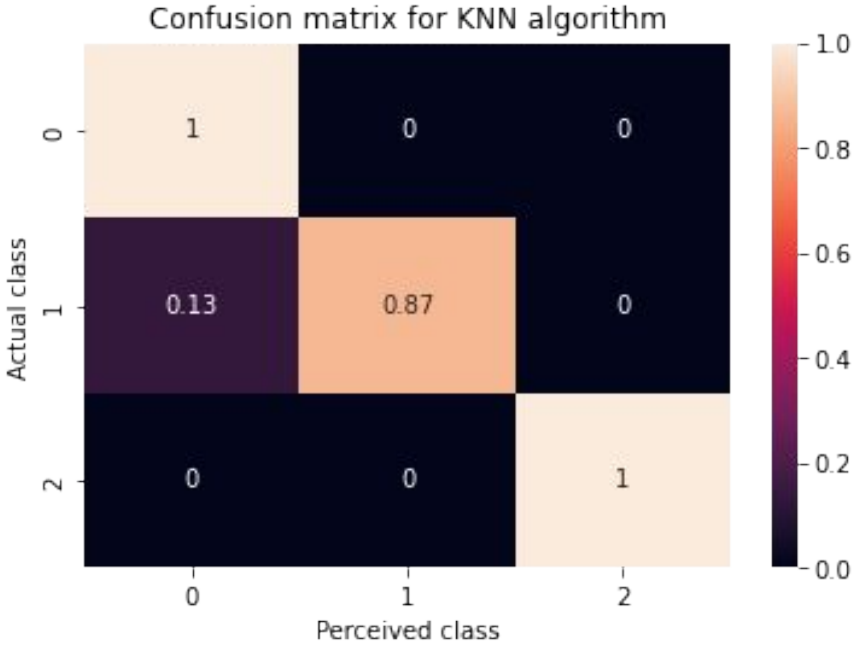

R
G
B
Frequency
Apple
0
10
20
30
40
50
500
600
700
800
900
1000
1100


R
G
B
Frequency
Orange
0
5
10
15
20
25
30
35
500
1000
1500
2000


R
G
B
Frequency
Lemon
0
10
20
30
40
50
60
70
750
1000
1250
1500
1750
2000
2250


RGB Ratio
R_ratio
G_ratio
B_ratio
Frequency
Apple
Orange
Lemon


Feature Correlations Heatmaps
Apple
Orange
Lemon
R
G
B
R_ratio
G_ratio
B_ratio
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75



0 - Apples, 1 - Oranges, 2 - Lemons