Bachelor's project

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Vebjørn Johansen Nilsen, Jon Elias Moen,
Katherine Asha Brox-Saidi and Eivind Kohmann

Non-destructive determination of kiwi quality parameters using hyperspectral imaging

Bachelor's project in Dataingeniør
Supervisor: Sony George
May 2021

NTNU
Kunnskap for en bedre verden

Vebjørn Johansen Nilsen, Jon Elias Moen, Katherine Asha Brox-Saidi and Eivind Kohmann

Non-destructive determination of kiwi quality parameters using hyperspectral imaging

Bachelor's project in Dataingeniør
Supervisor: Sony George
May 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

Food waste is a significant problem globally, and it is important to reduce it to obtain a sustainable future. Quality analysis of kiwis is mainly performed by destructive methods, which contributes to food waste. Destructive methods are also time-consuming as kiwis must be destroyed to be physically measured.

A relatively new, non-destructive, efficient, and sustainable way of performing quality analysis of fruits is hyperspectral imaging (HSI). Our group reviewed the current literature on the determination of kiwi ripeness and decided to contribute with state-of-the-art research.

In this bachelor's thesis, a high-quality dataset was collected by capturing hyperspectral (HS) images and performing physical measurements such as firmness, sugar content, pH, salt, size, and weight of kiwifruits. This dataset was pre-processed and used to develop regression and machine learning models for determining the sugar and firmness of "Hayward" kiwi fruits. The best model at determining sugar and firmness was UVE-PLS, which performed moderately (RMSE= 2.804 N; $R^2 = 0.434$) at predicting firmness and good to excellent (RMSE= 0.777°Brix; $R^2 = 0.759$) at predicting sugar content.

Sammendrag

Matsvinn er et stort problem i verden, og det er viktig å redusere det for å oppnå en bærekraftig fremtid. Kvalitetsanalyse av kiwier utføres for det meste ved destruktive metoder, noe som bidrar til matsvinn. Destruktive metoder er også tidkrevende ettersom kiwier må ødelegges for å bli målt fysisk.

En relativt ny, ikke-destruktiv, effektiv og bærekraftig måte å utføre kvalitetsanalyse av frukt på, er hyperspectral avbilding. Gruppen vår gjorde en gjennomgang av dagens vitenskapelige artikler om vurdering av kiwis modenhet og tenkte at vi kunne bidra med vår forskning. I denne bacheloroppgaven ble et datasett av høy kvalitet samlet ved å fange hyperspectrale bilder og utføre fysiske målinger som fasthet, sukkerinnhold, surhet, salt, størrelse og vekt på kiwi. Dette datasettet ble forhåndsbehandlet og brukt til å utvikle regresjons- og maskinlæringsmodeller for å bestemme sukker og fasthet av "Hayward" kiwifrukter. Den beste modellen for å bestemme sukker og fasthet var UVE-PLS, som presterte moderat (RMSE = 2,804 N; R^2 = 0,434) på å forutsi fasthet og bra (RMSE = 0,777°Brix; R^2 = 0,759) på å forutsi sukkerinnhold.

Contents

Figures

Tables

Abbreviations

ANN Artificial neural network. ix, 39, 40, 88, 98

BAMA BAMA Gruppen AS- ledende selskap i Norge innen fersk frukt, bær, grønnsaker, salater og poteter. 54

DM Dry Matter Content. 5, 47, 102

EW efficient wavelength. viii, xiii, 29, 92–94, 96, 97, 99, 101, 102

FAO Food and Agriculture Organization of the United Nations. 7

GA Genetic algorithm. 31, 32, 86, 87, 94, 99

HS hyperspectral. iii, xi, xii, 2, 5, 6, 8, 15, 19, 20, 22, 25, 56, 58, 63, 65, 66, 73, 101, 102

HSI hyperspectral imaging. iii, xv, xix, 1–6, 11, 13, 15, 16, 19, 27–29, 40, 53–56, 62, 73, 77, 101, 102

KNN K-nearest neighbor. ix, 38, 88, 93, 95, 96, 98

KPCA Kernel principal component analysis. 87, 88, 92, 93, 95, 96, 98, 100

KS Kennard-Stone. 32

MLP Multilayered Perceptron. 39, 40, 98

MLR Multiple linear regression. viii, ix, 33, 85, 94, 95, 99, 101

MSC Multiplicative scatter correction. 28, 99, 100

MSE Mean squared error. 36, 41, 65, 82, 88, 98

NIPALS Nonlinear Iterative Partial Least Squares. 36, 38

NIR Near-infrared. 5

NTNU Norwegian University of Science and Technology. xix, 8, 13

PCA Principal component analysis. 36, 37

pH Potential of hydrogen i.e Sourness. ix, xv, 4, 59, 61, 70, 71, 98, 101, 102

PLS Partial least squares. viii, ix, xi, 30, 32, 36–38, 86, 87, 93, 94, 96, 99, 101

RFE Recursive feature elimination. 31, 87

RMSE Root mean squared error. iii, xv, 41, 46–50, 85, 86, 92–94, 96, 101

SG Savitzky-Golay. xiii, 28, 29, 66, 82, 83, 88, 89, 94, 97–101

SNR Signal to Noise Ratio. 21, 22, 28, 62, 63

SNV Standard normal variate. xi, 27, 28, 82, 99–101

SPA Successive projections algorithm. 30, 85, 94, 95, 99, 101

SPXY sample set partitioning based on joint x-y distance. viii, 32

SSC soluble solids content. viii, ix, xv, 5, 43, 45–47, 51–53, 59, 60, 67, 69, 70, 83, 84, 92, 96, 97, 99–101

SVR Support vector regression. ix, 34, 87, 88, 100

SWIR short-wave infrared. xiii, xv, 5, 6, 8, 22, 56, 57, 62, 74–82, 93–97, 99, 102

TSS Total content of soluble solids. viii, 47, 97

TTO NTNU Techonology transfer organisation. 4, 13

UN United Nations. 7

UVE Uninformative Variable Elimination. 30, 86, 87, 93, 94, 96, 101

VNIR visible to near-infrared. xiii, xv, 5, 6, 8, 29, 56, 57, 62, 74, 75, 77–82, 93–97, 102

Preface

This is our bachelor's thesis at The Norwegian University of Science and Technology (NTNU), department in Gjøvik. The task was performed for the computer engineering program (BIDAT) at the Department of Computer Science (IDI).

The thesis is about determining the quality parameters of kiwifruits using hyperspectral imaging. The purpose of the thesis is to learn and explore new methods that can contribute to a more sustainable future and lead to changes in non-destructive determination in different kinds of fruits. Investigations were carried out in laboratories at NTNU in Gjøvik.

Thanks

We want to thank everyone who has helped us complete the bachelor's thesis. A huge thanks to our supervisor **Sony George**, associate professor at NTNU in Gjøvik, who supervised us during the whole project period.

Thanks go to **Binu Melit Devassy**, senior engineer at NTNU in Gjøvik. He helped us with the laboratory equipment and tests in the starting phase.

We would also like to thank **Hilda Deborah**, senior researcher at NTNU in Gjøvik, and our company's contact, who has helped us with answering our questions through the bachelor's thesis process.

Thank you very much for your help and cooperation!

Chapter 1

Introduction

Food permeates significant aspects of our lives. It provides us nutrients and energy to live and is essential for our physical and mental health. The increased availability of new information has changed people's perception of eaten food. Research has shown that there is a clear link between food quality, customer satisfaction, and loyalty. That is why fruit and vegetable suppliers always look for more accurate and efficient ways to determine their products' important external and internal features. This way, they can deliver higher quality food and improve customer loyalty, which provides higher revenue. Traditional inspections evaluate either the external structure of agricultural products or interior features using destructive techniques. There is, however, ongoing research on performing non-destructive determination of the internal components because the internal parts provide more valuable information, and using it would significantly reduce food waste. Computer vision technology is used to assess these interior features, which provides much more information than regular RGB imaging. In our bachelor's project, we focus on determining kiwi ripeness using a non-destructive method called hyperspectral imaging (HSI).

This bachelor's thesis allowed us to explore the boundary of what has been studied before and use advanced laboratory equipment. We found this very compelling and were prepared to put in the work to learn something entirely new for us. It included how to operate the equipment, use it to collect our data, and analyze it.

1.1 Significance of the study

There are many reasons why we decided to work with determining kiwi ripeness using a non-destructive method. We wanted to address a global issue, and we wanted to work on an innovative research project and obtain and master new skills that we can need in our future careers.

1.1.1 Food waste

In our bachelor's thesis, we wanted to do be innovative and contribute to an important global issue in our society- the low quality of available fruits and vegetables. 32% of the Norwegian consumers stated that they do not consume enough fruits and vegetables, giving a poor quality of these as their main reasons[1]. Studies also show that the amount of unnecessary food waste in developed countries is vast. One reason for this is that some fruits and vegetables are mistakenly marked as low-quality in manual sorting processes. They are then thrown away despite actually having high quality.

Nowadays, the food industry predicts the quality of fruits and vegetables primarily based on destructive tests. These tests are performed on a few samples from a given batch and then averaged to represent the batch as a whole. Although destructive methods are helpful, they might not accurately portray the quality of individual produces inside the set as there is variance within the agricultural samples. Although the average achieved from all destructive tests shows that a batch should be disregarded, individual produce can still satisfy the quality criteria. The non-destructive test that examines all samples can solve this problem. In addition to this, the non-destructive manner has other advantages: it is less time-consuming, does not damage tested products, and has higher prediction accuracy.

As explained above, destructive tests lead to higher food waste and do not accurately predict internal qualities. That is why we aim in this bachelor's thesis to develop a technique that will efficiently determine the ripeness of fruits non-destructively using hyperspectral imaging (HSI). Reducing food waste and increasing the prediction accuracy of agricultural products is very motivational for our group as we strive for a sustainable future.

1.1.2 Non-destructive method

The HSI market is something that is still very new in the world. Not many people have heard about it, and even fewer could experience it. We found it very valuable for our future perspectives to gain experience in working with hyperspectral (HS) cameras and analyzing HS images. This fact is highly motivating to touch on techniques and methods that might never have been tried before, and precisely using HSI. This technology is expected to grow drastically in upcoming years[2], technology that today is costly and mainly available only for doctoral students and researchers. Using HSI gives us the chance to innovate. However, the number of publications with multispectral imaging on food and agriculture products has skyrocketed since 2009(1.1B). This research area is still in need of new findings, and we hope to contribute to it.

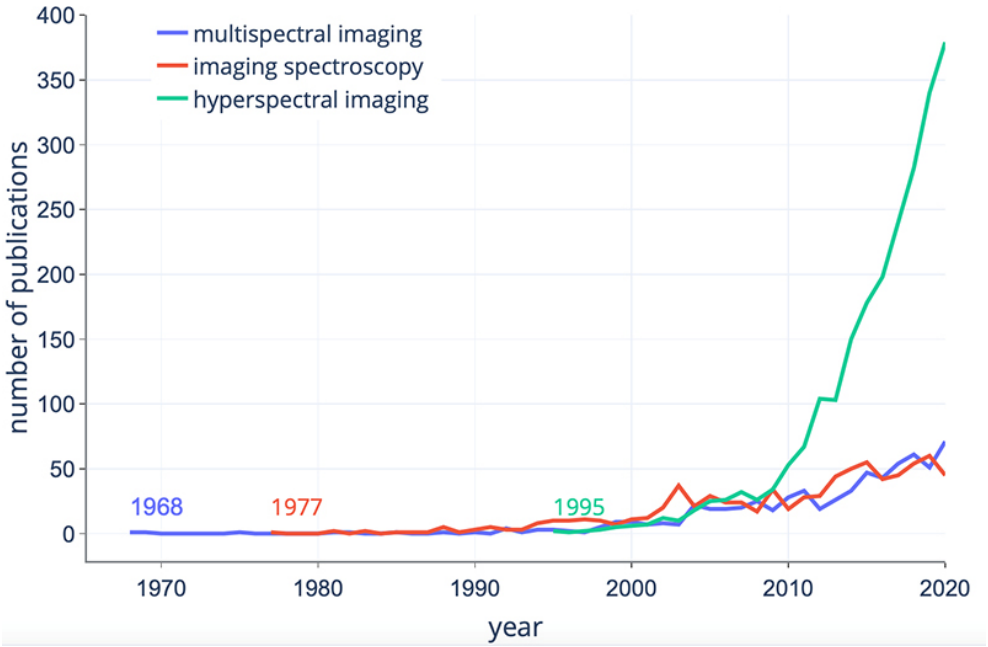


Figure 1.1: Publications on spectral imaging in recent years. The colored years inside the plot show the first publication (Image source [3]).

1.1.3 Working with kiwifruits

We wanted to study a popular product in Norway and whose ripeness is challenging to understand just by visual inspection. Among many possible fruits and vegetables to research, our choice fell on kiwifruits. Surveys show that the average Norwegian eats many kilograms of kiwis per year [4]. To ensure that chosen kiwifruits are ripe, people test them by pressing on them[5]. This tendency, unfortunately, damages fruits.

Kiwifruit is as well one of the fruits that are richest in nutrition benefits. It is rich in vitamin C, a good source of folate, potassium, and dietary fiber, and has many health benefits: lowering blood lipid levels and alleviating skin disorders. There are also investigations stating that its antioxidant and anti-inflammatory actions might help prevent cardiovascular disease, cancer, and other degenerative disorders[6]. So this fruit is not only well-liked through Norwegians but can also have a positive effect on their health.

1.2 Problem description

In this project, we are trying to determine the internal quality parameters of "Hayward" kiwi using hyperspectral imaging (HSI). Our goal is to make a fast and reliable way of determining the ripeness of kiwis without doing destructive measurements.

By using HSI, we can non-destructively collect sensory information of kiwis in a wide range of electromagnetic frequencies and use regression and machine learning models to find a correlation between spectral images and the internal quality parameters of the kiwifruit.

1.3 Target group

This project targets two different types of groups, researchers and food distributors.

1.3.1 Product target group

We provided results on HSI on "Hayward" kiwis for food distributors to take more informative commercial decisions regarding adopting the technology.

1.4 Research aims

The research questions to be examined in this thesis are:

1. To investigate the potential use of HSI to be used for non-destructive assessment of firmness, sugar content, and pH for kiwi of type "Hayward"?
2. Study the correlation between sensory information and the internal qualities for kiwi of type "Hayward"?
3. What techniques and models provide the best results for the determination of kiwi's ripeness?
4. Which wavelengths are most relevant for determining internal quality parameters of kiwi?
5. Can the process of predicting ripeness of kiwis be automated?

1.5 Contributions

This bachelor's thesis has provided both scientific and societal contributions, which are further discussed below.

1.5.1 Scientific contribution

This project was done in collaboration with Bama AS and NTNU Techonology transfer organisation (TTO), where Bama AS provided information about the kiwi market.

Bama AS is interested in having a portable device with different functionalities helpful in analyzing their products. This bachelor's thesis is considered as a step towards reaching that goal. There has been developed a non-destructive kiwi

	Portable device	Sugar	Firmness	pH	400-2500nm	Automatic Spectra Collector	RGB	Peels
Attempted	Partially	[7]	[7]	[7]	350-2500nm [8]	Yes But not "Hayward" [9]	Not done	Sugar [10]
Our Contribution	X	X	X	X	X	X	X	X
Future Work	X			X			X	X

Table 1.1: Broad terms where this bachelor's thesis made contribution.

	SVR	PLS	MLR	ANN/MLP	KNN	Tree regressor
Attempted	Limited	Yes	Yes	Not done	Not done	Not done
Our Contribution	X	X	X	X	X	X

Table 1.2: Our model contribution

quality meter by Felix instruments[1], which uses Near-infrared light to predict the Dry Matter Content and sugar content in the kiwi.

All the devices that claim to be portable and available in the market (for example, Felix) have several drawbacks, such as size, time of determination, weight, and specific range.

Most of the scientific papers published on the topic HSI which try to determine the ripeness of "Hayward" kiwi are limited to the visible to near-infrared spectral range. This project extends this range into the short-wave infrared, which has only been reported in two papers[7][8]. Although these papers cover the same spectral range, their datasets have covered only three kiwi parameters. In comparison, we provide two additional quality parameters (salt, core firmness) and more sensory information (RGB imaging and peels).

Previous research has limited data collection and machine learning models tested. Most of them measure sugar content (soluble solids content (SSC)) and flesh firmness; however, we measure additional quality parameters such as pH level, salt, weight, temperature, circumference, and core firmness. Our project also further innovates by implementing machine learning models that have rarely or never been tried before on "Hayward" kiwi.

Our bachelor's thesis captures HS images of kiwi peels, which shows the spectral information inside the kiwi. This has been done before to predict sugar content in kiwi with 1-MCP[10] chemical growth regulator. Providing a dataset with this information is of high scientific value.

Our project also explains in detail how our automatic spectra collector works. Which there is little detail about in the scientific articles.

In addition to HSI, our project provides RGB images captured with different

[1]https://felixinstruments.com/food-science-instruments/portable-nir-analyzers/f-751-kiwi-quality-meter/

illumination to contribute towards the future goal of determining kiwi ripeness using a regular camera.

All of these contributions is summarised in Table 1.2 and 1.1.

Further information about previous research on the topic is discussed later in **Chapter 4: State of the art**.

Our main scientific contributions follow:

- **Extended spectral range:** Our research explores spectral information in between 400 to 2500 nm, and in current literature, there is a limited number of papers on kiwifruit that exceed 1600 nm. This implies that we are contributing with innovative research within the shortwave infrared range beyond 1600 nm.
- **Creation of dataset:** We created a big dataset with 495 "Hayward" kiwifruits[2]. It contains the following data for each sample: 3 different modes of color images taken by a phone camera, VNIR and SWIR HS images of kiwifruit and a slice of the inside, measurements of firmness, sugar level, salt level, pH level, temperature, weight, and size (circumference). There has not been created a similar dataset with all these details. To create this set, we have prepared, learned about methods, and experimented until we were sure about the required skills on 200 different samples of kiwifruits bought from local stores.
- **Algorithm for automatic spectral extraction:** We created a semi-automated program for extracting spectra[3] of each kiwi sample from the HS images. This is usually time-consuming and crucial for creating large datasets.
- **Commercial application:** We have provided models, methods, and results of non-destructive quality parameter prediction of "Hayward" kiwifruits that possibly can be used in commercial applications and might be applied to other fruits as well.
- **Methodology for analysis:** We have tested models that have not been researched thoroughly on kiwifruit as of current literature and experimented with many different combinations of pre-processing techniques to enhance these models.
- **Publication in a scientific journal:** Our project contributes to the scientific field by new results and methods that have not been reported before for kiwi type "Hayward." We will publish our work in a scientific journal and hopefully encourage more research in this field.
- **Thorough description of methods:** All methods that were used are described thoroughly. Our bachelor's thesis can be reused as a study guide for those who want to further investigate "Hayward" kiwifruit with HSI.

[2]The "Hayward" kiwi is the main green variety produced for the world's markets and dominates production in most growing areas. It is moderately hairy and more rounded than other kiwi varieties[11]

[3]A spectrum is a graph that shows the intensity of the light being emitted or reflected

1.5.2 Societal contribution

In this subsection, an attempt is made to show the current situation where immature kiwifruits contribute to already enormous food waste contrary to the United Nations(UN) sustainability development goals.

Food waste has become a topic of societal concern and a focus of much research lately. In 2011, the Food and Agriculture Organization of the United Nations (FAO) stated that about one-third of the food produced for human consumption was wasted. This resulted in around 1.3 billion tonnes of food lost in the world [12]. This figure was estimated to about 88 million tonnes of food in the European Union, corresponding to around 173 kg per capita. In economic terms, this meant a loss of 143 billion euro per year. Although estimations from 2019 an improvement showed an improvement [13], there is still considerable room for future changes. The scientific research dedicated to food waste has more than doubled from 2011 until 2017 [14]. These studies looked both at the consumer, the retail level, and there is work attempting to explore food waste from an overall system perspective.

Statistics, methods of measurements, and definitions of food waste can vary according to the different sources one considers. It is, however, rarely doubted that lowering waste level is valuable for nature. The enormously high amount of thrown away food has a massive impact on our environment. Food losses affect the use of resources, such as fresh water, cropland, and fertilizers. According to a Swedish study from 2015, waste minimization by 35% could result in the reduction of greenhouse gas emissions of 800–1400 kg/tonne[15]. Food waste and climate action are two of the aims in UN sustainable development goals [16]. By lowering food waste level, we will be beneficial for social, economic, and environmental reasons. Inaccurate methods of analyzing quality and ripeness levels of fruits and vegetables in general, and in our case of interest- kiwis, play a prominent role of the contraries to UN goals. Consumers' perceptions of the quality of fresh and mature produce at the point of purchase and point of consumption play an important role in decisions about what to buy, eat, and discard. If they buy a product that does not satisfy their expectations, they will most likely throw it away. On the other side, there are available destructive ways of estimating kiwis' ripeness, leading to environmental instability. After the destructive measurements have been carried out, the kiwis are thrown away, meaning potential high-quality fruits will be wasted.

In our thesis, we contribute to achieving 2 of UN sustainability development goals by trying to partially resolve this enormous problem by attempting to find new or more efficient methods for analyzing kiwifruits ripeness. If we can ensure well-matured fruits in grocery stores, we will get satisfied customers [17], and thus there will be less waste.

We want to mention that we have not wasted any kiwifruits during the whole period of our tests. The remaining kiwis from destructive tests had been processed and used to make jam and smoothies.

1.6 Project limitations

The project has a deadline on 20th of May, giving us just above three months to review the literature on this topic, learn to use the complex imaging system, create a dataset, conduct pre-processing, analyze the data and prepare the report. The limitation of time has made us specify the task more clearly, as we do not have time to cover a study as broad as we would like. We had to prioritize firmness and sugar content as the processing of models takes much time.

Because the project was bound to start in January and end in May, we would not have access to fruits coming straight from harvesting due to the project not being in sync with the harvesting season. Also, the distributor does not have all the details of the harvesting and storage. Therefore, we do not know the exact harvesting date, how the fruits were stored, the temperature of storing, whether or not there were some pesticides used. All these factors would have some impact on the results as they bring uncertainty into the equation. Other similar scientific papers have more exact information about the kiwis used and ensured bio-variability using kiwi for different farms and harvest months.

In addition to this, we are limited to the equipment NTNU owns and are available for us to use. Currently, there are two HS cameras, one in the VNIR range and the other one in the SWIR range. We also had to take other students or professors into consideration and could not use the equipment at any time we would like. Another project was carried out at NTNU using the kiwis in parallel with ours. After reviewing research in the field, our group decided with our supervisor that looking at only kiwi would be enough for a bachelor's thesis. Kiwi was the fruit with little research among other candidates like mango and avocado.

Because we collected a large, high-quality dataset, we did not have enough time to analyze all the kiwi quality attributes from the data we collected in this thesis. Several quality parameters and sensory information were analyzed, and many models were tried. The data captured and models used had to be prioritized. Running the models used much computational power and often take a long time to run, even on high-end computers.

1.7 Thesis structure

- **Chapter 1: Introduction**, explains our contribution, project limits and general information about the project.
- **Chapter 2: Background**, explains the background of the group and scientific field.
- **Chapter 3: Theory**, explains the theory that the project builds upon, and ensures that the reader can follow the terms used later in the report.
- **Chapter 4: State of the art**, mentions previous research done within both destructive and non-destructive methods, and show results of different relevant papers.

- **Chapter 5: Data acquisition**, shows and explains the data acquisition period, materials used and further discusses this part of the project.
- **Chapter 6: Data processing**, shows equipment used and explains the automation of spectra collection, visualization, pre-processing and how the models were trained.
- **Chapter 7: Results and discussion**, shows the results of our models and discusses various aspects of the models.
- **Chapter 8: Conclusion**, concludes our report and explains improvements needed for future work and our learning outcome.

The project is split into two main parts, data acquisition, and modeling. Therefore, **Chapter 5: Data acquisition** has its discussion and structure as this fits better there. We did this to keep the main focus of **Chapter 7: Results and discussion** on the results of our models and reduce the size of this chapter to make it easier to follow for the reader.

Chapter 2

Background

In this chapter the following information is discussed.

- Group Background
- Contributors
- Scientific field of the project

2.1 Group Background

The computer vision Course (IMT 3017) is very relevant for this project. This course teaches how images are captured and filters that can be applied to images and to extract information from the image so that a computer can make decisions based on it. Eivind, Vebjørn, and Katherine have completed this course. They have shared relevant information from the course with Jon.

During this course, Vebjørn and Eivind had a project working with HSI of potatoes to detect bruises early. Because of this, they have the most experience with HSI. Vebjørn and Eivind were, because of this reason, put to work most closely with the camera.

Everyone except Katherine had the course Artificial Intelligence (IMT 3104). This course teaches how to make machine learning models so that a computer can make good decisions based on given information. Jon's background comes from software development and has experience with management from being class representative and leading lab activity in physics (REA2021).

2.1.1 Project Roles and Responsibility Management

Every group member has contributed to project creation, but there were also divided special responsibilities assigned to different project periods and main tasks. The responsibilities were split according to table 2.1. Even though responsibilities were divided among the group, everyone contributed to almost all the tasks. We made sure everyone got a good learning experience from each other. The total hours spent per period is shown in Figure 2.1, and per member on Figure 2.2.

Task	Main responsibility	Other contributors
Project management	Jon Elias	Katherine, Vebjørn
Report writing	Katherine	Everyone
Literature review	Katherine	Everyone
Data-acquisition	Eivind	Everyone
Data pre-processing and Modelling	Vebjørn	Eivind, Jon Elias

Table 2.1: Distribution of responsibilities.

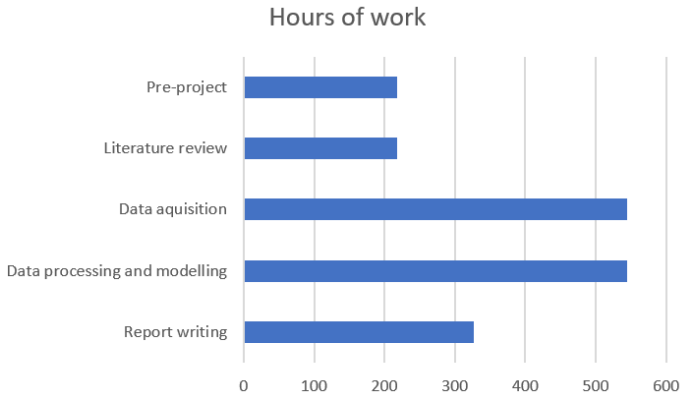


Figure 2.1: Distribution of total hours spent on each period. (total hours 1857)

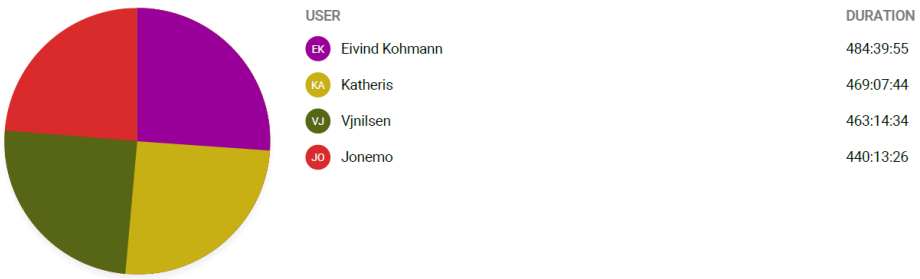


Figure 2.2: Distribution of total hours between members (See appendix B for more info).

2.1.2 Project Contributors

1. **Sony George** is an associate professor at NTNU, and he is our supervisor.

2. **Hilda Deborah** is a research scientist at NTNU and group's company contact. She has done a master thesis on HSI, so she will take part in the learning process and be at the disposal of any questions.
3. **BAMA-Gruppen AS** is Norway's largest private distributor of fruit and vegetables, and it will deliver needed fruits for student's bachelor's thesis project.
4. **Tom Røise** is a bachelor's thesis supervisor.
5. **NTNU Techonology transfer organisation (TTO)** provided project support and logistics related with fruit handling.

2.1.3 NTNU technology transfer organization (TTO)

NTNU Technology Transfer is an organization within NTNU that focuses on commercializing ideas that come from NTNU's employees, teams, and Central Norway Regional Health Authority (Norwegian: Helse Midt-Norge). These ideas are supposed to create a positive community impact. NTNU Techonology transfer organisation (TTO) explains its mission as: "Together with the teams at The Norwegian University of Science and Technology and The Central Norway Regional Health Authority, we create products and services that benefit society." [1]

2.2 Expected knowledge

We performed a survey on computer engineering students at Norwegian University of Science and Technology (NTNU). The results showed that none of them had heard about HSI before studies at NTNU, and students that attended about it during their studies took a course in Computer Vision. Students and many IT professionals are not aware of HSI which shows that the field's potential is tremendous. Because of this reason, this thesis provides an understandable introduction to HSI and spectral imaging.

2.3 Scientific field

This project touches many scientific fields, which are further described below to give the reader a short introduction to the topics. These are spectral imaging and machine learning which are combined into chemometrics.

2.3.1 Spectral imaging

Spectral imaging is an extension of traditional RGB imaging as it captures additional images in more wavelengths. Spectral imaging uses all wavelengths possible to acquire with image sensors, even outside of the visible spectrum. As a result, new information can be obtained that otherwise would not be captured in regular

[1] TTOs homepage: https://www.ntnutto.no/home/

RGB. This information obtained is often referred to as the spectra or spectrum to the object of analysis.

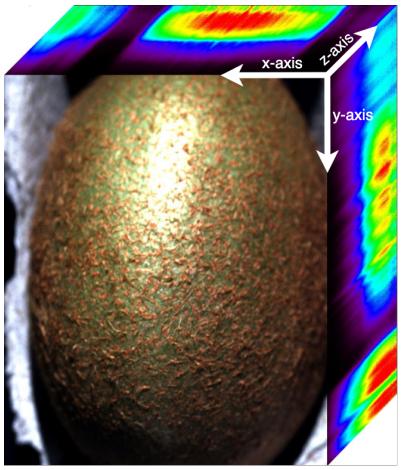
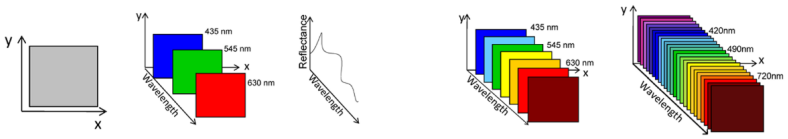


Figure 2.3: Spectral data (datacube) of a kiwi where the x- and y-axis representing the spatial dimensions and the z-axis the spectral dimension.



	Monochrome	RGB	Spectroscopy	Multispectral	Hyperspectral
Spatial information	Yes	Yes	No	Yes	Yes
Band numbers	1	3	From several dozens to hundreds	3 to 10	From several dozens to hundreds
Spectral information	No	No	Yes	Limited	Yes

Figure 2.4: Comparison of how different amounts of images are stacked together and named (Image source [18]).

Spectral images are stored as datacubes (also called image cubes). An RGB image can be expressed as a datacube containing only three layers; red, green, and blue, shown in figure 2.4. A datacube has two spatial dimensions and one spectral dimension explaining the wavelengths (images or bands).

The prefixes multi, hyper, and ultra are used to describe the number of bands and the wavelength spacing between them, seen in figure 2.4. This distinction is made to explain what information can be explained from the spectral images.

2.3.1.1 Multispectral imaging

Spectral imaging is usually classified as multispectral if the images captured are between 3 and 10. However, there is no exact boundary, but according to the

standard IEEE P4001[2], multispectral is defined as having a spacing of 20nm or more between the images.

Multispectral imaging captures images from specific areas of the electromagnetic spectrum and gives rise to many applications. It is much used for remote sensing like astronomy and as a tool for mapping details like vegetation and environmental changes on the Earth. It has also shown applications related to food, paintings, forensic sciences and archaeology, and many more in the last years.

There are advantages and disadvantages to multispectral imaging as it contains fewer bands. Because there are few bands, less image processing is required compared to hyperspectral imaging (HSI). It provides more flexibility as a smaller, more portable capturing device can be used. Depending on the spacing between the spectral bands, it can obtain some spectral information, but it would not form a continuous spectrum. This discontinuity is solved in hyperspectral imaging (HSI).

2.3.1.2 Hyperspectral imaging

Like multispectral, there is no exact boundary for how many spectral bands are needed to classify as HS. The number of bands can range from tens to hundreds, and the spacing between each band is defined as less than 20nm by the standard IEEE P4001 [2].

hyperspectral imaging (HSI) has many of the same real-world applications as multispectral imaging but can also be used where more spectral precision is needed. Some of the applications are agriculture, food processing, health care, surveillance, chemical changes, and astronomy. HSI provides high-quality spectral information required in specific use-cases where precision is crucial. However, it is worth noting that by researching topics with HSI, the applications may later be applied with multispectral imaging choosing the essential wavelengths to reduce redundant computation and costs.

The use of HSI in the food industry has increased in recent years due to its ability to evaluate food quality. Qualities such as flavor, freshness, ripeness, and defects (like bruises and fungi) can be assessed in rapid succession[19].

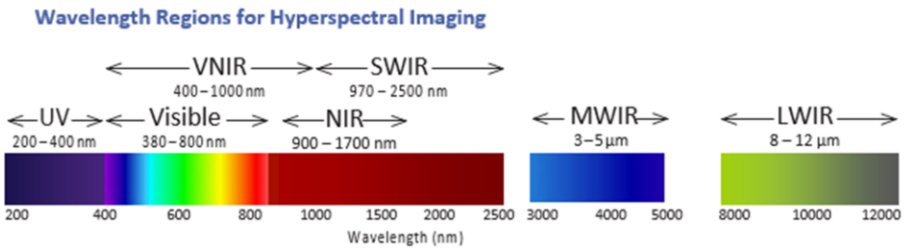


Figure 2.5: Names of different electromagnetic wavelength regions. (Image source [20])

[2] https://ieeexplore.ieee.org/document/8900295

Due to the satisfactory spectral resolution in HSI, it is possible to study the spectral changes with high accuracy compared to multispectral imaging. This is useful for research purposes as the high spectral resolution allows for analysis by spectroscopy.

HSI cameras are limited to specific spectral ranges (see figure 2.5) because of technical challenges and are expensive to manufacture. As the field grows and the technology advances, cameras with broader spectral range might become more common in the future.

HSI is used mainly in laboratories instead of commercial applications because the devices are expensive and complex with small efficiency. Making HSI cheaper, more user friendly, and more compact are some of the significant challenges in making it more widespread.

An imaging system with a higher spectral resolution than hyperspectral is called ultraspectral imaging. It has more than 500 bands.

2.3.2 Machine learning

Machine learning is the combination of statistics and computer algorithms to model relationships between data. One significant aspect of machine learning is its ability to model data where it is difficult or unfeasible to construct algorithms by conventional methods. Depending on the data to model, machine learning is split into Supervised and Unsupervised learning.

Unsupervised learning aims at finding relationships in data without human intervention. This involves finding patterns and unknown relationships like cluster analysis, seeking to separate the data into clusters.

While Unsupervised learning finds hidden relationships, Supervised learning aims at making predictions. These predictions can either be continuous (regression) or discrete (classification). Supervised learning is more relevant for this project because we are making a regression on spectra to predict kiwi ripeness.

Supervised learning can be further divided into two main categories, linear and non-linear. These categories describe how the relationship between two or more variables is connected. The difference between linear and non-linear models is noticeable. A linear model is used when a linear function can be used to predict the dependent variable. While the non-linear model is more complex, it uses a non-linear function to predict the dependent variable. Non-linear relationships are common in nature and physics [21] and are tricky to model perfectly, but it is still possible to find good approximations.

2.3.3 Chemometrics and multivariate analysis

Chemometrics is a field that inherits methods from computer science and statistics to solve descriptive and predictive problems in biochemistry, chemistry, medicine, and compounds engineering [22]. It uses machine learning models and multivariate analysis to derive chemical information from the spectra of the object of study.

This is possible because the spectra of the object change when the chemical properties change. Each periodic element has its electromagnetic radiation pattern, which can identify the periodic elements present and their concentrations. Using this fact, this can be further applied to a molecular and physiological level; it can be, for example, be sugar, water content, pH level, or even firmness.

Chapter 3

Theory

This project builds upon theory about hyperspectral imaging, pre-processing techniques, and several machine learning models. As a result, the theory chapter is quite large.

3.1 Hyperspectral Imaging

HSI system can be set up in many different ways. There can be other cameras, filters, and acquisition modes. These different ways to set up a HSI system are further explained below.

3.1.1 Types of cameras

There are multiple ways of capturing HS data, and each has its pros and cons. Some are easier to manufacture, and others may be more suitable for the specific situation. The different techniques are split into spectral, spatial, and non-scanning methods.

- A wavelength scanner(figure 3.1 C) is a spectral scanning method in which the whole object is captured at once by one wavelength/band at a time. Such a camera is usually designed by having different wavelength filters on a turntable wheel. The data cube is constructed by stacking each image on top of the other. It has the advantage of being able to pick and choose which wavelengths to use, and each 2D image has a direct representation of the actual scene. A weakness is that if the scene moves during the scan, it is impossible to do spectral correlation unless each band is realigned.
- Snapshot is a non-scanning method, meaning it captures the whole object and all the bands simultaneously, creating the data cube in the exact moment, just like a regular camera. Since these cameras are difficult and expensive to manufacture, they are nowadays mainly used in astronomy.
- Line and point scanners are Spatial scanners. They capture the full spectra at once, but only parts of the image at the time. A line scanner scans one

row of the image and needs the object or camera to move over the scene. Since it flips like a push broom mopping the floor, it is also called a push broom scanner.
A point scanner differs from a line scanner as it only captures all of the wavelengths for one pixel at a time. This scanner requires the object or camera to also move along the second spatial dimension. Because it scans like a whiskbroom, it is called a whiskbroom scanner. The line scanner is The advantage of spatial scanners is that they can provide high spectral resolution over a wide spectral range [23]. They are helpful in situations where the scene moves, e.g., imaging the sky using planes or satellites. However, with motion comes the complexity of needing advanced hardware to allow for imaging at different speeds, making it more complex and expensive to produce.

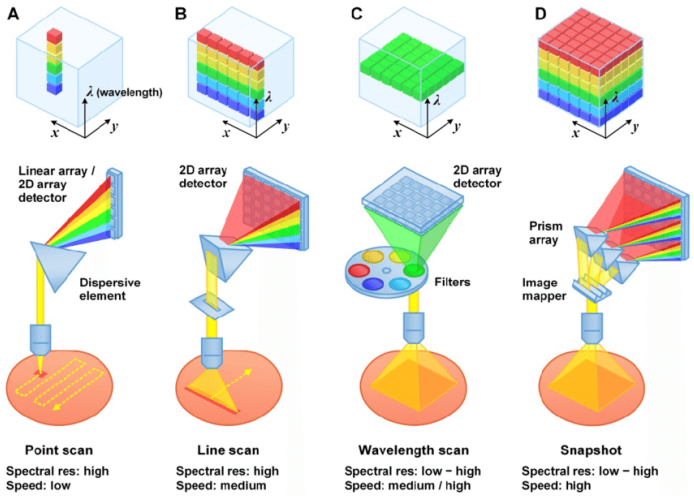


Figure 3.1: Different types of HSI cameras (Image source [24]).

3.1.2 Acquisition modes

HS cameras can be set up in different acquisition modes to satisfy other requirements. In these various modes, the camera and light source has different positions. It can be set up in either transmission, reflection, or interactance mode. Transmission and reflection are the two most basic modes to use, while interactance is more complicated. Transmission measures how much light the object does not absorb, while reflectance measures how much light the thing reflects (not absorb).

With reflection mode, the light and the camera are at the same side of the measured object (as shown on figure 3.2A), measuring the object's reflection. Reflection mode is helpful when evaluating the external quality of fruits and vegetables.

Transmission measures how much light is passed through the object. This

measurement is done by having the light source and camera on the opposite side of the object (as shown in figure 3.2B). The transmission mode is more useful when evaluating internal concentrations; however, little light is carried over to the camera.

It is possible to have a setup with both reflected and transmitted light; this is called absorbance mode. In this mode, the light source and the camera are on the same side of the measured object. Unlike in reflection mode, there is a wall between the light and the camera (as shown in figure 3.2C). This setup results in some light reflected from the object, and some are transmitted through the object[25].

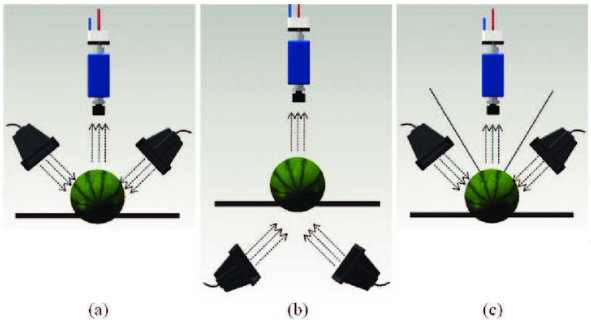


Figure 3.2: (A) reflection mode, (B) transmission mode (C) Absorbance mode (Image source [25]).

3.1.3 Camera noise

Hyperspectral cameras are particularly susceptible to noise due to their complex sensor that needs to be sensitive to many wavelengths. Since this noise influences each pixel and the overall quality of the image, it is important to reduce its effects. This is mainly done during the acquisition and after by preprocessing.

Then explaining noise during imaging, the term Signal to Noise Ratio (SNR) is often used and refers to the amount of noise compared to the measured signal. By having a relatively high signal compared to the noise, the signal is less influenced by it. Thus the signal is more reproducible and does not fluctuate between measurements.

A common technique to increase the SNR during imaging is to capture the same scene (in push broom scanners, this would be a line) multiple times and then average the results. This reduces the noise by a factor of $\sqrt{N}$, where N is the number of captured lines[26]. Another method of increasing the SNR when imaging is using an equalization filter.

3.1.3.1 Equalization filter

A filter for the quantum efficiency (sensitivity) of silicon-based sensors are weak in the spectral region around 400nm and 900nm. More light in this region is needed

for getting a low SNR. However, increasing the light results in saturating in the other spectral region where the light is most effective. Because of this, reason equalization filter aims to flatten this sensitivity bypassing more light onto the area where the quantum efficiency (sensitivity) is lower and suppresses where it is higher.

3.1.3.2 Radiometric calibration

Radiometric calibration involves four different steps to increase spectral accuracy. These steps involve dark current correction, sensor calibration, illumination correction, and spectral calibration.

The dark current is the signal that the sensor produces when the sensor does not capture any photons. This noisy signal increases with the temperature of the camera and the integration time of the camera. An image is captured with the HS camera with the lens closed to remove this signal. Then the dark current is removed from the resulting reflectance. The dark current calibration is the only thing done by the radiometric calibration that changes during different operating conditions[26].

Spectral sensors can have high variability in pixel sensitivity. Therefore a sensor calibration has to be carried out. The manufacturer of the camera does this with an integrating sphere.

Since the manufacturer already does the sensor calibration, the only thing explicitly done to the image acquired is removing the dark current.

3.1.4 Reflectance

HS cameras are dependent on the reflectance of the surfaces in the scene to obtain images. Reflection is the effectiveness of how well a surface reflects the radiant energy emitted on the surface. In our study, this radiant energy is in the form of electromagnetic radiation from the visible to shortwave infrared spectrum (SWIR). By measuring the amount of radiation (number of photons) emitted and then reflected onto the camera sensor, we obtain spectral information that allows us to study the changes within and outside a kiwi. As explained in section Chemometrics and multivariate analysis, different chemical compounds reflect different amounts of photons at different frequencies, meaning it is possible to correlate the spectral reflections with physiological parameters.

