Indian Institute of Technology Kanpur

Perception & Intelligence Lab

SURGE INTERNSHIP REPORT

"Machine Learning-Based Fruit Ripening Prediction: Assessing Ripeness Levels Over Time"

Student Name	**Student ID**
Ujjwal Bisaria	221154

Guide:

Dr. Tushar Sandhan

Submission Date : 20/06/2024

1 Introduction

This project aims to develop a machine learning method based on a classifier and regression model to predict fruit ripeness and sweetness using a non-destructive method. This will be very beneficial for the food supply chain, as well as for harvesting fruits at the optimal time.

In this experiment, we are dealing with images, including hyperspectral images with different wavelengths and the reflectance of the fruit, meaning how much light is reflected at different stages of the fruit. This helps us understand the changes that take place when a fruit goes from unripe to ripe to rotten. Using the images from a previous experiment on the same topic, we will carry out a simple technique of pixel difference, where the image pixel with the maximum condition will be subtracted to obtain more information about the interior differences between the two images. Phase 1 included work with reflectance, and Phase 2 included work with hyperspectral images.

Hyperspectral images are a type of imaging that captures light across many narrow, contiguous spectral bands, providing detailed information beyond the visible spectrum. Unlike traditional color images with only three bands (red, green, and blue), hyperspectral images can have hundreds of bands, allowing for precise identification and analysis of materials based on their unique spectral signatures. This advanced imaging technology is widely used in agriculture, environmental monitoring, mining, medicine, defense, and food quality control, offering non-invasive, detailed analysis. Despite its benefits, hyperspectral imaging generates large data volumes and requires significant processing capabilities.

In Phase 1 of the experiment, we collected the reflectance data of the fruit using a previously built machine with 3 different categories of wavelength(410-535 nm, 560-705 nm, 730-940 nm). This was done over 10 days, during which 4 different sensors captured the readings. We then distributed the data into 3 categories (ripe, overripe, rotten). After collecting the data, we needed to analyze it and draw possible conclusions. For that purpose, I built a KNN classifier model and plotted the raw data in Excel using a scatter plot after normalizing the data.

In Phase 2 of the experiment, we will be collecting data from the hyperspectral camera and then I will build a custom classifier model to determine the possible outcomes for the specified labels.

Lastly, we will try to combine the probabilities of both models and then, with the help of my custom regression model, I will assign weightage to each of them. This process will provide us with the final output.

2 Objectives and Learning Outcomes

1. To optimize the supply chain using this method and minimize the loss of food during storage, we will involve extensive work on both data and models. Therefore, we will try to:

   (a) Collect accurate and diverse data, including both reflectance and hyperspectral images, to develop a robust model with good variability.

   (b) To study how the inner chemical composition of the fruit changes over time as it goes through different phases of ripening.

2. To build a robust model aligned with our project's goals and identify the crucial parameters and features that affect the overall ripeness of the fruit.

3 Equipments

- Multispectral Camera, Different varieties of fruit, Reflectance measuring device.

4 Methodology

4.1 Structural change in Fruit

4.1.1 Part 1

1. Inital I studied some already published research paper related to my project theme and jot down the important points to be taken in consideration while analysing photos.

2. For this we took the images hyperspectral data that was captured earlier by other group of people

4.1.2 Part 2

1. Firstly, convert the images to greyscale and then resize the images such that the orientation of both the images are same.

2. To obtain this, I simply computed the absolute differences between two images and then applied a binary threshold between them.

