HYPERSPECTRAL IMAGE SEGMENTATION AND ANALYSIS TOWARDS AUTOMATED NON-DESTRUCTIVE MEASUREMENTS OF OLIGOSACCHARIDES AND OTHER TRAITS

HYSACS PROJECT

WEDNESDAY, JULY 15, 2020

CONTENTS

MOTIVATION

Problem Statement:

Screening of large populations(greenhouse and field).

Grain composition relies on labor and cost intensive wet lab analysis.

Solution:

Using light as a non-destructive analytical tool to predict grain composition.

Early determination(before maturity) to save time in greenhouse or on the field.

Approach:

AI-based prediction models based on hyperspectral data from developing region.

An image of the plants being screened at the hyperspectral plant imaging station at APPF.(right)

Hyperspectral images interrogate the plant physiology at a fine spectral resolution.

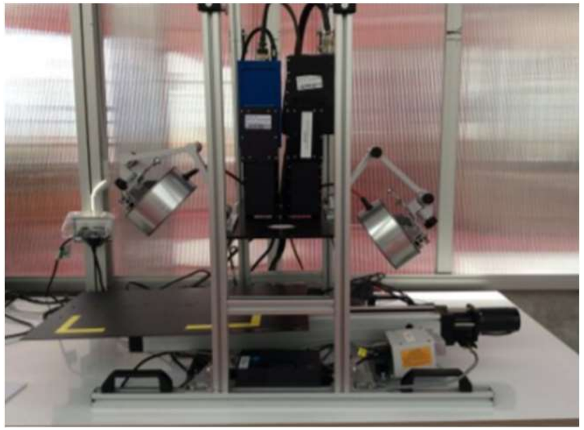


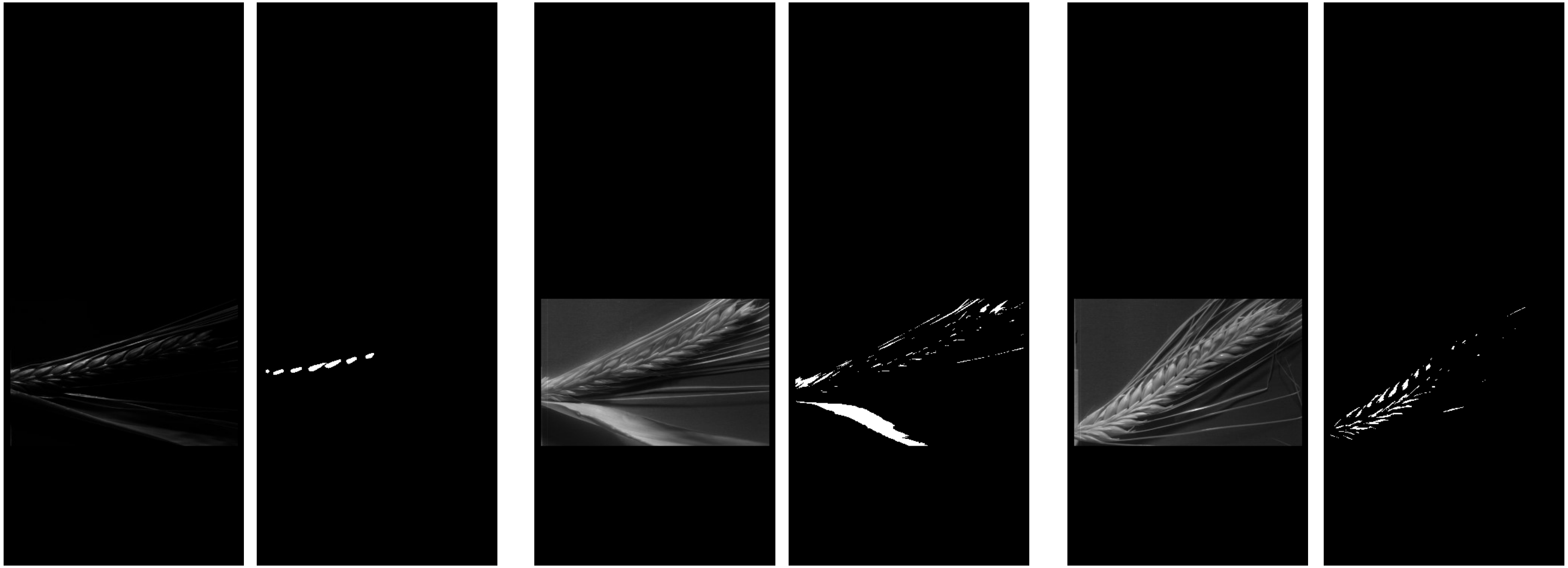
IMAGE SOURCE:HTTPS://WWW.PLANTPHENOMICS.ORG.AU/