We measure the reflectance R, of a surface by dividing the total reflected radiance flux, Φ_R, by the total emitted radiance flux, Φ_E.

$$R = \frac{\Phi_R}{\Phi_E} \tag{3.1}$$

There are a few scenarios when the reflectance value can exceed 1 (100%). There may be an external radiance source like a lamp that is not accounted for, or maybe the surface itself emits radiation, e.g., the material is fluorescent. The

scenario we experienced the most was due to the geometric shape of the object being imaged.

3.1.4.1 Geometric and material limitations

From section 3.1.4 we know what reflectance is and how to calculate it given the emitted and reflected radiance values. Even though we can measure both of the components to calculate the reflectance, difficulties with consistency appear when dealing with different shapes and surface qualities.

In general, there are two different types of reflections. The first one is specular reflection, where all emitted light is reflected at an equal angle on the opposing side from the surface normal. In physics, this is often explained by stating that the angle of incidence is similar to the angle of reflection. This reflection often appears when the reflection surface is clean and smooth, which can be reflected by a mirror. The other one is diffuse reflection, where the emitted light is reflected at different angles from the reflection point. This reflection happens when the reflection surface is rugged, for example, matte or satin finish.

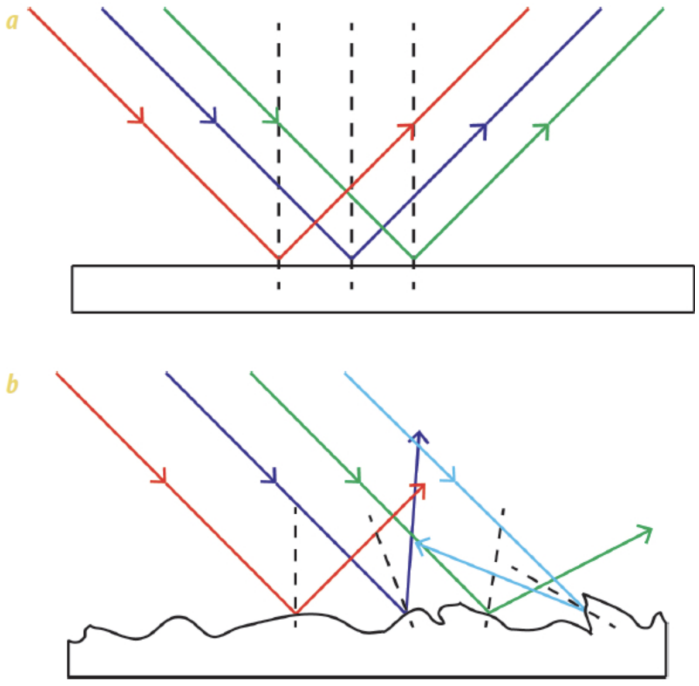


Figure 3.3: a) Specular reflection b) diffuse reflection (Image source [27]).

The surface of the reflectance target is flat and rough, which reflects diffusely and therefore gives constant reflection to the camera sensor. This is preferred because it avoids concentrating the light in one place, which would quickly lead to overexposure on the camera sensor depending on the angle of the incident light.

The shape of the kiwis is oval, which means the reflection surface is hemispherical. This hemispherical property is not uniform across different kiwis and serves inconsistencies as some are flatter, while others are more round. It leads to differences in the specular reflection in different kiwi areas, and this difference is therefore not uniform across all kiwis. Combined with the hemispherical inconsistencies, we do not know the exact properties of the different kiwi surfaces, leading to minor differences in perceived reflection for the camera sensor.

3.1.4.2 Normalization

When dealing with spectral images, it is helpful to transform the pixel values in the image into reflectance, called normalizing. Reflectance tells us how much of the emitted light got reflected up into the camera and is easier to interpret as it can be compared between different amounts of illuminant radiance. With normalized data, we can look at the spectral properties of objects and obtain information about the object's composition of elements. This is possible because different parts reflect light differently, as explained in section 2.3.3Chemometrics and multivariate analysis. However, by transforming the data into reflectance, we make the values more comparable to other datasets independent of different light sources used.

To transform the spectral images into reflectance we use equation 3.3 explained in 3.1.4, but with modified variable names for convince of this section. Since the radiance emitted from the laps, Φ_E in equation 3.2 is unknown; we need a way of calculating it. This is solved by having a surface in the scene where the reflectance value is already known and then constructing a new equation 3.3 to solve for Φ_E. A surface where the reflective properties are known is called a reflectance target[1].

Variables:

- Φ_E = Radiance emitted from the lamps
- R_{ref} = Reflection of reference object
- I_{ref} = Radiance of reference object
- R_{im} = Reflection value
- I_{im} = Radiance in the image

The equation for transforming each pixel value in an image to reflectance values. I_{im} is the intensity of one pixel and R_{im} its new value.

$$R_{im} = \frac{I_{im}}{\Phi_E} \tag{3.2}$$

$$\Phi_E = \frac{I_{ref}}{R_{ref}} \tag{3.3}$$

[1] `https://www.labsphere.com/labsphere-products-solutions/materials-coatings-2/targets-standards/test-child/`

If we combine the two equations above we get:

$$R_{im} = \frac{I_{im}}{I_{ref}} * R_{ref} \tag{3.4}$$

Figure 3.4: Each HS image must contain a spectralon reflectance target with known reflectance in order to transform the data to reflectance.

3.2 Pre-processing

Pre-processing techniques aim at preparing the data for the machine learning models. This section explains the theory behind these techniques.

3.2.0.1 Image morphology

Image morphology is the operation performed on an image to change the structure of the image. Binary morphology is the structural change in binary images based on the properties of the kernel used. The two main kernel functions are erosion and dilation.

- Erosion transforms pixel with value 1 to 0 if any of the other pixels in the kernel is 0, which causes the total amount of pixels with value 1 to decrease and the number of pixels with value 0 to increase.
- Dilation does the opposite of erosion. Pixels with value 0 change to 1 if any of the pixels in the kernel is 1.
- Opening is the operation of erosion to remove weak links between objects and then to do dilation to increase the size of the objects back to original.
- Closing is the operation of dilation to increase the size of the links between objects, then doing erosion after to reduce the size of the objects back to normal.

These operations are relevant for making binary masks in the automatic spectra collection algorithm we use for saving time and effort on a collection of spectra.

3.2.1 Spectral pre-processing

After collecting the spectra that will be used for creating prediction models and analysis, it is almost essential to apply some pre-processing techniques. It is primarily done to eliminate unrelated effects that have nothing to do with the sample's chemical nature. These effects may arise due to the different geometries between the samples.

3.2.1.1 Hanning window filter

A Hanning filter is a filter used for smoothing out the spectrum, often to reduce noise. The filter only operates at a portion of the spectrum at a time called the window, and the user specifies its size. The window size influences how the smoothness of the spectrum; a larger window size would result in more smoothing. The function is defined as $h(i) = 0.5 * (1 - \cos(\frac{2 * \pi * i}{N}))$, were N defines the window size and i specifies the ith element inside the window. Using the weights defined by the Hanning function and noisy spectra y, the resulting smoothed signal is defined as $R_j = \sum H(i) * y_i$. It is repeated until the moving window has convolved over the entire spectra. This is done for all the elements inside the window [28].

The function is defined from and including 0 to N, and to exactly hit y=0 at the lower peaks. The function can be seen from Figure 3.5.

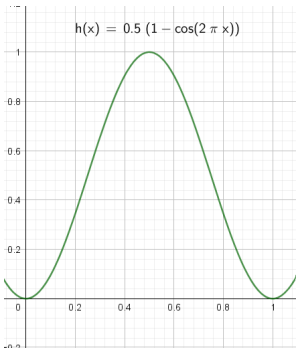


Figure 3.5: Plot of Hanning window with $N = 1$.

There is little research on using the Hanning window filter as a pre-processing technique. However, it has been shown that it works well compared to other pre-processing techniques for HSI [29]. Braun ([30]) has shown that the Hanning filter is "almost always" recommended when working with spectra.

3.2.1.2 Standard normal variate (SNV)

Standard normal variate (SNV) tries to correct for scattering effects by normalizing each spectrum. By subtracting the mean and dividing by the standard deviation for each spectrum, the spectra would all obtain a mean of 0 and have most of the values between $[-1, 1]$.

The normalization for a single spectrum X can be expressed as:

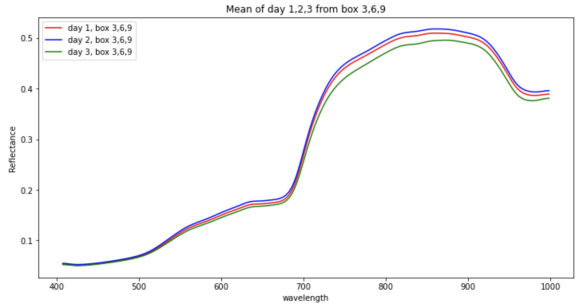
$$X = \frac{x_i - \bar{x}}{\sigma} \tag{3.5}$$

Where x_i is the reflectance value of a wavelength, $\bar{x}$ is the average reflectance of all wavelengths, and σ is the standard deviation.

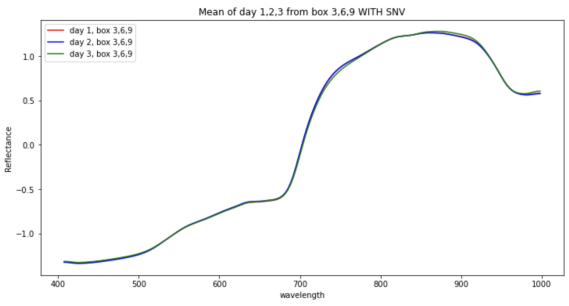
Equation 3.5 can be generalized to correct for all of the spectra at once by the following equation:

$$X_{i,j} = \frac{x_{i,j} - \bar{x}_i}{\sqrt{\frac{\sum_{j=1}^{N_{bands}} (x_{i,j} - \bar{x}_i)}{N_{bands} - 1}}} \tag{3.6}$$

SNV is effective at removing or decreasing multiplicative scattering effects that may occur during image capturing. It also corrects for the global intensity differences that might arise between different scans. As a result, SNV pre-processing can interpret spectra easier as it causes decreased dependency on scattering and global intensity differences[31].



(a) Mean of set of spectra **before** SNV pre-processing



(b) Mean of set of spectra **after** SNV pre-processing

Figure 3.6: Example of before and after SNV pre-processing, where each graph represents box (3,6,9) from a day (notice that the graphs are centered around 0 on the y-axis on b).

3.2.1.3 Multiplicative scatter correction (MSC)

Multiplicative scatter correction (MSC) is quite similar to SNV but requires a reference spectrum to start with. This reference spectrum should ideally be free from noise and scattering effects, but the spectra' mean spectrum is usually used. MSC corrects spectra by assuming that each spectrum can be approximated using a linear combination of the mean spectra, X_{mean}, expressed in equation 3.7.

MSC and SNV is widely used in HSI and on quality inspectrion of agricultrual products[32]. Therefore these are are highly relevant for this project.

$$X_i \approx a_i + b_i X_{mean} \tag{3.7}$$

Where a_i is a constant, b_i is a coefficient and X_i is a single spectrum, i represents a sample.

By solving for a_i and b_i using ordinary least squares, the corrected spectra can be calculated by using equation 3.8, where X^{msc}_{mean} is the corrected spectrum.

$$X^{msc}_{mean} = \frac{X_i - a_i}{b_i} \tag{3.8}$$

3.2.1.4 Savitzky-Golay (SG) filter

The SG-filter aims to increase the SNR ratio from a signal in the time domain. The filter works differently from standard data smoothing filters, which often is a function of a moving window (for example, Hanning 3.2.1.1). The filter aims to make a new point g_i in the middle of an moving average window (with size n) using polynomial least-squares fitted to the noisy signal points $f_w = [f_{i-\frac{n}{2}}, ..., f_i, ..., f_{i+\frac{n}{2}}]$ inside the window. All the points $g = \forall g_i$ is after completion the smoothed version of the signal $f = \forall f_i$. This polynomial fitted over f_w is often either of quadratic or quartic order; however, it can have any m order. For simplicity we define the window size $n = n_l + n_r$, where $n_l = n_r$ which is the number of points to left and right of $f_i \subset f_w$.

$$g_i = \sum_{k=-n_l}^{k=n_r} c_k f_{i+k} \tag{3.9}$$

Doing the least-squares technique for all the moving window combinations can take time. However, it is possible to use pre-calculated convolution coefficients (c_k) to estimate the smoothing operation. This is done according to equation 3.9 which defines the moving window with filter coefficients $c = \forall c_k$. By using these filter coefficients, the derivation of the noisy signal can also be provided. These filter coefficients vary for the window size, polynomial order, and the derivation order [33].

The filter coefficients are pre-calculated using standard least squares regression, with a unit vector instead of f values. More details on how these coefficents are calculated can be found in Press and Teukolsky[33].

SG-filter for spectral pre-processing

There are significant benefits of using the SG filter for spectral pre-processing. It is a better way of derivation on the spectra since it considers multiple points when computing the derivative. Since regular derivation also amplifies the noise that exists in the data, it is necessary to apply smoothing. The SG-filter does this and also preserves higher movements in the spectral data[33].

Different values of both window size and polynomial order give a different smoothed signal which enforces different smoothed versions of the spectral data. According to [34], the exact window size is related to one specific component of the spectra and recommends exploring the variables independently.

The SG-filter is widely used in HSI and has been applied to agricultural product inspection multiple times [35]. Therefore this filter is of high importance for our project.

3.2.2 Efficent wavelength (EW) selection

The dimensionality of the spectra is far too large to be used directly in the regression models as it would make poorly designed models. A regression model should consist of only the essential wavelengths that contribute to the prediction and show a real relationship. This is where effective wavelength (EW) selection comes in.

No technique works best for selecting EWs in every situation as it depends on the size of the dataset and the required accuracy of the prediction.

Wavelength selection algorithms are much better than randomly or brute-forcing the wavelengths to make the best model. If we assume that the model will pick 20 wavelengths from the 186 wavelengths captured (from VNIR), there are many possible combinations to choose from. To be more exact, this can be calculated using the binomial coefficient: $\binom{186}{20}$ which is roughly equal to $3.5 * 10^{26}$ combinations.

3.2.2.1 PLS - Uninformative Variable Elimination (UVE)

This section involves Partial least squares (PLS) which is explained later in section 3.3.3.

The coefficients in the regression line explained by PLS tell us how much the independent variable affects the measurements, Y. A more significant coefficient, regardless of its negative or positive, means that the independent variable is more important in explaining the relationship between X and Y. This can be used to eliminate uninformative variables as they tend to reduce the accuracy of the model. It is done by creating a PLS model, taking the absolute value of each coefficient, then removing the independent variable associated with the lowest coefficient. For the next iteration, the new PLS model is created by one less independent variable, and the same elimination process continues. The elimination process happens until a minimum number of variables specified by the user is reached.

3.2.2.2 Successive projections algorithm (SPA)

One of the simpler variable selection algorithms is SPA. It finds the N most unique wavelengths based on the spectra given (independent variables), where N is the number of wavelengths to be selected. It does so by starting with one wavelength, then incorporates a new wavelength for each iteration until N number of wavelengths is reached. The essence is to select those wavelengths whose content is minimally redundant and to reduce multicollinearity.

Initially, the algorithm needs one wavelength and the number of wavelengths to be selected (N). With the initial wavelength it creates a column vector from the independent variables (spectra) $X_{i,j}$, that we call $x_{initial}$. For all of the remaining column vectors, it calculates the projection onto the orthogonal subspace using the equation 3.10. The subscript j represents a wavelength, and x_j its column vector. P is the projection operator.

$$Px_j = x_j - (x_j^T x_{initial})x_{initial}(x_j^T x_{initial})^{-1} \tag{3.10}$$

The wavelength corresponding to the most extended projection, Px_j, is to be selected. This is calculated by comparing the length of each projected column vector, $||Px_j||$, and selecting the biggest one. For the next iteration, $x_{initial}$ is the column vector of the newly selected wavelength. It iterates until $N-1$ has been selected. $N-1$ because the user-specified the first one.

Because SPA is dependent on two initial values, it is often combined with a prediction model, e.g., MLR 6.5.2, and a loop that iterates through each wavelength and each N. The best wavelength subset can thTest all candidatesen be obtained by comparing how well each model performs.

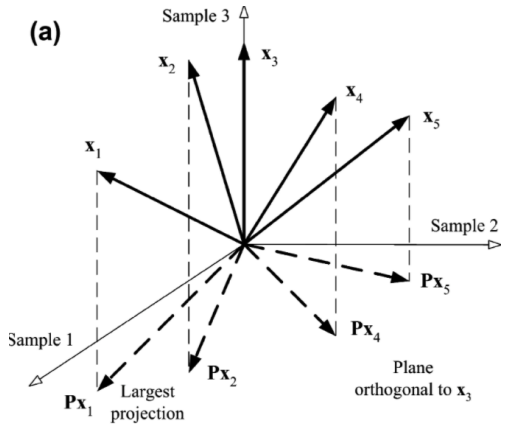


Figure 3.7: Simplified example of 5 measurements each containing 3 independent variables (called samples in this illustration) forming 5 unique column vectors. By selecting x_3 as the initial vector, the projection operations lead to the selection of x_1, as it presented with the longest projection(Image source [36]).

3.2.2.3 Recursive feature elimination (RFE)

The RFE eliminates features recursively until N variables are selected. The variables are scored using regression coefficients. If cross-validation is used, this is repeated for all parts of the cross-validation.

1. Select all features.
2. Calculate importance using a regression model.
3. Prone features with least importance.
4. Repeat until N variables are selected..

Since the best variables are kept and restored, these get a much higher score than the rest. If N is small, some variables have much higher importance values than the rest. This gives spikes in the band-importance plot.

3.2.2.4 Genetic algorithm (GA)

The Genetic algorithm (GA) implements Darwinistic perspective on variable selection. In other words the methodology is made to simulate evolution. Several candidates are tested, with different variables used then the best ones are selected for further processing. The ones selected are used to make new candidates to be tested. The procedure of the algorithm is also explained in the list below.

- Test all candidates.
- Select the best candidates to breed.
- Perform crossover and mutation.

A number of candidates to test each iteration is selected. Each candidate randomly selects bands to use then the fitness of each candidate is measured through

a machine learning model. For all of these fitness scores, the best ones are selected to breed into a new iteration; this cycle is repeated until an acceptable fitness is achieved.

The breeding process involves both crossover and mutation. With two selected candidates to breed, the children inherit 50/50 from each parent. Mutation on the variables selected by the children is performed to include variety, this involves randomly including or excluding variabels.

GA is used in one other paper performing non-destructive ripness determination of kiwi [9]. It has been proven to be a effective wavelength selection method combined with Partial least squares (PLS) on spectral data [37].

3.2.3 Data split using SPXY

When training machine learning models, a train test split has to be applied to the dataset. This is necessary as evaluating the performance of a model using the same data previously trained can lead to biased conclusions. By having a separate dataset for testing, we can be sure that the model can perform well on new data and is not overfitted.

A random train test split is not always the best way to split the data. The data in the test set does not necessarily provide a representation of the real-world scenario. Without further involvement, random test splits would give different accuracy depending on how the shuffle was done. This is why algorithms to pick the best data points for training and testing consistently have been developed. When working with spectral data, the Kennard-Stone (KS) algorithm[38] calculates a Euclidean distance matrix. This matrix contains the euclidean distance between all of the independent samples in a matrix x. This distance matrix is then used to make a test set representing the dataset as a whole. The training set is selected to represent the contour of the data feature space, while the test set its content.

It does this by selecting two samples in x with the highest distance between them into a subset S, which will eventually become the samples selected for training. Then the minimum distance between all the rest of the samples in x and the selected S is calculated. From these minimum distances, the maximum distance is selected and added to the subset s. These steps are repeated until k samples are selected into S, depending on the test size.

While the normal KS algorithm uses the regular Euclidean distance between all the samples, sample set partitioning based on joint x-y distance (SPXY) also includes the dependent samples y. The normal euclidean distance used by the KS is shown in equation 3.11, where M is the number of variables in each sample and $d_{KS}(i, j)$ is the distance between sample i and j. The variable t goes through all the measurements in the sample.

Equation 3.12 shows the SPXY distance formula. This formula implements the euclidian distance used in KS on both x and y, and divides these by the maximum distance, which results in a maximum value of 2[38].

$$d_{KS}(i,j) = \sqrt{\sum_{t=1}^{M} \left(x_{it} - x_{jt}\right)^2} \tag{3.11}$$

$$d_{SPXY}(i,j) = \frac{\sqrt{\sum_{t=1}^{m} \left(x_{it} - x_{jt}\right)^2}}{\max_{i,j\in[1.n]} \sqrt{\Sigma_{t=1}^{m} \left(x_{it} - x_{jt}\right)^2}} + \frac{\sqrt{\Sigma_{t=1}^{m} \left(y_{it} - y_{jt}\right)^2}}{\max_{i,j\in[1.n]} \sqrt{\Sigma_{t=1}^{m} \left(y_{it} - y_{jt}\right)^2}} \tag{3.12}$$

This data-splitting algorithm was used because it selects the best samples for training and testing. Also, it has been used before to perform ripeness determination of kiwi [39]. Its accuracy has however been doubted by Birba ([40]). However, this study seems not to use spectral data.

3.3 Multivariate models

Multivariate models are models that use multiple independent variables to predict one outcome. Since most of the changes in the real world are dependent on more than one variable/event, it is necessary to include multiple variables to be able to have good accuracy in the model.

There are many multivariate models to chose from, but the ones we will introduce here are standard chemometric methods and have shown good predictive power regarding spectral data. These are MLR, PLS, and SVR, but we will also introduce artificial neural nets and K nearest neighbor.

3.3.1 Multiple linear regression (MLR)

Multiple linear regression is a statistical technique and a generalization of simple linear regression to take on more than one independent variable. It assumes that the relation between the independent X and dependent Y variables is linear and tries to fit the best line to the data[41]. This fitting is done by estimating a coefficient β_p for each independent variable in X and an error term ϵ. The error term tries to correct for unobserved noise between the linear relationship between X and Y. For one dependent variable y_i, the equation would look like this 3.13:

$$y_i = \beta_0 + \beta_1 X_{i,0} + \beta_2 X_{i,1} + ..\beta_{p-1} X_{i,p-1} + \epsilon_i \tag{3.13}$$

Where i is one of the samples ($i = 1, .., n$) and p an index number for each independent variable ($p = 0, .., m+1$), m is the number of independent variables. An equation that covers all of the samples can be written using matrices. The shape of each matrix would look this:

$$Y = \begin{bmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \dots & X_{1,p-1} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \dots & X_{n,p-1} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (3.14)$$

Which allows all of the equation to be written together as matrices.

$$Y = X\beta + \epsilon \quad (3.15)$$

To find the best β coefficients, we want to minimize the Sum of Squared Errors (SSE). By using the fact that any one-dimensional vector has this property, $v^T v = ||v||^2$, we can write the SSE for matrices as follows:

$$\epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta) \quad (3.16)$$

By taking the derivative on both sides with respect to β, setting the equation to 0 (a vector of zeros), and solving for β, we obtain the minimum value[2] for 3.16.

$$\beta = (X^T X)^{-1} X^T Y \quad (3.17)$$

3.3.2 Support Vector Regression (SVR)

In ordinary linear regression, the goal is to minimize the sum of the squared error to get the best fitting regression function. That is achieved with a constraint of 0, meaning all points that do not lay on the fitted line are defined as an error.

$$MIN(\sum_{i=1}^{N} (y_i - w_i x_i)^2) \quad (3.18)$$

Constraint:

$$|y_i - w_i x_i| = 0 \quad (3.19)$$

y_i is the actual value, w_i is the coefficient and x_i is the feature.

In Support vector regression (SVR), we want to define a margin of error, ϵ, for the regression function to decrease how strictly the hyperplane fits the data points. As we most likely cannot include all data points within the margin of error, we need to include another variable, ξ, to measure the error of all points outside the margin of error. This works well in specific datasets as there is often a margin of error in real-world data, and the points outside the margin of error can be viewed as noise or outliers. With the constraints ϵ and ξ, we can define our boundaries for the hyperplane:

$$|y_i - w_i x_i| \leq \epsilon + |\xi_i| \quad (3.20)$$

[2]More details can be found in `https://www.stat.purdue.edu/~boli/stat512/lectures/topic3.pdf`[42]

So with the constraints in mind, we want to minimize the Euclidean distance of the data points to the margin of error. Before we do that, we need to express the orthogonal distance between the closest points of different samples by a vector w.

First, we need a unit vector, v, that is orthogonal to the margin of error on the opposite side. Then we need a vector, v, that expresses the distance between the samples on the margin of error. Vector v is found simply by subtracting the vector from the origin of the red sample by the vector from the origin to the green sample in figure 3.8. Now, by taking the dot product of u and v, we get a projection w that expresses the orthogonal distance between both sides of the margin of error. For a better visualization, see figure 3.8.

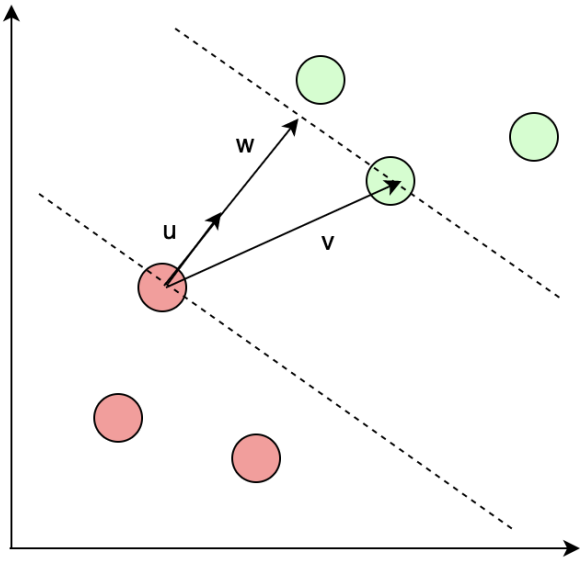


Figure 3.8: The orthogonal distance, vector w, is found by projection of unit vector, u, dotted with vector v.

Now we have an expression $\frac{2}{||w||}$ that we got by dotting unit vector u and vector v, which represents the width between the samples. We can maximize this expression to get a hyperplane that separates the features the most. Since w is in the denominator, maximizing the expression is the same as minimizing $||w||$, which allows us to rewrite as $MIN(\frac{1}{2}||w||^2)$.

Finally, we can minimize w and add a constant C to scale the dependence of the sum of errors of the points outside of the margin of error:

$$MIN(\frac{1}{2}||w||^2) + C\sum_{i=1}^{n}\xi_i$$

There are different methods of tuning the scale of the constant C, and more

explanation of SVR in detail in this footnote[3].

There is a large margin of error between different kiwis and scanning days in our project, making SVR interesting compared with ordinary linear regression. The margin of error and weight of points outside the margin of error can be tuned; it brings more hyperparameters that can influence results in various ways, making SVR a good candidate model.

3.3.3 Partial least squares (PLS)

Partial least squares (PLS) works quite differently from regular linear regression. Instead of working directly on the independent and dependent variables, PLS uses the components derived from Principal component analysis (PCA) to predict the dependent variable. PCA transforms the variables into latent variables called components, which reduces the dimensionality of the data. All of the components will explain the original data; however, only some components are needed to explain the majority, reducing the dimensions needed. These components consist of two matrices called scores and loadings used by PLS to make a prediction[39] [43].

PLS is a popular regression model as it is often used in chemometrics. This is because it is more robust than linear regression, and the model parameters do not change significantly when more data is used for calibration. It also works well when there are more independent variables than dependent. It also works well where there exists multicollinearity, meaning there is a correlation between independent variables[44]. Collinearity can often make a case for spectra because a lot of the wavelengths vary with each other.

The number of components used is compared towards a performance metric (often MSE or R^2) in order to decide on how many components to use in PLS regression.

3.3.3.1 Eigenvectors

An eigenvector can be called the characteristic vector of a matrix (A). This is because it does not change if matrix A is scaled or transformed; only the eigenvalue associated with that eigenvector changes. Formally this can be written as $T(v) = \lambda * v$, where v is the eigenvector, λ its eigenvalue and $T(v)$ is a transformation done on v.

An eigenvector of a matrix A can be solved as $|A - I * \lambda| = 0$ where I is the identity matrix.

The components calculated by the PCA algorithm can be seen as the eigenvectors of the covariance matrix. It can help understand the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm, which PLS algorithm implements.

[3] https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

3.3.3.2 NIPALS

To compute the principal components (scores and loading), the NIPALS algorithm is used. The algorithm works iteratively until convergence for each desired component.

All the measured independent variables are measured in a matrix X. This matrix is mXn long with n number of samples and m number of independent variables. All the data in X are mean-centered.

Firstly the algorithm takes a column (independent variable/ wavelength) of X called x_j. Then the covariance between that independent variable and the rest of the dataset is computed into a variable p called loadings, which can be seen as a new dimension of the data. To compute where the points are on this new p dimension, the covariance between p and X is calculated into variable t (scores). the calculation of t and p is repeated (step 2-3) until t is equal to t used in step 2 with some error.

1. $t = x_j$
2. $p = covar(t,X) = \mathrm{X^T t} / \left\|\mathrm{X^T t}\right\|$
3. $t = covar(X,P) = X * p$

These are the steps to calculate 1 component of X. Once one component is calculated, the component's data explains in X must be removed to calculate the next component. This is done by subtracting the component from X by $X_{new} = X - p * t^T$.

Once all of the components are retrieved, X can be written as the sum of all the components $X = \sum_i^N t_i * p_i^T$, where N is the number of components, and i is the i-th component.X can also be written as $X = t_0 * p_0^T + e$, where e is the sum of all the other components and N is the number of components.

It is shown that the NIPALS algorithm gives the same results as the eigenvector calculation on the covariance matrix of the data [43].

3.3.3.3 PLS Theory

As mentioned earlier, PLS tries to do a regression based on the components, not the values themselves. It consists of an inner relation linking individual scores from X and Y and an outer relation that looks at the entirety of the component. This can be shown illustrated graphically in figure 3.9.

Just doing the NIPALS algorithm for both Y and then X, and then compare the components in both of the variables is an option. This is done by setting $X = t * p^t$ and $Y = UQ^t + F$. t and p are the scores and loadings of x, U, and Q are the same for Y. While F is defined as the error. Then U is regressed by t through $U = b * t$, where b is the regression coefficient. This can be solved by $b = \frac{u^T * t}{t^T * t}$.

This is called the inner relation. However, the outer relation looks at the scores and loading in X and Y dotted together (also known as PCA component). The goal of PLS is to minimize the error F in Y while keeping the inner relations between the scores and loadings [46].

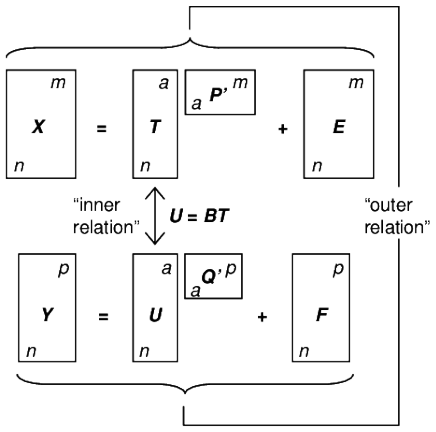


Figure 3.9: PLS has both inner and outer relation (Image source [45]).

A better way is to do the NIPALS algorithm on X and Y at the same time. This is the same as doing them after each other; however, the components for X are done towards a column of Y. Also, the components for Y are done towards the t the loadings component of X. The steps for the combined NIPALS algorithm are shown below [47].

1. $u = y_j$
2. $p = covar(u, X) = \mathrm{X^T u} / \left\| \mathrm{X^T u} \right\|$
3. $t = covar(X, P) = X * p$
4. $q = covar(t, Y) = \mathrm{Y^T t} / \left\| \mathrm{Y^T t} \right\|$
5. $u = Y * q$

The components retrieved from Y are almost equal to the regression method used before $U = b * t$, while the components for X are done towards one Y column.

PLS is highly relevant for this project because there exists much multicollinearity in the spectra, which other methods can struggle to cope with.

3.3.4 k-nearest neighbors algorithm (KNN)

K-nearest neighbor (KNN) uses "K"-neighbors in multidimensional space to make a prediction. It finds these neighbors by calculating the euclidean distance between the variable to be predicted and all the samples in the training set. Then it takes the mean of the dependent variable in these neighbors, which is the resulting prediction[48].

It seems like KNN has not been used for this application before. However, since it was originally developed for classification, it has been used to classify agricultural mold products. For example Xie *et al.*[49] used KNN to classify healthy and gray mold in tomato leaves.

3.3.5 Artificial neural network (ANN)

Artificial neural network (ANN) is made of many layers of nodes with interconnected connections with weights. Inside the nodes, there is an activation function. The input of this function is the sum of all the connections to that node times their weights. The output of that node for the next layer is the value the activation function produces.

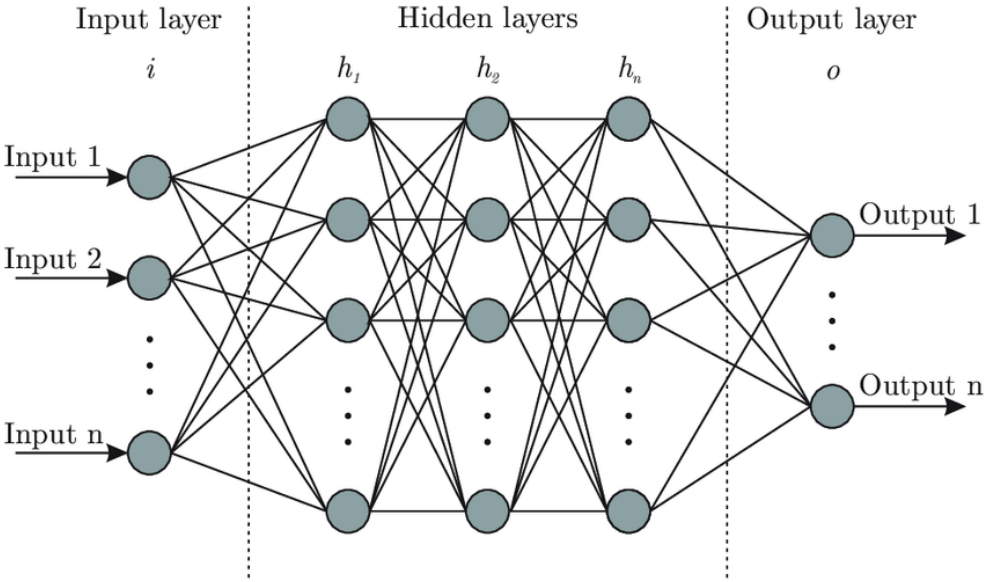


Figure 3.10: Architecture of a standard Artificial neural network (ANN) (Image source [50]).

Backpropagation is used to blame each weight for the total error. Once each connection is blamed with an error, gradient descent is performed to correct for that error. The gradient decent is performed through a loss function, which it tries to minimise.

If the higher the weight, the stronger the connection is. This is a sign of overfitting. There are techniques available to reduce high weights and overfitting.

The dropout layers would randomly drop out nodes from the neural network while training. The nodes left are then more responsible for the connectivity because more and more nodes are removed [51].

L2 regularization adds a penalty term to the loss function, which prevents large value weights. This term is a constant L2 times the error of the network applied the weights[52].

A callback can stop the training before the number of set epochs is finished (EarlyStopping). This is to stop over-training the model, stopping when the loss value has not improved.

Multilayered Perceptron (MLP) is a feed-forward artificial neural network. It means that each node is connected to the nodes above itself in the layer.

MLP and ANN has been widely used to perform classification on HSI. For example ElMasry *et al.* [53] used MLP and to detect chilling injury in Red Delicious apples. However it seams like it has not been applied to regression problems on agricultural products using HSI.

3.3.6 Tree regressor

Tree regressor performs regression with several decisions trees. Each measurement in one sample is mapped as input to one decision tree. This means that when using a tree regressor, there will equally as many decision trees as measurements in one sample.

These decision trees works similarly to ANN, however it is shaped like a tree and has inequalities inside the nodes instead of activation functions. The inequality inside the node follows this equation: $a \geqslant w * x \leq b$. These coefficients are then fitted using gradient decent and l2 regularizer[54].

3.4 Model evaluation

The most common way to evaluate the performance of a model is to divide the dataset into three subsets; calibration, validation, and verification. The calibration set is used to train the model; then, the validation set is used to find which parameters give the best result. After the best model parameters are selected, the verification set is used to verify the model's performance.

By testing the model on different datasets, we decrease the chance of overfitting as the model adapts to the general trend. By increasing the number of subsets, the quality decreases as there is less data in each subset. The balance between quality and quantity in each subset must be found to create a good model with low bias.

3.4.1 Evaluation metrics

In order to evaluate how well a model is explaining the data and how it performs when predicting new values, it is necessary to establish good evaluation metrics.

3.4.1.1 R-squared

R-squared (R^2) is a measurement of how accurate a prediction is. It is defined as the amount of total variation the model can recreate from the independent variable[55]. Its maximum value is one meaning all the variation can be explained from the independent variable. The lowest value possible value is 0, meaning the independent variable explains no variation.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i^N (y_i - f(x_i))^2}{\sum_i^N (y_i - \bar{y})^2} \tag{3.21}$$

In equation 3.21, x_i is the independent variables, y_i is the dependent variable, and $f(x_i)$ is the predicted value of x_i. $\bar{y}$ is defined as the mean of the variable y The variable i goes through all the N samples. The squared sum of residuals (SS_{res}) is defined as the squared error of the prediction. The total squared sum (SS_{tot}) calculates the standard deviation of the variable y.

When the prediction is perfect, the SS_{res} term would be 0 because $\forall_i : y_i = f(x_i)$, which results in the R^2 term to be 1. The more the $f(x_i)$ is not equal to y_i the closer R^2 would be to zero.

However, using R^2 for evaluating multiple regression results can lead to biased conclusions[55]. This is because R^2 provides no information if the model includes useless variables or even fits the data. Because of this, it is important to be aware of how many independent variables the model uses and draw conclusions based on the combined result from multiple metrics.

3.4.1.2 Mean squared error

Mean squared error (MSE) is defined by similar terms used above (Equation 3.22). MSE calculates the average squared difference between the predictions and ground truths. The error is squared to get positive values, and to have a larger effect on bad predictions. The absolute value of the error is also often used (called mean absolute error), but is not as mathematically convenient as MSE. Root mean squared error (RMSE) which is defined as the root of the MSE (equation 3.23) is however pretty close to the mean absolute error.

Compared to R^2, the value of MSE depends on the standard deviation of the dependent variable it predicts. This metric can not be compared to a model that transforms the features to change the standard deviation.

In the equation 3.22, N is the number of points, y_i is the ground truth while $f(x_i)$ is the prediction of y_i.

$$MSE = \frac{SS_{res}}{N} = \frac{\sum_i^N (y_i - f(x_i))^2}{N} \tag{3.22}$$

$$RMSE = \sqrt{MSE} \tag{3.23}$$

3.4.2 Cross-validation

Cross-validation is a model evaluation method where the calibration set is divided into K subsets, where the model is then trained and tested K times. Each subset is used once as a validation set, and after K evaluations, the average of all iterations is used as a model training result. See figure 3.11 for a better visualization.

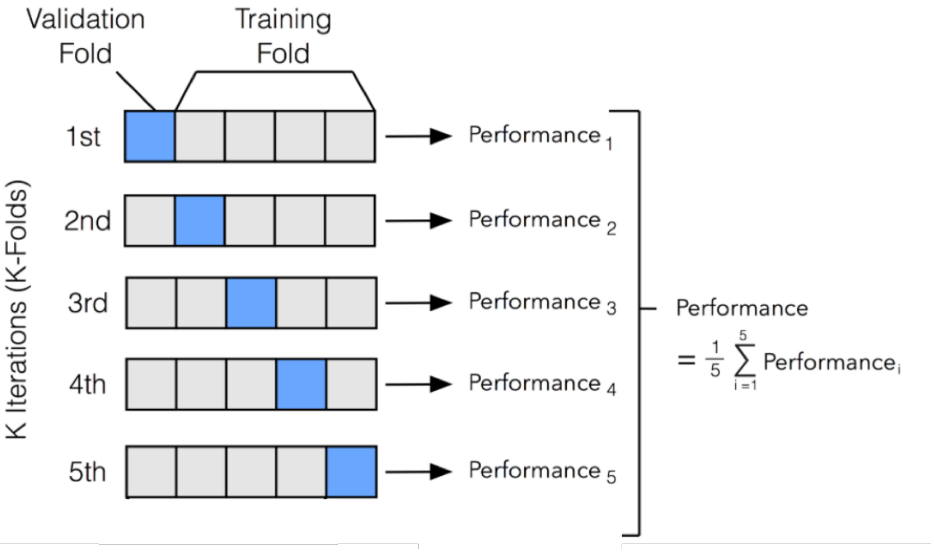


Figure 3.11: 5-fold cross validation (Image source [56]).

3.5 Fruit material

The kiwifruit "Hayward" originates from the species "Actinidia deliciosa" and is just one of many cultivars. Because of its many characteristics, such as its large size, good taste, and long-term storage potential, it sets the industry standard compared to the other cultivars. It became widely exported around the globe [57].

3.5.1 Difference between maturity and ripeness

Maturity and ripeness are commonly used as synonyms; however, they have different definitions. Maturity is used in the period before the fruit or vegetable is harvested. On the other hand, ripening starts when the color, flavor, and texture are fully developed (matured). After achieving ripeness, the produce starts its spoilage process [4].

3.5.2 External quality parameters

Many different types of kiwi are cultivated, and each of them has a different composition of pigments that affect their color. Researches[58] showed that in "Hayward" kiwi type, the color is almost the same- green at all stages of development, maturation, ripening, and its outer- and inner- pulp does not change much when fruit becomes ripe. When it comes to kiwi's weight, it is also very different for each kiwi type. Commercially produced "Hayward" kiwifruits typically weigh 95–115g.

[4] http://eagri.org/eagri50/HORT381/pdf/lec03.pdf

3.5.3 Internal quality parameters

There are many different physiological and internal parameters that can affect on kiwi's quality. In our project, we mainly focused on SSC, firmness, pH.

Firmness is a physical property that is one of the most important indices of ripeness and quality of many fruits and vegetables. In general, the firmness of fruits decreases moderately as they become more mature, and the firmness decreases quickly as they ripen. Overripe and damaged fruits become soft. Chen and Sun(1991) in a review of available methods for quality evaluation and sorting of agricultural products, concluded that the industry accepts firmness as a criterion for sorting agricultural products into different ripeness groups[60].

SSC is another critical quality parameter used for quality predicting. It indicates the sweetness of kiwifruits in laboratories for research and industry. It is measured in degrees Brix, which represents the sugar content of fruit juices.

pH measures the relative amount of free hydrogen and hydroxyl ions in the water or other liquid solutions. It says about how acidic a liquid is, and it has a range of pH that goes from 0 til 14, where seven is neutral. A level of less than seven indicates pH, whereas a pH level greater than 7 indicates a base. When it comes to fruits, each fruit has a different pH level, and it can change during ripening.