4.2 Change in Reflectance of fruit

1. I collected the data of 2 different fruits(apple and oranges) for continues of 10 days with the help of the apparatus provided

2. Then I built a KNN classifier with the help of data using the PCA technique for reducing the dimension in the input.

3. For more clarity, I also plotted the data on Excel and manually tried to get the possible observations from it.

4.3 Dry matter in Leek

1. We collected the data from the research done by other people on the fruit leek which comprised the values from wavelength 400-1700 nm and also consisted of dry matter content in the fruit over time
2. Then we pre-processed the data using techniques like MinMaxScaler, removing null values analyzing, and plotting heat-map to identify the wavelength highest affected by fruit content
3. Then we assigned the label for each day according to the dry matter content present in it while researching the limit for each label.
4. Finally after doing this, I built a custom LSTM model over this data, where we tried to add predict the label assigned over previous sessions

5 Result

5.1 Pixel based difference in the images

5.1.1 Model Result

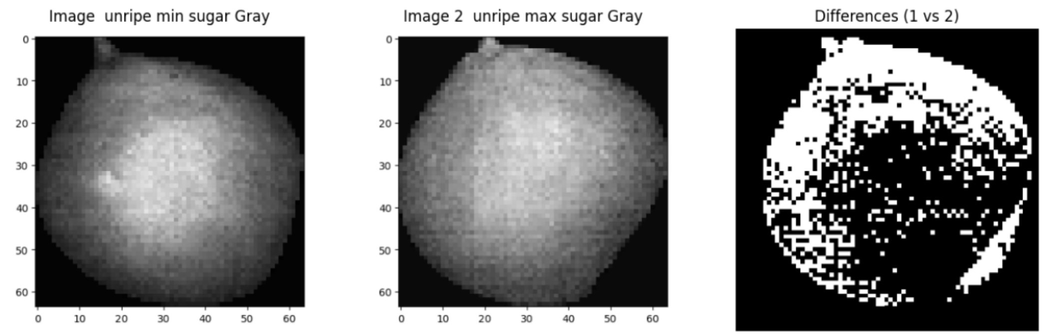


Figure 1

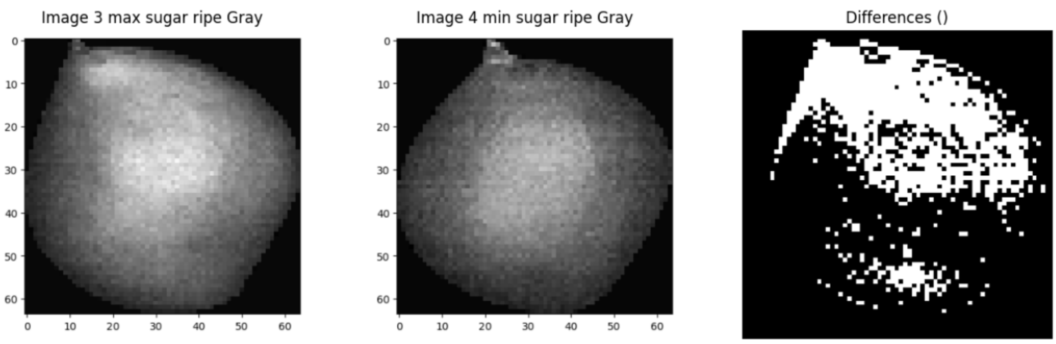


Figure 2

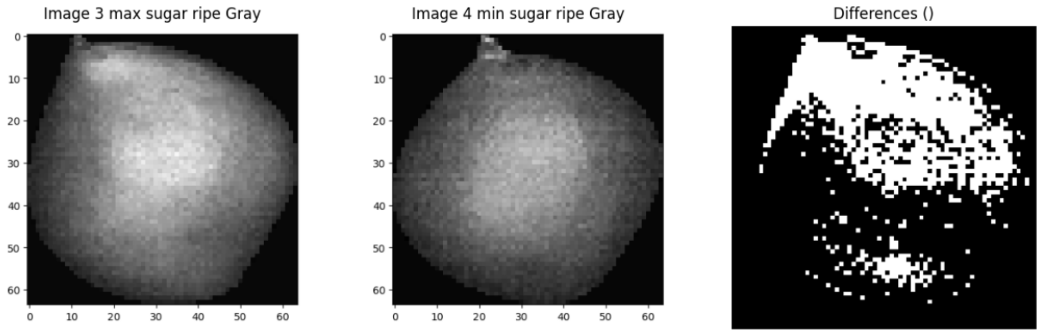


Figure 3: Pixel based differnces

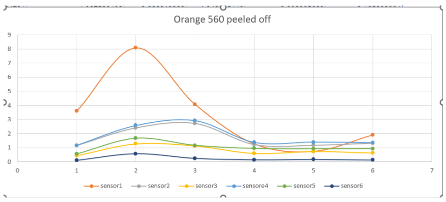
5.2 Reflectance Measurement

5.2.1 Experimental Result

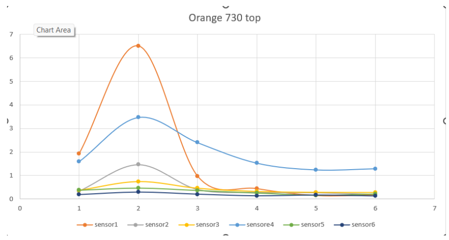
Sensor1	Sensor2	Sensor3	Sensor4	Sensor5	Sensor6	Result
0.724872588	0.303457378	0.245909738	0.368294876	0.274907073	0.279563854	ripe
1.897234523	0.895948717	0.357589143	1.45489961	0.366944139	0.15433572	ripe
1.678001504	0.249911621	0.255121909	1.129008091	0.35307052	0.175394263	ripe
0.411285631	0.310384445	0.281986272	0.258495498	0.304898243	0.333890118	ripe