DATA EXPLORING

- Data collected over the measurement during 2014 and 2015.
    - 2014: 488 recordings VNIR (400-1000 nm) and SWIR (1000-2500 nm)
    - 2015: 399 recordings
    - 2014: ~746 GB compressed spectral data
    - 2015: ~477 GB compressed spectral data
- 156 genotypes with 3 replicates each
- ENVI file format (header file + binary data)
- 16 bit unsigned integer/ 32 bit float

READING AND VISUALIZING IMAGES:

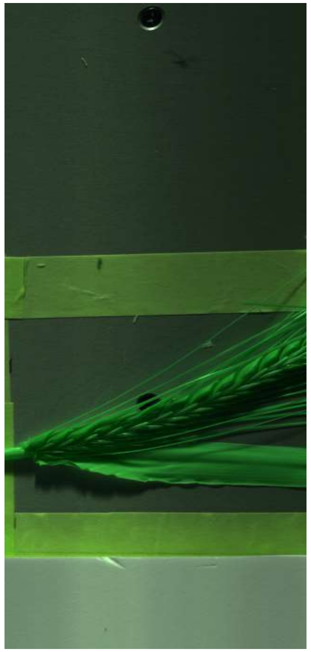
Results based on single channel



READING AND VISUALIZING IMAGES:

Results based on SWIR recommended channels in header file: 80th, 200th and 48th

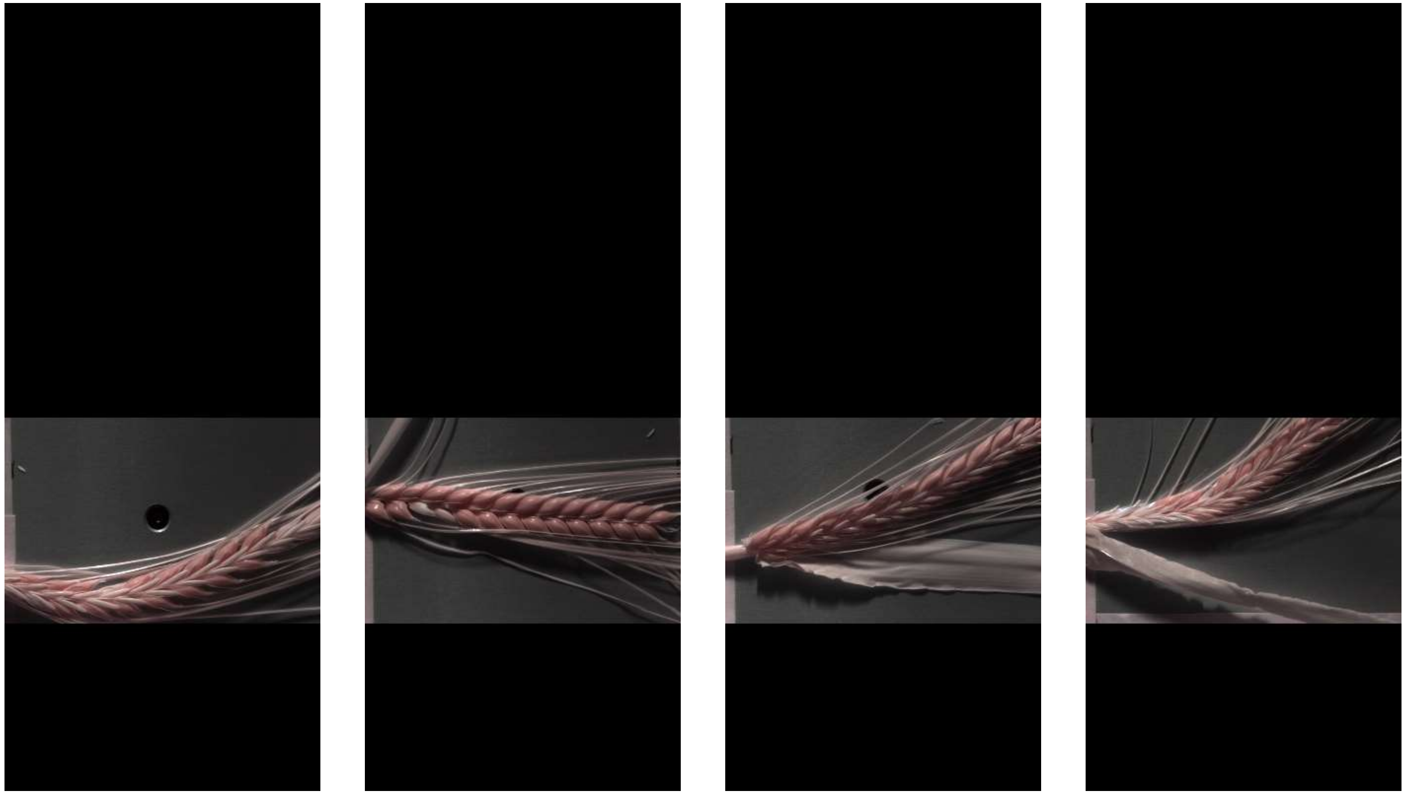




READING AND VISUALIZING IMAGES:

Using three Channels (SWIR)

- Only region where spikes located are kept for processing.
- Channels used: 55, 41, 12
- Still many other combinations possible.