3.5.4 Harvest and storage

Harvest time for kiwi sort "Hayward" is in Europe from October until April, whereas in New Zealand, Chile, and Australia, it varies from May until September[61]. Since SSC is the most used ripeness feature for kiwifruit, type "Hayward" is usually harvested at a 6.2°Brix[62] before they have finished maturing. By doing this, the fruit will remain in good condition while in storage. When this minimum maturity standard is achieved, all the kiwifruits can be harvested in one picking. This process is done by handpicking. When awaiting transport, kiwis need to be shade and cool as quickly as possible. Doing this maximizes storage. Later kiwifruits are transported to destinations and left for softening. Softening of crops depends on the temperature at which they are stored. Kiwis soften three times faster at 5C than kiwi at 0C. Kiwifruits should be stored at 90 to 95% relative humidity and at temperatures as neat to 0C as possible, but not below that, because these fruits are sensitive to freezing injury. Even when held in these conditions, they are maturing. About one-third to one-half of the remaining flesh firmness can be lost per month for type "Hayward." "Hayward" fruits can be stored from 3-6 months under proper storage conditions [63].

Chapter 4

State of the art

In this chapter, the papers that present previous work done in detecting internal and external features of kiwifruits are presented. Each sub-chapter contains a short introduction to the quality metric and its predictions done with spectroscopy. The following can be found:

- soluble solids content (SSC)
- Firmness
- pH
- Other quality metrics used for kiwis

4.1 Soluble solids content (SSC)

The relationship between SSC measured by refractometer, kiwifruit ripeness, and eating quality was investigated by Ford in 1971[64]. He found that "Hayward" kiwifruit harvested with SSC that was less than 6.0°Brix did not give a good flavor when ripe. Because his conclusion was not based on many different chemical tests on kiwis from different orchards- he stated that this relationship needs further investigation. That is why other researchers started to work on this subject to investigate previous results further. Few years later in 1977 Reid[65] studied "Hayward" kiwis from various places. The findings were similar to the previous and showed that kiwis harvested too early would never reach a final SSC as high to achieve pleasant flavor. In addition to that, these kiwis will have an overall poor quality. Subsequently, Harman(1981)[66] drew the same conclusion about low taste-panel scores for harvesting at below 6.0°Brix and recommended a minimum SSC value of 6.2°Brix for "Hayward" kiwis. This recommendation was accepted by the New Zealand kiwifruit industry and led to regulation that harvesting of "Hayward" kiwi can be done when these have reached a minimum of 6.2°Brix.

Mentioned 6.2°Brix level standard has, however, showed later not to be an optimum. Hopkirk *et al.*(1986)[67] concluded in the study that there is a considerable variation in SSC that occurs between individual fruits in orchards at any time. In addition to that, SSC value also varies along the length of kiwis. Because

of these reasons, there is a possible sampling error involved, which can lead to difficulties in establishing the generalized relationship between ripeness index at harvest and final eating quality based just on Brix refractometer value.

4.1.1 SSC as a predictor

Mitchell together with other co-workers ((Mitchell and Mayer 1987),[68],(Mitchell, 1990)[69]) found that SSC level in kiwifruit "Hayward" at the time of harvest was not a reliable technique for predicting SSC. This is because the huge amount of carbohydrates are still in the starch form at harvest. The fact observed by Reid *et al.*(1982)[70] showed that kiwifruit SSC continues to increase for weeks after the harvest and that this increase leads to increases in the concentrations of fructose, glucose, and sucrose. They showed as well that there occurs a fall in starch content. Mitchell *et al.* (1989)[71] confirmed these facts and stated that SSC of ripe kiwifruits could be used as a good predictor of ripeness and eating quality of kiwifruits.

4.1.2 Spectroscopic predictions of SSC in VNIR spectral range

There are available several types of kiwifruits, and each type has a different SSC level. Even the same kiwi type, but grown in a different country or region, can differ from others. Macrae *et al.*(1989)[72] compared kiwi "Hayward" from two orchards. They founded that these were not similar and that the major differences were in total starch concentration. As said previously, starch in kiwifruits changes with time into SSC and this fact may affect both how the fruit ripens after harvest and the final sugar level. This may explain the fact that there are different R^2 values achieved in previously done studies.

McGlone and Kawano(1998)[73] did a big study on "Hayward" kiwis. They used one large sample set of kiwis and four smaller ones. Each of them had a different orchard origin. The biggest set, called KFRM, consisted of 393 New Zealand kiwifruit, which had a large final firmness range. The other four sets were as follows: KOLD(61 Japanese kiwifruits), KMKT(86 New Zealand kiwifruits from a retail market), KSSZ(43 smaller sizes New Zealand kiwifruits), KLSZ(41 larger New Zealand kiwifruits). For data analysis, they used three datasets. They combined sets KOLD, KMKT, KSSZ, and KLSZ to one set-KMIX, consisting of 231 kiwifruits. They combined also all 5 sample sets into a KALL set, that consisted of 624 kiwifruits. All these sets were scanned by NIRSystem 6500 spectrophotometer that had the spectral range 400-1100nm. Modeling was done using the partial least squares method (PLS) with a random four-way cross-validation. Results of prediction were different for each of these sets. Dataset KFRM had R^2=0.88,RMSE=0.37, KMIX had R^2=0.85, RMSE=0.47 and KALL had R^2=0.90, RMSE=0.42.

In the same year- 1998 Martinsen and Schaare[74] presented their results on the 650-1100 nm spectrum range. They also used kiwi "Hayward," but exclusively from New Zealand. There were 1800 kiwis used for that study. Six hundred

of these were harvested at very low SSC, down to 4.7°Brix, and 1200 were collected during the normal commercial harvest 6.2°Brix. The other ones were allowed to ripen for three weeks at room temperature. Each day SSC of end slices from 6 kiwifruits were measured by refractometer. When there was an increase of 2°Brix founded at five kiwifruits, 30 kiwis were placed into cool storage at 0°C to await further analysis. Then the analysis was completed within five days from that time. Interval Partial Least-Squares Regression (iPLS) algorithm was used to make a conclusion based on SSC values and achieved spectrum. All this resulted in R^2=0.834 and RMSE = 0.39. These results were worse than ones presented above by McGlone and Kawano(1998)[73].

Osborne *et al.*(1999)[75] tested also a low cost NIR spectrometer on 322 kiwis from New Zealand to estimate SSC and DM at reasonably high speed. PLS regression in conjunction with wavelength selection provided models with RMSE of 0.27°Brix for SSC. This error was below previously published results by McGlone and Kawano(1998)[73]. Authors meant that their findings would allow the grading of kiwifruit by SSC into several classes with distinctly different taste attributes.

4.1.3 Spectroscopic predictions of SSC and TSS in SWIR spectral range

SWIR spectral range is not as widely used as VNIR. However, SWIR was used by some researchers, and good results were achieved.

Lee *et al.* in 2012[7] used a spectral range of 408 to 2492 nm for predicting quality features of 1530 "Hayward" kiwifruits that came from three different farms. Kiwifruits were ripened at 20°C on the day of harvest, and from the next day, 90 fruits from each of the three farms were evaluated by SWIR spectrum every five days. All this lasted 25 days. Prediction based on achieved results was calculated using the modified PLS regression method. Very good results were achieved for SSC. The prediction was: a standard error of prediction(SEP)[76]= 0.49°Brix and $R^2 = 0.98$.

Later in 2017 Li *et al.*[8] published also their work that used similar spectral range: 350-2500 nm. Their results were not as good as these of Lee *et al.* They used 1040 "Hayward" kiwifruit samples for prediction of Total soluble solids concentration(TSS)[1]. Li2017Jun used both PLS and SVM for prediction. Best results were achieved by SVM giving $R^2 = 0.83$ for TSS and RMSE = 1.02°Brix.

4.2 Firmness

The firmness and ripeness of fruits at harvest time, storage, and shelf life vary greatly. This is the case also for kiwifruits. At harvest, "Hayward" kiwifruits are in range 60-110N[77], and when at "eating ripe" approximately 3.9–7.8 N as Stec *et al.* founded in 1989 [78].

The ripe kiwifruits produce much more ripening hormone than unripe ones, so if the soft fruits are not separated from the firm ones, the increased ripening

[1]Total soluble solid means amount of total soluble solid present in the unit volume of solution

reduces the storage and shelf life of the whole batch. That why it is vital to could accurately measure their firmness and separate them.

There are many different methods for measuring firmness, both destructively and non-destructively. The most common one is by using a penetrometer. Some others methods were developed, but there are very rarely used today (Chen and Sun(1991)[59]).

4.2.1 Spectroscopic predictions of firmness in VNIR spectral range

In the study of McGlone and Kawano from 1998[73], mentioned already in 4.1.2, researchers tested also prediction of firmness on their sets. The combined KMIX set that consisted of four subsets with a total number of 231 kiwifruits achieved the worst results of R^2=0.42, RMSE=11.8 N. The results were better for a set KALL that consisted of five subsets with together 624 kiwifruits. R^2 resulted in 0.66 value and RMSE= 7.8 N. The best result was on the set KFRM- a set of 393 New Zealand kiwifruit of a large final firmness range. Within this narrow, in terms of orchard origin and size, the kiwifruit dataset was produced a better firmness model giving a R^2=0.76 and RMSE= 7N. The authors stated in conclusion that: "this model performed poorly against independent datasets, suggesting the influence of secondary correlations due to fruit characteristics that are not directly related to fruit firmness"[73]. For all these sets, a PLS was used as a model applied on the spectral range of 400-1100 nm.

The results from more recent studies were different. Berardinelli *et al.* in 2019[79] used a prototype based on a NIR sensitive camera and a Xenon lamp to scan "Hayward" kiwis. Their prototype was set up and used to capture 8-bit grayscale images of the radiation passed through the kiwifruits. Then the sum of the pixels with distinct gray tone was used to build models. Researchers used 116 kiwifruits with firmness values measured by a penetrometer in the range of 0.8N to 87N. They used different techniques for prediction, but the best result was achieved by the PLS algorithm that allowed prediction of the firmness with R^2=0.777 and RMSE = 13N.

4.2.2 Spectroscopic predictions of firmness in SWIR spectral range

Lee *et al.*(2012)[7] and Li *et al.*(2017)[8] have used SWIR range for predictions of firmness. Lee *et al.* achieved SEP of 3.32 N and $R^2 = 0.88$ using "modified" PLS on the spectral range 408-2492 nm. Their results was much better than those of Li *et al.* These researchers used both PLS-r and SVM-r on the spectral range 350-2500 nm. Models were tested on 2125 "Hayward" kiwifruits and resulted in $R^2 = 0.6$ and RMSE= 3.92 N for prediction.

4.2.3 Non-destructive measurements of kiwi's firmness in the field

In 1989 in Japan there was developed and commercialized a device- HIT-Counter I and later an improved version of it- HIT-Counter II[80] that could non-destructively

measure kiwis firmness. It evaluates the firmness by applying non-destructive compressive force. The machine applies the force within the range of elasticity to not damage an object, and no bruising occurs. It uses deformation as an index for evaluation.

Researchers carried out conventional, destructive fruit firmness tests to develop this device and compared it to those achieved by non-destructive machines. The relations (R^2= 0.927) between the HIT-Counter values and the internal quality features, including sugar and acid contents, were investigated. Findings showed that HIT-Counter was highly correlated with a destructive firmness tester. The correlation between its values and the pH values of kiwifruits was high (R^2= 0.821). No clear relation between the sugar content and device' values was observed. By this, developers concluded that HIT-Counter could be used as an index of the internal quality evaluation based on the pH.

The HIT-Counter II device did not require any technical skills and is easy to operate. There is a possibility to use it to evaluate ripeness, prime eating condition, quality control in shipping, quality evaluation during storage. Although the device was produced commercially over 30 years ago, it was not widely used in the industry.

In 1999 Peleg[81] developed a commercial fruit firmness sorter for various fruit types and shapes. The machine provided a system that allowed physical contact of the inspected fruits by a sensor on the line and, based on the achieved results to sort fruits into a firm or soft without producing any damage. This sorter was successfully calibrated and tested on "Hayward" kiwifruits. For the experiment, two samples were sorted by this machine. The first sample had 246 kiwis that were from a cold storage room. The second sample had 213 kiwis that were stored before sorting in a controlled atmosphere. After sorting, their firmness values were compared to those achieved by a regular penetrometer. Results showed larger mean firmness readings of the fruits stored in a controlled atmosphere, but the machine successfully separated kiwifruits into correct firmness ranges.

4.3 pH

The industry standards for quality control are SSC and firmness. However, there were also attempts to use pH for this. Although there are not many studies trying to predict pH level in kiwis, some researchers have gotten good results using spectroscopy for prediction.

4.3.1 Spectroscopic predictions of pH in VNIR spectral range

In 2010 Moghimi *et al.*[82] used Vis/NIR spectroscopy and chemometrics to predict pH of kiwifruit. They obtained transmission spectra of 100 "Hayward" kiwifruits in the wavelength range from 400 to 1000 nm. Their prediction models were developed using PCA and PLS. They resulted in quite high correlation coefficient(R)[83]. It was R=0.943 and RMSE 0.076, which are better than later find-

ings of Zhu *et al.*[9]. These authors have used HSI for predicting pH of kiwifruits based on variable selection algorithms and chemometric models. There were 133 kiwis of 3 different types: "Xuxiang," "Hongyang," "Cuixiang," and an HSI push-broom reflectance imaging system covering 450–1670nm spectral ranges was used for acquiring images of them. From all of the tried models (Genetic Algorithm–Partial Least Square) GAPLS [84] and (Least-Squares Support-Vector Machine)L S-SVM [85] gave the best results for R= 0.88 and RMSE=0.0152.

4.3.2 Spectroscopic predictions of pH in SWIR spectral range

As mentioned before, there are, in general, not so many studies that were done on pH. The only study that was available on the SWIR spectral range for "Hayward" kiwifruits was the one did by Lee *et al.* in 2012[7]. By applying a spectral range of 408 to 2492 nm on 1530 "Hayward" kiwifruits, they have gotten better results than researchers that used the VNIR range described in the section above. pH prediction was calculated using the modified PLS regression method and resulted in a SEP of 0.28% and $R^2 = 0.91$.

4.4 Other quality metrics used for kiwis

4.4.1 Skin color as predictor

There are around 60 species of kiwi that are cultivated[86], and each type has a different composition of pigments. This affects color as well as health-promoting effects of them as founded by Nishiyama *et al.* (2005)[87].

During fruit maturation, flesh color can vary, and most fleshy fruits are green only during the earlier stages of development. In kiwis used in this project, Actinidia deliciosa "Hayward" kiwis, which are also sometimes called Actinidia chinesis "Hayward," it is different. Researchers Montefiori *et al.* (2009)[58] compared these kiwis to three types of A. chinensis: "Wuzhi No. 3", "Hort16A" and "Jinfeng". These results are shown in 4.1. At the beginning of their study the pericarp tissue[2] of four varieties had a hue angle[3] of h° >110°. All these kiwis were recognized as being green. After an almost three-month period during which samples were harvested, there was only a slight drop in the hue angle for "Hayward" kiwis. "Hayward" kiwifruits appeared green at all stages of development: maturation, ripening, and its outer and inner layers retained their chlorophyll[4] Figures nr: 4.2, 4.3 show this. Because of the lack of changes, the ripeness level of kiwi type "Hayward" can not be predicted just by skin color and hue angle.

[2]In fleshy fruits, the outer layer is the pericarp. The pericarp- "is the tissue that develops from the ovary wall of the flower and surrounds the seed to protect it in environments apart from the parent plant"[88]

[3]Hue is one of the main properties of color; typically represented by a single number, often corresponding to an angular position around a central or neutral point or axis[89].

[4]Chlorophyll is a green pigment found in plants.

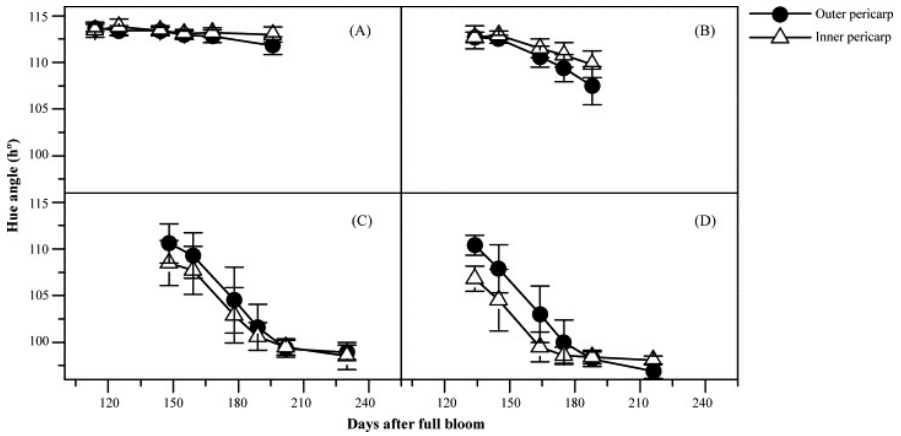


Figure 4.1: Outer and inner pericarp colour in fruit of Actinidia deliciosa "Hayward" (A) and A. chinensis: "Wuzhi No. 3" (B), "Hort16A" (C) and "Jinfeng" (D) developing on the vine. Data points are means ± standard errors. (Image source [58]).

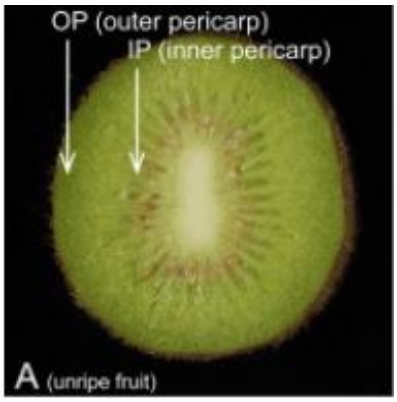


Figure 4.2: Unripe "Hayward" kiwifruit.(Image source [58]).

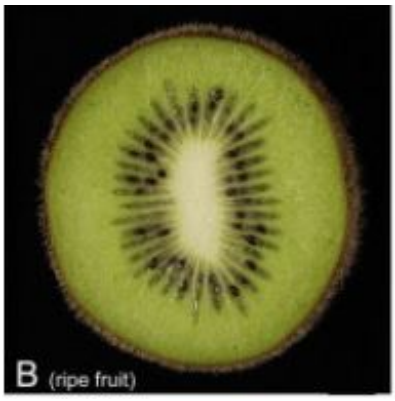


Figure 4.3: Ripe "Hayward" kiwifruit. (Image source [58]).

4.4.2 Density as a predictor

The density of a given substance is its mass per its volume; it varies with temperature and pressure, but this variation is typically small for solids and liquids[90], so it can be neglected. Density has been used as a good estimator for quality evaluation for vegetables, fruits[91], [92] and as well as for evaluating of ripeness of "Hayward" kiwifruits (Jordan *et al.* (2000)[93]).

4.4.3 Summary of previous research

Most of the papers done on "Hayward" kiwi have been on SSC and firmness determination. This was not surprising because both SSC and firmness are used as

standard parameters for the determination of quality and ripeness. Although there are much fewer studies on the SWIR range, they gave better results than VNIR. In fact, all of the best results were achieved on SWIR spectral range 408-2492 nm. Additionally, these results were produced by the same researchers Lee *et al.* (2012)[7] using "modified" PLS model. For SSC they achieved $R^2 = 0.98$, for firmness $R^2 = 0.88$ and for pH $R^2 = 0.91$(table 4.1).

There has been some confusion about the accuracy achieved by Zhu *et al.* (2017)[9]. Review articles on the subject have assumed that this article used R^2. However, this is not the case. This article used R and not R^2, and we recalculated their accuracy into R^2 so that they are more comparable to other results.

	Reference	Kiwi variety	Spectral range (nm)	Model	$RMSE_p$	R^2	Samples	Year
SSC	Lee *et al.*(2012)[7]	"Hayward"	408-2492	"modified" PLS	N/A	0.980	1530	2012
	McGlone *et al.*(2002)[91]	"Hayward"	300–1140	PLS	0.320	0.940	360	2002
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	0.390	0.900	624	1998
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	0.370	0.880	393	1998
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	0.470	0.850	231	1998
	Martinsen and Schaare(1998)[74]	"Hayward"	650-1100	iPLS	0.390	0.834	1800	1998
	Berardinelli *et al.*(2019)[79]	"Hayward"	NIR	PLS	1.400	0.550	116	2019
	Guo *et al.*(2016)[39]	"Xuxiang", "Huayou"	865–1711	FS-LSSVM	0.589	0.971	200	2016
	Schaare and Fraser(2000)[94]	Yellow-fleshed	300-1100	PLS	N/A	0.930	1000	2000
	Osborne *et al.*(1999)[75]	not stated, from New Zealand	300-1140	PLS	0.270	N/A	322	1999
TSS	Li *et al.*(2017)[8]	"Hayward"	350-2500	SVM-r	0.660	0.830	1040	2017
Firmness	Lee *et al.*(2012)[7]	"Hayward"	408-2492	"modified" PLS	N/A	0.880	1530	2012
	Berardinelli *et al.*(2019)[79]	"Hayward"	NIR	PLS	13.0	0.777	116	2019
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	7.00	0.760	393	1998
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	7.80	0.660	624	1998
	Li *et al.*(2017)[8]	"Hayward"	350-2500	SVM-r	3.04	0.600	2125	2017
	McGlone and Kawano(1998)[73]	"Hayward"	400-1100	PLS	11.8	0.420	231	1998
	Zhu *et al.*(2017)[9]	"Xuxiang", "Hongyang", "Cuixiang"	450–1670	LS-SVM	17.9	0.930	133	2017
pH	Lee *et al.*(2012)[7]	"Hayward"	408-2492	"modified" PLS	N/A	0.910	1530	2012
	Moghimi *et al.*(2010)[82]	"Hayward"	400-1000	PCA, PLS	0.076	0.889	100	2010
	Zhu *et al.*(2017)[9]	"Xuxiang", "Hongyang", "Cuixiang"	450–1670	GAPLS-LS-SVM	0.0152	0.823	133	2017

Table 4.1: Literature review of kiwifruits.

Chapter 5

Data acquisition

In this chapter the following can be found:

- Fruit Sampels
- Hyperspectral system
- RGB imaging
- Physiological measurements.

The data acquisition lasted five weeks, where two of these weeks were used for a pilot project. We learned during it about the scanning environment and procedures. Data acquisition also involved HSI, RGB imaging, and physiological measurements. Among these physiological measurements were salt, SSC, firmness, and pH level. Both the kiwi and kiwi peels were captured with HSI. During each week, the data collection took place for three days with one day in between, e.g., Monday-Wednesday-Friday. This gave the kiwi time to ripen, as all of the kiwis were gradually obtained with the same ripeness.

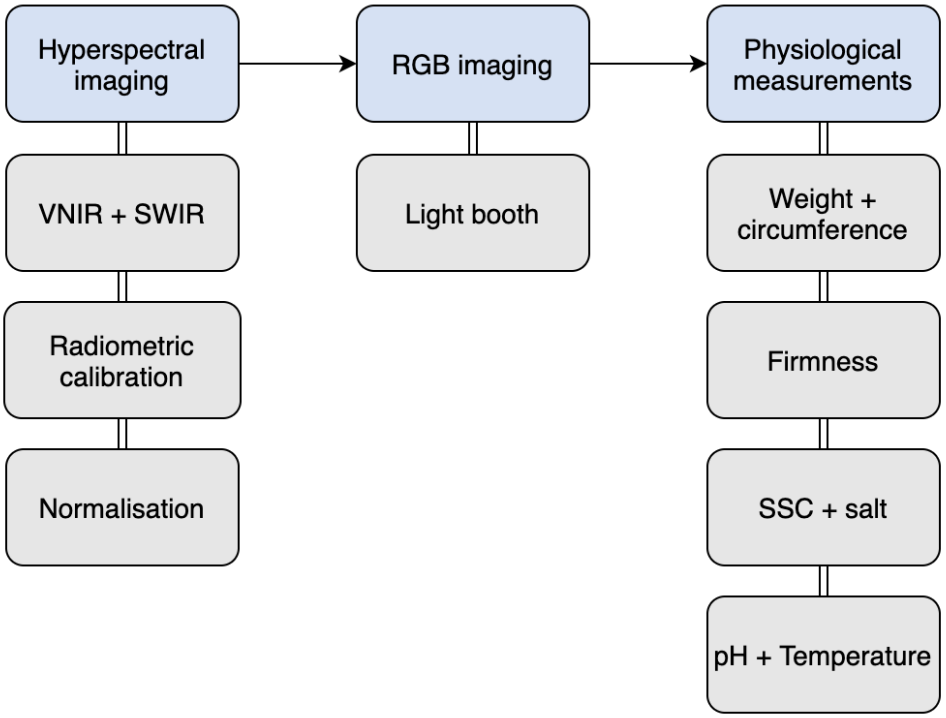


Figure 5.1: Acquisition workflow.

5.1 Fruit samples

Kiwifruits (Actinidia deliciosa "Hayward") were harvested in Italy and delivered to us through BAMA at the beginning of February 2021. They provided us with boxes (1-9), with each having 55 fruits, giving us a total of 495 kiwis. Upon arrival, the boxes were stored in the refrigerator at 4°C to prevent them from ripening at a fast pace.

Each particular day in the data acquisition weeks included HSI of multiple boxes of kiwi, except on day 3, but only one box to go through RGB imaging and physiological measurements, see table 5.1. The HSI of the same boxes throughout the week was additional data needed for tracking the spectral changes in kiwi. At the start of each week, three kiwi boxes were carried out from the refrigerator and labeled. Then the measurements began.

The first measurement was the weight and circumference of each kiwi; the circumference was measured using measuring tape along the longest axis. After they were measured, kiwis in batches of 8 were placed into egg cartons and imaged using HSI. The egg cartons worked as a stabilizer to keep the kiwis from wiggling. When the imaging of every kiwi in a box was complete, it got either placed into a refrigerator for storing at 8°C (seen on figure 5.2) or sent to RGB imaging and physiological measurements. Table 5.1 illustrates which boxes are being imaged and which box is being measured (RGB and physiological measurements) each

		Imaging (HSI)	Physiological measurements and RGB imaging
Week 1	Day 1	Box 1, Box 2, Box 3	Box 1
	Day 2	Box 2, Box 3	Box 2
	Day 3	Box 3	Box 3
Week 2	Day 1	Box 4, Box 5, Box 6	Box 4
	Day 2	Box 5, Box 6	Box 5
	Day 3	Box 6	Box 6
Week 3	Day 1	Box 7, Box 8, Box 9	Box 7
	Day 2	Box 8, Box 9	Box 8
	Day 3	Box 9	Box 9

Table 5.1: Weekly plan of which boxes are being scanned by HSI and what box are being measured destructively for each particular day.

day.

Figure 5.2: Kiwis in 8°C storage.

5.2 Hyperspectral configuration

The HSI system consisted of two different HS line-scanner cameras (VNIR[1] and SWIR[2]), a laboratory rack with conveyor belt, two full-range light sources and a computer. The cameras have the following specifications:

	HySpex VNIR-1800	**HySpex SWIR-384**
Sensitivity range (nm)	400-1000	930-2500
Spatial pixels	1800	384
Spectral sampling (nm)	3.26	5.45
Spectral channels	186	288
Bit depth	16	16
Sensor type	CMOS	MCT
Field of view	17°	16°
Max speed (fps)	260	400

Table 5.2: Important specifications of VNIR-1800 and SWIR-384.

Our hyperspectral setup had the cameras mounted vertically with one lamp for each camera. There are several reasons why we chose to use this setup.

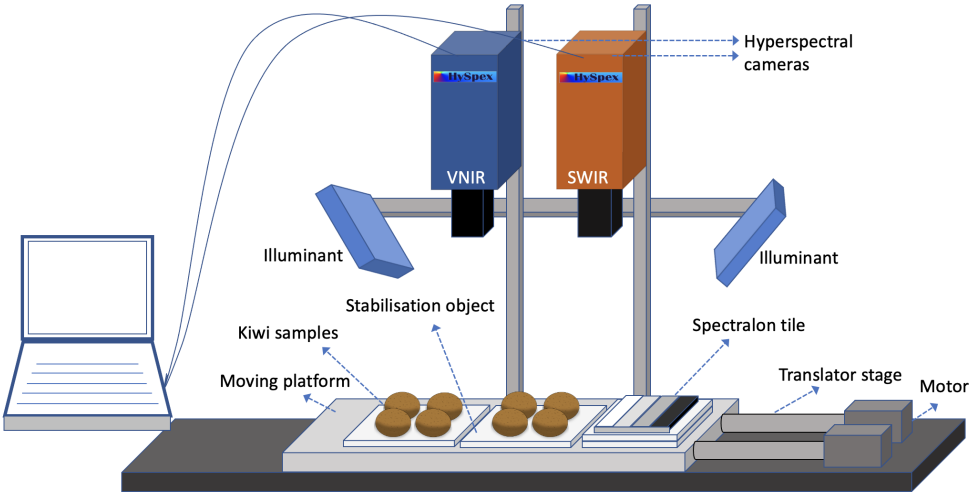


Figure 5.3: Our setup with two hyperspectral (HS) cameras, two illuminants, conveyor belt with moving platform, spectralon tile, and kiwis on a stable surface.

It was possible to scan with both cameras at the same time while not needing to change the lamp geometry by using this setup. This would have been the case if both lamps were used for each camera. Using one lamp for each camera, we halved the scanning time and exposed the kiwis for less heat.

[1] https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-vnir-1800
[2] https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-swir-384

Although the cameras could have been slightly angled towards a common field of view, we discarded this solution as the images would have different viewing angles, which would make them harder to compare and possibly merge.

Even though our setup meets our main demands, it still has some drawbacks. By assigning one lamp to each camera, the light casts shadows differently as the kiwis do not have a flat surface. This makes it harder to compare the same points of a kiwi in both VNIR and SWIR.

5.2.1 Hyperspectral imaging

The camera software and image acquisition settings (camera height, focus, illuminants) were tuned before imaging to ensure the scene and image capturing remained as equal as possible between days and scans. The kiwis were put on the conveyor belt 30 cm away from the lens of the hyperspectral cameras as this was in the depth of field. The reflectance target was adjusted so that it was the same distance away from the cameras.

There were some last procedures before scanning took place. We ensured that the lights were aligned with the line scanner to give close to uniform light distribution at the focus point. External lights were turned off to not interact with the imaging process.

The setup was optimized for speed as the lamps heated the kiwis. This was done by using a lower frame period (6000μs and 26000μs for vnir and swir respectively) to make the conveyor belt move faster while scanning. A fan was also set up about 80cm away from the kiwis for further cooling.

5.2.1.1 Parameters and filters

Integration time is a given time the sensor uses to measure a signal, while frame period is how long the shutter is open each second second[3]. The camera parameters were adjusted to allow for rapid imaging of kiwi. This meant that the frame period and the integration time were short (integration time equal to 6000μs) and that equalization and polarization filters were not applied. This compromise was made as we tried to minimize the heating of the kiwis when imaging. Both the equalization and polarization filter would have increased the image quality, but we decided not to use them due to the cameras needing a longer exposure time. This does not mean the image quality was poor, but filters could have improved it. The same holds for not using frame averaging.

[3] `https://www.ophiropt.com/laser--measurement/knowledge-center/article/8003`

5.2.1.2 Reflectance target

The reflectance target (5.4) consists of four different panels that reflects light differently. The target reflects diffusely, as explained in section 3.1.4.1. These four surfaces are constructed to reflect 99%, 50%, 25%, and 10% of the incident light. Since the surfaces do not reflect uniformly for all wavelengths, the manufacturer provided a table for all precise reflectance values.

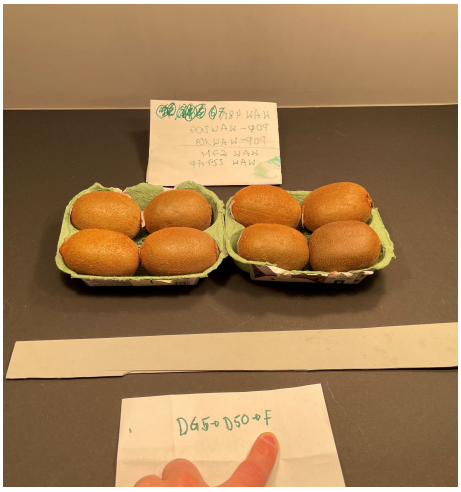
Figure 5.4: A multi-step reflectance target (Image source [95]).

5.3 RGB imaging

In addition to HS images, different RGB images in a light booth were also captured. The light booth (Veri Vide CAC 60-5) that can simulate four different lighting conditions; D65, D50, UV, and F (Fluorescent light, but only D65, D50, and F), was used for RGB images. The images were captured using an iPhone 12 Pro Max with the primary lens. The specifications [4] of the phone are listed below:

- Primary: 12 MP sensor (1.4 micron photosites), 26 mm-equivalent f/1.6-aperture lens, phase detection autofocus(PDAF), optical image stabilizer(OIS)
- Resolution: 4032x3024 pixels
- Compression: HEIC

(a) Veri Vide CAC 60-5 light booth for simulating lighting conditions



(b) Sample RGB image in fluorescent light

Figure 5.5: A light booth and an example of a RGB image.

[4] https://www.dxomark.com/apple-iphone-12-pro-camera-review-great-smartphone-video

5.4 Physiological measurements

After the kiwis were imaged in the light booth, the destructive physiological measurements were taken- this involved measuring firmness, sugar content (SSC), salt, and pH of each kiwi. More detailed information about the measurements carried out can be found in Appendix D.

5.4.1 Firmness

Before measuring the firmness of a kiwi, we gently peeled a strip of the skin using a fruit peeler on both sides. Then we measured the kiwi flesh firmness (measured in Newton) at where they were peeled using a stand-mounted penetrometer (Vaiseshika Digital Fruit Pressure tester) mounted to a fixed frame with an 8mm diameter tip. The speed at which we did the measurement was 5mm/s, as higher speeds could give improper results [96]. The depth at which we penetrated was 8mm, the same as the diameter of the tip. The penetrometer used can be seen in figure 5.6.

Figure 5.6: Penetrometer used for the firmness measurements.

After both sides had been measured, we cut the kiwi into four parts, two slices, and two halves. This was done by the following procedure: cut the kiwi in the middle to make one stem and one bottom half. Then cut a 3mm slice from each half, as illustrated in figure 5.7. A firmness measurement of the core of each half was then measured. We used the same penetrometer setup as the first one.

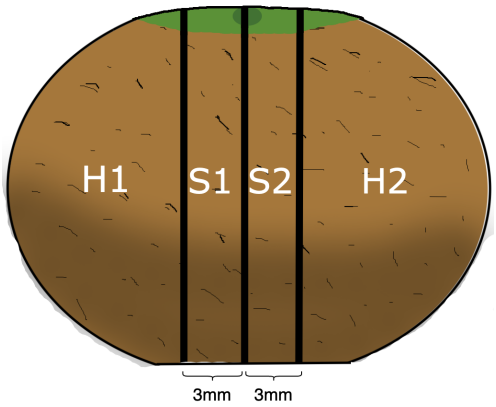


Figure 5.7: Slicing of the kiwis. H1 and H2 represents kiwi halves and S1 and S2 slices 3mm thick.

5.4.2 Brix refractometer

One of the slices (S1 or S2 in figure 5.7) was used for measuring sugar (SSC) using the Brix refractometer (Digital Atago Pal-1 Digital pocket refractometer shown in figure 5.8.

The slice was put into a garlic squeezer and squeezed so that both 3-4 drops of kiwi juice entered the device. After the device had been measured, we clean it using water and soft tissue so that no kiwi juice was remaining.

A Brix refractometer measures the sugar content (Measured in soluble solids content (SSC)) in an aqueous solution by refraction. This solution could be from fruits, vegetables, juices, soft drinks, wine, beer, or other plant-based foods [97]. When light enters a liquid at an angle, it bends differently depending on the density of the liquid. Using the fact that water and water with dissolved sugar have different densities, the refractometer can estimate how much sugar is in a solution/liquid. The measurements are in Brix degrees(°Bx), where one degree Brix corresponds to 1 gram of sucrose in a 100 gram solution. However, there are other dissolved solids than sucrose in the solution. The Brix refractometer only approximates the dissolved solids content (known as SSC)[98].

Figure 5.8: Digital Atago Pal-1 Digital pocket refractometer.

5.4.3 Salt meter

The same slice and procedure from measuring SSC were used to measure salt with the salt meter (Atago Pal-Salt seen in figure 5.9)[99].

The salt meter uses the conductivity of the solution as a predictor for how much salt it contains. The conductivity is found by transmitting a small electric charge through the solution. The Pal-Salt can measure the salt concentration in the range from 0.00 to 10.0% (g/100mL) with an accuracy of ±0.05% for the salt concentration of 0.00 to 0.99%.

Salt meter is widely used in the food industry to check salt content in various products [100]. It is not as often used in fruit measurements because these values are not so significant in fruits.

Figure 5.9: Pal-Salt: Digital hand-held salt meter by Atago.

5.4.4 pH and temperature

One of the two halves (H1 or H2 in figure 5.7) was also used for measuring pH and temperature. Both measurements were done by inserting a rod into the kiwi half and waiting for the value to converge. The pH rod was cleaned and put into a jar of water between each measurement.

A pH meter is an instrument that determines the acidity or alkalinity expressed as [5]. The range measured layed between 0-14 pH and $\pm 1000 mV(redox)$.

(a) pH measurement of the kiwi

(b) Portable pH meter, pH 7 model

Figure 5.10: The pH meter used and how it was operated.

[5] In chemistry, pH (denoting "potential of hydrogen" or "power of hydrogen") is a scale used to identify the pH or basicity of an aqueous solution[101]

5.4.5 Kiwi slices

The other slice was put on a tray to be imaged using hsi. This HSI took place during the physiological measurements and used the same camera parameters as the other measurements. The slices were imaged at an equal distance as the kiwis, 30cm from the lens.

5.5 Normalization

Because the images captured by the VNIR and SWIR cameras were raw image files, they included some noise. Radiometric calibration was applied to remove noise, increasing the Signal to Noise Ratio (SNR), and then data was transformed into reflectance, called normalizing. This is because reflectance gives information about the object in the scene and not just the radiance captured by the camera. It also makes the data easier to interpret.

5.5.1 Radiometric calibration

Since the raw images contained dark current noise[6], it was removed by using software provided by HySpex. The program only subtracted the dark current from the raw image. The raw image files were stored as backup.

5.5.2 Transforming to reflectance

The images were transformed into reflectance by using the darkest shade (10%) on the reflectance target. This shade was manually selected on both the VNIR and SWIR images. However, since the reflectance target stayed in the same place for all images, it was possible to automate.

We normalized the image by transforming each column into reflectance one at a time. This was to reduce artifacts from influencing the normalization process and because the radiance along the spatial points of the image is not uniform. The points used as reference values in each column were also averaged by taking the mean of 20 pixels below and above. This was to compensate for noise and unwanted particles on the reflectance shade.

Further calculations were made to retrieve the reflectance of the kiwi. Since the actual reflectance of the reflectance panel is known, light emitted from the source can be calculated. This was done according to Equation 3.3. Equation 3.2 was then used on all columns for all bands.

Hanning window filter with window size 11 was applied to the spectra after normalization to increase the Signal to Noise Ratio (SNR) ratio. However, several smoothing parameters were evaluated before coming to this conclusion; among tried smoothing filters were median and mean filters.

[6]Section explaining dark current noise 3.1.3.2

Figure 5.11 shows the results before and after applying the hanning window. From the figure it is possible to see that the SNR, measured with standard deviation in the spectra has decreased.

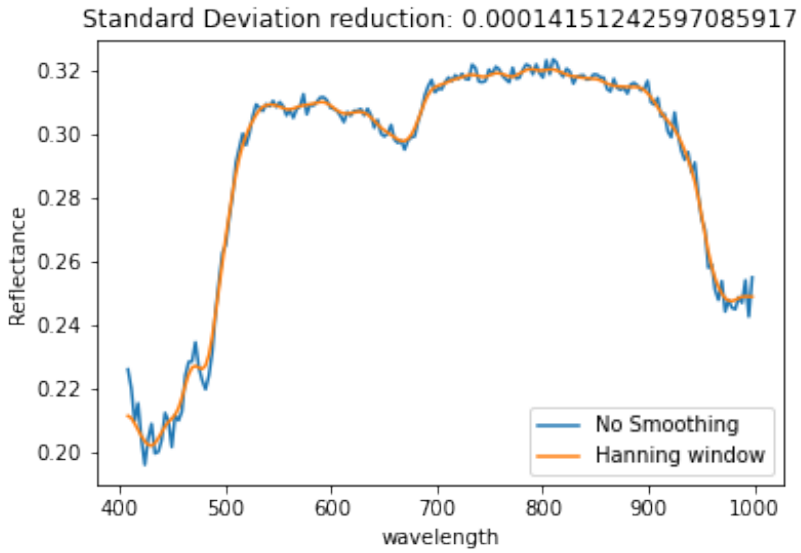


Figure 5.11: Spectra of kiwi white core before and after applying the Hanning window smoothing filter with size 11. Wavelength is in nm.

When smoothing spectra, there is a trade-off between noise reduction and data retention. If the spectra is smoothed too much, fewer data about the kiwi is retained. Therefore it is necessary to experiment and find a suitable smoothing filter that reduces noise and does not erase much spectral information.

5.5.3 Verification of normalization

In order to verify that the normalization transform was correctly done, we compared the normalized values on a color checker and reflectance target with the ground truth. To make these values comparable with the HS images normalized for kiwi, they were retrieved using the same camera parameters and normalization patch (10%).

The X-rite passport ColorChecker (seen in figure 5.12) was used and had its ground truth color values provided by the manufacturer. The shades used to compare normalization values were the top row of shades seen on the ColorChecker in figure 5.12. These shades are called Black, N3.5, N5, N6.5, N8, and white. From these shades, Black towards N6.5 was used. The number after "N" represents the whiteness of a shade (Munsell value), which correlates with its average reflectance value. The average reflectance values were retrieved from a technical review [102].

Figure 5.12: X-rite passport ColorChecker

To calculate the average reflectance values of these shades, wavelengths from 380 to 730 nm were used; this was the same as from the technical review. The values from these wavelengths were then averaged and calculated to be in percentage.