0.726468877	0.744586832	0.577515305	1.08219576	0.652475829	0.190420297	ripe
0.797440372	0.390543526	0.488095974	2.386000581	0.379413493	0.15231818	ripe
0.7	0.40590743	0.386420683	0.315186673	0.427945709	0.47919872	ripe
3.533195021	1.162809582	1.092745984	1.538301354	0.931217217	0.258477445	ripe
4.695435685	0.685038571	0.785893574	3.689126878	0.616882064	0.233552834	ripe
0.095266322	0.129754265	0.045445497	0.06071715	0.080888457	0.132044433	ripe
0.218477433	0.353811238	0.862383756	0.407406347	1.201127781	0.13567057	ripe
1.808620544	1.062607607	2.44848278	12.94778898	3.431938286	1.160363675	ripe
0.652193523	0.425833994	0.367140719	0.334109015	0.328218448	0.351282755	over-ripe
1.36852392	1.016849663	0.493426191	1.569031128	0.561378623	0.183512828	over-ripe
0.948792735	0.327452661	0.357768159	1.598598297	0.365137632	0.202496913	over-ripe
0.420135106	0.377329293	0.374391935	0.35147413	0.380025829	0.442179084	over-ripe
0.157441708	0.251724805	0.149514808	0.506600811	0.232716315	0.066514839	over-ripe
0.173676182	0.186472371	0.381650914	2.118601177	0.339733104	0.162672161	over-ripe
0.271539516	0.231185557	0.214462916	0.227513601	0.25002527	0.253907441	over-ripe
0.476199178	0.50732825	0.335224843	0.816929273	0.435206712	0.130262528	over-ripe
0.35678392	0.234899064	0.277880749	1.419628347	0.213939149	0.106055989	over-ripe
0.384611702	0.294933255	0.261231107	0.223246962	0.248757845	0.295305687	rotten
0.264387628	0.511468397	0.538674756	0.945395989	0.703669805	0.173748089	rotten

Figure 4: Data for KNN model as well as for excel plot

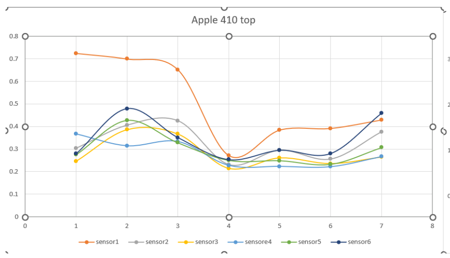
5.2.2 Excel Plots



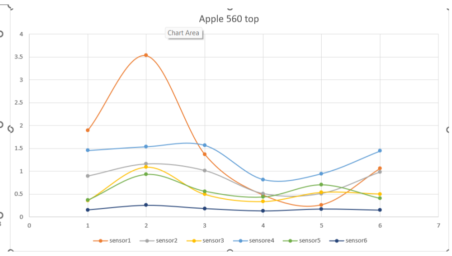
(a) orange at wavelength 560 nm



(b) orange at wavelength 730 nm



(c) apple at wavelength 410 nm



(d) apple at wavelength 560 nm

Figure 5: Excel plot for above data for various sensor on different days marked as points