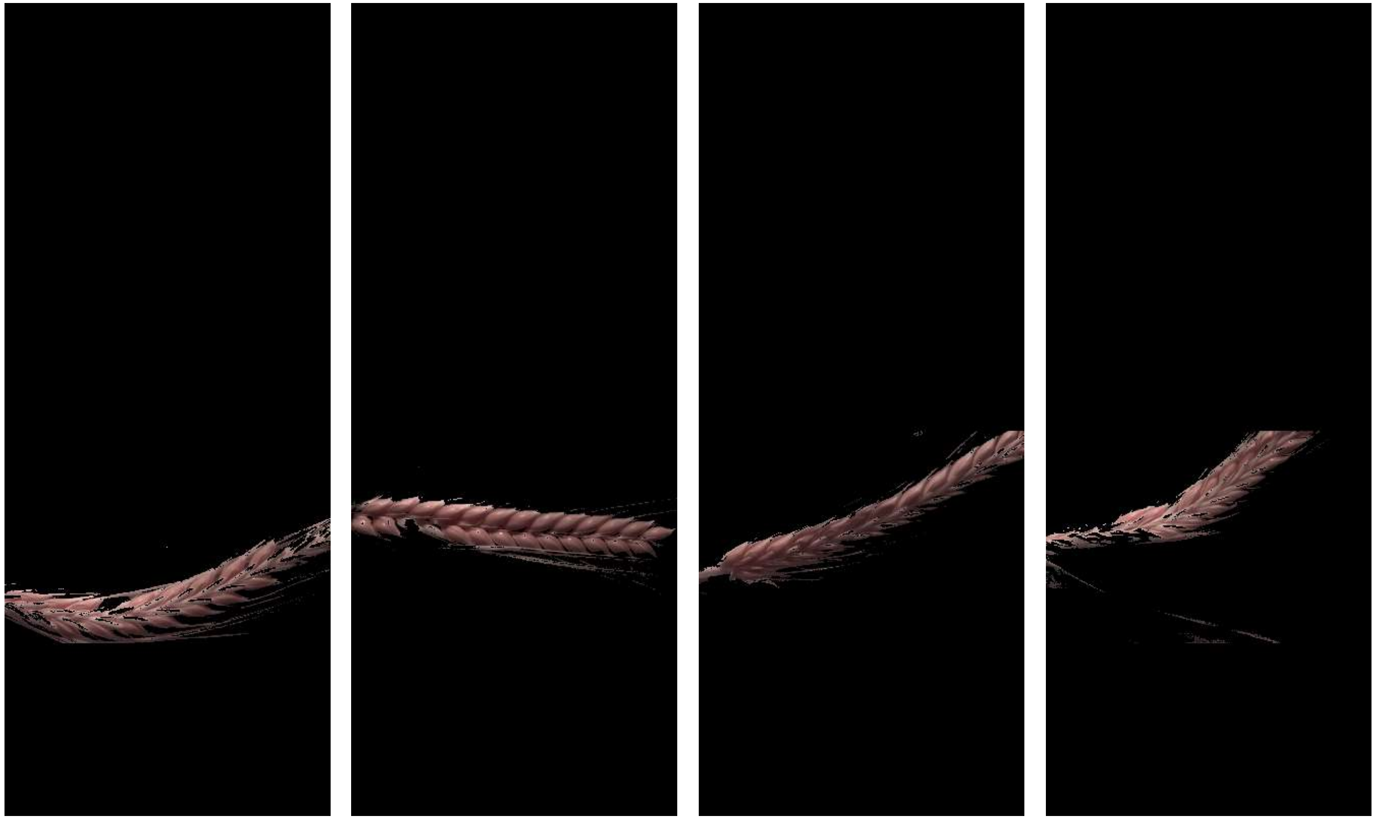
SEGMENTATION USING OPENCV

- Changing Color-space
  - Color Conversion
    - We use the function cv2.cvtColor(input_image, flag) where flag determines the type of conversion (e. g. cv2.COLOR_BGR2HSV).
- Applying mask
  - We use the function cv2.inRange(HSV_converted_image, lower_bounds, upper_bounds)
  - For HSV, Hue range is [0,179], Saturation range is [0,255] and Value range is [0,255]. Different softwares use different scales. So need to normalize these ranges.

RESULT OF APPLYING MASK

- Lower bounds:
  (0, 60, 50)
- Upper bounds:
  (100, 255, 255)