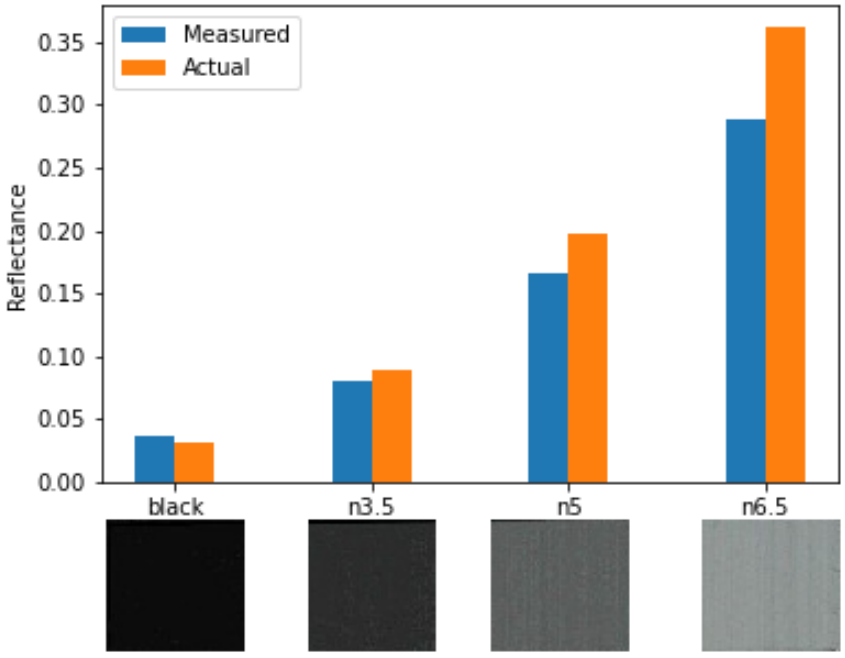


Figure 5.13: Measured ColorChecker reflectance values.

The x-axis on figure 5.13 shows the different colors from black to N6.5. It was found that the brighter the reflectance, the more inaccurate the captured reflectance values were. The same conclusion can be made from the spectra of the colors retrieved from [103].

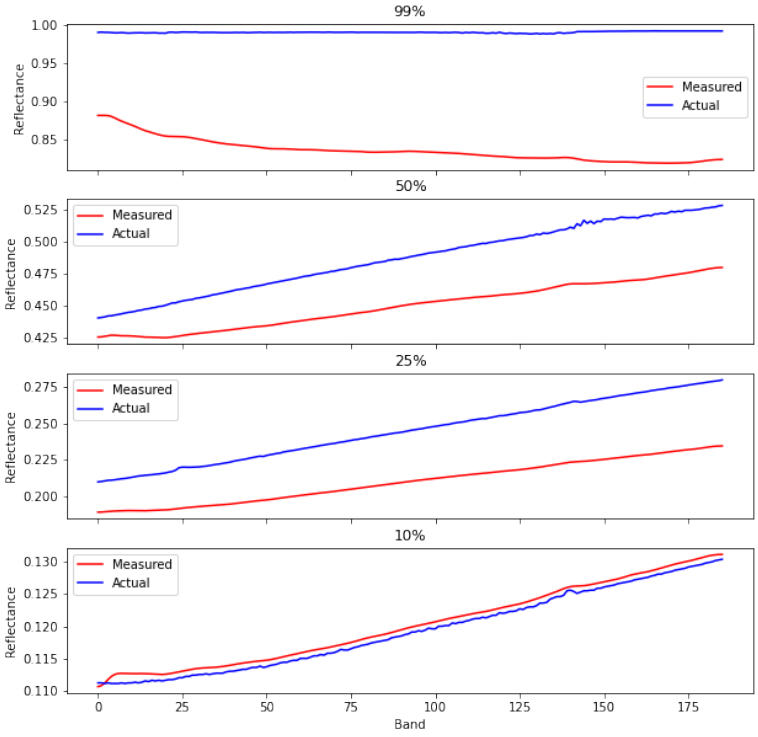


Figure 5.14: The ground truth compared to the captured reflectance of the 4 different shades (99%, 50%, 25%, 10%) on the reflectance target used in the data acquisition.

Figure 5.14 shows the difference between the captured reflectance values from the reflectance target and the ground truth. The same conclusion from the ColorChecker can also be made here. As the reflectance of the ground truth increased, the more significant the difference between the captured reflectance values and ground truth.

The reflectance values captured from the 10% patch on Figure 5.14, is very close to the ground truth. This is because the HS image was normalized using this patch. If the normalization was done using a different patch, the same would be observed only for that patch. The same would also happen on the ColorChecker if it was normalized using the 99% patch, the whitest shade would be closest to the ground truth.

The MSE between the real and captured reflectance percent from the ColorChecker was around 0.0016 and 0.0067 for reflectance target. This is a tiny reflectance error. Therefore the normalization is correct. However, there are reflectance values captured from the kiwi closer to the ground truth, and some of

them will be further away.

An attempt was made to find the best SG-filter parameters using the ColorChecker. However, this was without success. The parameters found did not work well when used as pre-processing for the models.

This verification was also done on a HS image that was not smoothed with the Hanning window. This showed that with the Hanning window, the accuracy was 40% better on the ColorChecker and about 1% better on the reflectance target.

5.6 Physiological measurement visualization

After we collected the physiological measurements, it was necessary to visualize the data to see any trends. We used boxplots to visualize how the data varied from day to day because it compares the data value distribution well. Using boxplots, we understood what could be explored in the dataset and its primary qualities. This was important as we could prioritize the dataset parameters that made the most sense to utilize in our models.

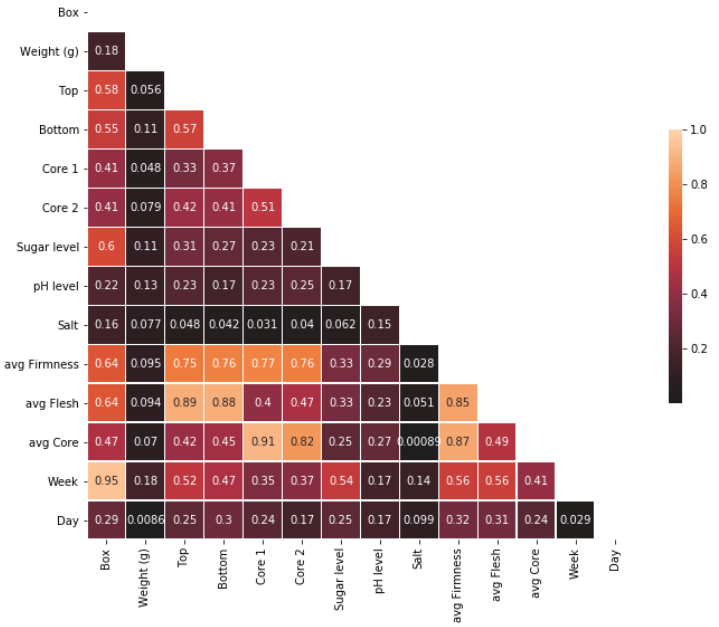


Figure 5.15: Correlation matrix of cleaned dataset.

We also used a correlation matrix to see how the dataset correlated within its internal parameters (see figure 5.15). This strengthened our confidence in what parameters were the most valuable for determining the ripeness of the kiwi. By looking at the matrix (figure 5.15), we can see that ssc and firmness has the highest correlation, which indicates that these parameters have the best trends.

Even though we have a large dataset that includes many of the internal parameters of the kiwi, there is still some uncertainty as we have little information about the harvesting and transportation of the kiwis. As a result, we do not know how much the kiwis had ripened before we received them. Firmness and SSC have different periods of rapid change, and if our data does not include this stage, the data might be skewed more towards fully ripe. This means that we could have fewer data points where the kiwis are unripe, making it harder to determine the ripeness by using models. Our limited amount of samples strengthens this uncertainty. Four hundred fifty-five kiwis is a lot and takes a long time to acquire measurements and images for. However, a larger dataset in the specific rapid change periods would represent a customarily distributed dataset.

There is also a possibility of human error when writing down measurements. Even though we tried our best not to make any mistakes, the error is probable when inputting around 4000 values manually.

5.6.1 Firmness

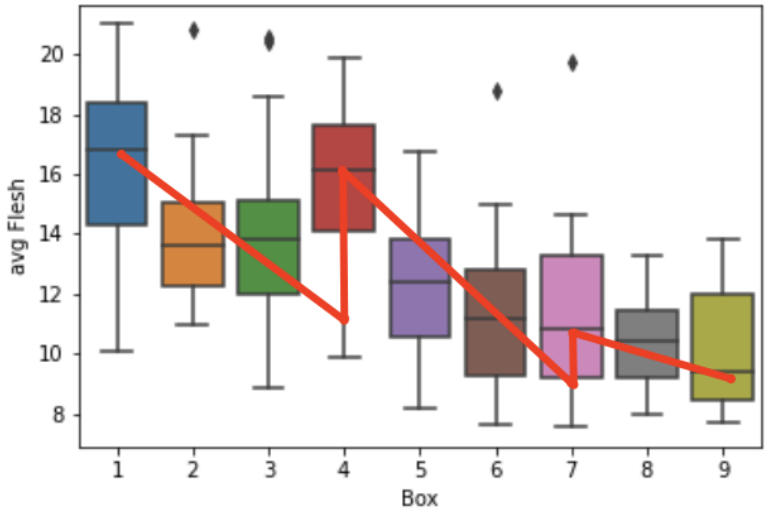


Figure 5.16: Flesh firmness from the 9 days of measurements, where each box (day) has its own color.

Figure 5.16 shows the average flesh firmness of each box (day) of measurements, and we can see a downward trend as the kiwis ripen. This is expected because the firmness decreases as the kiwi ripens. As mentioned in section 5.1, three boxes were taken out of 4° storage each week and placed in 8° storage between scanning days. This caused boxes 2 and 3 to ripen faster than box 4, which was stored at a lower temperature. The same applies to box 5 and 6 compared to box seven, as we can see as a "staircase" effect in figure 5.16. This did not affect the results of our models as the spectra and measurements are being

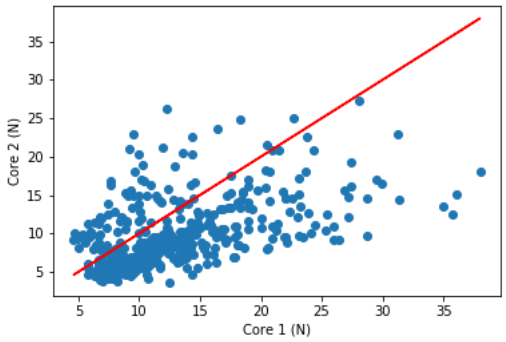
compared. However, it is an interesting finding that the kiwi might ripen faster at two weeks stored at 4° than one week stored at 8°.

There were several sources of error while measuring firmness. Firstly, The kiwi samples vary in shape, size, and firmness, which impacts the results. Secondly, manual measurements cause some variation in speed of penetration with the penetrometer, which ideally should be constant for all samples. This human error was somewhat reduced as the penetrometer was mounted but could not be eliminated. The peeling was also done manually, which caused variation in the depth of peeling. This might impact the firmness measurements as some kiwis might be softer towards the core of the pulp.

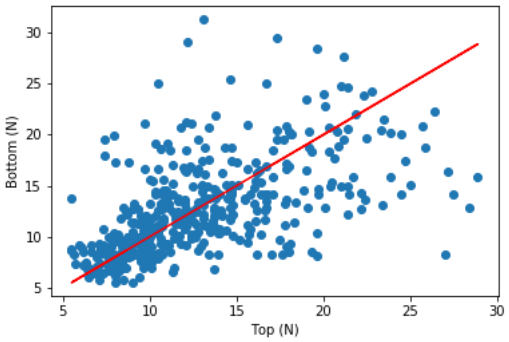
	Week 1	Week 2	Week 3
Monday	-	19.7 °C	23.5 °C
Wednesday	19.8 °C	24.4 °C	23.8 °C
Friday	18.3 °C	23.0 °C	22.7 °C

Figure 5.17: Kiwis' temperature during destructive measurements.

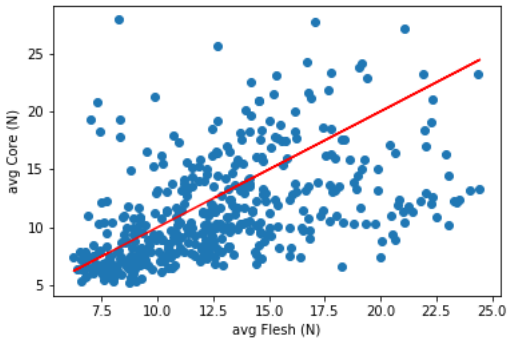
In the kiwi industry, there is a standard that states that the firmness of kiwifruits should be measured at temperatures between 0 and 5°C-ideally at 0°C[104]. It is because the temperature of fruit is known to affect firmness measurements[105]. Our firmness measurements were performed at 20° (see figure 5.17) and could vary between boxes and batches as there were delays and other contributing factors. This temperature variation might also cause some errors in the kiwi firmness measurements.



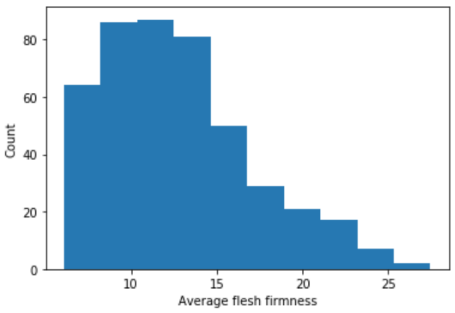
(a) Core 1 vs Core 2 firmness similarity, 0.51 correlation.



(b) Top vs Bottom flesh firmness Similarity, 0.57 correlation.



(c) Flesh vs Core firmness similarity, 0.49 correlation



(d) Histogram of the distributions on the average flesh firmness ((top + bottom) /2) values.

Figure 5.18: All the firmness measurements (top/bottom flesh and both sides of the core) and the distribution of values.

5.6.2 SSC

The boxplots of SSC show that there is an upwards trend of sugar content in the kiwis (see figure 5.19). The first two boxes have the most significant leap as we did not let the Brix refractometer fully converge the first two days of scanning. Even though the measurements on the first two boxes are less accurate compared to the rest of the boxes, there is still a clear trend of increase in SSC.

Since we are not sure if all the measurements on boxes 1 and 2 are accurate, we cannot assume that the change in SSC is as rapid as the boxplot shows. This further enhances our uncertainty of exactly what stage the kiwis were at when we received them. Another source of error is the temperature of the kiwis as we measured with the Brix refractometer. The temperature should ideally be 20°, but as we had many samples to measure, this was not always the case, as seen in figure 5.17.

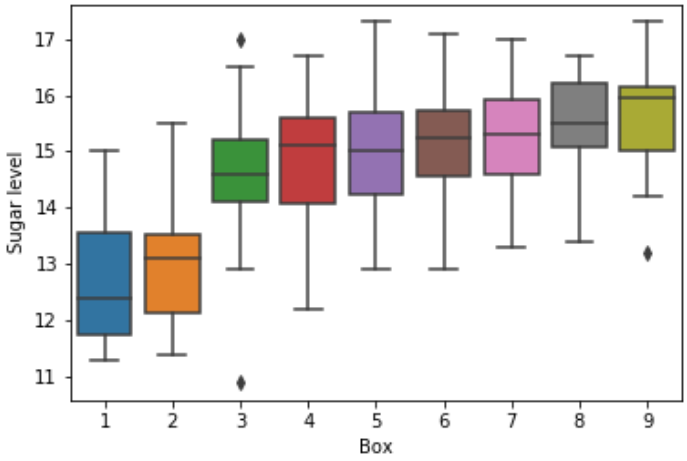
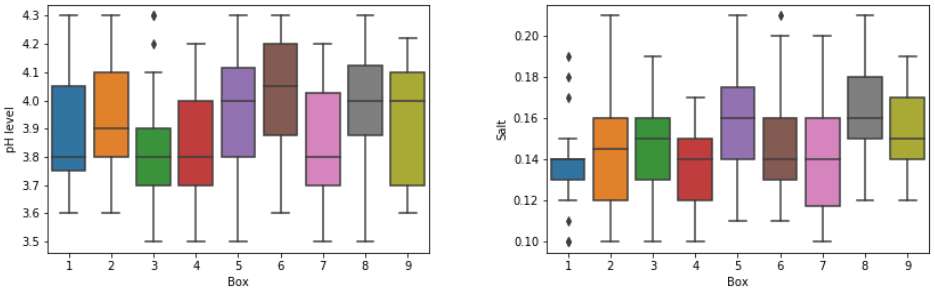


Figure 5.19: SSC from the 9 days of measurements, where each box (day) has its own color.

As with firmness, the variation between samples can also sway the results as we do not have many samples in each box. It is mainly the sugar gradient (difference in SSC from the core versus the edge of the pulp) within the kiwis that vary and depending on how significant this gradient is and which part of the pulp was used when squeezing the sugar mixture, the result might change a lot [94].

5.6.3 pH and salt



(a) pH level development **(b)** Salt level development

Figure 5.20: ph and salt measurements over 9 days (boxes), where each color represents a day/box.

As we can see in figure 5.20, there is no evident trend for salt or pH in our dataset. This might be why very few papers have attempted to use these parameters for the quality assessment of kiwi fruits. In the correlation matrix figure 5.15, we can see that salt has a low correlation with the other parameters. The rows and

columns are colored blacks with a correlation value mostly less than 0.1. pH on the other hand, correlates more with firmness and sugar, which indicates that it could have potential as a quality parameter even though there is no evident trend in the boxplot.

For both parameters, there is some source of error in the measuring equipment. This source of error is larger on the equipment used for measuring salt but cannot be neglected for pH. Both measurements were done manually and can therefore include some human error.

The salt meter used in this project does not work optimally with mixed contents, impacting the results. How much of an impact the salt meter had is hard to speculate on, but better equipment for measuring salt would increase the accuracy of the measurements.

For pH, the values were rounded up by one decimal, which might skew the data more towards larger values and hide some of the smaller trends.

Chapter 6

Data pre-processing and modeling

Hyperspectral images are composed of many data (e.g., one HSI image of 7000x1800 pixels and 186 spectral bands contains more than 2 billion individual pixels), which makes them challenging to work with. In order to handle data at this scale, it is necessary to reduce it into something more manageable by removing uninformative and redundant data. In addition to reducing the data, pre-processing also involves cleaning it from artifacts like random/systematic noise and physical phenomena.

After the pre-processing, the spectral data is combined with data from the physiological measurements to find relationships between them using different machine learning and regression models. Since not all wavelengths are relevant for modeling this relationship, dimensionality reduction is used to locate the relevant ones.

This chapter will cover the pre-processing steps used for extracting and creating spectral datasets, exploring spectral data, and cleaning it using three different methods before it is used for modeling.

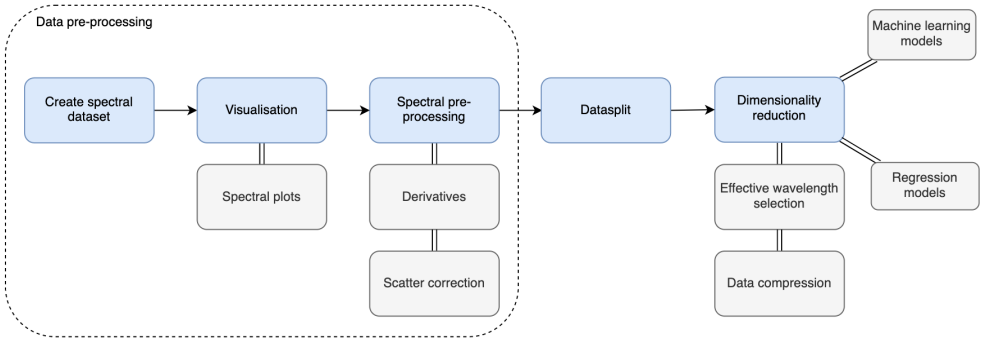


Figure 6.1: The steps involved in pre-processing HS data before using it to train models.

6.1 Programming languages and libraries

For analyzing the collected data, the programming language Python was used. This is because Python provides easy access to comprehensive libraries for scientific purposes. Here is a list of libraries used:

1. Pandas (CSV file handling) [106]
2. Numpy (Numerical calculations) [107]
3. Spectral python (Hyperspectral file handling) [108]
4. Sklearn (Machine learning algorithms) [109]
5. Tensorflow (Artificial neural network)[110]
6. Seaborne & Matplotlib (Visualizations)[111][112]
7. Opencv (Morphological operations on images) [113]
8. XGboost (Tree regressor) [114]
9. Statmodels (Statistical models, variable selection) [115]
10. SciPy (Signal processing, Function minimalisation) [116]

6.2 Automatic spectra collection

We automated the process to save time, make new spectral datasets with ease, and reduce errors caused by human mistakes like false labeling. The program ran under human supervision to ensure no mistakes were made. This was necessary as some variations in kiwi size, geometry, and the kiwi's position in the frame were necessary.

As we had four months to plan and execute the project, automating the collection of spectra was the only way to go. We estimated that with three weeks, six boxes to be scanned each week with 55 kiwis in each box, two camera types, and two datasets would require labeling of 3960 samples.

$$3 * 6 * 55 * 2 * 2 = 3960$$

When labeling, it takes time to load the images and create the mask of each kiwi; thus, we assumed that, on average, the time to label each kiwi would be between 1 and 2 minutes.

$$3960\,samples * [1,2]\,minutes \quad = \quad [3960, 7920]\,minutes \quad = \quad [66, 132]\,hours$$

The main differences between VNIR and SWIR images when collecting the kiwi spectra was kernel size for morphological operations and threshold values. The kernel was scaled because of higher spatial resolution in VNIR images compared to SWIR.

Creation of kiwi mask

To represent the position of each individual kiwi, a binary mask was created as this makes calculating center points in the next section more accessible.

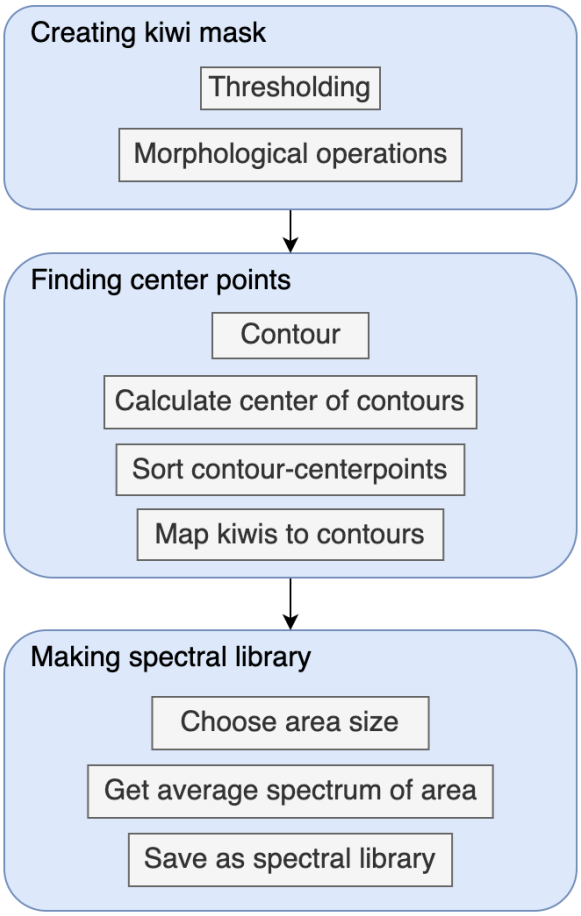


Figure 6.2: Flow of the automatic spectra collection algorithm.

The first step in creating the mask was to do simple thresholding on a grayscale representation of the image. Through experimenting, it was found that the best thresholding bands were 130 and 34 for VNIR and SWIR respectively. In these bands, the kiwis were the brightest, making it easier to separate the kiwis from the background. These bands also turned out to be the global maxima for each kiwi spectrum.

After thresholding, unwanted information was removed using the opening (erosion followed by dilation) morphological operation multiple times, but with different kernels, as shown in figure 6.3. The opening operation was preferred as it mostly preserved the size of each kiwi, allowing for accurate separation of kiwi and background.

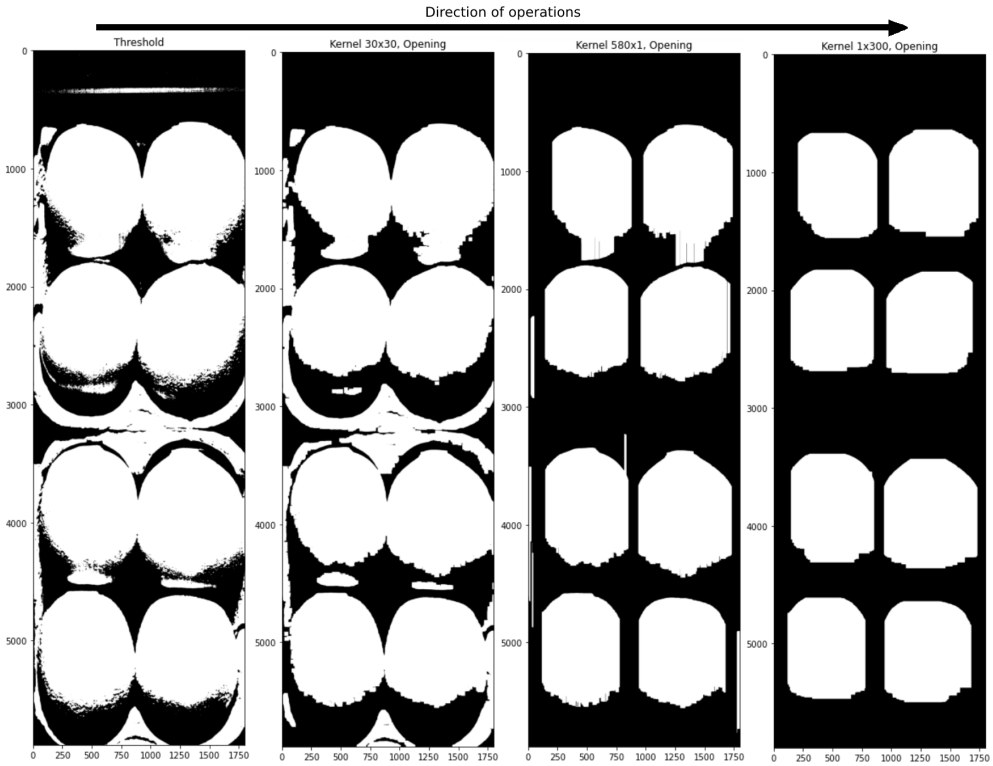


Figure 6.3: Example of how the morphological operations shape the binary image. Going from left to right, 1) Thresholded image, 2) Opening with kernel 30x30, 3) Opening with kernel 580x1, 4) Opening with kernel 300x1.

Finding center points

Since the binary mask constructed in the previous section consisted of distinct, separated shapes, a border-following algorithm was used. It detected shapes and provided information about how big and where its center was located. The algorithm is described in further detail by (Satoshi Suzuki [117]) but is implemented as a function in OpenCV library called findContours [1].

Sorting kiwis

As the function findContours returned all centroids in the incorrect order, it was necessary to map each centroid to the corresponding kiwi. It was done by sorting the kiwis by row and then sorting by column.

Generation of the spectral library

Finally, a spectrum for each kiwi center was obtained and stored in a spectral library by calculating the mean spectrum for an area of 30x30 for SWIR and 150x150

[1] `https://docs.opencv.org/master/d4/d73/tutorial_py_contours_begin.html`

for VNIR. There was also an option to use the mean spectrum of the whole kiwi. This was achieved by calculating the average spectrum for each kiwi for every spectrum in the kiwi mask.

6.2.1 Combining the spectra

An attempt at stitching the VNIR and SWIR spectra to one continuous spectrum was make but showed to be in general very difficult[118]. The difficulties arise due to several factors like spatial, geometric, our HSI setup, and normalization differences that each introduces small changes. Therefore, implementing a proper solution to combine them is outside the scope of this thesis, but implementation for illustration purposes was made.

The combined spectra was created by exploiting the small overlap between the VNIR and SWIR in the range 930-1000nm. VNIR was used as ground truth, and SWIR was moved up or down until the spectral lines overlapped. Because of slight differences in reflection values between the two spectra, a weighted average for making the transition smooth was used.

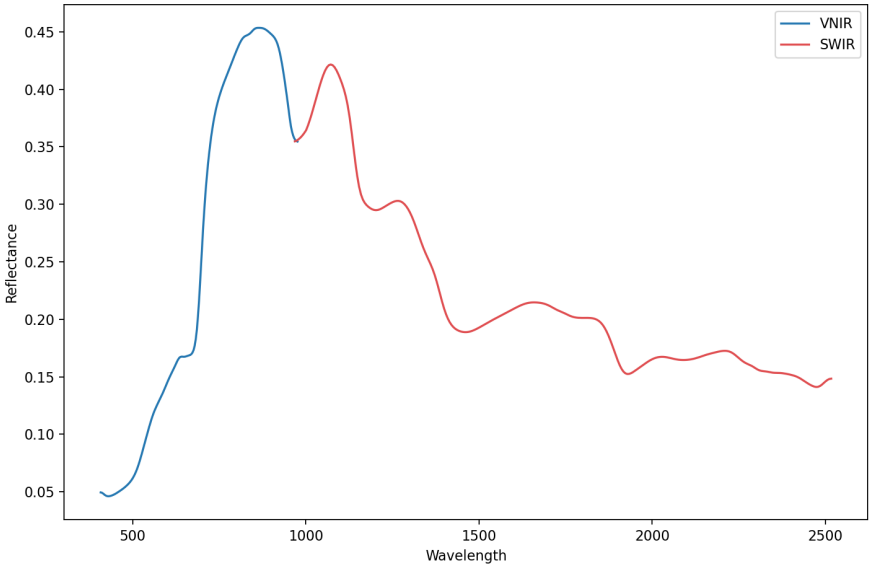


Figure 6.4: Full kiwi spectrum ranging from 400 to 2500nm where the VNIR and SWIR range are colored differently.

6.3 Visualization of spectral data

Compressing an image of a kiwi into a single spectrum is difficult as the spectrum needs to be representative and provide good spectral information. Since the decisions in doing so will influence the predictive power of the models, it is important to do this step properly by experimenting.

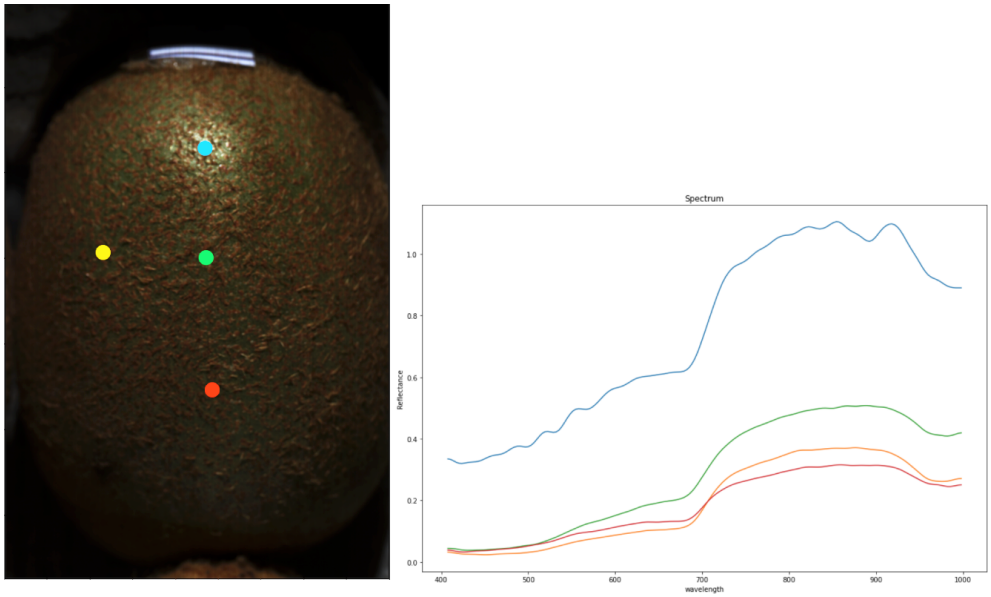


Figure 6.5: Different regions on a kiwi give different spectra. 4 unique points on the kiwi and the corresponding spectrum is plotted in the same color. Each spectrum was the average of an 5x5 area.

The spectra of a kiwi can vary a lot depending on where it was picked from. Figure 6.5 illustrates this by showing four different points from different illuminations and surface curvatures. Notice how the blue spectrum takes on values beyond 1. This is due to specular reflection, which causes the spectra to include more noise due to overexposure. The red and yellow spectra was picked from darker regions than the green point, which made them smoother and generally have lower reflectance values.

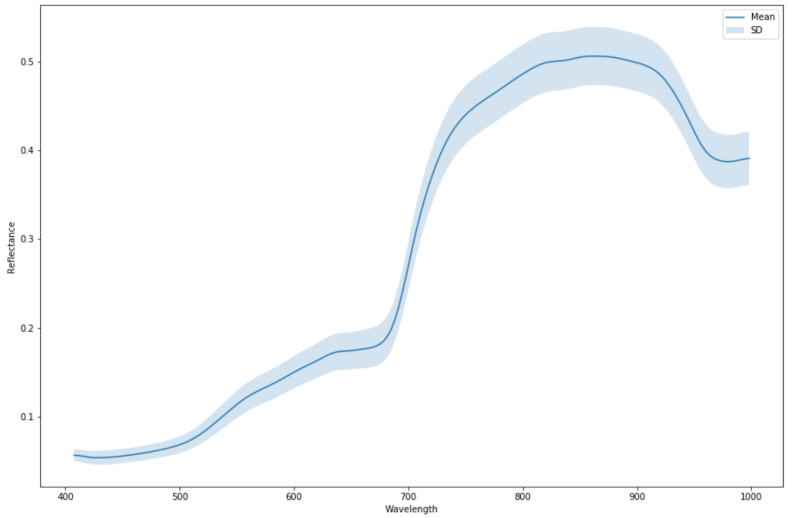
The green point was picked for creating a dataset because it was not too smooth, was consistent on the other kiwis, not hard to find, which made the process automatable and in-camera focus. This spectrum for the center point was averaged over a 30x30 and 150x150 area for VNIR and SWIR respectively. The areas were of this size to reduce noise, and other papers had used a similar size.

A second dataset was also made with the average of the whole kiwi. This is because it is difficult to know if the spectral information needed for predicting specific quality parameters in kiwi are present in the spectrum or not. Having a second dataset is helpful as it provides a bit of different spectral information, which is useful in concluding if the pre-processing methods were any good.

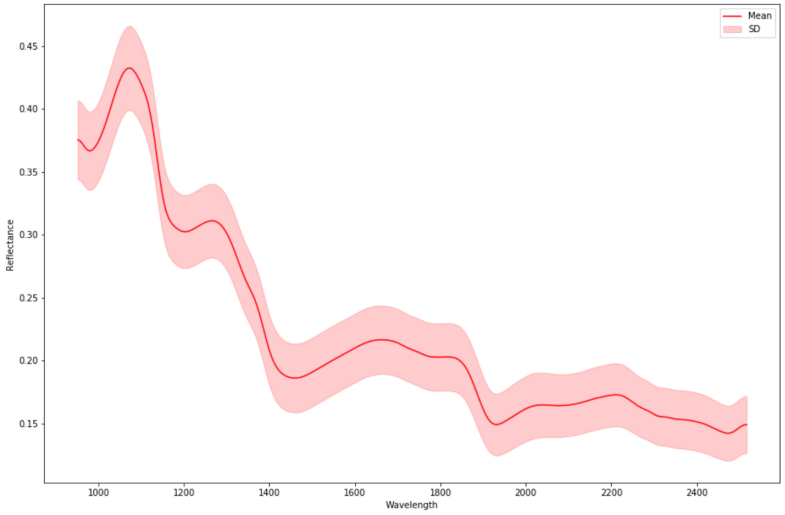
6.3.1 Spectral trends

After the spectral datasets are created, it is necessary to visualize them and understand how the data is formed. In our case, we want to look for spectral changes and trends that change between the days and weeks.

This was done by making a program to select and visualize the different spectra. Since the spectra from the kiwis are a bit different from each other, but a box of kiwis was to represent kiwis with the same ripeness, the average spectrum for each box was used as an estimate. The variations of spectra in a box is visualized in figure 6.6 in both spectral ranges. These variations are expected before the spectra are pre-processed and on a common baseline. Since pre-processing of spectra makes them harder to interpret when visualized, the visualizing was without pre-prepossessing.



(a) Mean VNIR spectra and the standard (SD) deviation.



(b) Mean SWIR spectra and the standard deviation (SD).

Figure 6.6: The variations in spectra of 55 kiwis (box 3 on day 3) in the VNIR and SWIR spectral ranges.

The average spectrum of box 3,6,9 (these boxes were imaged three times each

particular week in the acquisition period) was plotted for each of the three days. These boxes made it possible to track the same 55 kiwis and their spectral changes shown in figure 6.7.

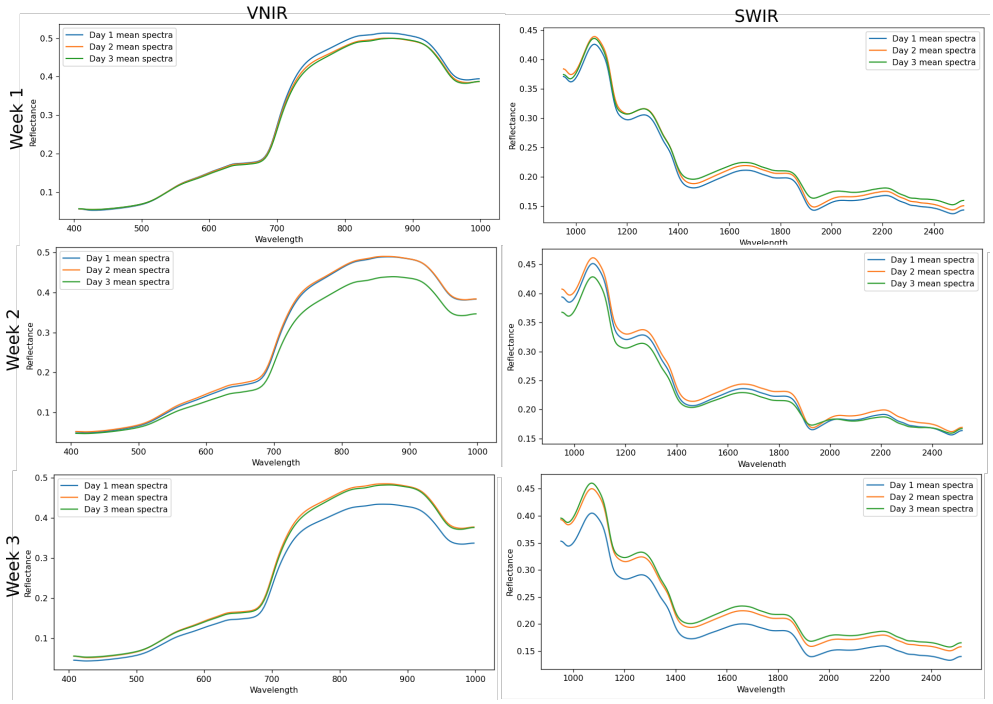


Figure 6.7: Average spectral plots of 55 kiwis (one box) for each day, each week for both spectral ranges. The boxes used was box 3,6,9 for week 1,2,3 respectively.

The plots reveal little change between the days in week 1 as all spectra lay on top of each other. There are some slight height differences at the maxima and local minima, but not enough to conclude anything. Week 2 shows more differences as the spectrum tends to drop for each day and was what we expected to see; however, the final week is the opposite.

This concludes that the general trend is not visible in these six spectral plots, but that does not necessarily mean there is no trend. The changes may not be considerable differences in height but small changes in slopes or bumps, which are hard to pick up visually. It is common to do some spectral pre-processing to find these differences, but that makes the visualization harder to interpret as the changes are much more minor. To find these differences, more advanced techniques or modeling has to be used.

However, there is undoubtedly a connection between the two spectral ranges as SWIR follows the VNIR trend, even though two different cameras captured them. The conclusion that can be drawn from this is that each week has its own trend and is unlikely to be data acquisition or preprocessing errors as the same pattern arises each week.

In addition to comparing the spectra to each other to look for trends, we also

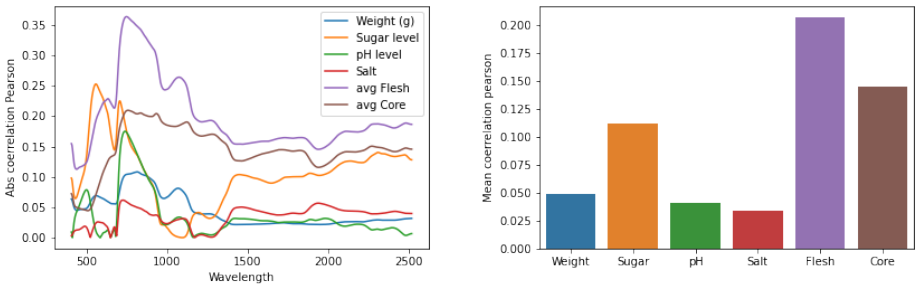
compared the changes in specific wavelengths to the changes in the physiological measurements.

6.3.2 Physiological correlation with spectra

To see if there were any correlation between the physiological and spectral changes, we calculated the correlation between each band to change in every quality parameter. The full spectral range (VNIR and SWIR) was used to measure the correlations. The VNIR range goes to around 1000nm, the SWIR range continues from there until 2500nm. The spectra used is the average spectra from the whole kiwi mask (Full).

Using the spectra of the kiwi and the physiological measurements, the Pearson correlation coefficient between these two was calculated. In short terms, the Pearson correlation is a normalized version of the covariance matrix. This method would tell us what wavelength correlated most with what feature and a prediction can be made on what feature will be predicted the best. To do this, all the features and wavelengths were looped through in a double for loop, then the absolute correlation coefficient between the feature and the wavelength was calculated.

Lastly, given all the Pearson correlation between all the features and a wavelength, the mean of these correlations was retrieved for each feature and plotted in a bar chart.



(a) Physiological measurements correlation with each wavelength (nm).

(b) Physiological measurements and their maximum correlation with a band. Notice that the maximum on the graph is not 1.

Figure 6.8: Feature correlation with spectra.

As seen from the bar chart on Figure 6.8b, average flesh correlates most with the captured spectra of the kiwis. As seen from (a) on the same figure, most of the features correlate best in the VNIR range, at around 740-760nm. In the case of sugar, it has big peaks around 500 nm but increases in the SWIR region.

Using the bar chart, we can expect the models to work best on flesh firmness, core firmness, and sugar. Less performance is expected from pH, salt, and weight. Lastly, the VNIR range is expected to perform better than the SWIR range.

However, there is less oscillation in the SWIR range than in VNIR, which can

be a sign that this range is better for predictions. The VNIR range goes a lot up to and down while the SWIR range is flatter with minor variations. These slight variations can explain more of the data with a machine learning model than the VNIR range.

6.4 Spectral pre-processing

The spectral pre-processing algorithms used were SNV, MSC, and derivatives, which are regularly used when working with spectral data. In order to find what pre-processing worked best for the separate models and quality parameters, each model was tested with and without and a combination of SNV or MSE with different Savitzky-Golay derivatives. This was tested on many of the models by brute-forcing a wide range of different parameters and combinations.

6.4.1 Scatter correction

The scattering correction algorithms SNV and MSE were experimentally used with and without derivatives to remove artifacts that interfered with the spectra. Because it is difficult to know if the spectral information is enhanced or reduced when combining them with derivatives, they were tested in various ways with the models.

6.4.2 Derivatives

As the reflection offset values in the spectra varied, derivatives using Savitzky-Golay were used to make a baseline for each spectrum. As seen in figure 6.9, the amplitude and slope differs, while the baseline is almost the same.

It was found that brute-forcing the SG-filter parameters before training the models worked best. The optimal SG-filter parameters changed depending on the physiological measurement to be predicted and the type of machine learning model used.

Regarding the Savitzky-Golay derivatives, the parameters tested were window size from $7-51$, polynomial orders from $2-7$, and 1st-3rd derivatives.