5.2.3 KNN Model Plots

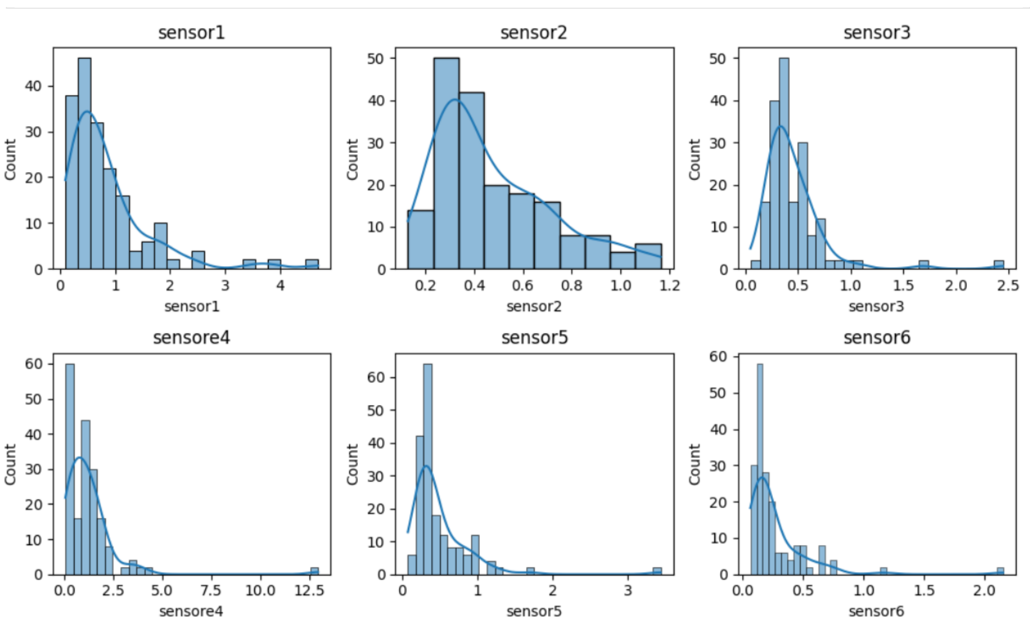


Figure 6: Difference in value with different sensors

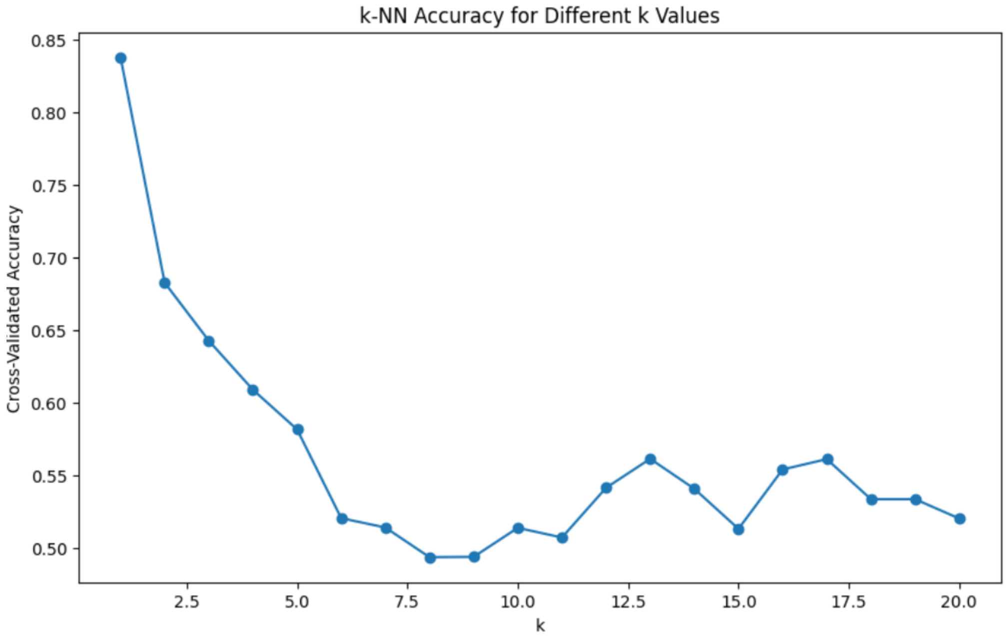


Figure 7: KNN accuracy for different K values

5.2.4 LSTM Model plots

Leek	DRY MATTER	397.66	400.28	402.9	405.52	408.13	410.75	413.37
L1	0.168033	0.415485	0.383244	0.351456	0.326875	0.308858	0.296541	0.288141
L2	0.171717	0.417281	0.383864	0.350971	0.326197	0.307104	0.294010	0.285275
L3	0.145000	0.424143	0.392674	0.361699	0.337758	0.319835	0.308165	0.300610
L4	0.160656	0.421169	0.392193	0.362329	0.338916	0.320957	0.310502	0.302020
L5	0.147436	0.424819	0.392754	0.360710	0.335537	0.317260	0.304983	0.296590

Figure 8

1691.59	1695.17	1698.75	1702.33	1705.91	1709.49	1713.07	1716.65	Label
0.220452	0.223861	0.228166	0.233664	0.240525	0.248254	0.256775	0.263427	Rotten
0.214919	0.218351	0.222681	0.228466	0.235393	0.243194	0.251689	0.258660	Rotten
0.202916	0.206433	0.210867	0.216627	0.223640	0.231442	0.240114	0.247030	Overripe
0.198973	0.202568	0.207206	0.212854	0.219782	0.227738	0.236556	0.243064	Rotten
0.219864	0.223265	0.227586	0.233124	0.239957	0.247428	0.256028	0.262627	Overripe