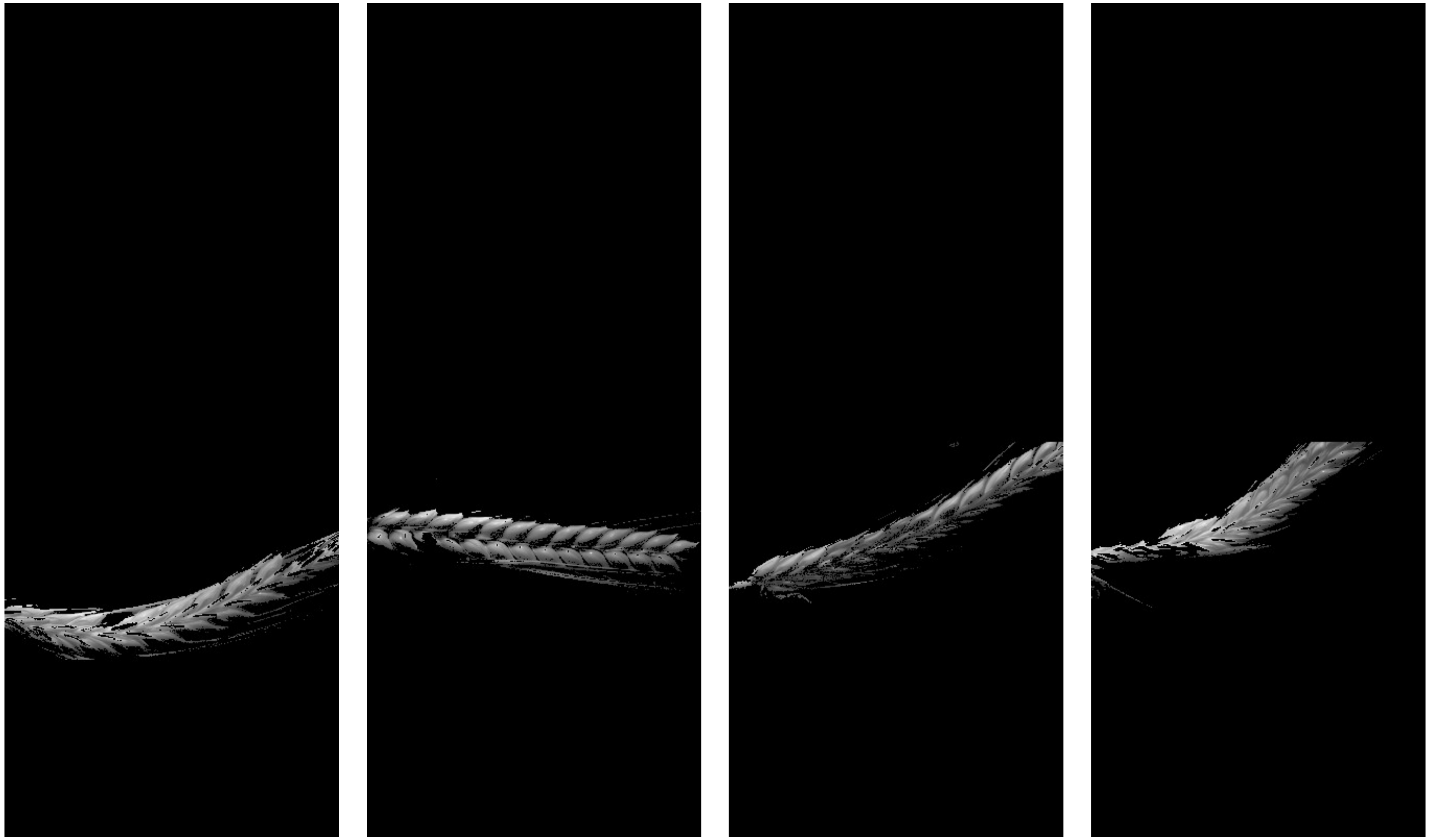
Leaves = removed



SEGMENTATION - SWIR

Although most of leaves are eliminated, the image still contains some noises around grains.

Our focus was to make the image data cleaner by trying to remove this noise as well.