As mentioned in theory, it reduced noise in the spectra and should give better performance [34].

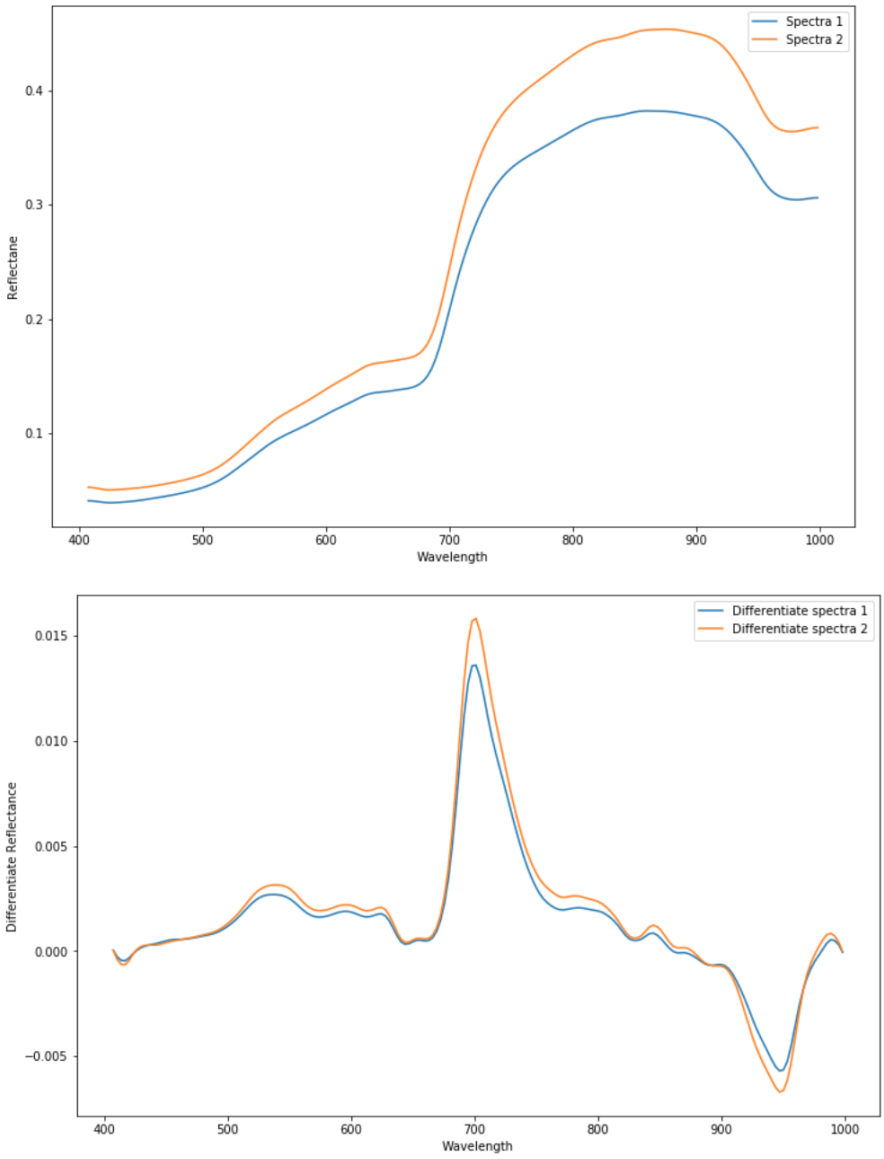


Figure 6.9: Difference before and after differentiation of two spectra using Savitzky-Golay with window size 5, 3rd order polynomial and 1st derivative.

6.5 Modeling

Several different regression and machine learning models were developed in an attempt to combine the trends found in the physiological measurements with the spectral changes. The models implemented were MLR, PLS, ANN, Tree regressor, and KNN. MLR and PLS models were implemented because of their good performance in predicting SSC and firmness in other kiwi cultivars. ANN, Tree regressor, and KNN were implemented as little research regarding these had previously been

done. All models were combined with different efficient wavelength selection algorithms and/or pre-processing techniques, but all models used the same data splits.

All models were combined with pre-processing techniques and/or efficient wavelength selection algorithms to search for the optimal, but all models used the same data splits.

6.5.1 Dataset split

In order to ensure that our models would be able to predict a wide range of values, we used the SPXY algorithm to split the dataset into 4 smaller ones. Before the split, we discarded samples where the physiological measurement value layed outside a 99% confidence interval, reducing the number of kiwis in the dataset down to 444. This was according to the error specified by the manufacturers for the physiological measurement equipment. The SPXY algorithm was first used to create a calibration and a prediction set with a 7 : 3 distribution as illustrated in figure 6.10. The prediction set was further split into validation and verification using the same algorithm but with a 1 : 1 distribution. It is important to note that the dataset split was applied after any spectral pre-processing. The calibration and verification datasets have values spanning over the most extended range since they are being used to train and test each model.

Sample Sets	Samples	Quality parameter	Range	Mean	SD
Calibration	311	Firmness (N/cm^2)	6.1-27.45	12.83	4.42
		SSC (°Brix)	9.5-18.4	14.56	1.65
		pH	3.3-4.8	3.95	0.30
Prediction	133	Firmness (N/cm^2)	6.25-22.25	11.83	3.47
		SSC (°Brix)	10.3-16.6	14.71	1.20
		pH	3.4-4.5	3.93	0.20
Validation*	67	Firmness (N/cm^2)	7.15-21.8	11.33	3.03
		SSC (°Brix)	13.1-16.6	15.03	0.86
		pH	3.6-4.3	3.92	0.17
Verification*	66	Firmness (N/cm^2)	6.25-22.25	12.3	3.79
		SSC (°Brix)	10.3-16.5	14.39	1.39
		pH	3.4-4.5	3.93	0.23

Table 6.1: Statistics of quality parameters in the different datasets without any spectral pre-processing. (* is 50% of the prediction dataset).

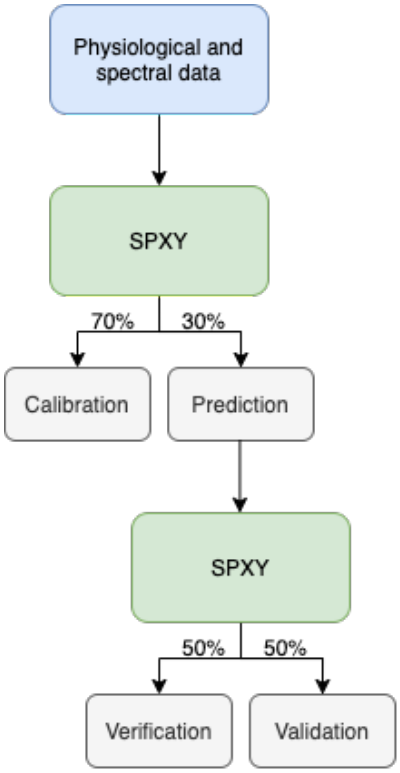


Figure 6.10: The construction of the different datasets using SPXY.

6.5.2 Multiple linear regression (MLR)

The wavelengths selection was carried out by finding the subset of wavelengths that gave the best predictions when used in an Multiple linear regression (MLR) model. This was done iteratively by testing subsets containing up to 100 wavelengths. For each subset testing, every possible initial wavelength for SPA[2]. For each subset created by SPA, an MLR model was trained on calibration data and evaluated using the Root mean squared error (RMSE) of the validation data (Root mean squared error (RMSE)V). The subset of wavelengths corresponding to the MLR model with the smallest RMSEV was then selected as a basis for the first wavelength elimination step.

The MLR models were constructed using the Successive projections algorithm (SPA) for wavelength selection and three elimination procedures to exclude unimportant wavelengths further.

The first elimination step involves evaluating candidate subsets of the selected subset. Candidate subsets with L number of wavelengths are created by taking the L first wavelengths from the subset. Each of the candidate subsets is then used to create new MLR models, and the best candidate subset is the one with the smallest

[2], the two initial values for SPA are the wavelength for calculating the first projections and how big a subset of wavelengths to return

RMSEV.

The second elimination step involves a backward elimination process to remove wavelengths that do not efficiently contribute to the model's predictive ability. This included making a relevance index for each wavelength belonging to the subset selected in the previous step. This index is calculated by multiplying the absolute value of the regression coefficient by the standard deviation, calculated by the calibration data, of each wavelength. The wavelengths are then sorted in descending order according to the relevance index, and then the same procedure used in the first elimination step is carried out.

The final step looks at the differences between the minimal RMSEV and the RMSEV values of the smaller subsets obtained from the second step.

Suppose the difference between the minimal RMSEV and a subset with fewer wavelengths has a slightly larger RMSEV. In that case, the subset with the slightly larger RMSEV is chosen due to the difference not being significant.

The final step involves selecting the smallest subset so that the RMSEV is not significantly larger than the subset corresponding to the minimum RMSEV. This is calculated using an F-test with a significance level $\alpha = 0.25$, which has been used elsewhere [119].

6.5.3 Partial Least Squares (PLS)

Partial least squares was combined with two different efficient wavelength selection algorithms, Uninformative Variable Elimination (UVE) and Genetic algorithm (GA). These algorithms are aimed to exclude wavelengths that do not contribute a significant amount to the prediction.

6.5.3.1 UVE-PLS

PLS was combined uninformative variable elimination (UVE) to reduce the number of wavelengths used. This worked by iteratively applying PLS to the dataset and calibrating it using cross-validation. The model that performed the best is selected and has the wavelength that contributes the least removed. This was done by extracting the regression coefficients from the model, then taking the absolute value and removing the lowest one. The elimination of uninformative wavelengths happens until a user-specified amount of wavelengths is reached.

6.5.3.2 Variable scoring for GA-PLS

A score matrix was set up to score the wavelength importance before training the GA-PLS models. The max score was set to normalize the score given by the different variable selection methods. Then the score given was assigned accordingly.

1. Regression model (P-values)
2. Variance Threshold from "Sklearn" (Variable variance)
3. Select Percentile (Mutual information)

4. Recursive feature elimination (RFE)

The first three steps are often referred to as Univariate variable feature elimination. Which uses statistical tests to test the performance of each variable. The last step: RFE implements UVE recursively.

The Select percentile punishes variables that have mutual information. This metric was derived from information theory, which calculated how information could be retrieved from one wavelength by observing another. This is relatively equal to the covariance matrix.

After going through all of these variable scoring techniques, the Genetic algorithm was kick-started. The resulting score matrix was thresholded between 80 and 25% to get the initial population to be scored and mated.

GA used PLS to score each of the candidates, then selected the best four candidates from a population size of 8 to mate. Once two parents were selected, they would merge 50-50 to create new candidates. The mutation was also introduced, which randomly changed the usage of bands.

6.5.4 Support Vector Regression SVR

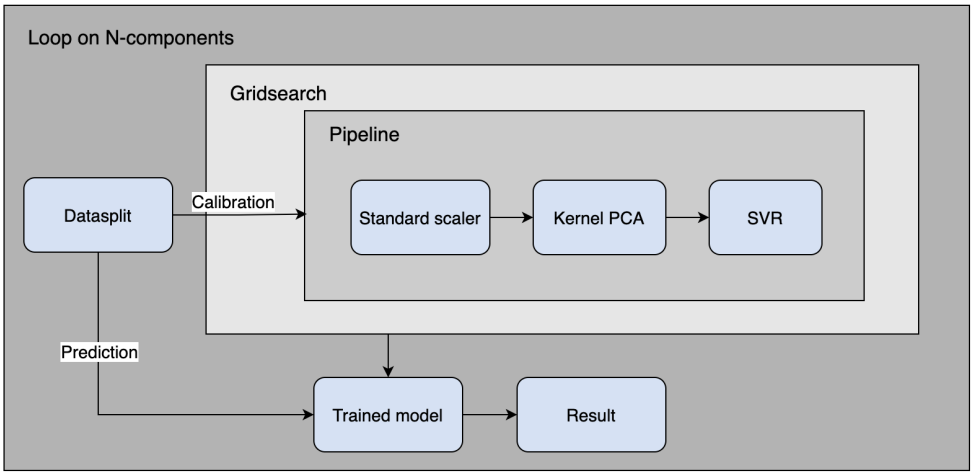


Figure 6.11: SVR model workflow.

The tools used in the model were imported from the library "sklearn"[109].

The model was created as a pipeline with StandardScaler, Kernel principal component analysis (KPCA), and SVR, and used cross-validation with GridSearchCv to find the optimal hyperparameters.

GridSearchCv is used for permuting a set of hyperparameters on the model to find the optimal values, which would be very time-consuming to do manually.

The spectra and features were scaled down to have a mean of 0 and standard deviation of 1 by using StandardScaler in the pipeline to decrease scale importance. Then KPCA was used for dimensionality reduction, which reduced the

features by 3-4x while keeping the essential information in new feature vectors. This improved both the speed and accuracy of the model. Lastly, in the pipeline, the feature vectors were passed into a SVR model trained with a calibration data split.

GridSearchCv was used on the pipeline to find the optimal value of the hyper-parameters C (which as discussed earlier in section 3.3.2) and $gamma$ (to control the variance or spread of the RBF kernel function).

This process was looped from 5 to 60 components for the KPCA parameter "N-components" to find which number of components gave the best result while also reducing dimensionality (see figure 6.11).

6.5.5 K-Nearest neighbor KNN

A function to minimize was made. Given SG-parameters and the number of components to use in KPCA, the data dimensionality was reduced accordingly. Then GridSearchCv was used to find the optimum number of neighbors for the model. The GridSearchCv module was set to use the validation dataset to test the models with different neighbors. The function returned the MSE value of the validation dataset.

To minimize this function, scipy-differential evolution was used. It uses gradient descent to find the minima of a function.

6.5.6 Artificial Neural Network ANN

To compile and train the neural network, the python library Tensorflow was used.

The spectra and features were scaled down to have a mean and variance of 1 (using StandardScaler). Then there was experimentation with the number of neurons and activation functions. As loss function MSE was used.

The layers was setup like the list shows below.

- Input Dense 512 Sigmoid, l2=0.001
- Dropout 0.2, l2=0.001
- Hidden 512 Relu, l2=0.001
- Dropout 0.2, l2=0.001
- Hidden 512 Softmax, l2=0.001
- Output 1 linear

There were some techniques used to prevent overfitting. An early stopping callback was introduced; this stopped the training if the loss function did not decrease after two epochs. Then a l2 regularizer with l2=0.001 and a few dropout layers were used.

The output layer had to be linear because the output is a regression and not a classification.

6.5.7 MLP

Using GridSearchCv, different layer sizes, activation functions, and learning rate were tried. This was then put inside a brute-force for SG-filters.

6.5.8 Tree regressor

The GridSearchCv module from "Sklearn" was used to test out different parameters. The parameters were built upon some initial guesses. Then, random parameters close to the initial guess were added. The GridSearchCv module then picked the parameters that worked best, which was then used to build new random parameters.

Chapter 7

Results and discussion

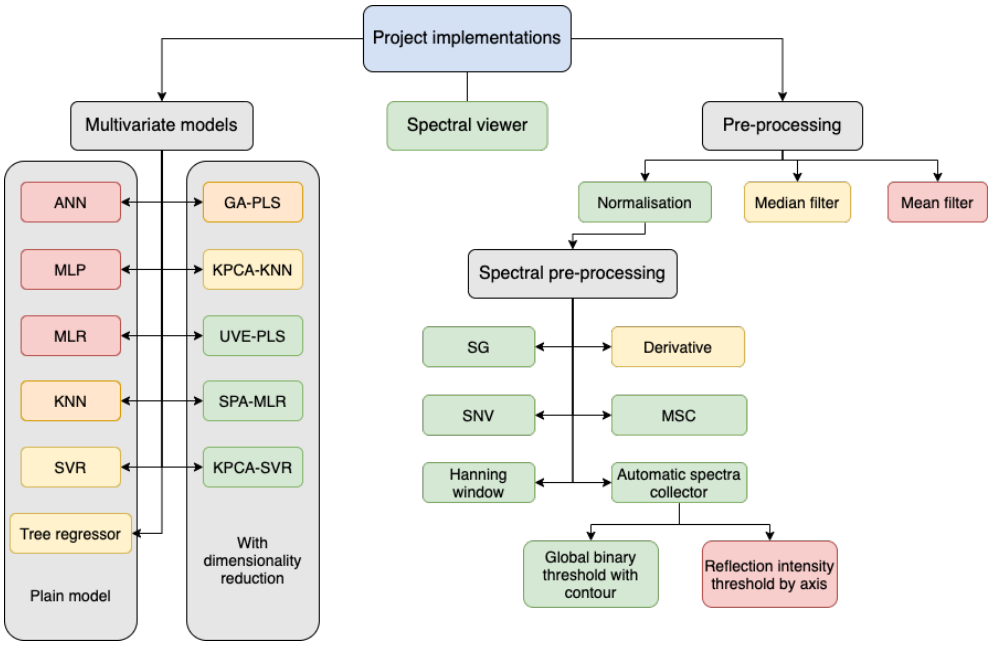


Figure 7.1: Tree of all implementations made in the project. Green indicates best performance, yellow second best performance, orange ok performance, red did not work.

In this chapter, the project arrangement and the predictive accuracy of the machine learning and regression models are discussed. Figure 7.1 shows all the implementations that have been attempted in this thesis and color graded by how well they worked. It gives a good overview of our efforts in this part of the project. However, it does not tell the whole story as some of the red or orange parts could be improved if we had more time to experiment.

7.1 Project priorities

Due to several limitations, our priority layed according to Figure 7.2. The RGB images and peels were only collected and not looked at because of the lack of time. Our main priority was SSC and firmness, using the models listed in the Figure.

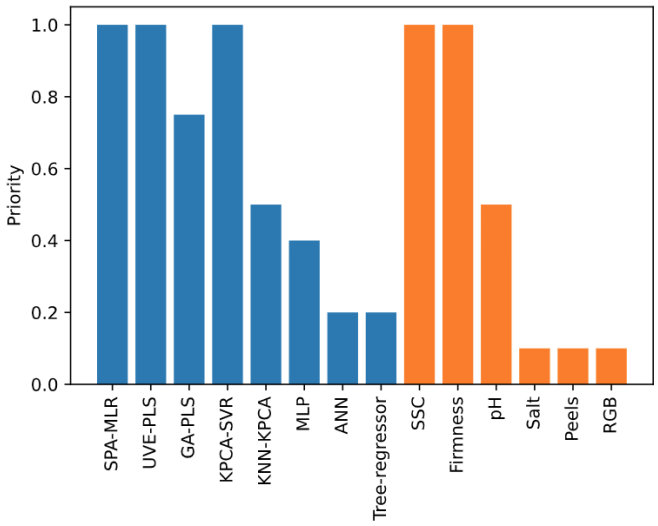


Figure 7.2: The prioritization of models and quality parameters during the project work.

7.2 Model selection

The different wavelength selection methods SPA, UVE, and GA were used to reduce the number of wavelengths into a smaller portion that predicted the quality parameters. They do not work independently but were fused with either MLR or PLS, creating SPA-MLR, UVE-PLS, and GA-PLS. For models not using conventional regression, the dimensionality reduction method Kernel principal component analysis (KPCA) was used.

The best predictions of firmness and SSC by MLR, PLS, SVR, and KNN using the wavelength mentioned above selection and dimensionality reduction methods are displayed and compared in table 7.1 and 7.2 respectively.

Each model was tested with several different spectral pre-processing and internal parameters, which produced variation in performance. The best performing model was chosen based on several evaluation metrics and the number of EWs used. Good performance was defined by having a low RMSE and high R^2 for both the calibration and verification dataset with slight differences between them and a minimum number of EWs. This assured that the model was not overfitted to

one particular dataset and included wavelengths primarily with good predictive power.

7.3 Firmness

UVE-PLS combined with the SWIR spectral range seemed to give the best performance at predicting firmness with overall the lowest RMSE (2.804) on the verification set and fewest EWs (8). These EWs and its measured-predicted graph can be found on figure 7.3

Notice that KPCA-KNN uses StandardScaler on the physiological measurements; for this reason, the RMSE values of this model cant be compared to the rest of the models. Therefore the focus is turned to the R_v^2 when comparing it to other models.

Our R_v^2 results are lower than what has been achieved on "Hayward" kiwi with 0.88 [7]. This article had a much bigger dataset and higher bio-variability than we had access to. They had 1530 kiwi samples retrieved from 3 different farms. They also stored the kiwis for ripening at 20°C instead of 8°C, like in our project. Doing this in our project would have provided a much higher bio-variability among the kiwis, which could have improved our predictive power.

As mentioned earlier in Chapter 5, the kiwis should be around 5°C when measuring the firmness using the penetrometer. This was not accounted for when measuring firmness during the Data acquisition. Therefore our measurements could have a more significant uncertainty than wanted.

The EWs selected by the best regression model for predicting firmness in SWIR were around 1611-1622, 1649-1655, 2347-2353nm. Li *et al.*[8] found EWs around 1050, 1395, 1450, 2100, 2300nm, and several EWs around 1700-1900nm. It makes sense that we did use many of these EWs, since their accuracy is higher than ours. In the VNIR range these EWs have also been reported: 555, 1000, 638, 503, 452, 700, 822. This was done by Zhu *et al.*[9], which looked at another kiwi variety. Because of little overlap between our EWs and other reported wavelengths, we can not verify that our EWs is strongly correlated with firmness.

7.3.1 VNIR vs SWIR

Since there is much research done in the VNIR range, one would expect this spectral range to perform best. This makes sense because the VNIR range is sensitive to most chemical, organic bonds[39]. However, our results show the contrary.

For firmness, the most successful models were in the SWIR range (As seen in figure 7.4 C). The average R_v^2 using this range was around 0.26 while for VNIR this was around 0.22. Also, the best prediction using UVE-PLS was using this range. This shows that using the SWIR range instead of the VNIR range to predict firmness is a better choice.

As Figure 7.4D the both datasets worked much better in the SWIR region. The improvement in SWIR region using the full dataset is much less than for region.

Firmness

Camera	Dataset	Model	Pre-processing	EWs	$RMSE_C$	$RMSE_P$	R^2_C	R^2_V
VNIR	Full	SPA-MLR	SavitzkyGolay(17,5,1)	56	3.282	3.425	0.478	0.315
		UVE-PLS	MSC SavitzkyGolay(33,5,2)	24	3.539	3.092	0.385	0.357
		GA-PLS	SavitzkyGolay(43,2,2)	86	2.333	1.842	0.019	0.181
		KPCA-SVR	StandardScaler	comp=45	3.355	3.277	0.609	0.272
		KPCA-KNN	SavitzkyGolay(21,4,2)	9	0.891*	0.864*	0.283	0.269
	Region	SPA-MLR	SavitzkyGolay(7,5,1)	7	4.184	3.61	0.168	0.167
		UVE-PLS	MSC SavitzkyGolay(27,5,1)	25	3.884	3.025	0.307	0.249
		GA-PLS	SavitzkyGolay(47,3,2)	86	2.266	1.928	0.054	0.085
		KPCA-SVR	StandardScaler	comp=55	3.479	3.291	0.511	0.234
		KPCA-KNN	SNV, SavitzkyGolay(41,6,0)	36	0.988*	0.988*	0.085	0.084
SWIR	Full	SPA-MLR	SNV	48	3.488	3.19	0.404	0.397
		UVE-PLS	**SNV SavitzkyGolay(21,5,1)**	8	3.376	2.804	0.452	**0.434**
		GA-PLS	SavitzkyGolay(11,2,1)	116	19.67	13.797	0.062	0.049
		KPCA-SVR	StandardScaler	comp=45	3.377	2.848	0.629	0.371
		KPCA-KNN	SNV, SavitzkyGolay(21,2,0)	14	0.938*	0.923*	0.17	0.182
	Region	SPA-MLR	SNV SavitzkyGolay(33,5,1)	24	3.718	3.222	0.31	0.398
		UVE-PLS	MSC SavitzkyGolay(21,5,1)	22	3.646	3.07	0.344	0.381
		GA-PLS	None	89	4.632	3.385	0.016	0.024
		KPCA-SVR	StandardScaler	comp=30	3.479	3.291	0.603	0.22
		KPCA-KNN	SavitzkyGolay(21,5,1)	15	0.93*	0.844*	0.243	0.193

Table 7.1: The best performances of each model for evaluating firmness in the VNIR and SWIR range separate. Both spectral ranges was tested on two different datasets, Full (average of whole kiwi) and Region (150x150 and 30x30 averaged area for VNIR and SWIR respectively). Every model used the same calibration (70%) and verification (15%) set. (* KPCA-KNN uses StandardScaler, therefore the RMSE value of this model is not directly cis not directly comparable)omparable)

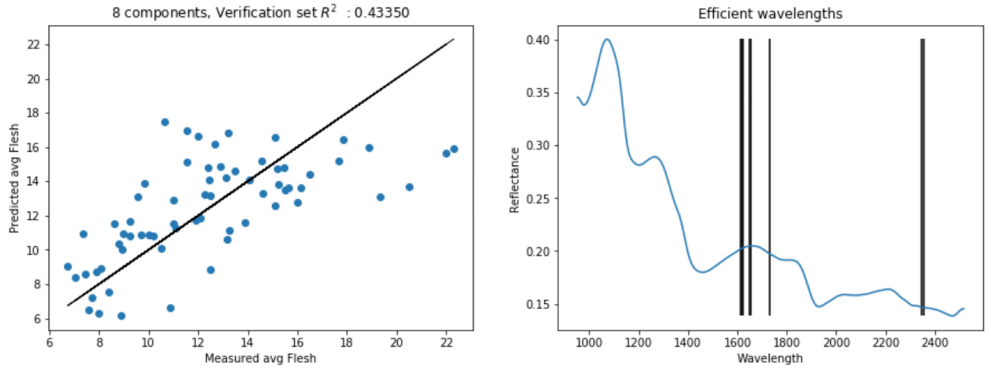
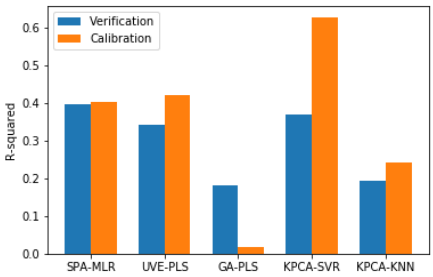
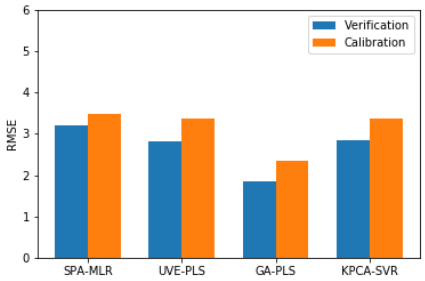


Figure 7.3: Performance and EWs of UVE-PLS using SG filter with window size 21, 5th polynomial and 1st derivative in the SWIR spectral range. The selected EWs are: 1611, 1616, 1622, 1649, 1655, ..

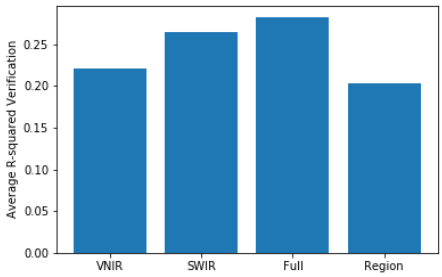
SPA-MLR and UVE-PLS had a substantial accuracy increase using region in SWIR than in VNIR, while the other models only had a slight increase. GA-PLS goes to the contrary compared to the models: the accuracy was better in VNIR when using region. This can be because the SWIR region is in some way less susceptible to noise than VNIR, or because the different ways the models functions.



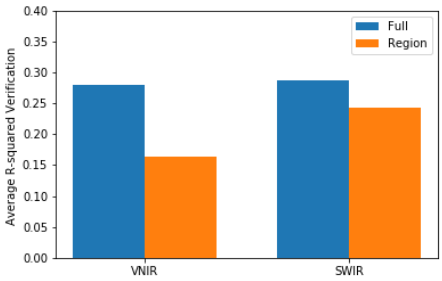
(a) Best firmness R^2 values for each model.



(b) Best firmness RMSE values for each model (KPCA-KNN excluded).



(c) Comparison between VNIR, SWIR, Full and Region dataset.



(d) Full/region dataset breakdown for VNIR and SWIR.

Figure 7.4: Comparisons of how well the models performed and the differences in performance between the two datasets between VNIR and VNIR.

7.3.2 Full vs Region

In most papers on ripeness determination of "Hayward" kiwi, what area on the kiwi they used to collect the spectra is not specified. However, Zhu *et al.* ([9]) showed that using the full dataset gave the best results, but this was not used for "Hayward". It has also been shown that the selection of the region of interest is crucial for the results [120].

In our dataset, the Full region improved the accuracy significantly. The average R^2_v for the Full region was about 0.28 while for the region, about 0.20 (As seen on figure 7.4C). For example, for SPA-MLR on VNIR the accuracy increased from 0.167 to 0.315 when using the Full region. For almost all the models, the predictive power was increased when using the Full region. The exception to this rule is SPA-MLR and KPCA-KNN for SWIR.

The Full dataset performed better because the area of average reflectance is larger. This reduces sources of error like specular reflection and unwanted reflection from the kiwi hairs. The quality attributes can also be different in various regions on the kiwi, therefore using the Full dataset reduces these differences.

7.4 SSC

Many of the same conclusions drawn from firmness can be made here as well. SWIR performed much better than VNIR and Full performed much better than Region. There is an exception here as well, for VNIR KPCA-KNN performed best with Region. KPCA-KNNs predictive power seems to vary significantly depending on what to be predicted, camera, and dataset.

The best R_v^2 made on SSC on "Hayward" kiwi was 0.98[7], this is the same paper that had the best results for firmness. Our best R_v^2 was 0.759 with UVE-PLS, which is far away from the best result which has been done before. However, our R_v^2 of 0.759 is considered to be moderate to good. It is worth noting that the perception of what an excellent R^2 value is depends on the situation. Its generally accepted that a R^2 value above 0.95 is considered to be significant.

SSC

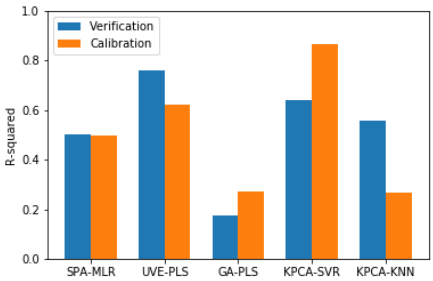
Camera	Dataset	Model	Pre-processing	EWs	$RMSE_C$	$RMSE_V$	R_C^2	R_V^2
VNIR	Full	SPA-MLR	SNV SavitzkyGolay(27,5,1)	21	1.343	1.11	0.397	0.405
		UVE-PLS	SNV SavitzkyGolay(27,5,1)	16	1.308	1.047	0.428	0.39
		GA-PLS	SavitzkyGolay(45,2,2)	86	2.783	1.669	0.097	0.156
		KPCA-SVR	StandardScaler	comp=41	1.109*	1.141*	0.682	0.24
		KPCA-KNN	MSC, SavitzkyGolay(31,2,0)	178	1.082*	0.779*	0.068	0.062
	Region	SPA-MLR	SNV SavitzkyGolay(21,5,2)	33	1.364	1.131	0.376	0.371
		UVE-PLS	SNV SavitzkyGolay(27,5,1)	20	1.352	1.1	0.381	0.314
		GA-PLS	MSC SavitzkyGolay(49,2,1)	86	1.078	0.989	0.231	0.111
		KPCA-SVR	StandardScaler	comp=59	0.878*	1.24*	0.813	0.216
		KPCA-KNN	SNV, SavitzkyGolay(39,5,2)	17	0.952*	0.939*	0.171	0.191
SWIR	Full	SPA-MLR	SNV SavitzkyGolay(33,5,2)	24	1.186	1.11	0.495	0.503
		UVE-PLS	**SavitzkyGolay(11,5,1)**	**15**	**1.016**	**0.777**	**0.621**	**0.759**
		GA-PLS	SavitzkyGolay(49,6,2)	131	2.202	1.825	0.27	0.173
		KPCA-SVR	StandardScaler	comp=49	0.745*	0.851*	0.866	0.638
		KPCA-KNN	SNV SavitzkyGolay(45,2,1)	6	1.056*	0.812*	0.265	0.556
	Region	SPA-MLR	SNV SavitzkyGolay(7,5,2)	11	1.533	1.547	0.163	0.174
		UVE-PLS	SavitzkyGolay(27,5,2)	15	1.222	0.977	0.494	0.493
		GA-PLS	SavitzkyGolay(21,6,2)	136	2.873	2.05	0.042	0.018
		KPCA-SVR	StandardScaler	comp=59	0.993*	1.062*	0.747	0.403
		KPCA-KNN	SNV, SavitzkyGolay(45,2,2)	6	0.964*	0.815*	0.427	0.513

Table 7.2: The best performances of each model for evaluating SSC in the VNIR and the SWIR separate. Both spectral ranges was tested on two different datasets, Full (average of whole kiwi) and Region (150x150 and 30x30 averaged area for VNIR and SWIR respectively). Every model used the same calibration (70%) and verification (15%) set. (* KPCA-SVR and KPCA-KNN uses StandardScaler, therefore the RMSE value of this models is not directly comparable).

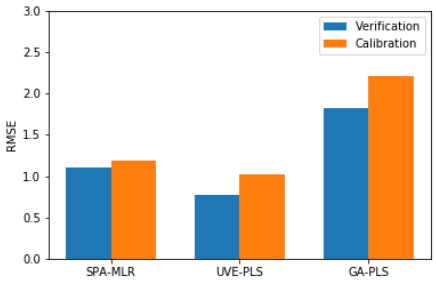
However, as seen from the box-plots in data acquisition, the SSC of the first two boxes deviates a lot from the rest of the data. It was discovered that excluding these boxes from the dataset decreased the RMSE, but the R_v^2 did not increase. This is probably because the number of kiwi samples decreases by $55 * 2 = 110$. If we had measured more kiwis with greater accuracy and higher bio-variability, the predictive power of our models would also increase.

The EWs that was included in the best SSC prediction model was in the range 957, 984-995, 1398-1415, 1616-1627 and 2069-2074nm. This showed to be quite

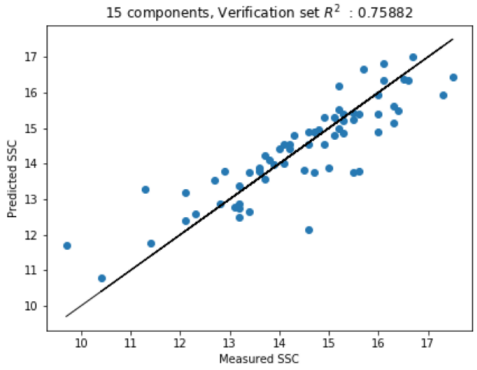
similar to what Li *et al.*[8] observed, which reported 950, 1150, 1200, 1400 and 1650nm (but this was for TSS). Zhu *et al.* also reported that 945, 960 and 987nm served as EWs after investigating the VNIR spectral range on another kiwi variety. Because we share common EWs with other publications, the wavelengths can be justified of having a strong connection with SSC in kiwi.



(a) Best SSC R^2 values for each model.



(b) Best SSC RMSE values for each model (KPCA-KNN/KPCA-SVR excluded).



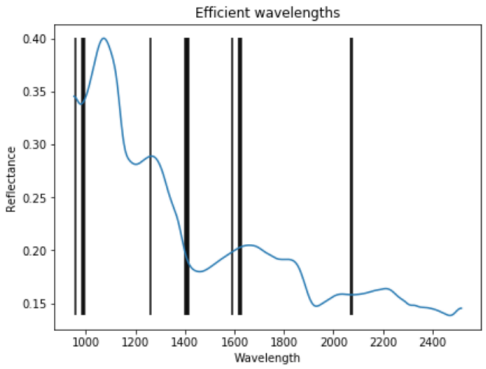


Figure 7.6: Performance and EWs of UVE-PLS using SG filter with window size 11, 5th polynomial and 1st derivative in the SWIR spectral range. The selected EWs are: 957, 984, 989, 995, 1262, 1398, 1404, 1409, 1415, 1589, 1616, 1622, 1627, 2069, 2074.

7.5 Other models and pH

Not all models were prioritized, which are listed in the table below. Also, an attempt at predicting pH was made. Because of the lack of priority, not all performance metrics are listed. Salt did not show any great results. Therefore this feature is not listed.

Some of these metrics show great potential, but because of the lack of time it was not further investigated.

Other results (VNIR full)

Model	Pre-processing	Feature	$RMSE_p$	R^2 - adj_V	R^2_V
MLP (adam, relu)	MSC SavitzkyGolay(25,2,2)	pH level	0.039	0.14	N/A
KNN	SavitzkyGolay(15,6,1)	pH level	0.047	0.139	0.165
Treeregressor	None	Firmness	2.339	N/A	N/A
ANN	StandardScaler	Firmness	N/A	N/A	0.003
RFE-MLP	SavitzkyGolay(43,3,0)	Firmness	2.051	N/A	N/A

Table 7.3: Results for other models and (pH).

7.6 Model discussion

A discussion about some of the models used, is made here. A justification is made about why the model performed well or poorly.

7.6.1 Artificial neural network (ANN)

The neural network never got below 0.9 MSE on firmness, however, the data was scaled so the error was in reality much larger. The R^2 score was around 0.09, which shows there was little success with the neural network. The neural network winded up only guessing the mean of the training data, which was a problem since the data was scattered.

Scaling the value range so that the ANN would be able to do a prediction was troublesome. Optimally the value range should be between -1 and 1, scaling the spectra and measurements to this range was hard.

For the ANN to work properly, methods to scale the data has to be applied and much more samples needs to be collected. Also ANN works better for classification than for regression models, and has been applied accordingly.

The MLP showed greater potential than the regular neural network. Achieving a 0.14 R^2 adjusted on pH.

7.6.2 KNN

At first KNN was used without KPCA, only with and without SG-filter parameters. By using KPCA the curse of dimensionality was avoided and much better predictions where made.

7.6.3 SPA-MLR

SPA-MLR was tested with several pre-processing techniques such as MSC and SNV in combination with derivatives using Savitzky-Golay. These improved the overall model performance and resulted in using fewer EWs. The results in both SSC and firmness was competing against the best model as it performed quite well.

7.6.4 GA-PLS

GA-PLS seemed to have trouble converging towards fewer EW. Usually, it selected way to many variables and even with modifications it seemed to only converge against many EWs.

The GA was warm-started through different variable scoring techniques through "sklearn". Without this warm-start the GA would take an exceptionally long time to find any reasonable results. An example of how the variable scores would look like is shown in Figure 7.7.

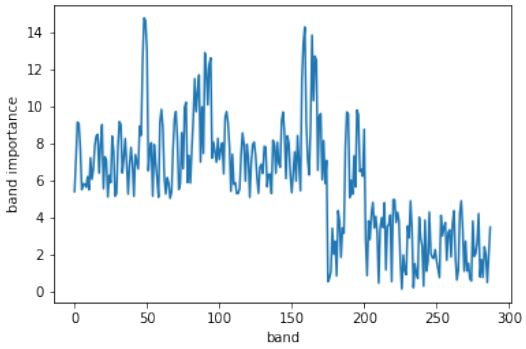


Figure 7.7: Importance of each band for flesh firmness in SWIR.

Many scientific papers have used GA-PLS in combination with other models to achieve better results. For example Zhu *et al.* (Zhu *et al.*) achieved 0.95 R^2 on flesh firmness with GA-PLS-MLR. Because of the time limitation, this was not tried out.

7.6.5 UVE-PLS

The maximal number of components was set to 25 as the search for a good model with few independent variables was preferred. It was tested with MSC, SNV and derivatives using Savitzky-Golay with different window sizes but only fitting using the 5th polynomial. Only this polynomial was thoroughly tested to save time due to the algorithm being computationally expensive. The model generally performed well in predicting both kiwi parameters but can be more tested using other orders of the polynomial fitting.

7.6.6 KPCA-SVR

SVR was tested with pre-processing techniques like SNV, StandardScaler and SG filter, both with and without dimensionality reduction by KPCA. The results with pre-processing and dimensionality reduction combined were lower than just using dimensionality reduction.

7.7 Evaluation of spectral pre-processing

Standard normal variate (SNV) and multiplicative scatter correction (MSC) improved the performance of most of the models at predicting both firmness and SSC. If these were combined with derivatives, they did not always seem as necessary as the models did not get any significant performance boost.

The use of derivatives for spectral pre-processing showed to be advantageous as it generally increased the predictive power of each model. The 1st derivative showed to give better results at predicting firmness, but the 2nd derivative for predicting SSC.

When differentiating the spectra, the height differences are eliminated, and the spectra becomes more comparable. By doing this, it forces the models to look more at the contours of the spectra instead of the spectra itself. By using the SG-filter to retrieve the derivatives, noise in the spectra is also reduced.

When differentiating spectra, the height variations are eliminated, and the small variations are amplified. This makes it easier for models to find the small differences, which may be essential for predicting different quality parameters. By this observation, we can address that most models for predicting both SSC and firmness are dependent on the slight differences.

Each model performed differently on the same pre-processed data. This means that there are no general best pre-processing techniques as it depends on the data and the models used.

Chapter 8

Conclusion

The results obtained in this project have not been reported before, so that we will publish our work in a scientific journal/conference.

The SSC accuracy achieved is moderate to good (R^2=0.74, $RMSE_v = 0.777$). This was achieved using UVE-PLS with SNV and SG-filter. This is very good with the geometrical and material limitations that comes with using HS cameras.

The firmness accuracy was however harder to predict (R^2=0.43, $RMSE_v$=2.80). This was done using UVE-PLS with the SG-filter.

8.0.1 Answer to research aims

1. To investigate the potential use of HSI to be used for non-destructive assessment of firmness, sugar content and pH for kiwi of type "Hayward"?

   HSI can be used to perform non-destructive assessment of "Hayward" kiwi. Especially Sugar content (SSC), however firmness and pH is harder to predict.
2. Study the correlation between sensory information and the internal qualities for kiwi of type "Hayward"?

   This has been looked at through our machine learning models. However, more research is needed to draw a more accurate conclusion.
3. What techniques and models provide the best results for the determination of kiwi's ripeness?

   Conventional methods like UVE-PLS and SPA-MLR works well for non-destructive determination of kiwi ripeness.
4. Which wavelengths are most relevant for determining internal quality parameters of kiwi?

   The wavelengths our models used to regress, in combination with those provided by state of the art, the EWs for firmness, and SSC have been retrieved.
5. Can the process of predicting ripeness of kiwis be automated?

   Yes, through our automatic spectra collector, pre-processing techniques, and machine learning models, we have seen the potential of

automating the process successfully in the future.

8.1 Future work

- **SWIR range:** More work is needed at using the SWIR range with HS cameras since our results has shown that using this range gives better predictions.
- **Efficient wavelengths:** EWs for quality parameters like pH and DM needs to be found. This applies in both SWIR and VNIR ranges with HS cameras.
- **Portable device:** Once the EW are further specified, the possibility of using multispectral imaging to make a small portable device that can determine kiwi ripeness.
- **RGB imaging:** In the wide future, research needs to be conducted on the possibility of using RGB imaging to determine kiwi ripeness.
- **Hayward kiwi type:** There are very few studies on "Hayward" kiwi. More research needs to be done on this kiwi variety.
- **Combining VNIR and SWIR ranges:** Future research needs to look at combining the spectra from both VNIR and SWIR as this can possibly contain more information and provide better results.