Figure 9: Dry Matter content and 400-1700 nm wavelength data with labels

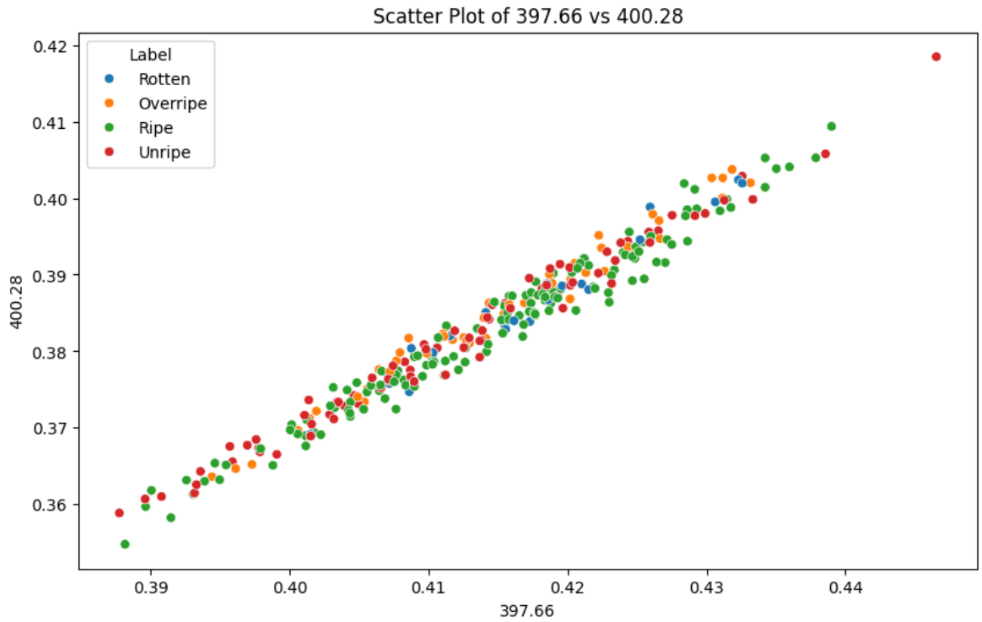


Figure 10: Wavelength dependency on each other

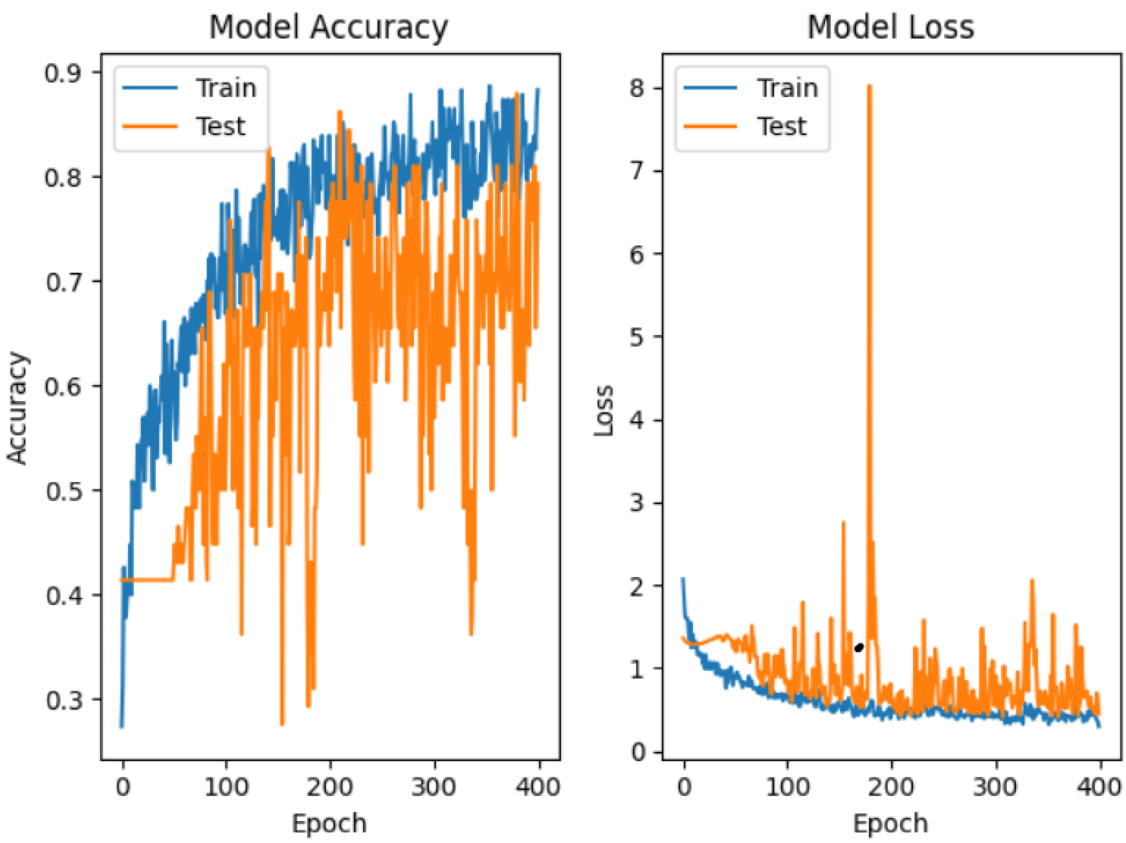


Figure 11: Loss Vs Epoch

6 Discussion

6.1 Part1

The hyperspectral imaging analysis reveals significant insights into the internal structural changes of fruit at different ripeness stages. The pixel brightness differences between the images of unripe, maximum sugar content, and ripe fruit highlight the varying internal transformations. These changes are particularly pronounced as the fruit transitions from unripe to high sugar content, indicating substantial structural evolution during this phase. The differences become less pronounced once the fruit reaches ripeness, suggesting a stabilization of internal changes. This analysis demonstrates the efficacy of hyperspectral imaging in detecting subtle internal variations, making it a valuable tool for monitoring fruit ripeness and quality control in the food industry, ultimately helping to optimize harvesting times and reduce waste.

The hyperspectral imaging analysis, while insightful in revealing internal structural changes in fruit at different ripeness stages, has its limitations. A slight change in the orientation of the images can significantly affect the results, making it challenging to distinguish whether the observed differences are due to chemical composition changes or other structural alterations not relevant to the current study. Additionally, numerous factors can contribute to variations in the images, casting doubt on the reliability of this method alone for our model. Consequently, we transitioned to using reflectance data measurements, which provided more consistent and reliable insights into the chemical composition changes in the fruit, thus offering a more robust foundation for monitoring ripeness and quality control.

6.2 Part2

The Excel plots show six different lines for each sensor, with dots marking the data points over ten days. These plots help us see how the reflectance changes as the fruit ripens and eventually rots. When the fruit is ripe, the reflectance values are stable, indicating a consistent surface. As the fruit becomes over-ripe, the reflectance values vary more, suggesting changes in texture and moisture. Finally, in the rotten stage, the reflectance values are very erratic, showing that the fruit's structure is breaking down. These patterns indicate that the sensors can effectively track the ripeness of the fruit, making it possible to develop automated systems to detect when fruit is ripe, over-ripe, or rotten.

The KNN model, applied to the reflectance data from six sensors, demonstrates the potential for accurately classifying fruit ripeness stages. The distribution plots of sensor values show distinct patterns, highlighting the variability and uniqueness of the data captured by each sensor. This variability is crucial for the KNN model, which leverages these differences to differentiate between ripe, over-ripe, and rotten stages. The model's accuracy, as depicted in the k-value plot, is highest with smaller k values, indicating that fewer neighbors yield better classification performance due to the distinctiveness of the sensor readings. Overall, the successful implementation of the KNN model on this dataset underscores its effectiveness in developing automated systems for monitoring and classifying fruit ripeness based on reflectance measurements.

6.3 Part3

The data preprocessing steps involve transforming image data into feature matrices suitable for the LSTM model. Initial data visualization indicated that multiple parameters influence the sugar content in the fruit, making it evident that a single parameter would not suffice for building a comprehensive model.

The model achieved an accuracy of around 80% in predicting the ripeness stages of the fruit. However, during the training process, large spikes in the data indicated a high dependency of each wavelength on others. This was further supported by the heat map, which failed to provide a clear visualization of the dependencies. The model was trained for approximately 300 epochs, and the loss vs. epoch plot suggested that the model's performance could be improved by considering the interdependencies of various wavelengths and possibly exploring different modeling approaches.

7 Conclusion

Hyperspectral imaging analysis reveals significant internal structural changes in fruit at different ripeness stages, particularly during the transition from unripe to high sugar content. However, its sensitivity to image orientation and other factors limits its reliability. Reflectance data measurements offer more consistent and robust insights, effectively tracking ripeness and supporting quality control. The KNN model applied to this data demonstrates high accuracy in classifying ripeness stages, leveraging the variability captured by the sensors. This underscores the potential of using reflectance data and machine learning models for developing automated ripeness detection systems, optimizing harvest times, and reducing waste in the food industry.
The LSTM model achieved around 80% accuracy in predicting fruit ripeness but exhibited high dependency among wavelengths, as indicated by training spikes. Future improvements should focus on refining the model by exploring additional features and optimizing parameters to enhance accuracy and reliability in monitoring fruit ripeness.

8 References

[1] Github link: https://github.com/Ujjwalb2/Surge_files.

[2] Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning: https://arxiv.org/pdf/2104.09808.

[3] Sugariness prediction of Syzygium samarangense using convolutional learning of hyperspectral images: https://www.nature.com/articles/s41598-022-06679-6?fromPaywallRec=false.

[4] How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method: https://pubmed.ncbi.nlm.nih.gov/27719929/