.

8.2 Learning outcome

To be able to complete this project, much knowledge was required. However, the members have been able to acquire this knowledge and use it to achieve good results. The members have learned about "Hayward" kiwi, HSI with its complex equipment and acquisition techniques, machine learning, pre-processing techniques, dimensionality reduction, and data collection. For HSI this is very useful knowledge to have in the current growing market.

The group also learned how to work in teams to coordinate efforts to achieve a common goal. We further learned how to apply programming to solve real-world problems.

This research project gave us an insight on how to properly do a literature review and write a scientific report that contributes to state-of-the-art research.

8.3 Contribution

By combining the knowledge from the spectroscopic devices[8][7], and our results from the HS cameras, a small portable HS camera that can assess kiwi ripeness is closer to reality. The specific wavelengths selected for each quality parameter retrieved from the spectroscopic devices and our HS camera can be put onto a

portable multispectral camera. By using this small multispectral camera, the ripeness of "Hayward" kiwi can be evaluated quickly and remotely.

A table that shows our contributions is listed below. Light green color indicates our main contributions, while the light yellow color indicates less significant contribution.

Contribution	**Comment**
Portable Device	Our project has further shown what wavelengths and models are relevant for SSC and Firmness. By further specifying this, the future goal of having a portable ripeness assessment device is one step closer.
SSC	Our predictions made on SSC has further improved its predictability
Firmness	Our predictions made on firmness has further improved its predictability
pH	pH was harder to predict. However a attempt was made to predict this, using MLP and KNN Futher work needs to be done to predict this quality parameter.
400-2500nm	A dataset of hyperspectral images of kiwi covering the spectral range 400-2500nm. This has high scientific value. What wavelengths and models work in this range is also of high importance.
Automatic spectra collector	Automatic spectra collection has been done on kiwis before However our results show that this is also possible for "Hayward" kiwi, and that the full dataset gave the best results.
RGB imaging	By taking RGB images of the kiwis, we have created a dataset that can be used in the future to predict ripeness. This dataset has high scientific value because it sets the foundation of what could be possible in the future.
Peels	The collected spectral information of the peels has opened doors for future work
Models	We have investigated the predictability of new and common models for determining kiwi ripeness. UVE-PLS and SPA-MLR worked best for sugar and firmness, while MLR showed potential to predict pH level. Tree-regressor and ANN showed poor performance this application

Bibliography

[1] M. Wandel, 'Dietary intake of fruits and vegetables in Norway: Influence of life phase and socio-economic factors,' *Int. J. Food Sci. Nutr.*, vol. 46, no. 3, pp. 291–301, Sep. 1995. DOI: 10.3109/09637489509012561.

[2] *Hyperspectral Imaging Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026)*, [Online; accessed 5. May 2021], May 2021. [Online]. Available: https://www.mordorintelligence.com/industry-reports/hyperspectral-imaging-market.

[3] *The hype in spectral imaging*, [Online; accessed 11. May 2021], May 2021. [Online]. Available: https://www.spectroscopyeurope.com/article/hype-spectral-imaging.

[4] *Dette er den mest populære frukten*, [Online; accessed 5. May 2021], Feb. 2014. [Online]. Available: https://www.godt.no/aktuelt/i/g7On7A/dette-er-den-mest-populaere-frukten.

[5] wikiHow, 'How to Tell if a Kiwi Is Ripe,' *wikiHow*, Feb. 2020. [Online]. Available: https://www.wikihow.com/Tell-if-a-Kiwi-Is-Ripe.

[6] K. Singletary, 'Kiwifruit: Overview of Potential Health Benefits,' *Nutrition Today*, vol. 47, no. 3, pp. 133–147, May 2012, ISSN: 0029-666X. DOI: 10.1097/NT.0b013e31825744bc.

[7] J. S. Lee, S.-C. Kim, K. C. Seong, C.-H. Kim, Y. C. Um and S.-K. Lee, 'Quality Prediction of Kiwifruit Based on Near Infrared Spectroscopy,' *Horticultural Science & Technology*, vol. 30, no. 6, pp. 709–717, 2012, ISSN: 1226-8763. DOI: 10.7235/hort.2012.12139.

[8] M. Li, R. R. Pullanagari, T. Pranamornkith, I. J. Yule and A. R. East, 'Quantitative prediction of post storage 'Hayward' kiwifruit attributes using at harvest Vis-NIR spectroscopy,' *J. Food Eng.*, vol. 202, pp. 46–55, Jun. 2017, ISSN: 0260-8774. DOI: 10.1016/j.jfoodeng.2017.01.002.

[9] H. Zhu, B. Chu, Y. Fan, X. Tao, W. Yin and Y. He, 'Hyperspectral Imaging for Predicting the Internal Quality of Kiwifruits Based on Variable Selection Algorithms and Chemometric Models,' *Sci. Rep.*, vol. 7, no. 7845, pp. 1–13, Aug. 2017, ISSN: 2045-2322. DOI: 10.1038/s41598-017-08509-6.

[10] W. Hu, D.-W. Sun and J. Blasco, 'Rapid monitoring 1-MCP-induced modulation of sugars accumulation in ripening 'Hayward' kiwifruit by Vis/NIR hyperspectral imaging,' *Postharvest Biol. Technol.*, vol. 125, pp. 168–180, Mar. 2017, ISSN: 0925-5214. DOI: 10.1016/j.postharvbio.2016.11.001.

[11] *Hayward - Kiwi - Good Fruit Guide*, [Online; accessed 22. Apr. 2021], Apr. 2021. [Online]. Available: https://goodfruitguide.co.uk/product/hayward.

[12] FAO, *Global food losses and food waste – extent, causes and prevention*, 2011.

[13] *UN: 17% of all food available at consumer levels is wasted*, [Online; accessed 13. Apr. 2021], Jul. 2021. [Online]. Available: https://www.unep.org/news-and-stories/press-release/un-17-all-food-available-consumer-levels-wasted.

[14] K. Schanes, K. Dobernig and B. Gözet, 'Food waste matters - A systematic review of household food waste practices and their policy implications,' *J. Cleaner Prod.*, vol. 182, pp. 978–991, May 2018, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2018.02.030.

[15] A. Bernstad Saraiva Schott and T. Andersson, 'Food waste minimization from a life-cycle perspective,' *J. Environ. Manage.*, vol. 147, pp. 219–226, Jan. 2015, ISSN: 0301-4797. DOI: 10.1016/j.jenvman.2014.07.048.

[16] *Sustainable Development Goals | UNDP*, [Online; accessed 12. Apr. 2021], Apr. 2021. [Online]. Available: https://www.undp.org/content/undp/en/home/sustainable-development-goals.html.

[17] L. Sabir, *Customer Satisfaction Parameters for Fruits and Vegetables Retail-An AHP Approach*, [Online; accessed 17. May 2021], May 2021. [Online]. Available: https://www.academia.edu/23519078/Customer_Satisfaction_Parameters_for_Fruits_and_Vegetables_Retail_An_AHP_Approach.

[18] Q. Li, X. He, Y. Wang, H. Liu, D. Xu and F. Guo, 'Review of spectral imaging technology in biomedical engineering: achievements and challenges,' *J. Biomed. Opt.*, vol. 18, no. 10, p. 100 901, Oct. 2013, ISSN: 1083-3668. DOI: 10.1117/1.JBO.18.10.100901.

[19] G. Özdoğan, X. Lin and D.-W. Sun, 'Rapid and noninvasive sensory analyses of food products by hyperspectral imaging: Recent application developments,' *Trends in Food Science & Technology*, vol. 111, pp. 151–165, 2021, ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2021.02.044.

[20] *What is hyperspectral imaging? - Hyperspectral Imaging Cameras and Systems | Middleton Spectral Vision*, [Online; accessed 4. Mar. 2021], Aug. 2019. [Online]. Available: https://www.middletonspectral.com/resources/what-is-hyperspectral-imaging.

[21] J. F. de Canete, C. Galindo and I. Garcia-Moral, *System Engineering and Automation*. Berlin, Germany: Springer, 2011, ISBN: 978-3-64220230-8. [Online]. Available: https://books.google.no/books?id=h8rCQYXGGY8C&q=most+systems+are+inherently+nonlinear+in+nature&pg=PA46&redir_esc=y#v=snippet&q=most%5C%20systems%5C%20are%5C%20inherently%5C%20nonlinear%5C%20in%5C%20nature&f=false.

[22] Wikipedia contributors, *Chemometrics — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=Chemometrics&oldid=991875113, [Online; accessed 2-May-2021], 2020.

[23] M. Abdo, V. Badilita and J. Korvink, 'Spatial scanning hyperspectral imaging combining a rotating slit with a dove prism,' *Opt. Express*, vol. 27, no. 15, pp. 20 290–20 304, Jul. 2019. DOI: 10.1364/OE.27.020290.

[24] Y. W. Wang, N. P. Reder, S. Kang, A. K. Glaser and J. T. C. Liu, 'Multiplexed Optical Imaging of Tumor-Directed Nanoparticles: A Review of Imaging Systems and Approaches,' *Nanotheranostics*, vol. 1, no. 4, p. 369, 2017. DOI: 10.7150/ntno.21136.

[25] X. Li, R. Li, M. Wang, Y. Liu and J. Zhou, *Hyperspectral Imaging and Their Applications in the Nondestructive Quality Assessment of Fruits and Vegetables*. Aug. 2018, ISBN: 978-1-78923-290-5. DOI: 10.5772/intechopen.72250.

[26] R. Pillay, J. Y. Hardeberg and S. George, 'Hyperspectral imaging of art: Acquisition and calibration workflows,' *Journal of the American Institute for Conservation*, vol. 58, no. 1-2, pp. 3–15, Apr. 2019, ISSN: 0197-1360. DOI: 10.1080/01971360.2018.1549919.

[27] *Considerations for Diffuse Reflection Spectroscopy*, [Online; accessed 7. Apr. 2021], Apr. 2021. [Online]. Available: https://www.americanlaboratory.com/914-Application-Notes/190743-Considerations-for-Diffuse-Reflection-Spectroscopy.

[28] Contributors to Wikimedia projects, *Hann function - Wikipedia*, [Online; accessed 6. May 2021], Jan. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Hann_function&oldid=1001711522.

[29] B. Regeling, B. Thies, A. O. H. Gerstner, S. Westermann, N. A. Müller, J. Bendix and W. Laffers, 'Hyperspectral Imaging Using Flexible Endoscopy for Laryngeal Cancer Detection,' *Sensors (Basel).*, vol. 16, no. 8, p. 1288. Aug. 2016, ISSN: 1424-8220. DOI: 10.3390/s16081288. eprint: 27529255.

[30] S. Braun, 'WINDOWS,' in *Encyclopedia of Vibration*, Walthm, MA, USA: Elsevier, Jan. 2001, pp. 1587–1595, ISBN: 978-0-12-227085-7. DOI: 10.1006/rwvb.2001.0052.

[31] D. R. M. Zeaiter, 'Comprehensive chemometrics,' *Science Direct*, vol. 3, 2009. [Online]. Available: https://www.sciencedirect.com/topics/mathematics/standard-normal-variate.

[32] Y.-Y. Pu, Y.-Z. Feng and D.-W. Sun, 'Recent Progress of Hyperspectral Imaging on Quality and Safety Inspection of Fruits and Vegetables: A Review,' *Compr. Rev. Food Sci. Food Saf.*, vol. 14, no. 2, pp. 176–188, Mar. 2015, ISSN: 1541-4337. DOI: 10.1111/1541-4337.12123.

[33] W. H. Press and S. A. Teukolsky, 'Savitzky-Golay Smoothing Filters,' *Comput. Phys.*, vol. 4, no. 6, p. 669, 1990, ISSN: 0894-1866. DOI: 10.1063/1.4822961.

[34] B. Zimmermann and A. Kohler, 'Optimizing Savitzk Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy,' *Appl. Spectrosc.*, vol. 67, no. 8, pp. 892–902, Aug. 2013, ISSN: 0003-7028. DOI: 10.1366/12-06723.

[35] B. M. Nicola1, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron and J. Lammertyn, 'Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review,' *Postharvest Biol. Technol.*, vol. 46, no. 2, pp. 99–118, Nov. 2007, ISSN: 0925-5214. DOI: 10.1016/j.postharvbio.2007.06.024.

[36] S. F. C. Soares, A. A. Gomes, M. C. U. Araujo, A. R. G. Filho and R. K. H. Galvão, 'The successive projections algorithm,' *TrAC Trends in Analytical Chemistry*, vol. 42, pp. 84–98, 2013, ISSN: 0165-9936. DOI: https://doi.org/10.1016/j.trac.2012.09.006.

[37] R. Leardi, 'Application of genetic algorithm–PLS for feature selection in spectral data sets,' *J. Chemom.*, vol. 14, no. 5-6, pp. 643–655, Sep. 2000, ISSN: 0886-9383. DOI: 10.1002/1099-128X(200009/12)14:5/6<643::AID-CEM621>3.0.CO;2-E.

[38] *Kennard-Stone algorithm and SPXY algorithm with python implementation | HXHC Blog*, [Online; accessed 14. Apr. 2021], Mar. 2021. [Online]. Available: https://hxhc.github.io/post/kennardstone-spxy.

[39] W. Guo, F. Zhao and J. Dong, 'Nondestructive Measurement of Soluble Solids Content of Kiwifruits Using Near-Infrared Hyperspectral Imaging,' *Food Anal. Methods*, vol. 9, no. 1, pp. 38–47, Jan. 2016, ISSN: 1936-976X. DOI: 10.1007/s12161-015-0165-z.

[40] D. E. Birba, *A Comparative study of data splitting algorithms for machine learning model selection*. 2020. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1506870&dswid=-85.

[41] K. A. Marill, 'Advanced statistics: Linear regression, part ii: Multiple linear regression,' *Academic emergency medicine*, vol. 11, no. 1, pp. 94–102, 2004. [Online]. Available: https://www.schmidheiny.name/teaching/ols.pdf.

[42] K. Jennings, 'Statistics 512: Applied linear models,' *Internet.[Google Scholar]*, 2012. [Online]. Available: https://www.stat.purdue.edu/~boli/stat512/lectures/topic3.pdf.

[43] P. Geladi and B. R. Kowalski, 'Partial least-squares regression: a tutorial,' *Anal. Chim. Acta*, vol. 185, pp. 1–17, Jan. 1986, ISSN: 0003-2670. DOI: 10.1016/0003-2670(86)80028-9.

[44] Contributors to Wikimedia projects, *Partial least squares regression - Wikipedia*, [Online; accessed 26. Apr. 2021], Mar. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Partial_least_squares_regression&oldid=1014028663.

[45] E. Chow, D. Ebrahimi, J. J. Gooding and D. B. Hibbert, 'Application of N-PLS calibration to the simultaneous determination of Cu2+, Cd2+ and Pb2+ using peptide modified electrochemical sensors,' *Analyst*, vol. 131, no. 9, pp. 1051–7, Oct. 2006, ISSN: 0003-2654. DOI: 10.1039/b604690h.

[46] H. Lohninger, *PLS - Partial Least Squares Regression*, [Online; accessed 19. May 2021], Mar. 2021. [Online]. Available: http://www.statistics4u.com/fundstat_eng/dd_pls.html.

[47] K. S. Ng, 'A simple explanation of partial least squares,' *The Australian National University, Canberra*, 2013.

[48] Contributors to Wikimedia projects, *k-nearest neighbors algorithm - Wikipedia*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=K-nearest_neighbors_algorithm&oldid=1021698330.

[49] C. Xie, C. Yang and Y. He, 'Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities,' *Comput. Electron. Agric.*, vol. 135, pp. 154–162, Apr. 2017, ISSN: 0168-1699. DOI: 10.1016/j.compag.2016.12.015.

[50] *Fig. 1. Artificial neural network architecture (ANN i-h 1-h 2-h n-o)*. [Online; accessed 10. May 2021], May 2021. [Online]. Available: https://www.researchgate.net/figure/Artificial-neural-network-architecture-ANN-i-h-1-h-2-h-n-o_fig1_321259051.

[51] *A Gentle Introduction to Dropout for Regularizing Deep Neural Networks*, [Online; accessed 24. Apr. 2021], Aug. 2019. [Online]. Available: https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks.

[52] *Neural Network L2 Regularization Using Python – Visual Studio Magazine*, [Online; accessed 24. Apr. 2021], Apr. 2021. [Online]. Available: https://visualstudiomagazine.com/articles/2017/09/01/neural-network-l2.aspx.

[53] G. ElMasry, N. Wang and C. Vigneault, 'Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks,' *Postharvest Biol. Technol.*, vol. 52, no. 1, pp. 1–8, Apr. 2009, ISSN: 0925-5214. DOI: 10.1016/j.postharvbio.2008.11.008.

[54] *How XGBoost Works - Amazon SageMaker*, [Online; accessed 20. May 2021], May 2021. [Online]. Available: https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html.

[55] K. Grace-Martin, 'Can a regression model with a small r-squared be useful?' *The Analysis Factor*, 2012. [Online]. Available: https://people.duke.edu/~rnau/rsquared.htm.

[56] *Cross Validation and its types! | Data Science and Machine Learning*, [Online; accessed 19. May 2021], May 2021. [Online]. Available: https://www.kaggle.com/discussion/204878.

[57] N. K. Given, 'Kiwifruit,' in *Biochemistry of Fruit Ripening*, G. B. Seymour, J. E. Taylor and G. A. Tucker, Eds. Dordrecht: Springer Netherlands, 1993, pp. 235–254, ISBN: 978-94-011-1584-1. DOI: 10.1007/978-94-011-1584-1_7. [Online]. Available: https://doi.org/10.1007/978-94-011-1584-1_7.

[58] M. Montefiori, T. K. McGhie, I. C. Hallett and G. Costa, 'Changes in pigments and plastid ultrastructure during ripening of green-fleshed and yellow-fleshed kiwifruit,' *Sci. Hortic.*, vol. 119, no. 4, pp. 377–387, Feb. 2009, ISSN: 0304-4238. DOI: 10.1016/j.scienta.2008.08.022.

[59] P. Chen and Z. Sun, 'A review of non-destructive methods for quality evaluation and sorting of agricultural products,' *Journal of Agricultural Engineering Research*, vol. 49, pp. 85–98, 1991.

[60] H. Li, P. Pidakala, D. Billing and J. Burdon, 'Kiwifruit firmness: Measurement by penetrometer and non-destructive devices,' *Postharvest Biol. Technol.*, vol. 120, pp. 127–137, Oct. 2016, ISSN: 0925-5214. DOI: 10.1016/j.postharvbio.2016.06.007.

[61] *Hayward - Kiwi - Good Fruit Guide*, [Online; accessed 26. Apr. 2021], Apr. 2021. [Online]. Available: https://goodfruitguide.co.uk/product/hayward.

[62] *An investigation of ripening and handling strategies for early season kiwifruit (Actinidia deliciosa cv Hayward)*, [Online; accessed 28. Apr. 2021], Apr. 2021. DOI: 10.1002/jsfa.2740470402.

[63] [Online; accessed 26. Apr. 2021], Feb. 2020. [Online]. Available: https://research.libraries.wsu.edu/xmlui/bitstream/handle/2376/7177/pnw507.pdf?sequence=1&isAllowed=y.

[64] J. Ford, 'Harvesting and maturity of chinese gooseberries,' *Orchardist NZ*, 1971.

[65] M. Reid, 'Maturation and storage of kiwifruit,' in *Proceedings of Kiwifruit Seminar, Tauranga*, 1977, pp. 3–52.

[66] J. Harman, 'Kiwifruit maturity [new zealand].,' *Orchardist of New Zealand*, 1981.

[67] G. Hopkirk, D. Beever and C. Triggs, 'Variation in soluble solids concentration in kiwifruit at harvest,' *New Zealand journal of agricultural research*, vol. 29, no. 3, pp. 475–484, 1986.

[68] F. Mitchell and G. Mayer, 'Relation of dry weight to 'hayward'kiwifruit maturity,' *Progress Report to the California Kiwifruit Commission*, 1987.

[69] F. G. Mitchell, 'Postharvest physiology and technology of kiwifruit.,' *Acta Hortic.*, no. No. 282, pp. 291–307, 1990, ISSN: 0567-7572. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/19920316257.

[70] M. S. Reid, D. A. Heatherbeli and H. K. Pratt, 'Seasonal patterns in chemical composition of the fruit of actinidia chinensis1,' *Annu. Rev. Plant Physiol*, vol. 25, pp. 135–166, 1982.

[71] F. Mitchell, G. Mayer and W. Biasi, 'Using total solids to measure kiwifruit maturity,' *Progress Report to the California Kiwifruit Commission*, 1989.

[72] E. A. Macrae, J. H. Bowen and M. G. H. Stec, 'Maturation of kiwifruit (Actinidia deliciosa cv Hayward) from two orchards: Differences in composition of the tissue zones,' *J. Sci. Food Agric.*, vol. 47, no. 4, pp. 401–416, Jan. 1989, ISSN: 0022-5142. DOI: 10.1002/jsfa.2740470403.

[73] V. A. McGlone and S. Kawano, 'Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy,' *Postharvest Biol. Technol.*, vol. 13, no. 2, pp. 131–141, Apr. 1998, ISSN: 0925-5214. DOI: 10.1016/S0925-5214(98)00007-6.

[74] P. Martinsen and P. Schaare, 'Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy,' *Postharvest Biol. Technol.*, vol. 14, no. 3, pp. 271–281, Nov. 1998, ISSN: 0925-5214. DOI: 10.1016/S0925-5214(98)00051-9.

[75] S. D. Osborne, R. Künnemeyer and R. B. Jordan, 'A Low-Cost System for the Grading of Kiwifruit,' *J. Near Infrared Spectrosc.*, vol. 7, no. 1, pp. 9–15, Jan. 1999, ISSN: 0967-0335. DOI: 10.1255/jnirs.229.

[76] R. Bro, Å. Rinnan and N. K. M. Faber, 'Standard error of prediction for multilinear pls: 2. practical implementation in fluorescence spectroscopy,' *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 1, pp. 69–76, 2005.

[77] E. MacRae, R. Redgwell and T. Wegrzyn, 'The whens and hows of fruit softening,' *NZ Kiwifruit*, vol. 64, pp. 15–17, 1990.

[78] M. G. H. Stec, J. A. Hodgson, E. A. Macrae and C. M. Triggs, 'Role of fruit firmness in the sensory evaluation of kiwifruit (Actinidia deliciosa cv Hayward),' *J. Sci. Food Agric.*, vol. 47, no. 4, pp. 417–433, Jan. 1989, ISSN: 0022-5142. DOI: 10.1002/jsfa.2740470404.

[79] A. Berardinelli, A. Benelli, M. Tartagni and L. Ragni, 'Kiwifruit flesh firmness determination by a NIR sensitive device and image multivariate data analyses,' *Sens. Actuators, A*, vol. 296, pp. 265–271, Sep. 2019, ISSN: 0924-4247. DOI: 10.1016/j.sna.2019.07.027.

[80] H. Takao and S. Ohmori, 'Development of device for nondestructive evaluation of fruit firmness,' *Japan agricultural research quarterly*, vol. 28, pp. 36–36, 1994.

[81] K. Peleg, 'Development of a Commercial Fruit Firmness Sorter,' *J. Agric. Eng. Res.*, vol. 72, no. 3, pp. 231–238, Mar. 1999, ISSN: 0021-8634. DOI: 10.1006/jaer.1998.0367.

[82] A. Moghimi, M. H. Aghkhani, A. Sazgarnia and M. Sarmad, 'Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit,' *Biosyst. Eng.*, vol. 106, no. 3, pp. 295–302, Jul. 2010, ISSN: 1537-5110. DOI: 10.1016/j.biosystemseng.2010.04.002.

[83] J. L. Rodgers and W. A. Nicewander, 'Thirteen Ways to Look at the Correlation Coefficient,' *American Statistician*, Feb. 2012. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1988.10475524.

[84] K. Hasegawa and K. Funatsu, 'Partial Least Squares Modeling and Genetic Algorithm Optimization in Quantitative Structure-Activity Relationships,' *SAR QSAR Environ. Res.*, vol. 11, no. 3-4, pp. 189–209, Aug. 2000, ISSN: 1062-936X. DOI: 10.1080/10629360008033231.

[85] H. Wang and D. Hu, 'Comparison of SVM and LS-SVM for Regression,' in *2005 International Conference on Neural Networks and Brain*, vol. 1, IEEE, Oct. 2005, pp. 279–283, ISBN: 978-0-7803-9422. DOI: 10.1109/ICNNB.2005.1614615.

[86] Contributors to Wikimedia projects, *Kiwifruit - Wikipedia*, [Online; accessed 28. Apr. 2021], Apr. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Kiwifruit&oldid=1019851690.

[87] I. Nishiyama, T. Fukuda and T. Oota, 'Genotypic Differences in Chlorophyll, Lutein, and beta-Carotene Contents in the Fruits of Actinidia Species,' *J. Agric. Food Chem.*, vol. 53, no. 16, pp. 6403–6407, Aug. 2005, ISSN: 0021-8561. DOI: 10.1021/jf050785y.

[88] Contributors to Wikimedia projects, *Fruit anatomy - Wikipedia*, [Online; accessed 28. Apr. 2021], Apr. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Fruit_anatomy&oldid=1019129433.

[89] Contributors to Wikimedia projects, *Hue - Wikipedia*, [Online; accessed 28. Apr. 2021], Mar. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Hue&oldid=1011624670.

[90] Contributors to Wikimedia projects, *Density - Wikipedia*, [Online; accessed 15. Apr. 2021], Mar. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Density&oldid=1014162510.

[91] V. A. McGlone, R. B. Jordan, R. Seelye and P. J. Martinsen, 'Comparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content,' *Postharvest Biol. Technol.*, vol. 26, no. 2, pp. 191–198, Sep. 2002, ISSN: 0925-5214. DOI: 10.1016/S0925-5214(02)00014-5.

[92] R. Cangi, E. Altuntas, C. Kaya and O. Saracoglu, 'Some chemical and physical properties at physiological maturity and ripening period of kiwifruit ('Hayward'),' *1.*, vol. 10, no. 27, pp. 5304–5310, 2011, ISSN: 1684-5315. DOI: 10.4314/ajb.v10i27..

[93] R. B. Jordan, E. F. Walton, K. U. Klages and R. J. Seelye, 'Postharvest fruit density as an indicator of dry matter and ripened soluble solids of kiwifruit,' *Postharvest Biol. Technol.*, vol. 20, no. 2, pp. 163–173, Sep. 2000, ISSN: 0925-5214. DOI: 10.1016/S0925-5214(00)00125-3.

[94] P. N. Schaare and D. G. Fraser, 'Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis),' *Postharvest Biol. Technol.*, vol. 20, no. 2, pp. 175–184, Sep. 2000, ISSN: 0925-5214. DOI: 10.1016/S0925-5214(00)00130-7.

[95] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*. Guilford Press, 2011. [Online]. Available: http://www.warsash.com.au/products/light-measurement/REMOTE-SENSING-CALIBRATION-TARGETS.php.

[96] H. Li, P. Pidakala, D. Billing and J. Burdon, 'Kiwifruit firmness: Measurement by penetrometer and non-destructive devices,' *Postharvest Biology and Technology*, vol. 120, pp. 127–137, 2016, ISSN: 0925-5214. DOI: https://doi.org/10.1016/j.postharvbio.2016.06.007.

[97] [Online; accessed 23. Apr. 2021], Apr. 2017. [Online]. Available: https://www.ble.de/SharedDocs/Downloads/EN/Nutrition-Food/Quality-Control/BestimmungFruechteEN.pdf?__blob=publicationFile&v=1.

[98] *What is a Brix Refractometer, and how do they work?* [Online; accessed 23. Apr. 2021], Apr. 2021. [Online]. Available: https://www.instrumentchoice.com.au/news/what-is-a-brix-refractometer-and-how-do-they-work.

[99] *ATAGO CO.,LTD.* [Online; accessed 25. Apr. 2021], Apr. 2021. [Online]. Available: https://www.atago.net/en/index.php.

[100] *Brix and Salt Hybrid Meter| ATAGO CO.,LTD.* [Online; accessed 25. Apr. 2021], Apr. 2021. [Online]. Available: https://www.atago.net/product/?l=en&f=products-bx-salt-top.php#salt-spec.

[101] Contributors to Wikimedia projects, *pH - Wikipedia*, [Online; accessed 25. Apr. 2021], Apr. 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=PH&oldid=1018700712.

[102] R. D. Myers, 'Colorchecker passport technical review,' *Robin Myers Imaging (www. rmimaging. com)*, vol. 3, 2010.

[103] *The ColorChecker Pages (Page 2 of 3)*, [Online; accessed 5. May 2021], May 2021. [Online]. Available: https://www.babelcolor.com/colorchecker-2.htm#CCP2_data.

[104] P. B. Jeffery and N. H. Banks, 'Firmness-temperature coefficient of kiwifruit,' 1994.

[105] J. A. Abbott and D. R. Massie, 'Nondestructive sonic measurement of kiwifruit firmness,' *Journal of the American Society for Horticultural Science*, vol. 123, no. 2, pp. 317–322, 1998.

[106] *pandas - Python Data Analysis Library*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://pandas.pydata.org.

[107] *NumPy*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://numpy.org.

[108] *Welcome to Spectral Python (SPy) — Spectral Python 0.21 documentation*, [Online; accessed 18. May 2021], Apr. 2020. [Online]. Available: http://www.spectralpython.net.

[109] *scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://scikit-learn.org/stable.

[110] *TensorFlow*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://www.tensorflow.org.

[111] *seaborn: statistical data visualization — seaborn 0.11.1 documentation*, [Online; accessed 18. May 2021], Apr. 2021. [Online]. Available: https://seaborn.pydata.org.

[112] *Matplotlib: Python plotting — Matplotlib 3.4.2 documentation*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://matplotlib.org.

[113] *Home - OpenCV*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://opencv.org.

[114] *XGBoost Documentation — xgboost 1.5.0-SNAPSHOT documentation*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://xgboost.readthedocs.io/en/latest.

[115] *Introduction — statsmodels*, [Online; accessed 18. May 2021], Mar. 2021. [Online]. Available: https://www.statsmodels.org/stable/index.html.

[116] *SciPy.org — SciPy.org*, [Online; accessed 18. May 2021], May 2021. [Online]. Available: https://www.scipy.org.

[117] K. Satoshi Suzuki, 'Topological structural analysis of digitized binary images by border following,' *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985, ISSN: 0734-189X. DOI: https://doi.org/10.1016/0734-189X(85)90016-7.

[118] J. K. Delaney and M. Picollo, 'JAIC special issue on "Reflectance hyperspectral imaging to support documentation and conservation of 2D artworks",' *Journal of the American Institute for Conservation*, vol. 58, no. 1-2, pp. 1–2, Apr. 2019, ISSN: 0197-1360. DOI: 10.1080/01971360.2019.1589001.

[119] R. K. H. Galvão, M. C. U. Araújo, W. D. Fragoso, E. C. Silva, G. E. José, S. F. C. Soares and H. M. Paiva, 'A variable elimination method to improve the parsimony of mlr models using the successive projections algorithm,' *Chemometrics and Intelligent Laboratory Systems*, vol. 92, no. 1, pp. 83–91, 2008, ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2007.12.004.

[120] K. Chao, C.-C. Yang and M. S. Kim, 'Spectral line-scan imaging system for high-speed non-destructive wholesomeness inspection of broilers,' *Trends Food Sci. Technol.*, vol. 21, no. 3, pp. 129–137, Mar. 2010, ISSN: 0924-2244. DOI: 10.1016/j.tifs.2009.08.001.

Appendix A

Meeting minutes

Meeting nr 1, with Sony: 02.01.2020

- Get harddrive by this week or next week
- Camera need the computer
- We can get the images next week or in the weekend
- Will be reasembeld this weekend
- We should book a couple of next week 1-2 days
- Get more familiar
- We should test with different filter
- With polarisers we need to change the integration time
- See the shadow formation
- Talked with tto yesterda
    - 3 week of february
    - ★ Need to know how much kiwi we need. BY TODAY
    - Between 150-200
    - 1 box week 3
    - 2 box week 4
    - See their opinion on delivering
    - How long can we keep it
- ★ Everytime we make a measurement we can weigh them
- ★ We get a thermal camera as well to see the temperature.
    - Way to reduce is to keep the lights appart from the camera.
    - Adjust the distance from the light to the fruit.

- The more parameters the better
- ★ Sort the kiwis so that all the scans have the same size
- ★ Note down so that we know how much time a scan takes for kiwi
- Once set just hit record and get the data.
- There are code available to do PLS
- Remember to learn and read about that happens on the inside.
- ★ Measure firmness from different sides on the kiwi and do the mean
- Everyone in the lab
    - Plan on who to do what
    - One to prepearre
    - Second measure firmness
    - Third measure sugar and ph
    - Assembly line.
    - Someone write readings from the measurements.
    - Kiwi numbers with the measurement
    - Identification
        - In every measurement
        - Define which one is 1 and 2
        - Have a sheet of paper with the numbers.
        - Take a picture phones of the kiwi.
        - Next week sony will be in campus.
        - Do the measurement on Monday, make sure data is correct.
    - Should not wait for long until pre-processing and sorting out
- Next week we can to the measurement with the other instruments as well.

- Plan for the meeting next Monday
    - Next Monday 2 pm
    - Book the camera.
        - Monday 2pm - 5pm
        - wedsneyday 2pm-5pm.
        - GET FAMILIAR WITH THE THINGS
- ★ Use the colorchecker for the measurements.
    - It is recommended
- ★ The image normalisation explain next week
-
-
-
- Det viktigste er å finne parametrene
- Sortere med størrelse
- Måle parametrene
- Koble bildet til kiwiene vi har scannet

Meeting nr 2, with Bama: 18.01.2021

- Sigvart jobbet med mye rart i bama
- Brukt optisk sortering
- Making a database and connect it to the internal/external parameters
- Kiwi
    - Might not be the correct season
    - Harvested 2 times a year
    - European
    - Harvested unmature, stored
        - Depends on the market
        - Matured in the store, elevating temperature
    - Norway buy from seppers, new zealand
    - Needs to be in the beginning of the season, is at end
    - Measured by softness, maturity
    - Last season problem
        - Get better at delivering a more even maturity of the kiwis
- Off season
    - Plums from south africa, stonefruits
    - Peaches start at end of april
- Likes to have a close reltionship with the work
- Took some pictures last week

They have a rig at place.
They would like some results.

Avocado has the biggest problem.
Avocado is something to work on

- Not able to make a uniform maturity for the customers
- Sorting company in holland.
- More or less equal maturity when you buy it
- Year around season
    - ripining in holland before sent to norway

    - Grown in europe.
- Rootstock
    - Not original rootsystem
    - Is cut with part of the tree
- Different orgins depens

What is done before?

- A lot of sorting equipment has gotten really far
- Different project
    - Aiming at a more handhold device
    - Needs to have multiple devices
    - Rather than one big machine

Mechanical damage is interessing

-

All berries are outsourced to holland
Pear

- Conference from belgum
- Harvested in october

We can take kiwi and avocado whenever we want
They are perfect fruits
Can be able to play with pre
Avocado has a big market share.
Kiwi is a good alternative. Bama is really of on the maturity compared to other firms.
Avocado price has gone up.
Have been trails to start growing in spain as well
Mafia problems around avocado in mexico
Keep ourselves, kiwi avocado and mango
What are available? Are they working with rootstock?
Cant get two variaties, and rootstock and not.

Use a penetrometer to test hardness.
See if the penetrometer coerrelates with the hardness.
Get fruit before 1. februar.
Third week of february to get fruit from bama
Datacapture
Maybee buy something from before store before bama deliverment.

Storage

- 4 degrees best
- 12 degrees 5 days kiwi ==> medium

Need to look at different paramaters and images
Need to do less or need to do more

Non disclosure agreement

We need to destroy the fruit at some point
Able to test firmness, ph, and sugar level
For bruise different time periods with fridge storement

Delivering the project plan is not a formal requirement
They had a plan and if they deviated from the plan.
If we change the plan we still need to have a plan.

This week next week, then first of februrary. To make a project plan
To be approved.
What to do, dont worry about people have done it before.

60% are scanned.
Processing speed, accuracy, different set of parameters, new dataset.
Document everything, zotero?
Mai 20 is deadline. Give atleast 1 week
Start report from april

- Making a structure

Leave 1 week for literature review in february.
Second week can then be used for testing.
Last week of february get from bama

- 1 or 2 week to test

Preprocessing march
Testing different methods
Have a list of techniqes
A list of spectral regions
Make roles in the group. Where should each contribute
How often to meet.
Arrangement

Data aquisition 1. 2. week february
Keep some milestones.
Think about the milestones

- Database
- Preprocessing
- Testing

Have a teams groups without tracing emails

Need a form of quality assurance for the camera
Some people are using the camera next week.
Do it by this week
8010
binu

Meeting nr 3, with Hilda: 20.01.2020

If we stick with one fruit it should be avocado
If it is avocado we need something to fall back to. which is kiwi
Hilda will try to be at the camera assistance
Bring kiwi and avocado

We are unsure about the main research problem
Bama wants to check maturity.

Its ok to only check maturity

- Even more important to have 2 fruits.
- Avocado skin is thick
    - Hyperspectral has only micro meters of depth
- Check with sony if he has water meter
    - Ph scale
    - Penetrometer etc.
- Pressure, water something measure, ph level acidity (kiwi)

Customers
Narrow it down to maturity

Project plan

- We want to check maturity with potential/feasability with hyperspectral camera
- Later check the coerrelation with hyperspectral camera with the traditional
- The biggest chuck of time is the data gathering
    - In therms of chucks it is doable
    - A big part of it is thinking.

We need to know how often to check the fruits with the camera.

Next week

- Try to buy (the night before or morning)
- A bit soft and also hard

There is a range for water abosbition. Which messes with the spectral imaging
Try to find water aborbition band

There is a function of moving reflection to abosbition.

Maturity checking
With kiwi and avocado
Plan to check the coerellation

Next week

- Figuring out the camer
- Look at the data
- With different methods

Check with sony of how general and detailed we should be.
How short notice does bama need

- There are workaround
- Dont worry so much. It is fine. Last meeting was a mes
- Hilda is open for technical images.

Meeting nr 4, with Sony 25.01.2021

- Show gant diagrams
- Ask about the different measurement instruments
    - Hyperspectral camera, vnir swir, sugar/salt brix meter,
    - PH meter acidity or basic, Penetrometer firmness
- Will bama provide fruit?
    - Still can do a really good project without bama
- Maturity kiwi
    - Not worrying about 1 fruit or 2 fruits
    - 1 fruit ==> more analysis
    - 2 fruit ==> more spead of time
    - While you do it, its easy to check for everything else
    - It is interessting to check for the rest of the parameters other than penetrometer (maturity)
    - Brix meter is very small.
    - Need to get storage parameters from bama
- Can we get harddrives?
    - Yes we get harddrives.
- NTNU shutdown?
    - Binu will help us this week
    - More people in lab ==> use mask
    - We apply to get an entry under shutdown
    - Can have atleast two of us
    - Scan and get data in initial status
    - Quarintine:
        - One of us goes and does the scanning
        - Or some engineer does the scanning for us.
    - We could be

- Try to learn from hilda
- Get data from kohmann
- Covid-19 is going over
- We need a confirmation with bama
- TTO technology transfer unit
    - Getting contact through them
    - Comeralisation project
    - Sony has talked with them about this
    - We need to sign an agreement
        - We sign
        - Sony sign
        - Head of department sign
        - About: How would we publish?
            - We can get a share.
    - We dont need an NGO

We need to book time for the camera.

- We can do far ahead of time

    - 3 week of february - first week of march
    - Remove the plan if we are not going to use it
    - Can see when it is available

  This week: learn the system

  Next week: wont be available in the lab. Going with the camera. More likely the camera will stay, next time is in mai ish
- How many fruits do we need, kiwi
    - 200+ kiwis
    - Before: 50 (not good enough)
    - We need to say to bama how much kiwis we need
- Literature review
    - Read heavily in the beggining
    - Change research to literature review
- Time tracking is important for the group
    - Plan for time usage each week
    - What roles should each one have
- Kanban is ok?
    - Kanban is ok!
    - We dont know exacltly what to use
    - And we may need more time than expected
    - Project plan should start from 1. feb
    - Think about the planning from february 1.
- Project plan does not need to be in the GANTt
- Seperate the management tools from the technical tools
    - Techinal
        - Computer harddrive camera
        - Libraries for learning algorithms
- Need a backupplan for dataloss and code.
    - Also mentioned in resources.
- Send to sony by Friday can get feedback during weekend.
- NTNU sponsores KIWIS we buy
    - Scan it and send it to SONY
- Keep the meetingnotes as docx

Meeting nr 5, with Sony: 02.02.2021

- Get harddrive by this week or next week
- Camera need the computer
- We can get the images next week or in the weekend
- Will be reasembeld this weekend
- We should book a couple of next week 1-2 days
- Get more familiar
- We should test with different filter
- With polarisers we need to change the integration time
- See the shadow formation
- Talked with tto yesterda
    - 3 week of february
    - Need to know how much kiwi we need. BY TODAY
    - Between 150-200
    - 1 box week 3

    - 2 box week 4
    - See their opinion on delivering
    - How long can we keep it
- Everytime we make a measurement we can weigh them
- We get a thermal camera as well to see the temperature.
    - Way to reduce is to keep the lights appart from the camera.
    - Adjust the distance from the light to the fruit.

- The more parameters the better
- Sort the kiwis so that all the scans have the same size
- Note down so that we know how much time a scan takes for kiwi
- Once set just hit record and get the data.
- There are code available to do PLS
- Remember to learn and read about that happens on the inside.
- Measure firmness from different sides on the kiwi and do the mean
- Everyone in the lab
    - Plan on who to do what
    - One to prepeare
    - Second measure firmness
    - Third measure sugar and ph
    - Assembly line.
    - Someone write readings from the measurements.
    - Kiwi numbers with the measurement
    - Identification
        - In every measurement
        - Define which one is 1 and 2
        - Have a sheet of paper with the numbers.
        - Take a picture phones of the kiwi.
        - Next week sony will be in campus.
        - Do the measurement on Monday, make sure data is correct.
    - Should not wait for long until pre-processing and sorting out
- Next week we can to the measurement with the other instruments as well.
- Plan for the meeting next Monday
    - Next Monday 2 pm
    - Book the camera.
        - Monday 2pm - 5pm
        - wedsneyday 2pm-5pm.
        - GET FAMILIAR WITH THE THINGS
- Use the colorchecker for the measurements.
    - It is recommended
- The image normalisation explain next week
-
-
-
- Det viktigste er å finne parametrene

- Sortere med størrelse
- Måle parametrene
- Koble bildet til kiwiene vi har scanneta

Meeting nr 6, with Binu, Sony and Hilda: 10.02.2021

- Remember to turn of the camera after usage
- Use can of air on the reflectance target if it is dusty
- We used the topp lid of 6 pk egg. Worked perfectly
- Need to be sure that the reflectance target and the kiwis are at the same height
- We increased the frame period so that the scan would og much faster.
- I bought a pippette for putting water on the spectrometer.
- We knew that we really needed distilled water for the spectrometer. However we thought normal water would be ok.
- We decided last time that eivind was going to have responsibility for the camera setup, right height etc..
- Vebjørn nilsen is responsible for the camera parameter changes for perfect image
- Me and katherine is responsible for the physiological measurements.
- We found out we had to cut a "lid" of a kiwi after the puncture meter. This will then be scanned again acording to the same IDS.
- It went 28 minutes, after scanning until all the kiwis was measuerd and made a lid of.
    1. However we did 3 different scans for the sugar measurement
    2. One from top, bottom, then lastly from the middle
    3. The middle was a lot higher than the rest.
- Sony told us that we should do all the scanning before we do any measurements.
    1. Scanning of kiwi
        1. Everyone should have eyes on the scan so that everything is done correctly
    2. Cut the kiwi
    3. Scan the lids
    4. 3 people is hands on the measurements
        1. Penetrometer
            1. Both sides on the middle of the kiwi
        2. Sugar meter
            1. Top, bottom, middle
    5. Last person notes down the numbers
        1. Maybee have a paper to write town stuff as well.

- We got info that the highest reflectance target should be around 70-80 saturation
    - Changing the frame period for the VNIR first then changeing the integration time of the other two afterwards.

- We winded up writing ids on the egg cartong that we used. Also on the spot where the lids are going

- Sony came in
    - Need to make a checklist for everything that we need in the lab
    - We will get a beefy computer to have at the scan site to do the normalisation after each scan day
    - We should do the normalisation of the scans the same day we do them
        - We should also keep a backup copy of the scan before each step of the normalisation
        - We also get a server at ntnu to store data
        - Should make a normalisation script that handels the normalisation automatically

Group meeting 15.02.2021

- Found out we had to cut a 5 mm slice to do the brix
- Penetrometer top bottom, middle
- 40 kiwier
- Gjøre d ordentlig

  Får 3-4 boxer med 50-40 kiwier
- Scan dag 1
- Ta bilde av alle kiwiene
- Measurements på en box.
- Lagre de andre
- Lagrer de andre boksene til dag 2.
- Skanner

- Kjøp 40 kiwier, altså 10 kiwier hver.
- Får kiwi neste mandag
- Jeg kjøper 3 pakker med egg. Katherine 4 stk.
- Kniv og skjærefjøl.

Får harddisker imorra.
Skanner 120 på en dag
Diskuterer om vi skal ha 10 eggkartonger for 40 kiwier eller ikke.
Kom fram til at vi må ha id system på eggkartongene.
Id system i boksene

- Er en linje i midten som har 7 og ikke 8 på lengde siden.
- Det blir da en scan med oddetall og ikke partall.

Do all the scanning first
Then do the measurements afterwards follow plan from sony
Tenker å

Week1	3 boxs x 40	
Day 1	HSI of all 3 boxs	Other measurements of box 1, store boxes 2 and box 3 at 8 degree celcious
Day 2	HSI of box 2 and box 3	Other measurements of box 2, store box 3 at 8 degree celcious
Day 3	HSI of box 3	Other measurements of box 3
Week2	3 x 40	
Day 1	HSI of all 3 boxs	Other measurements of box 1, store boxes 2 and box 3 at 8 degrees
Day 2	HSI of box 2 and box 3	Other measurements of box 2, store box 3 at 8 degrees
Day 3	HSI of box 3	Other measurements of box 3
Week3	3 x 40	
Day 1	HSI of all 3 boxs	Other measurements of box 1, store boxes 2 and box 3 at 8 degree celcious
Day 2	HSI of box 2 and box 3	Other measurements of box 2, store box 3 at 8 degree celcious
Day 3	HSI of box 3	Other measurements of box 3

Meeting nr 6, with Sony at lab: 17.02.2021

- We need to do all 40 before the week is ended
- We need to make sure that the report is innovative
    - That we tried something new
- There will be a identification system in place for the kiwis when they come out of the box
- They are thinking about haveing a sticker laber on each kiwi.
- There is another group getting the kiwis before us
- Keep measureing both of the cores.
- Should we measure the core on both of them
- Be gentle with the kiwi when moving it. Only touch the end parts.

TODO:

- Book the camera for tomorrow.

General information gained during scanning

- Before we began doing propper scanning with the kiwis from BAMA we did a lot of practise with the measurements
    - we where able to push down the scanning+measurement time to 15-20 min
    - Se timing sheet in drive folder
    - We got access to binus normalisation code to find acceptable integration time
    - We found out with sony (and binus norm code) that we could push the integration time a lot down making the scan a lot faster

        - this resulted in the kiwi not beeing in the 70-80% reflectance but it worked well
- We took pictures of our measurement procedure and put together a pdf and sendt to BAMA for verification
    - Bama said not to stab the kiwi as deep as we did. only as long as the mark on the penetrometer.
    - This was discovered after week1
    - The procedure pdf is uploaded to the drive.
- there was put a identification system in place
    - there was used double sided tape and paper
    - wrote ids from 1-55 inside each box
- we followed the scanning plan shown above
- Before a box was scanned another group also did scanning on the kiwis with nir
    - this was coordinated via messenger
- During week1 we discovered from bama that we didnt need to penetrometer that deep
    - as well as the brix meter converged after a few atempts
    - Because of this reason brix has a lot bigger variation in week 1
    - We also started measuring temperature of the kiwi given by the ph meter
        - also some temperatures from a cooking temp meter however this was dropped because of its inaccuracy
    - We started measuring circumference of the kiwis
    - Jon started writing on the introduction part of the report
    - We bought in equipment to do the measurements more properly
        - a fruit knife, wasing soap, cutting board, more egg containers
    - the lights was positioned weirdly one of the days, to they where setup properly again to take proper images.
    - We got a gopro to document the project
- during week1 Jon and Vebjørn coded the normalisation code together.
    - with help from Eivind Kohmann bugs was fixed and the code was working
- during week2 the normalisation code had experimental modifications
    - all the images was normalised again with median filter and hanning filter
    - after some time it was discovered with binu that only hanning filter was needed
- during week3 all the images was normalised again with only hanning filter.
    - Jon started coding a program to find roi of kiwis, using the values from a X line
        - this was a failiure in the end.
- after the scanning was finished
    - Eivind and Vebjørn joined in on doing a more masking approach to find roi
        - this apporach was successfull in the end
    - Katherine named the images taken with iphone under different light conditions.

Lynkurs rapportskriving: 4.03.2021

- Se pdf lynkurs_rapport.pdf fra undervisningsmaterial på blackboard (bacheloroppgave)
- Forside får man fra ntnu grafisk senter
- Etter man har skrevet ferdig bachelor må man fylle ut mal for sammendrag som må leveres inn
    - Må også skrives på engelsk
- Forord
    - Halvt a4 ark
    - Takk til veileder, oppdragsgiver osv
- Innholdsfortegnelse
    - Ha anker
- Innledning
    - 6-8 sider.
    - Her burde numereringen av sidetall begynne
    - Klipper kap 2 fra prosjektplanen blir kapittel 1.1 her
    - Problemområde + avgrensning (1-2 a4 ark)
        - Avgrensning er å snevre inn det store omeråde til det spesifikke oppgava handler om
        - Oppgavedefinisjon er spesifikt hva oppgava handler om.
    - Målgruppa
        - Hvem oppgava er laged for?
    - Egen bakgrunn og kompetanse
        - Hva hver av de i gruppa kan
        - Ikke for spesifikt av fag og alt
        - Men hva spesifikt som er kunnskap relevant for oppgava
    - Skal presenteres for medstudentene dine
        - Målgruppen har peiling på data.
        - Skal være skrevet for datakyndige folk
        - Nye begreper innenfor det spesifikke området må beskrives.
    - Terminologi
        - Finnes det et godt norsk ord for det det er snakk om så bruk det
- Hovedkappitler
    - Kravspek er hva som skal lages.
- Avsluttning
    - 4-10 sider (4 er litt lite) aka 5-10
    - Effektmål
        - Hva er det oppdragsgiver tjener på dette
            - Tid penger ressurser
        - Reultatmålet
            - Hva som er levert
        - Lærningsmål
            - Hva har vi lært av prosjektet

- Tegn opp gant diagrammet og evaluerer ifht til det.

- Konklusjon
    - Halvt a4 ark
- Litteraturliste, halvt a4 ark, sist besøkt
    - Nummersystemet ikke harvard
    - Url som fotnote
    - Relevant støtte/bakgrunnsoft
    - Hvem, hva hvor når?
        - Hva = hva det er som er hentet herfra kappittel?
- Appendix
    - Definisjoner

Generelle tips:

- Få folk til å lese over
- Nummerer figurer etter kappittel 3.1, 3.2 osv
- Ta med mail på aksept av bruk av spesifikke figurer
- Hele rapporten som 1 pdf. (1-2 filer skal opp på inspera)
- Kildekode som en zip fil.
    - IKKE BRUK GITHUB LINK
- Filene må navngis med etternavn - oppgavens kortnavn
    - Finn på noe annet for æøå

Meeting nr 7, with Hilda 17.03.2021

- We asked hilda about methods of finding roi
- we went through different methods in envi
- She came up with a good approach in the end
- otsu would be weird with the specular reflection
- Doing a computational distance of all the spectra compared to the kiwis. (use big area on kiwi to compare with)
    - then do a binary thresholding with otsu
    - opening/closing
    - then do a half-tranform to identify circles to find the middle of the kiwi
- Using spectral angle mapper gave a lot of shades
    - worked partially from band 100 and out
- Using Spectral divergence gave good results
    - then do some filters and square detection.
- spectrum could be stored as .npy and pandas frame as pickle.

Meeting nr 8, with group: 22.03.2021

Møtesammendrag					
Totalt antall	4				
Møtetittel	General				
Start tidspun	22.3.2021, 15:10:01				
Sluttidspunk	22.3.2021, 16:14:32				
Fullt navn	Tidspunkt fo	Tidspunkt fo	Varighet	E-post	Rolle
Jon Elias Mo	22.3.2021, 15	22.3.2021, 16	1 timer 4 mir	jonemo@ntı	Arrangør
Eivind Kohm	22.3.2021, 15	22.3.2021, 15	4 min 49 sek	eivinkoh@nt	Presentatør
Eivind Kohm	22.3.2021, 15	22.3.2021, 16	58 min 19 sel	eivinkoh@nt	Presentatør
Katherine Br	22.3.2021, 15	22.3.2021, 16	1 timer 1 mir	katheris@nt	Presentatør
Vebjørn Joh	22.3.2021, 15	22.3.2021, 16	59 min 51 sel	vebjorjn@nt	Presentatør

- Kohmann har klart å finne roi
- Bruker erosion og open etc
- Bruker 7x7 område på midten av kiwi (swir) 28x28 på vnir
- Lagret alle spectra
- Kan hente inn kiwi spectra ettersom man trenger det
- Får noen gang reflectance hvor dag 1 er høyere enn dag 2 i snitt
    - Kan derimot ikke fjerne disse boksene.
- Katherine blir ferdig med bachelorboka i dag.
    - Hun legger ut notatene til boka på gdrive
    - Vanskelig å få A
- Bli ferdig 2 uker før fristen med oppgava.
    - 44 dager med oppgava
- Ukemål
    - Koble sammen roi spectra til pandas dataframe.
    - Gjør metoder på dataframe
    - Skrive på oppgava, hver dag.
    - Lese
    - Rydde opp i filer som er feil og mangler

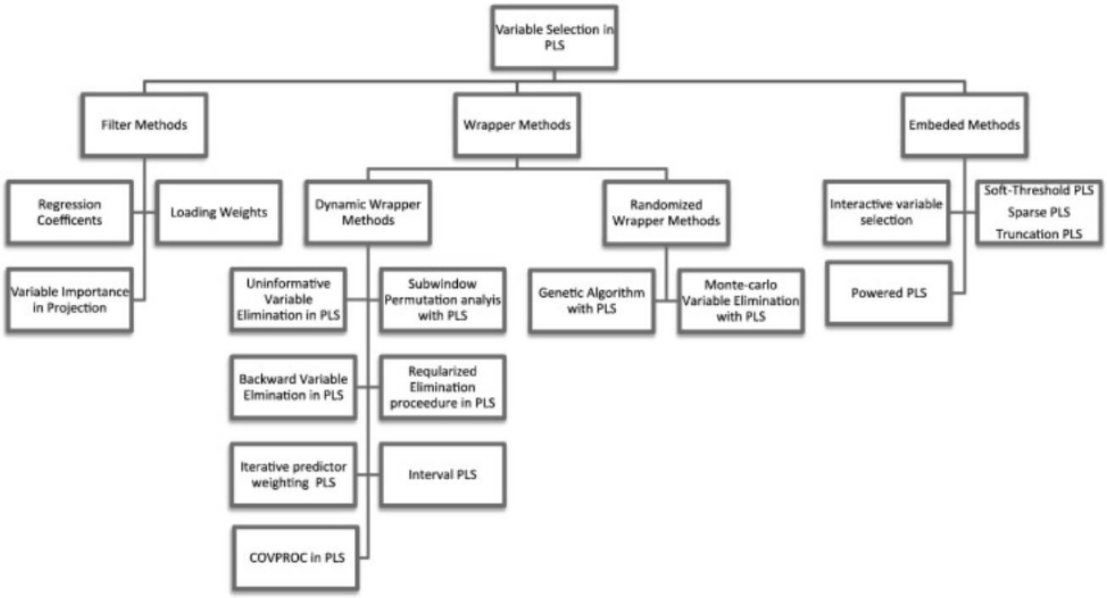
- Ha med en undersøkelse som viser hva folk vet om faget allerede.
- Man kan lage intervju med folk som er i IT, hvor mye de kan om prosjektet.
- Vi er litt due med prosjektskrivinga
- Alle må holde orden på metodene som er brukt og hva som fungerte og ikke fungerte.
    - Alle skrive ned metodene de gjør
    - Personlig dagbok ish.
- Sette numpy array inni en pandas dataframe
    - Få da dag1, dag2 dag3 som columns
- All spectra er 1.6mb så går ann å bare bruke pandas dataframe og spectra filene
- Ikke fjerne outliners på data
- Lage en funksjon som henter rett fysisk og rett spectra
- Mye lettere å søke på dag, uke batch enn unik kiwi id

- Lage en funksjon som søker med day week med spectra.

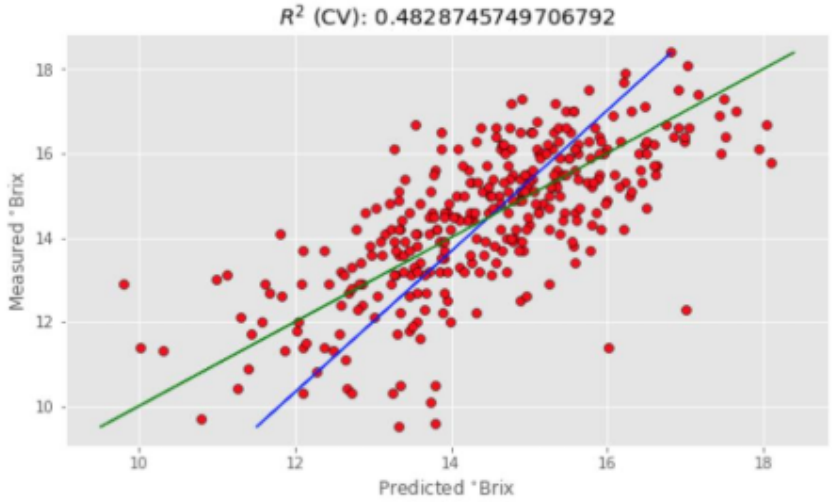
Meeting nr 9, with group: 07.04.2021

- katherine tar et kurs i akademisk kurs i skriving
    - fikse med amerikansk/britisk til slutt ⇒ katherine
- kohmann har vært oppslukt i metodene
    - fått 0.3 r^2
- vebjørn har lest om svr
    - men håper på å få fikse mer på dataen før vi gjør noe mer
- kohmann lagde nytt dataset
    - fikk bedre resultat med andre punkter.
    - flyttet sentrum litt ned fra det lyseste punktet på kiwien
    - han deriverte også spektrumet, og dobbelt deriverte.
        - kan prøve å normalisere.
- Jon Elias tenker å se på FFT
- De harde kiwiene forvirrer modellene
    - så det å fjerne de gir bedre resultater mest sannsynlig
    - men de er ikke representative for kiwier
    - derimot så var de faktisk så harde
    - Er noen som fjerner vanskelige verdier
    - skal vi ha en modell som er ok på alle eller bra på de i snittet.
    - Burde finne firmness som er ansett normalt (6-14 ish)
- Spørr sony om hva vi skal gjøre med rare verdier
- kohmann prøvde å satte sammen vnir og swir spectra
    - trekte swir ned slik at de traff hverandre
    - trekker fra en konstant på swir.
- jon skal prøve mer på å forskyve punktet for å passe bedre på vnir
    - swir 9 vnir 16
- vnir er best med vnir+swir er bedre.
- katherine skal skrive på metode (oppbevaring av kiwi)
    - strukturere noen av vedlegg.
    - ta med møtereferatene
- Spør sony
    - hvordan håndterer vi dataene
    - si til han hvordan vi går.
    - dele med han det vi har skrevet
    - se allerede igjennom det
- peels
    - se om det var noen sammenheng
    - om vi kunne bruke interne vs externe parameter i kameraet
    - backup hvis de eksterne parametrene ikke fungerer.
    - Validation
    - SID?
- Hilda delte SID koden med oss
    - kan brukes på peels. hvis vi ikke får til thresholding

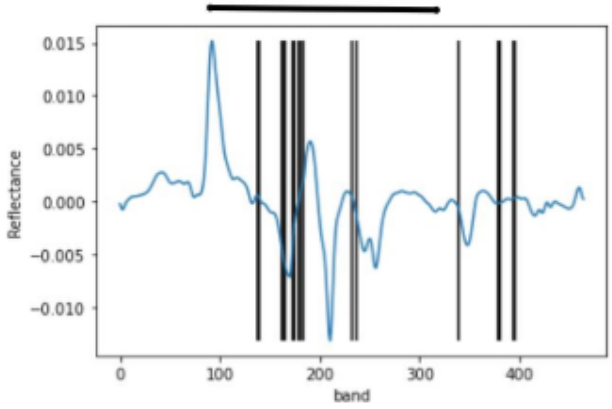
- kohmann har good stuff med variable selection
- Vi kan bruke både harvard og I.E.E på citations.
- https://nirpyresearch.com/variable-selection-method-pls-python/
-



```
Optimised MSEP  1.5365006840228501
R2 calib: 0.669
R2 CV: 0.483
MSE calib: 0.978
MSE CV: 1.527
```




```
R2 prediction:  0.3104870384438695
```



Meeting nr 10, with Sony: 09.03.21

- selecting points with specular reflection
    - seperate vnir and swir
    - it is challenging
- we are going to get some algorithms from sony
- he asked if he could have some data.
- make a full overview of what is happening and send to sony.
- send all the details to sony. then make a plan for the next 4 weeks
- because we need to make some choice
    - combine or not combine
- friday 1 pm, dont wait for these meetings.
- contributions
    - DATA IS VALUABLE
    - need some more output.

Meeting nr 11, with Sony 16.04.21

- do not combine vnir and swir
- non flat surfaces does not provide good results
    - is it possible
- Before monday
    - make 1-2 pages of how we take spectra.
- strawberry
    - average larger area
- Day 1 is lowest
- We should document everything.
  answers:
    - ANN: more info
    - How do we collect spectra?
        - send images of some kiwis
    - report everything!

Meeting nr 12, with Sony 21.04.2021

- we went through the report with sony:

Introduction is good

Urkund – to upload there and check for plagiat

Abstract- acknowledgment

Problem statement, motivation - Not too late

TTO can come little bit late

Problem field can come in the beginning

At the end of 1 chapter or 2nd is to say what we contribution

1st chapter

Name the section scientific contribution:

-huge dataset, we prepared couple of weeks, main contribution is data, you explore as much as possible, you can say that data is validated, not opened, this will be done in the near future

-algorithm to semi automatically algorithm to extract spectrum

-created a lot of data that will be useful with mobile, doing all this thing simultaneously makes meaning,

-not dataset available on this fruit

-3-4 contributions

-say that what we captured data for, but not parameters are relevant

After reading they should get an idea

Target group

Group background

2nd chapter

Theory:

-hyperspectral imaging main heading

Defining spectroscopy

Spectral imaging

Hyperspectral imaging – what is VNIR and SWIR

Picture of kiwi- cube image

-chemometrics

Chapter: previous research – state of the art

More than a page- cite a key paper, most related to our topic

Chapters- material

-materials,

-image acquisition with the phone

-hyperspectral acquisition

-destructive measurements, show picture

Add workflow with fail method

Add preprocessing

Add libraries,

Every method that you introduce should be introduced firstly

You should explain how it works, not super deep

Add details to photos, did you use polarizer, how many pixels are there

Add camera specification, range of photos materials

Details about distance to the object,

Question that may come -> focus depth

When we present something make sure that we have good quality rgb images

Discussion: add there things that didn't work, fails

Result and discussion can be together

We should assume that an examinator doesn't know anything even what the regression is

Add picture of spectrum variation

Plot sugar values

Firmness- 0.8-0.9 PLS

Sugar level - 0.5-0.6 PLS

Data firmness – remove end of the data, because there is less data

Try k-nearest regression

Find parameters, keep model outside

Prediction graph

Check if your model is good – not overfitting

Final chapter:

Say about all the methods

Conclusion and future work:

-we have data for many parameters, but we have decided to work with this later

-

Method: try to look for the difference in weeks, train only on week 1,2,3

Meeting nr 13, with group: 30.04.2021

- Katherine venter på mange artikler.
- Går ann å ta copy paste inn i et nytt prosjekt for ordentlig review
- viktig å ha med at teorien er viktig for vårt prosjekt.
- omstrukturering
    - eget kapittel previous research.
    - endre navn på spectral imaging.
- burde skrive mer om vnir og swir.
- har allerde en del om det i background og i slutten av theory
- lagde et nytt kappittel state of art.
- burde skrive enn plass i teori om hyperspectral kube.
- Spørr sony om han kan skjekke det programmet for plagiat.
- feilkilder
    - delays før ting ble satt i fryseren igjen
    - programmet hengde seg ofte under acquisition og rad stuff
- ulike silikon filtere.
- på equalization filter forklar bilde og forskjellen mellom sensorene.
- stor forskjellen mellom maturity og ripeness så det burde være forsikret mellom alle

- stor usikkerhet mellom om vi må spørre om bruk av bilder eller ikke. katherine spørr sony.
- se forskjellen mellom spectrum på samme box bare flere ganger ila uka
- Sony hvor innafor er det å bruke wikipedia som kilde
- Se over de punktene i den planen, marker selv det som er ferdig

Meeting nr 14, with Sony: 6.05.2021

- Problem describtion first chapter
- The UN part could be a subsection inside the contribuion
- One is scientific the other one is associated contribution
- Problem field could be in the background chapter
- Has to come before state of the art
- Background is diffferent from introduction
- Introduction whole basic idea
- Background is more into detail
- Introduction
- Ready early next week could give it to someone else to check it.
- The survey could be under motivation
- Bring the audience in, in the introductin
- Contribution is the answer to the problem
- Contribution
    - Innovation
    - Un
- A lot of recent detail in the area
- Next week is important to finalize
- Introduction have a good impression keep it until the end
- The next important part is theroy and
- All the people we
- Background: background needed for reading the report
- Background can be missleading. Project and group background
- Spectral imaging
    - Multiple bands
    - Beyond visible
- State of art
    - Name and year on citations
    - Also mention groups
    - Not leave reader to have to og to paper.
- Rapid monitoring determination'
- Check guy sony yee https://www.ucd.ie/sun/
- Contribution have a table
    - What is missing, what is our contribution
    - Use colors
    - What is done what is the future
    - For eksempel.

    - What is done, what is achieved what is the future work. Table
- Subsection conclusion about future work.
- Iphone can be used to do the determination.
- Use mobile phone instead of phone when not in materials.
- Use portable devices
    - These exists already.
- A lot of potential
- More portable device (bama)
- Introduction
    - What products portable/nonportable exsists for fruit/kiwi

Meeting nr 15, with Sony: 11.05.2021

- Datacube
    - 3d cube is oversized
    - Mark arrows for axis
    - Write that in caption
    - No wrap figure, no half page
- Acronym
    - Change to abreviation
    - It is not acronyms
    - Use fullform the first time
- Theory materials methods
    - Can be fussy
- Introduction
    - 2 sentence review of HSI
    - Then its applied to food and agricultural products.
    - Hsi has several advantages over RGB imaging
    - Look at any paper on hsi (introduction)
- contribution
    - Papers are limited to a specific range
    - Non of the studies covered the swir+vnir range
        - And all the different quality parameters
        - + peels+rgb
        - New methods (KNN,)
        - Complete dataset
        - Measured a lot more than other studies.
        - Compare the models?
- Has been studied for many other fruits, but for kiwi it has not been researched throughrly (extensively)
- Have many things but it is scattered
    - Beginning with hsi in intro
- Merge materials and method
    - Mention details when writing in the methods.

        - Not in large detail on how the equipment work'
        - Tell in the beggining (kiwi - fruit sampels)
        - Then data aquisition
            - Hyperspectral imaing (camera stuff, ref target)
            - Instruments for the destructive measurements

Methods and materials

1. Kiwi
2. Data aquisition
    1. HSI
    2. Destructive measurements
    3. RGB (in between)

Sony will take a look at the flow.
Ask hilda about author year in latex
Send report to hilda

Move the why we had this to under figure from vebjørn of the setup.
Why we have this setup
Title:

- Only firmness - use firmenss
- More parameters - quality parameters.
- Salt

We need to explain that we tried the different stuff. And what we prioritised
Subsection should drive the reader. The index should be understandable. It should be explainable from the index. "measurements, results"

The results is the actual model results then discussions.
Vi skriver litt om våre meninger i methods. Så ta med diskushon i methods

Contribution is not only results and discussion
Contributed also towards processing

Table (contribition)

- Dont include the accurcacy but include the citation
- Attempted instead of been done

Learning objective in the introduction
Outcome in conclusion or end of discussion

Conclusion table.

Portabledevice	What has contributed towards that goal
Ph	Etc…

With all the contributions in introduction. Written in text

Conclusion

1. Paragraph: All conclusions from discussion
    - This is more usefull
2. Table contribution
3. Future works
4. Learning outcome

Make sure there are no dupliate section titles

Pre-processing means a lot
Results and discussion
Firmness
Results for firmness

Meet hilda tomorrow

Spend least energy on state of the art
There are other sections more important

Have a time distrobution table.
In the presentation we should have a table that thows the time.
Depends on time but could also be in the report (introduction) or background with team experience stuff

Make a table of contents for tomorrow with hilda
Give another task to Katherine
Prioritize results and discussion

Contribution table for the models (any maybee spectral range)

Contribution

- Hsi kiwi
- Data subsection
- Method subsection

First: table of contents fix
Results and discussion
Talk with hilda tomorrow.
Share report early.

Meeting nr 16, with Hilda: 12.05.2021

- Redundant contents
    - Restructure
        - method

§ Intro, back, theory, method?
§ Method -> method + (data aq and measurements own chapter)
§ Norm and radiometric in data aq + measurements

- Flowchart for processing part
- Diagram/ flowchart of SVR, PLS, MLR where do they go, preprocessing etc. Show how they are used (5.6)
- Diagram of datasplitting
- Separate discussion for data aq and data processing
- Merge multi and hyperspectral
- Veb: flytte kiwi theory lenger opp? Bør diskuteres med gruppa
- 1.7: Thesis structure
- Conclusions
    - Summary of contribution
        - § Measurements
        - § Processing
    - Future works

Meeting nr 17, with Sony: 17.05.2021

- Then come to buisness
- Project priority
    - Confusing title
    - Not clear what we want to say
    - Move to the beggin
- Results and discussion
    - We focused on these two parameters
    - We focused on sugar and firmness.
- Firmness is a good as a main section
    - Pre-processing?
    - Under firmness could have firmness measurement from vnir
    - Then for swir
    - Vnir vs swir
    - Full vs region
- The same with 7.4 sugar
    - Pre-processing?
- Low priority models and features
    - Utbytt med salt of ph
- Time constraints in the introduction
    - Project limitations, and beginning of results and discussion.
    - Less priority for salt and ph
- Firmness
    - More discussion
    - Show data in different ways
        - Vnir vs swir vs full vs region bar plot?

        - Ews vs predictive power?
    - Talk about in more details about more of the models
        - Has been used before`?
        - Cite atleast 1 paper.
        - Imbalance between the citations
            - Un sustainability goals has to many citations, needs only one.
            - Models has not a single citation, broaden the scope if not found
    - Results and discussion
        - Check paper sony uploaded.
    - Not all about achieving something new, but also learning
        - This is a bachelor thesis.
        - We have achieved good r-squared.
    - Now it is in a good shape
    - Can have graph on spectral range and publications
        - However explain there is less research here than other fruits.
-
    - Biblographi, only need the doi and not the url
    - Focus resultsdiscussion and conclusion
    - Does not need "table shows" in caption
    - Easy for b and c, to get to A
        - ADRESS ALL THE DETAIL
          overall impression
    - Contribution tabels
        - Change attempted to attempted previously.
        - Caption on the tabels.
        - "contribution in this bachelor thesis"
    - Repeating the abbreviations many times.
        - Last minute things
    - Presentation 9. june
        - Change of karakter comes at after the precentations
        - Can be a game changer.
        - Cant represent everything in the report
            - Have the video in the presentation
    - Introduction
        - This bachelor thesis has a lot of research content.
        - Do not compare!
        - ¨very innovative
        - Unlikely other topics had less research content.
        - The other supervisors also want to feel they have done something very nice.
        - Choosen this topic because more research content, learning complex labratory instruments and so on
    - Rewrite bama meeting contribution
    - List contribution
        - Have a subsection for the contribution list.
        - Titles for each point

- Abstract is something we can improve
    - Missing what we have actually done.
    - Needs to be written after the conclusion
- Dataset is one word.
- pH or acidity decide.
- Abbreviation is missing many things
    - ANN, KNN, MLR, SNV, UVE
- Keep blob diagrams in the same form
- Chapter5 - text before the figure.
- Table in state of art
    - Ssc firmness tss should rotate 90 degress
    - Kiwi variety use short name
    - Spectral range move nm to title
    - "modified" pls should be just pls
- The group contribution needs to be changed.
    - Make a table?
    - Leading contributer, but everyone also contributed
- Theory
    - Why we used hanning window filter.
- Check SPXY
- Finished at 5pm
- Meeting 5:30 wedsnday. Og through everything
    - Submit early

Appendix B

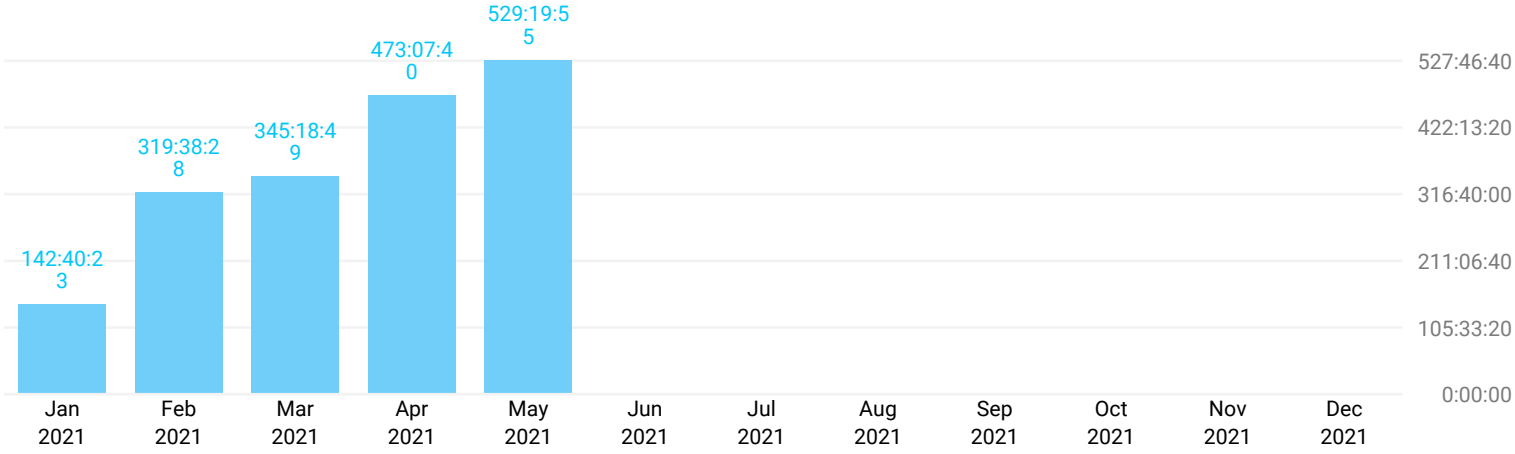
Time usage

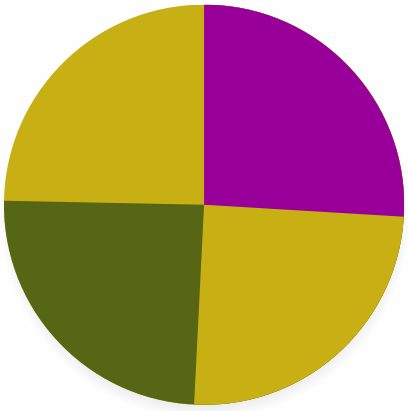
The pdf attached below shows how much time was spent completing this bachelor thesis. The software application toggl was used to keep track of this time usage.

Summary Report

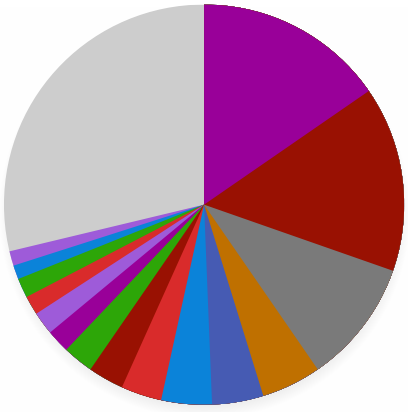
01/01/2021 – 12/31/2021

TOTAL HOURS: 1810:05:15





USER	DURATION
EK Eivind Kohmann	471:05:02
VJ Vjnilsen	447:54:00
JO Jonemo	445:36:29
KA Katheris	445:29:44



TIME ENTRY	DURATION
writing	278:25:15
bachelor	270:18:09
Without description	182:57:08
Lab work	86:31:10
Scanning	75:23:00
Writing	75:07:19
coding	58:32:10
SPA	51:00:34
Kiwi scan	45:58:57
meeting	33:29:12
python work	32:41:50
Python Work	29:31:28
chat and discussion	28:01:00
Meeting	21:36:17
Writing and meeting	19:32:56
Other time entries	520:58:50

USER - TIME ENTRY	DURATION
EK Eivind Kohmann	471:05:02
At the lab	2:00:15
coding	1:00:00
Coding and writing	3:09:54
Coding/writing	2:53:09
Combined spectra with measurements	1:41:43
Creating kiwi dataset	2:40:49
Creating spectra	4:21:06
Creating spreadsheet	0:31:32
EMSC	9:47:58
Group Meeting	2:38:29
Kiwi dataset	0:32:41
Kiwi discussion	1:00:02
Kiwi mask	7:30:40
Kiwi maske	1:00:17
Kiwi scan	45:58:57
Kiwi scan testing	7:00:01

USER - TIME ENTRY	DURATION
Kiwi scanning Day 1	8:00:01
Kiwi spectra	5:33:33
Looking into the draft of the project plan	1:55:23
Meeting	0:20:01
Meeting and preparations	3:03:21
Meeting and reading	2:20:42
Meeting and writing	13:01:52
Meeting with Bama and Sony	2:30:47
Meeting with hilda and coding	2:00:05
Meeting with sony and project plan	1:00:23
model prediction	14:18:34
Normalisation	1:55:14
PLS-weighted coefficients	7:20:18
Preperations for week 5,6,7	0:30:03
Programming	2:21:19
programming	2:01:24
Project meeting regarding Gantt	1:40:16

USER - TIME ENTRY	DURATION
Project plan	9:18:12
Project plan, reading about avocado and looking at GANTT	1:34:54
reading	2:08:46
Reading about normalisation	0:16:17
Reading about SPA	0:46:18
Reading papers	5:16:20
REading papers	0:04:02
Reading papers and analysing sample size	0:30:02
Review project plan	1:10:02
Scanning kiwi	9:00:38
Scanning procedure	0:30:01
Setting up scanning procedure	2:28:20
Skriving	1:33:45
SPA	51:00:34
SPA and spectra stitching	3:56:03
Spa and writing	5:29:44
SPA implementasjon	4:06:22

USER - TIME ENTRY	DURATION
Spectra + measurements	0:24:47
Spectral analysis	5:42:01
spectral stitching	0:32:37
Test scanning	4:30:02
wiritng	0:06:12
witing	2:34:33
wriitng	11:35:12
Wriitng	3:10:07
Writing	75:07:19
writing	54:17:06
Writing	5:17:55
writing and coding	14:43:49
Writing and meeting	11:05:41
Without description	19:06:32
JO Jonemo	445:36:29
Academic reading	7:15:45
Camera teaching	2:00:00

USER - TIME ENTRY	DURATION
Colorchecker	2:52:23
Data fi	0:23:31
datafix	3:31:58
Gant diagram	1:47:00
Group meeting	1:09:27
kiwi collection	3:39:42
Kiwi reading	6:41:15
Lab work	86:31:10
Lynkurs	1:09:07
Management	1:58:19
Meeting	17:01:44
meeting	4:26:53
Meeting and reading	2:56:17
Meeting and work	4:23:19
Meeting minute summary	0:45:00
Meeting sony	2:04:02
Meeting with bama and Sony	1:56:17

USER - TIME ENTRY	DURATION
Meeting with hilda and group	1:26:49
Meeting with Sony	1:32:35
oblig 2	0:47:26
oblig 3	4:22:30
Pomodoro Break	0:19:23
Project plan review	1:11:32
Projectplan	12:02:31
python work	32:41:50
Python Work	29:31:28
python work & writing	0:33:47
rapport	5:04:44
Rapportskriving	1:07:08
Reading	3:30:31
Reading and python	1:18:00
Reading kiwi	5:47:49
Report	2:45:15
report	0:25:46

USER - TIME ENTRY	DURATION
report writing	0:34:13
review	0:36:10
Seminar	3:59:00
software setup	0:33:10
Software setup and meeting	1:11:22
writing	135:08:09
Writing and meeting	8:27:15
writing at school	7:50:16
Without description	30:14:41
KA Katheris	445:29:44
bachelor	270:18:09
Kick off meeting	1:29:42
lab	3:45:00
Meeting	2:25:00
Meeting with Hilda	3:19:40
meeting with Sony	1:00:00
meeting with Sony, planning	1:25:00

USER - TIME ENTRY	DURATION
project planning	4:23:00
reading	2:23:41
Reading a Bachelor guide	1:05:00
Research	9:33:30
Research + writing	3:07:09
scanning	4:55:00
writing	2:43:58
Without description	133:35:55
VJ Vjnilsen	447:54:00
chat and discussion	28:01:00
coding	57:32:10
coding + writing	7:44:47
datastructure	1:57:00
experimenting with kiwidata	1:22:49
Group meeting + discussion	2:07:07
Kiwi and Avocado research + discussion	4:18:00
kiwi and normalisation research	3:59:01

USER - TIME ENTRY	DURATION
kiwi discussion	3:42:38
Kiwi dyrking/modning research	2:35:54
kiwi experiment	1:53:00
Kiwi experiment + kiwi research	5:26:47
kiwi research	4:11:35
Kiwi research	1:50:00
kiwi research and scan plan	4:42:34
kiwi weight and salt research	1:56:00
Make plan + discussion	2:26:06
meeting	29:02:19
Meeting	4:14:32
meeting + chat/discussion	10:08:03
Meeting + reading about kiwi	4:08:17
meeting and writing	13:06:16
meeting at school	4:06:00
Meeting company	1:48:00
Meeting with Bama	2:05:29

USER - TIME ENTRY	DURATION
Meeting with supervisor	1:00:00
Normalization + discussion	3:51:12
Normalization coding	6:25:15
Orientering kode og scan	4:33:52
Project plan	2:26:00
Project plan + image normalization + discussion	8:20:22
projectplan writing	1:52:29
Prosjektplan + discussion	2:40:06
Prosjektplanlegging	0:50:41
reading	9:23:01
reading normalisation correction	1:39:40
Scan plan	0:48:58
Scanning	75:23:00
scanning	3:30:00
Scanning kiwis	3:46:00
writing	86:16:02
writing + coding	9:17:45

USER - TIME ENTRY	DURATION
writing + meeting	7:27:44
writing normalisation theory	2:01:00
writing report	11:55:29

Appendix C

Project deal

Our project deal, signed by all the members.

Norges teknisk-naturvitenskapelige universitet

Prosjektavtale

mellom NTNU Fakultet for informasjonsteknologi og elektroteknikk (IE) på Gjøvik (utdanningsinstitusjon), og _Hilda Deborah_ (oppdragsgiver),

og Jon Elias Moen, Vebjørn Nilsen, Eivind Kohmann, Katherine Brox-Saidi (student(er))

Avtalen angir avtalepartenes plikter vedrørende gjennomføring av prosjektet og rettigheter til anvendelse av de resultater som prosjektet frembringer:

1. Studenten(e) skal gjennomføre prosjektet i perioden fra 1.1.21 til 20.05.21.

Studentene skal i denne perioden følge en oppsatt fremdriftsplan der NTNU IE på Gjøvik yter veiledning. Oppdragsgiver yter avtalt prosjektbistand til fastsatte tider. Oppdragsgiver stiller til rådighet kunnskap og materiale som er nødvendig for å få gjennomført prosjektet. Det forutsettes at de gitte problemstillinger det arbeides med er aktuelle og på et nivå tilpasset studentenes faglige kunnskaper. Oppdragsgiver plikter på forespørsel fra NTNU å gi en vurdering av prosjektet vederlagsfritt.

2. Kostnadene ved gjennomføringen av prosjektet dekkes på følgende måte:
   • Oppdragsgiver dekker selv gjennomføring av prosjektet når det gjelder f.eks. materiell, telefon, reiser
   og nødvendig overnatting på steder langt fra NTNU i Gjøvik.
   Studentene dekker utgifter for
   ferdigstillelse av prosjektmateriell.
   • Eiendomsretten til eventuell prototyp tilfaller den som har betalt komponenter og materiell mv. som
   er brukt til prototypen. Dersom det er nødvendig med større og/eller spesielle investeringer for å få gjennomført prosjektet, må det gjøres en egen avtale mellom partene om eventuell kostnadsfordeling og eiendomsrett.
3. NTNU IE på Gjøvik står ikke som garantist for at det oppdragsgiver har bestilt fungerer etter hensikten, ei heller at prosjektet blir fullført. Prosjektet må anses som en eksamensrelatert oppgave som blir bedømt av intern og ekstern sensor. Likevel er det en forpliktelse for

utøverne av prosjektet å fullføre dette til avtalte spesifikasjoner, funksjonsnivå og tider.

4. Alle bestått bacheloroppgaver som ikke er klausulert og hvor forfatteren(e) har gitt sitt samtykke til publisering, kan gjøres tilgjengelig via NTNUs institusjonelle arkiv NTNU Open.

Tilgjengeliggjøring i det åpne arkivet forutsetter avtale om delvis overdragelse av opphavsrett, se «avtale om publisering» (jfr Lov om opphavsrett). Oppdragsgiver og veileder godtar slik offentliggjøring når de signerer denne prosjektavtalen, og må evt. gi skriftlig melding til studenter og instituttleder/fagenhetsleder om de i løpet av prosjektet endrer syn på slik offentliggjøring.

Den totale besvarelsen med tegninger, modeller og apparatur så vel som programlisting, kildekode mv. som inngår som del av eller vedlegg til besvarelsen, kan vederlagsfritt benyttes til undervisnings- og forskningsformål. Besvarelsen, eller vedlegg til den, må ikke nyttes av NTNU til andre formål, og ikke overlates til utenforstående uten etter avtale med de øvrige parter i denne avtalen. Dette gjelder også firmaer hvor ansatte ved NTNU og/eller studenter har interesser.

5. Besvarelsens spesifikasjoner og resultat kan anvendes i oppdragsgivers egen virksomhet. Gjør studenten(e) i sin besvarelse, eller under arbeidet med den, en patentbar oppfinnelse, gjelder i forholdet mellom oppdragsgiver og student(er) bestemmelsene i Lov om retten til oppfinnelser av 17. april 1970, §§ 4-10.
6. Ut over den offentliggjøring som er nevnt i punkt 4 har studenten(e) ikke rett til å publisere sin besvarelse, det være seg helt eller delvis eller som del i annet arbeide, uten samtykke fra oppdragsgiver. Tilsvarende samtykke må foreligge i forholdet mellom student(er) og faglærer/veileder for det materialet som faglærer/veileder stiller til disposisjon.
7. Studenten(e) leverer oppgavebesvarelsen med vedlegg (pdf) i NTNUs elektroniske eksamenssystem. I tillegg leveres ett eksemplar til oppdragsgiver.
8. Denne avtalen utferdiges med ett eksemplar til hver av partene. På vegne av NTNU, IE er det instituttleder/faggruppeleder som godkjenner avtalen.

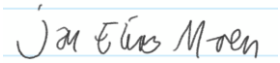
9. I det enkelte tilfelle kan det inngås egen avtale mellom oppdragsgiver, student(er) og NTNU som regulerer nærmere forhold vedrørende bl.a. eiendomsrett, videre bruk, konfidensialitet, kostnadsdekning og økonomisk utnyttelse av resultatene. Dersom oppdragsgiver og student(er) ønsker en videre eller ny avtale med oppdragsgiver, skjer dette uten NTNU som partner.
10. Når NTNU også opptrer som oppdragsgiver, trer NTNU inn i kontrakten både som utdanningsinstitusjon og som oppdragsgiver.
11. Eventuell uenighet vedrørende forståelse av denne avtale løses ved forhandlinger avtalepartene imellom. Dersom det ikke oppnås enighet, er partene enige om at tvisten løses av voldgift, etter bestemmelsene i tvistemålsloven av 13.8.1915 nr. 6, kapittel 32.

12. Deltakende personer ved prosjektgjennomføringen:

NTNUs veileder (navn): Sony George

Oppdragsgivers kontaktperson (navn): Hilda Deborah

Student(er) (signatur):

 dato 14.01.2021 15:58

 dato 14.01.2021 16:01

 dato 14.01.2021 16:07

 dato 14.01.2021 16:08

Oppdragsgiver (signatur)

 dato 20.01.2021 09:27

Signert avtale leveres digitalt i Blackboard, rom for bacheloroppgaven. Godkjennes digitalt av instituttleder/faggruppeleder.

Om papirversjon med signatur er ønskelig, må papirversjon leveres til instituttet i tillegg.

Plass for evt sign:

Instituttleder/faggruppeleder (signatur):
_____ dato _____

Appendix D

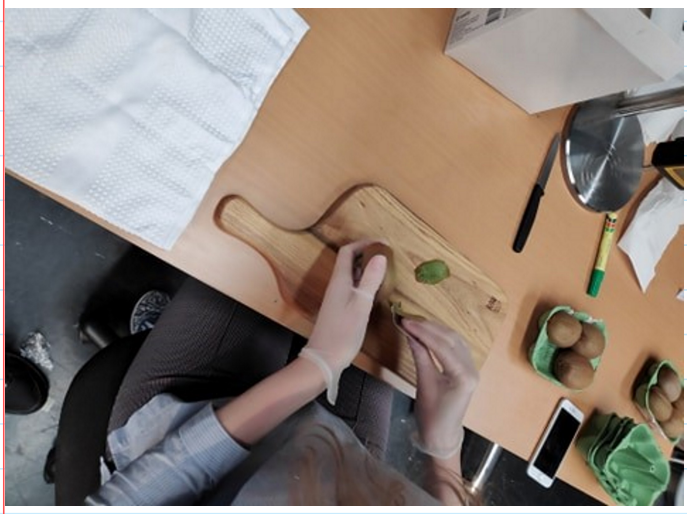
Physiological measurements detailed

Detailed information on how the physiological measurements were carried out.

Measuring procedure

torsdag 18. februar 2021 12:25

1. Kiwi is weighed before the scanning takes place
2. After the scanning the measureing process begins
3. One person sits by a computer noting down the values for each kiwi
4. While the rest is doing the measurement process.



The kiwi are peeled on both sides

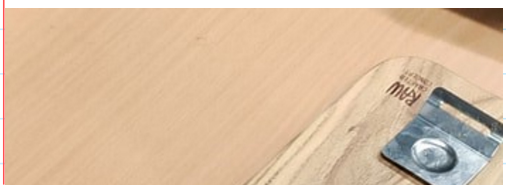
TOP SIDE

With the use of a penetrometer the firmness is measured on both sides of the kiwi.

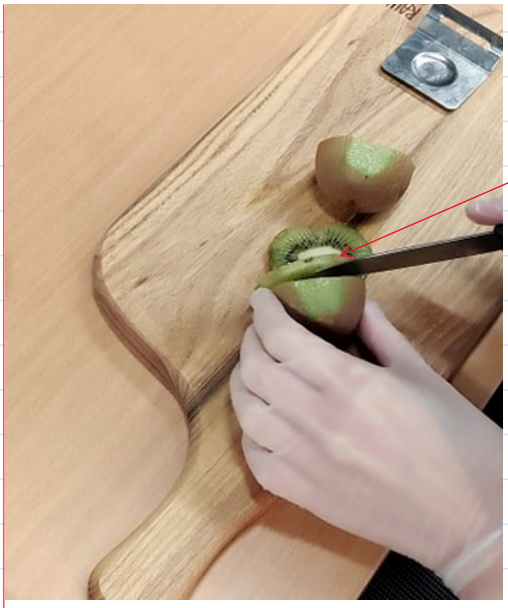
Bottom side

The kiwi is then sliced in half
Then two vertical slices of the kiwi is cut up
One slice to be used for 2. round of scanning
Another slice used for salt and brix meter

First slice

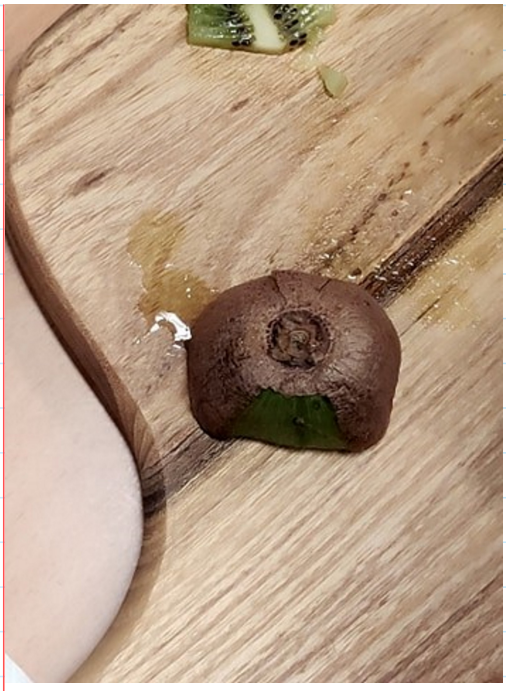


Second slice



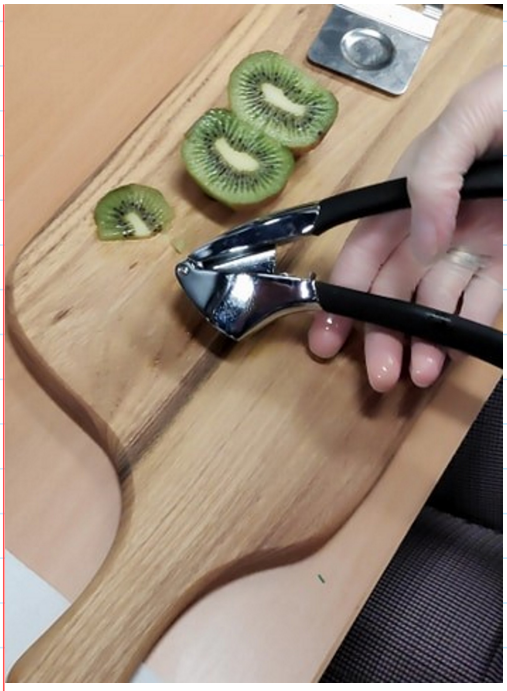
Second slice

Then the core is measured. Both on the stem side and the other half.
So we get two core measurements.



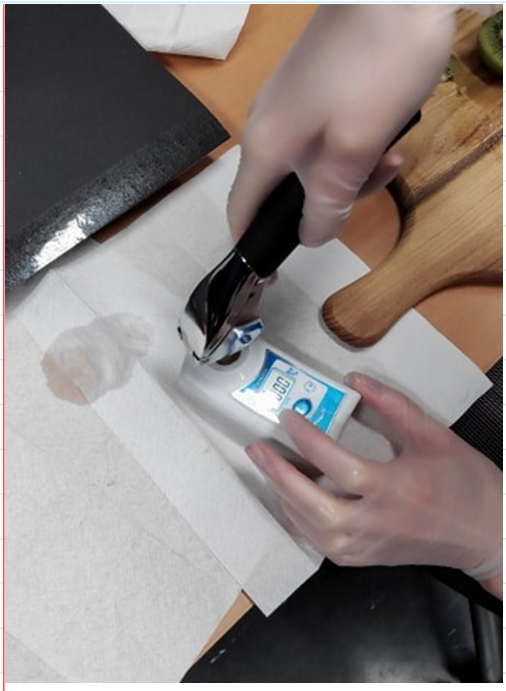
The stem side is measured first

A garlic squeeser is used to squeese the juice out of the kiwi.



One of the vertical slices is cut in half and layed in the squeeser

Water is put on the brix meter to calibrate the sensor

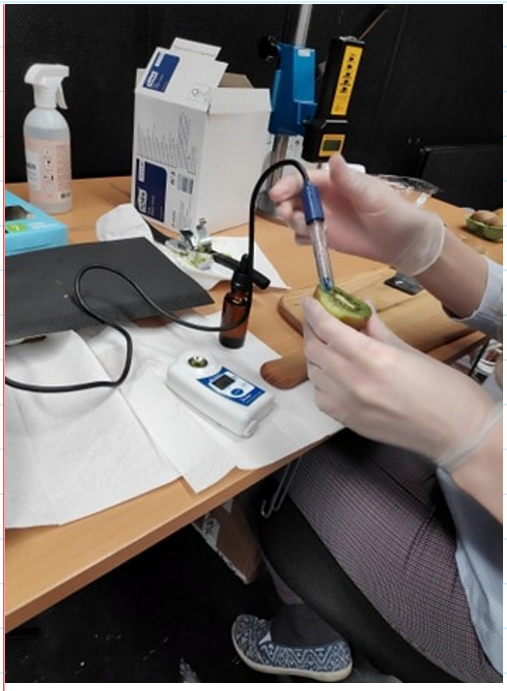


Then the juice is pressed on top of the brix meter.

A reading is measured from the brix meter

This squeesing and calibrating is also repeated for salt meter

A



Then the ph meter is put inside the kiwi
The ph meter stays in the kiwi until the ph meter is "normalised"
It shows a smileyface when the reading is correct.

A good reading from the ph meter

The last slice of the kiwi is put on a black paper to be scanned

Appendix E

Project plan

The project plan created for this bachelor thesis.

Contents

1 Goals and framework (introduction)

1.1 Background

The background for this project is a commercial interest from Bama- Norway's largest private distributor of fruit and vegetables and scientific interest from The Norwegian University of Science and Technology (NTNU), where we want to utilize hyperspectral imaging (HSI) in order to make analyzing and sorting processes more efficient. For Bama, this would mean more capital efficiency in terms of margins and saved labour, while providing higher quality products to customers.

Although Bama probably can benefit from our findings , it did not play a supervising company role in this bachelor thesis project. Bama contributed by providing fruits which were used for research. It was NTNU together with NTNU Technology Transfer (TTO) which decided to go through with this bachelor thesis proposal. This is the main reason why Hilda Deborah who works at NTNU was chosen to be a company primary contact person in this work. This would say that the project was internally driven and we were able to decide about possible project issues within HSI ourselves. The only requirements for these were that they had to be confined within borders given by Hilda.

Although hyperspectral imaging is a relatively new analytical technique and the full potential of it has not yet been realized, hardware and software suited to it are in the middle of intensive development world wide. These technology solutions are getting better and more applicable for all forms of companies and institutions. There are already known many benefits of using HSI. One of the its major assets is the fact that HSI cameras can measure objects in a non-destructive manner and provide back information that is invisible to the human eye. This information can potentially lead to enourmous improvments in many field. In this project, we wanted to take advantage of this fact and experiment with methods within hyperspectral imaging in order to find better non-destructive ways of characterizing parameters within kiwi.

1.2 Project goals

The goal of this project is to determine maturity of kiwi using hyperspectral imaging. To determine the maturity different internal properties are measured in each kiwi. These include acidity, saltiness, firmness and sweetness.

Learning goals

1. Acquire knowledge about the use of spectral imaging, how it works and how to interpret its data.
2. Learn how to work as a group in a more serious manner. Proper team management is carried out.
3. Learn new methods and algorithms in the fields of image processing and image analysis.

Result goals

The main goal of this project is to find hyperspectral data that can determine or correlate with the maturity of kiwis. We also want to find useful correlations between parameters like sugar content, water content, firmness and acidity in the kiwi fruit.

Effect goals

Our bachelor thesis can be categorized as a research project. Since we are not directly working with BAMA, it's very difficult to get into their data, costs and profits they have now. That's why it is impossible for us to predict now how our solution will impact their targets in the future. That's why our effect goals concern only our group and not external factors.

- Achieve a R2-score of our model to be equal or greater than 0.50 this will say the approximately half of the observed variation can be explained by the model's inputs by 10.05.2021
- Master usage of VNIR and SWIR hyperspectral cameras by 1.04.2021
- Discover new algorithms and methods that can be used to efficiently predict a maturity level by 1.05.2021
- Find internal parameters in kiwi that can be used to predict a maturity level by 1.04.2021
- Learn how to define, write and document a research problem by 20.05.2021
- Receive A grade on bachelor thesis by 1.07.2021
- Publish an scientific research about our results by 1.07.2021

2 Project Scope

2.1 Problem field

Kiwifruit is a popular fruit consumed all over the world for its healthy properties and taste. At present, kiwi is usually sorted manually or automatically based on its external quality features, such as shape, color, size, and lack of defects. These quality features could not provide any guidance to internal quantities. Traditional methods used to measure these internal features are destructive. The development of non-destructive measurements of inner features would be very useful for producers and distributors. It could help them to ascertain fast evaluation, save costs and gain many new satisfied customers. There have been done researches that aimed to produce a non-destructive models of predicting soluble solids content (SSC), pH level, sweetness and sourness in kiwifruits. These factors are very interesting for customers; however, distributors of fruits and vegetables are mostly interested in predicting the maturity of their products. Exactly this factor can help them hit higher targets. Herein lies the main part / challenge for our bachelor thesis project. Our goal is to create a non-destructive solution that will allow us to predict a maturity level of kiwifruits based only on internal factors. These internal factors will firstly be predicted by models based on hyperspectral information, and later compared to results gained by usage of destructive testing methods, for example with help of a penetrometer, Ph Meter, Brix Meter and a salt meter. Hopefully, by using internal factors and computer vision solutions, we can create more predictability for distributors and cause a forward step in current solutions.

2.2 Problem limitation

The project has a deadline the 20. May, giving us just above 3 months to collect the data and write a report covering our process and findings. The limitation of time has made us specify the task more clearly, as we do not have time to cover a task as broad as we like.

Because the project was bound to start in January and end in May, we would not have the access to fruits coming straight from harvesting due to the project not being in sync with their harvesting season. Also we are not sure how much information about the kiwifruits we can receive from distributor. There is a possibility that we would not be able to know the exact harvesting date, information on how the fruits were stored, the temperature of storing, whether or not there were some pesticides used. All these factors would have some impact on the results.

In addition to this, we are limited to the equipment NTNU owns and are avail-

able for us to use. Currently there are two hypespectral cameras; one in SWIR range and the other one in VNIR range. These is the only two ranges taken into account. For a complete list of the techincal tools used, see table 2.

2.3 Problem description

Currently, fruit companies are determining the maturity of kiwis by measuring its firmness using some sort of a penetrometer. This measurement poke holes in the fruit and measure how much force it could withstand. Since this method of determining firmness ruins the fruit, making it not an sellable item anymore and ends up as waste. This costs the fruit companies money and time, as this procedure is mostly manual labour.

In this project we shall explore the possibility of determining kiwi maturity by imaging them with hyperspectral cameras. If this seem to be possible, we would provide a non-destructive solution that could replace the current destructive method.

Not just helping the company throw away less completely healthy kiwis, but also improve the quality of them who reach the grocery store. Since kiwis are being touched and pressed in grocery stores in order to check their firmness and to see if the fruits have defects. When many people touch the same fruit, more damage is caused, which further decreases the quality of the fruit to a point were no one would like to eat it, making it end up in the trash. In the future, maybe all of the kiwis can be scanned and get stickers associated with its maturity level, providing more information to the customer. Leading to less touching and more customer confidence in the quality of the fruits.

3 Project Organisation

3.1 Project Roles and Responsibility Management

1. **Jon Elias Moen** acts as a project leader - he is responsible for taking notes during all meeting, following up meeting times, agreements and he can sign on behalf of the group. If he cannot attend a meeting or becomes sick his responsibility should be taken by Katherine Brox-Saidi. He also has main responsibility for the Analysis period together with **Katherine Brox Saidi**.
2. **Katherine Brox Saidi** is responsible for reading a book "Bacheloroppgaven-Skriveråd og regler for utformingen" written by Aage Rognsaa. She undertakes to share the most important points from this book with all group's members.

3. **Eivind Kohmann** is responsible for the data acquisition period. He also works as a technical advisor with **Vebjørn Nilsen**, since they have experience in the field.

4. **Vebjørn Nilsen** is responsible for the pre-processing period.

5. **Sony George** is an associate professor at NTNU and group's supervisor. He acts as well as a mediator in conflict situtions.

6. **Hilda Deborah** is a research scientist at NTNU and group's company contact. She has done a master thesis on HSI, so she will take part in the learning process and be at disposal of any questions.

7. **BAMA-Gruppen AS** is Norway's largest private distributor of fruit and vegetables and it's going to deliver needed fruits for student's bachelor thesis project.

8. **Tom Røise** is a bachelor thesis supervisor.

Even though the different project periods are divided among the group, everyone helps everyone.

3.2 Routines and Group Rules

1. All group members have agreed that they want to achieve the best possible grade (A) by submitting the bachelor thesis. Therefore, they undertake to follow all NTNU's guidelines for writing the bachelor thesis, as well as listen to feedbacks from the supervisor and company's contact person.

2. Each group member undertakes to work a minimum of 25 hours per week. The group members must have physical meetings - at least one physical meeting per week if the corona situation allows to do so.

3. Group members must be notified if there is a someone who does not have the opportunity to show up at the agreed time.

4. All hours used to work with the Bachelor Thesis, must be registered in Toggl.

5. If a group member is placed in corona quarantine or feels sick, he/she shall not attend physical meetings.

6. If one of the group members becomes ill during a spring semester the group should have a meeting with purpose to discuss occurred situation.

7. If one of the group members does not perform work in accordance with the rules, the routine in the event of "serious breaches of the rules" will be followed.

8. All conflicts must first be tried to resolve locally within the group.
9. All members must contribute throughout the process.

Work rules

1. Obligation to attend project meetings, meetings with the supervisor and meetings with the company contact person.
2. Exception from the obligation to attend may be granted for a good reason.
3. At best an agenda is mailed to the supervisor before the meeting. Giving the supervisor a chance to prepare.
4. Meeting minutes are written for all meetings with supervisor or company. They are made available for our supervisor.
5. All members undertake to follow the authority's recommendations regarding the coronavirus.
6. All members undertake to wear face masks where it is not possible to keep 1 meter distance.
7. Group members are obligated to inform if they think that something should be done differently.
8. All group members undertake to follow the rules for code standard, tool use and report writing.
9. After the bachelor thesis is submitted, the whole group must work on preparing the presentation. A 2-week preparation period is set for this.

Routine for solving serious problems

1. In all cases, the disagreement must first be discussed within the group.
2. Conversation between all group members about the problem. No one should judge others. The purpose of such conversations is to solve the problem.
3. Written or oral warning from other members which should include:
   (a) violations
   (b) requirements to compensate for lost work
   (c) explanation of the consequences of further breaches of the rules.
4. Conversation with the whole group and the supervisor – Sony George.

5. Conversation with the whole group and company's contact person – Hilda Deborah.
6. Conversation with the whole group and bachelor thesis supervisor – Tom Røise.

4 Planning, Follow-up and Reporting

4.1 Meeting plans

Meeting with supervisor Sony is planned to happen on Mondays at 3 PM. During these meetings we will go through the meeting agenda that was prepared and mailed to Sony beforehand. Clearing up any questions about the current position in the project.

Meeting with company Hilda Deborah is planned to happen on Wednesdays at 3:30 PM. During these meetings we will ask any questions about Hyper spectral imaging, also further clarifying any misconceptions.

5 Organisation of Quality assurance

5.1 Documentation, storage and source code

Every meeting with company or supervisor is summarised and stored as documentation.

The source code is going to be uploaded to GitHub. Working with Github also provides a local copy of the source code on every involved computer.

5.2 Data management

During the pre-processing time period, a lot of processing power is needed. Therefore a computer at NTNU with the needed processing capability is available. This computer is also possible to use during any other time-period of the project, for example during the Analysis.

After the pre-processing, each person in the group get their own harddrive with a copy of the hyperspectral images. In this manner every person can experiment with the data, on their own computer. The source code however is shared with everyone through a private repository at GitHub.

5.3 Tools

We have separated the tools used for project support (Table 1) and technical tools (Table 2). Technical tools also include the instruments to measure the different internal properties of the kiwis.

Name	**Type**	**Application**
Overleaf	Online tool for sharing and writing LaTeXdocuments	Project report
Instagantt	Online tool for making and using Gantt-diagrams	Project planning
Github	Version Control of code, and integrated Kanban board	Version Control
Google Drive	File Storage and File Collaboration	File sharing
Toggl	Online tool for time tracking	Time tracking

Table 1: Project tools

Name	**Type**	**Application**
Computer at NTNU	for ENVI and Pre-processing	Image analysis
Python	High level programming language	Coding
Spectral python	Python Library for Spectral images	Coding
Scikit-learn	Python Library for machine learning	Coding
GLIMPS and ENVI	Software for visualizing and analyzing spectral images	Data visualization and analyzing
HySpex VNIR-1800	Hyperspectral camera, 400-1000nm	Camera
HySpex SWIR-384	Hyperspectral camera, 930-2500nm	Camera
Ph Meter	Tool for checking acidity	Measuring instrument
Brix meter	Tool for checking sugar content	Measuring instrument
Salt meter	Tool for checking salt content	Measuring instrument

Table 2: Technical tools

5.4 Risk Analysis

There are different risks associated with this research project. A course of action to avoid each risk and or what to do if it happens.

In a perfect world BAMA will deliver kiwi that has different levels of maturity. If this is not the case, small changes in methodology will solve the issue.

Likelihood is categorised as very low, medium, high, very high, while impact is categorised as minor, moderate, major and catastrophic.

ID	Activity	Likelihood	Impact	Action
1	Bama not supplying with fruit	Medium	Moderate	If Bama does not provide us with fruit, we have to get it ourselves from a grocery store.
2	Fruit delivered by Bama has same maturity	Medium	Minor	This means we have to divide the fruits in different time batches. Where one batch is scanned immediately, the rest of them are left to rot for different amount of time before they are scanned.
3	Covid-19 complications	Very low	Moderate	If someone is in quarantine, this person works from home. If the case is that everyone is in quarintine during data acquisition week, we can get someone at NTNU to do the scan or postpone the scanning. If the case is that NTNU locks down during this week, we write a proposal to NTNU and request for permission to enter.
4	Other sickness	Medium	Minor	Work from home. If the person cannot work. Split up the tasks and distribute to the other team members
5	Loss of primary code or other important information/data	Very low	Major	Store the images on multiple hard-disks and keep the repository up do date. Retrieve local copy if needed.

6	Problems with external systems, hyperspectral cameras	Medium	Major	Seek assistance from someone else with higher knowledge of the system.
7	Unexpected amount of time spent on some of the parts of the project	High	Moderate	Make plans that have room for delays and follow up on scheduled work and revise if needed
8	The research problem ends up being too complex, and takes longer than planned	Medium	Moderate	Have regular meetings with supervisor and find solutions/workarounds
9	Cannot get access to the cameras when we need them	Very Low	Major	Reserve lab room at least one week in advance and ask supervisor when cameras are unavailable.

Table 3: Potential risks that will affect the project.

		Impact			
		Minor	Moderate	Major	Catastropic
Likelihood	Very low		5, 3	9	
	Medium	2, 4	1, 8	6	
	High		7		
	Very High				

Table 4: Risk distribution

6 Plan of execution

Id	Name	Start	Due
1	**Meeting with company reference:**	13/01/2021	21/04/2021
2	**Meeting with supervisor:**	12/01/2021	17/05/2021
3	**Research:**	07/01/2021	15/03/2021
3.1	Literature review	07/01/2021	15/03/2021
3.2	Image Normalisation research	07/01/2021	15/03/2021
4	**Test Experiments:**	21/01/2021	14/02/2021
4.1	Getting familiar with the camera	21/01/2021	14/02/2021
4.2	Test of scan procedure	27/01/2021	27/01/2021
5	**Data acquisition:**	14/02/2021	03/03/2021
5.1	Obtaining fruit from Bama	14/02/2021	20/02/2021
5.2	Take images	17/02/2021	03/03/2021
6	**Preprocessing:**	04/03/2021	16/03/2021
6.1	Image normalisation	04/03/2021	16/03/2021
7	**Analysis:**	17/03/2021	06/05/2021
7.1	Experiment with methods in python	17/03/2021	06/05/2021
8	**Report:**	03/02/2021	20/05/2021
8.1	Make first draft	03/02/2021	04/02/2021
8.2	Report writing	05/02/2021	07/05/2021
8.3	Receive Feedback from Supervisors	10/05/2021	19/05/2021
8.4	Deliver Report	20/05/2021	20/05/2021
9	**Presentation preparation:**	20/05/2021	10/06/2021
9.1	Preparing Bachelor Thesis presentation	20/05/2021	10/06/2021
9.2	Mock-up presentation with supervisor 1	24/05/2021	24/05/2021
9.3	Mock-up presentation with supervisor 2	31/05/2021	31/05/2021
9.4	Mock-up presentation with supervisor 3	07/06/2021	07/06/2021

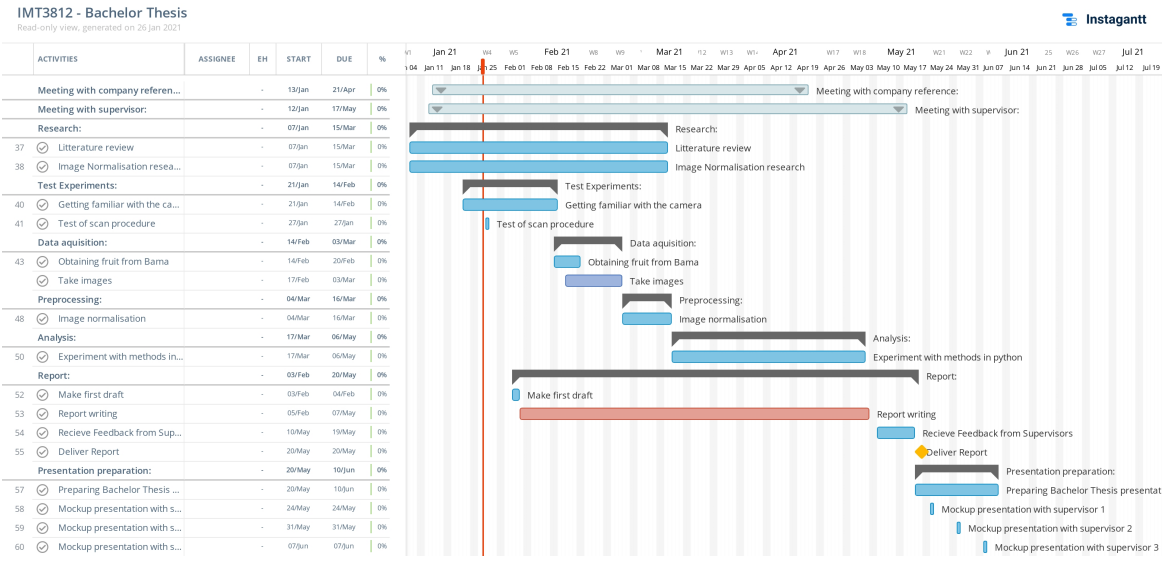


Figure 1: Gantt-diagram (instagantt)

Appendix F

Survey questions

The questions for the survey performed on basic knowledge about Hyperspectral imaging.

Basic knowledge about hyperspectral imaging

The goal of this survey is to identify whether or not students have heard/learned about hyperspectral imagining.
The survey is anonymous.

*Må fylles ut

What grade are you in? *

- 1
- 2
- 3

Are you an IT student? *

- Yes
- No

Have you heard about hyperspectral imaging before starting at your current university? *

- Yes
- No

Be om redigeringstilgang

Have you learned about hypersectral imaging during your studies at your current university? *

- Yes
- No

Have you ever seen a hyperspectral camera? *

- Yes
- No

Do you think that capturing photos with a hypespectral camera is much different than capturing with an ordinary digital camera? *

- Yes
- No

Do you know what hyperspectral imaging can be used for? *

- Yes
- No

If your answer was "yes" in the previous question, could you please name some applications of it?

Svaret ditt

Be om redigeringstilgang

Send

Send aldri passord via Google Skjemaer.

Appendix G

Email about images

Email about the use of different images. However this was not done for all the images, as it was discovered it was not necessary.

Introduction

Figure 2.5 is retrieved from the source [20]. With the courtesy of Gabor J. Kemeny from the company Middletonspectral. See email from figure G.1.

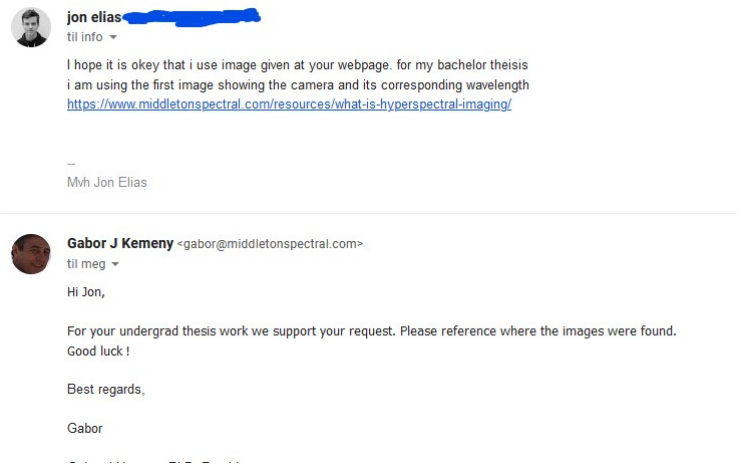


Figure G.1: Email from Gabor J. Kemeny

Theory

Figure 3.3 is retrieved from the source [27]. With the courtesy of Stacey Carrier from the American Laboratory. See email from figure G.2.

Figure 3.2 is retrieved from the source [25]. With the courtesy of Baohua Zhang. See email from figure G.3.

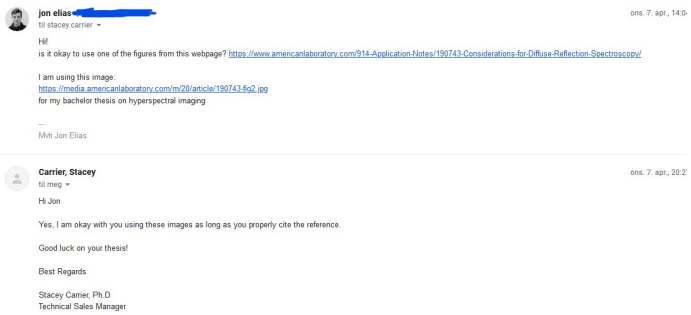
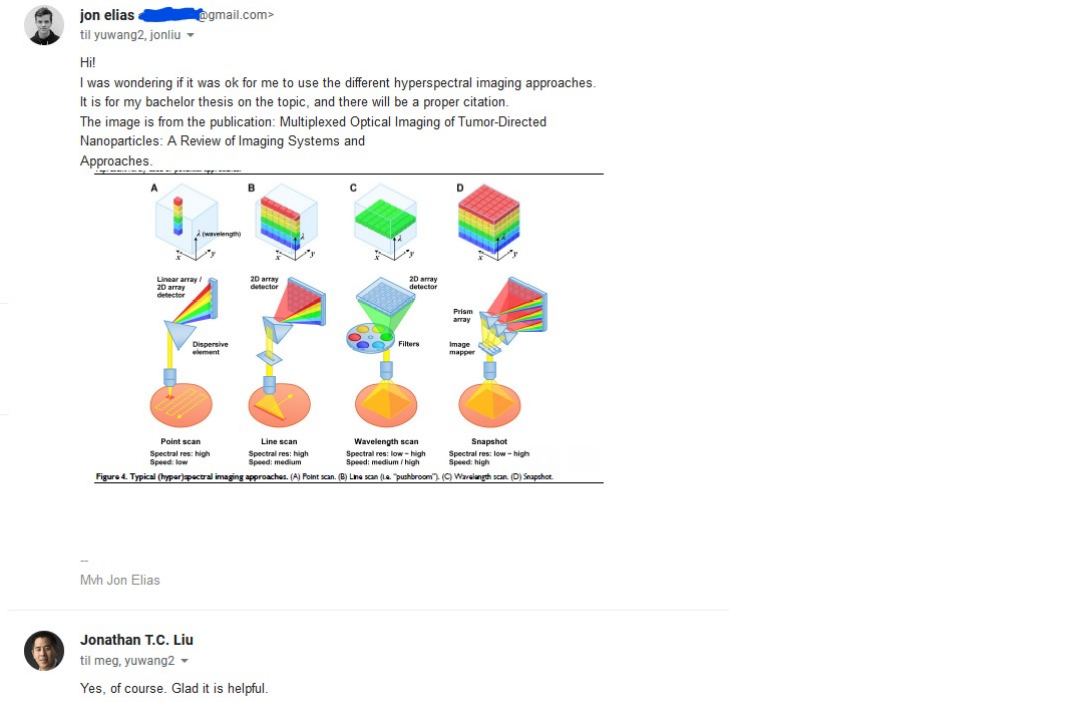


Figure G.2: Email from Stacey Carrier

Figure G.3: Email from Baohua Zhang

Figure 3.1 is retrieved from the source [24]. With the courtesy of Jonathan T.C Liu. See email from figure G.4.

Figure 5.4 is retrieved from the website [95]. With the courtesy of Derek Huxley from the company Warsash Scientific Pty Ltd. See email from figure G.5



jon elias [redacted]@gmail.com>
til yuwang2, jonliu

Hi!
I was wondering if it was ok for me to use the different hyperspectral imaging approaches.
It is for my bachelor thesis on the topic, and there will be a proper citation.
The image is from the publication: Multiplexed Optical Imaging of Tumor-Directed Nanoparticles: A Review of Imaging Systems and Approaches.

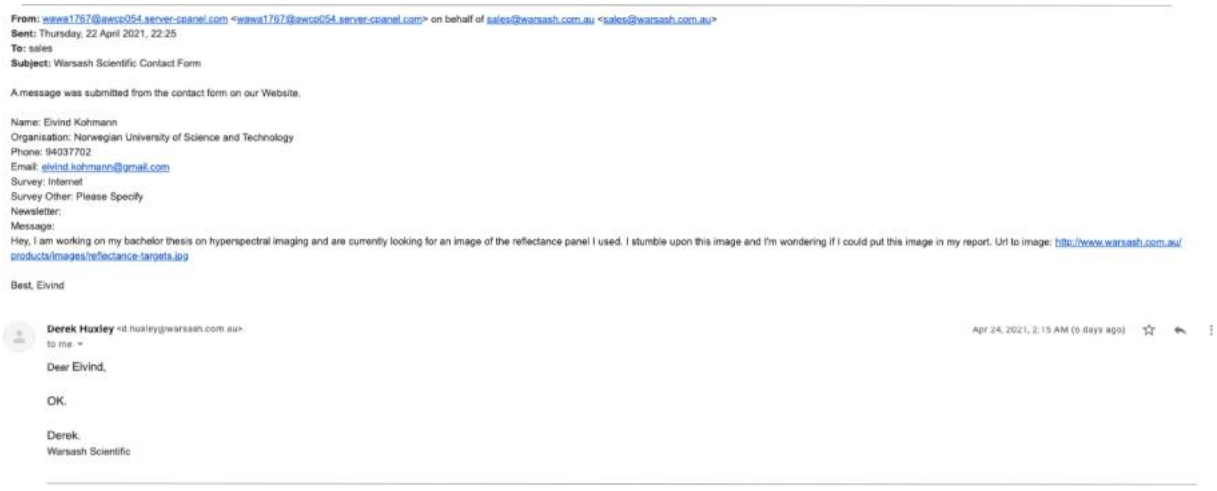
Figure 4. Typical (hyper)spectral imaging approaches. (A) Point scan. (B) Line scan (i.e. "pushbroom"). (C) Wavelength scan. (D) Snapshot.

--
Mvh Jon Elias

Jonathan T.C. Liu
til meg, yuwang2

Yes, of course. Glad it is helpful.

Figure G.4: Email from Jonathan T.C Liu



From: wawa1767@awcp054.server-cpanel.com <wawa1767@awcp054.server-cpanel.com> on behalf of sales@warsash.com.au <sales@warsash.com.au>
Sent: Thursday, 22 April 2021, 22:25
To: sales
Subject: Warsash Scientific Contact Form

A message was submitted from the contact form on our Website.

Name: Elvind Kohmann
Organisation: Norwegian University of Science and Technology
Phone: 94037702
Email: elvind.kohmann@gmail.com
Survey: Internet
Survey Other: Please Specify
Newsletter:
Message:
Hey, I am working on my bachelor thesis on hyperspectral imaging and are currently looking for an image of the reflectance panel I used. I stumble upon this image and I'm wondering if I could put this image in my report. Url to image: http://www.warsash.com.au/products/images/reflectance-targets.jpg

Best, Elvind

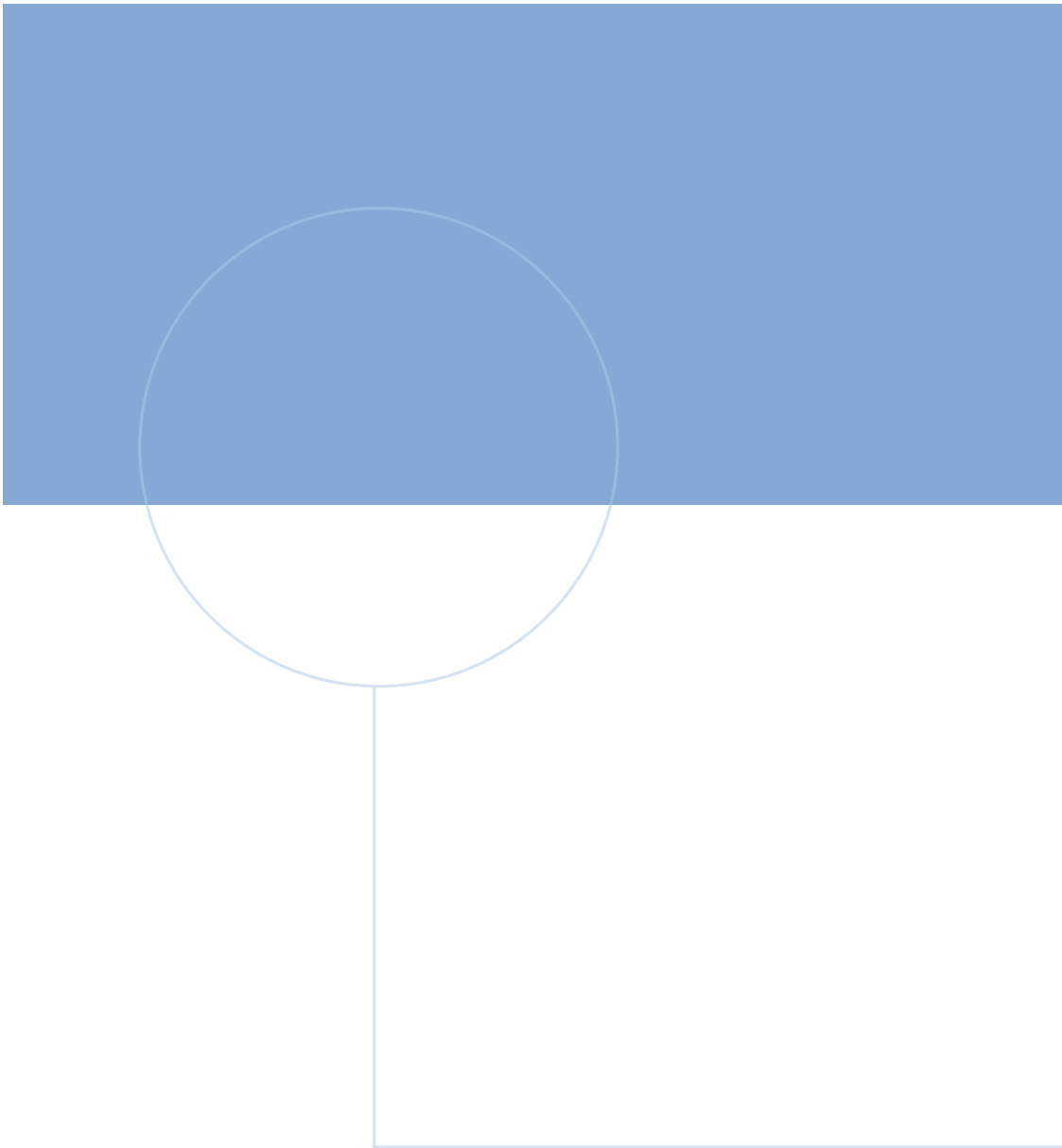
Derek Huxley <d.huxley@warsash.com.au>
to me
Apr 24, 2021, 2:15 AM (6 days ago)

Dear Elvind,

OK.

Derek.
Warsash Scientific

Figure G.5: Email from Derek Huxley



NTNU

Kunnskap for en bedre verden