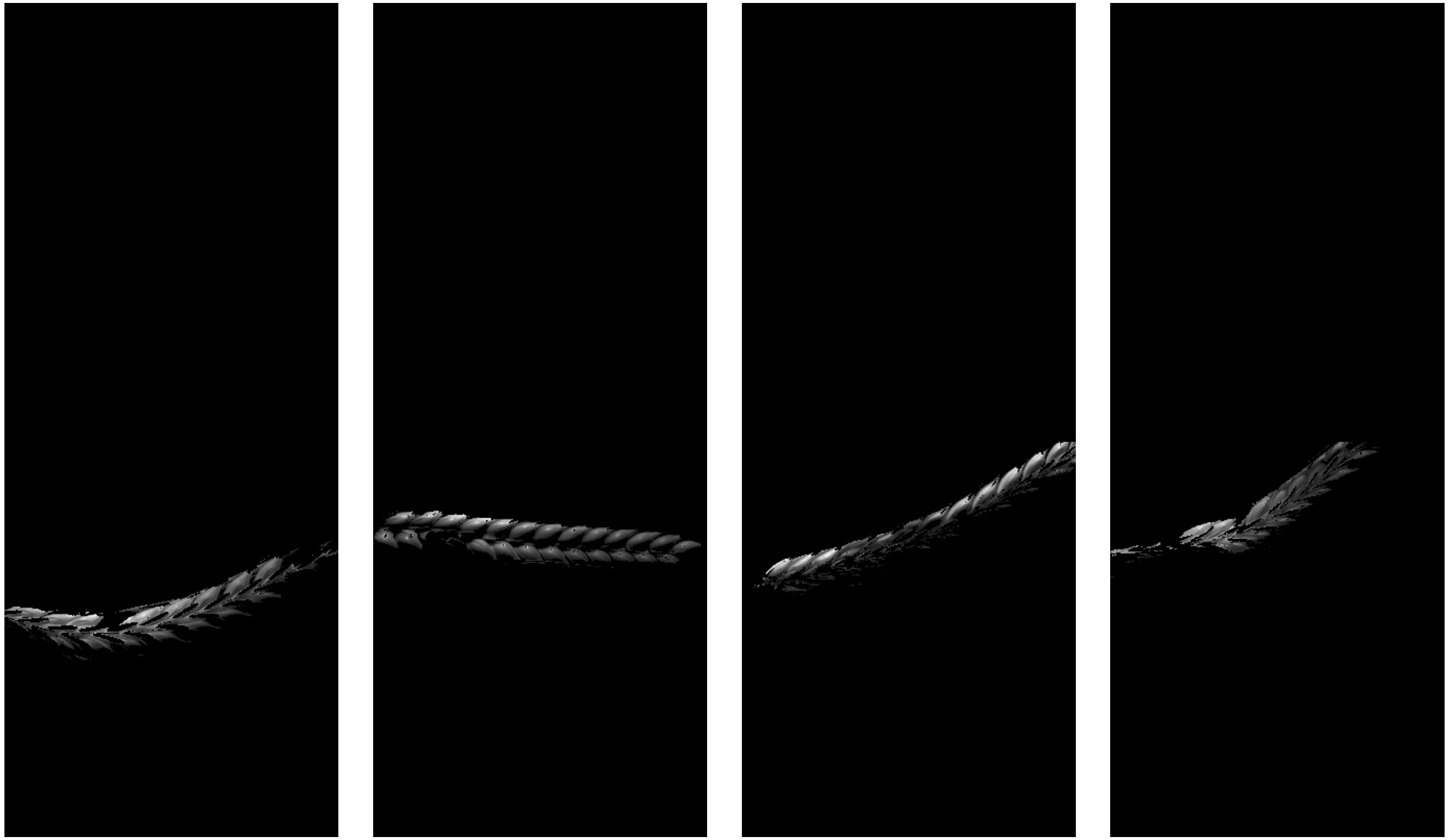


VARIABLE TRANSFORMATION

Goal:
keep grains well separated

Approach:
Quadratic transformation x^2 pushing back to the range of 0-255

Pros:
can pick a global threshold easier

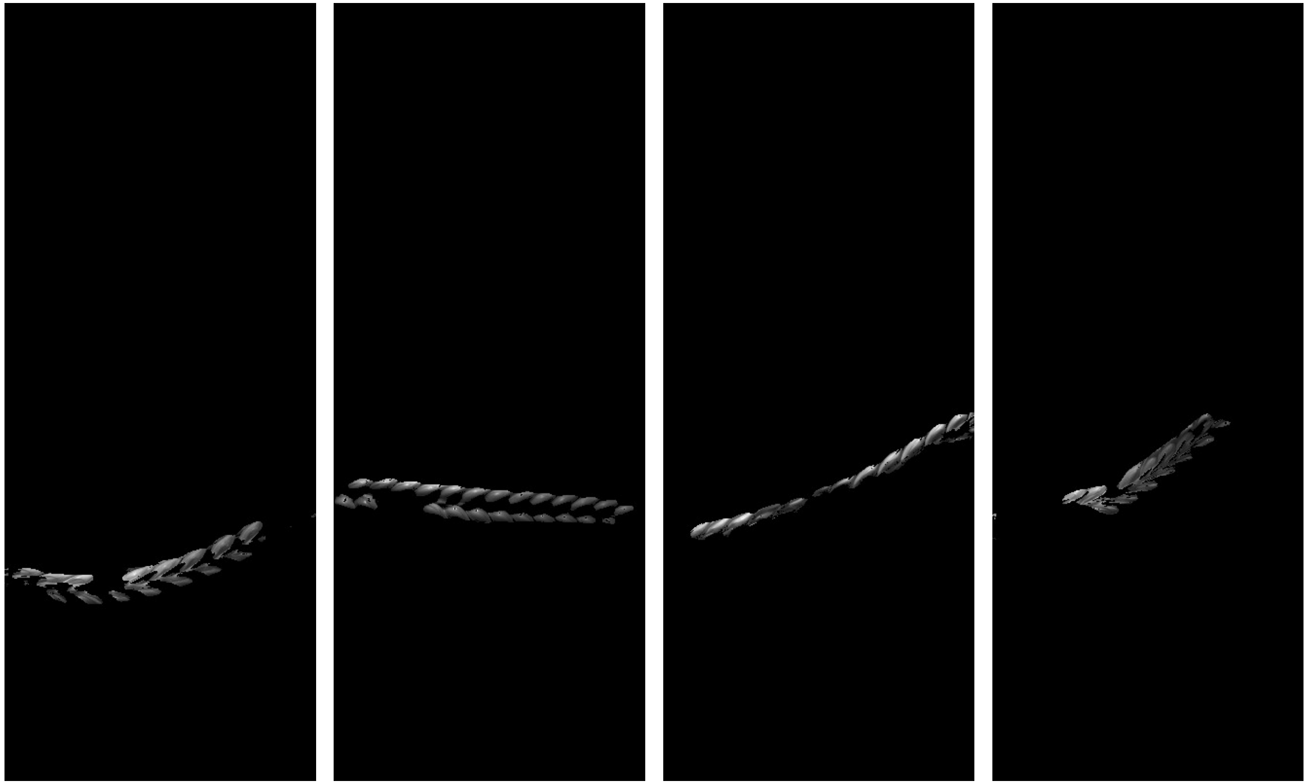


REMOVING NOISES:

Method : Structuring Elements with Radius = 5, Minimum points = 50

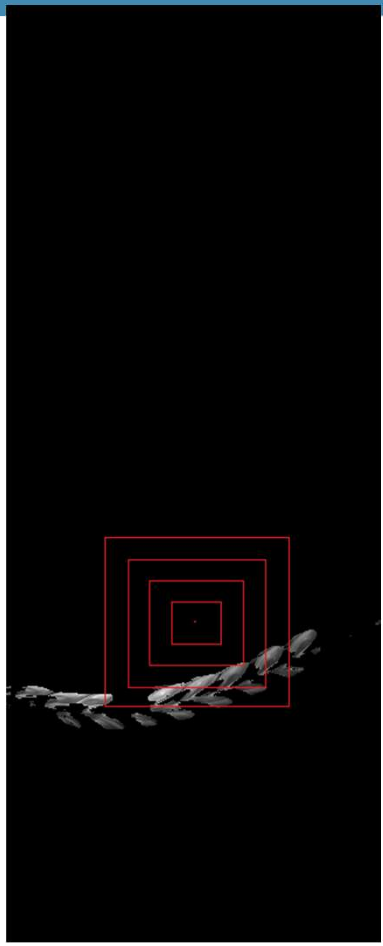
Radius is based on Moore neighborhood

For any pixel, if the number of neighbors being equal or larger than minimum points (50), its kept otherwise its value is changed to zero.

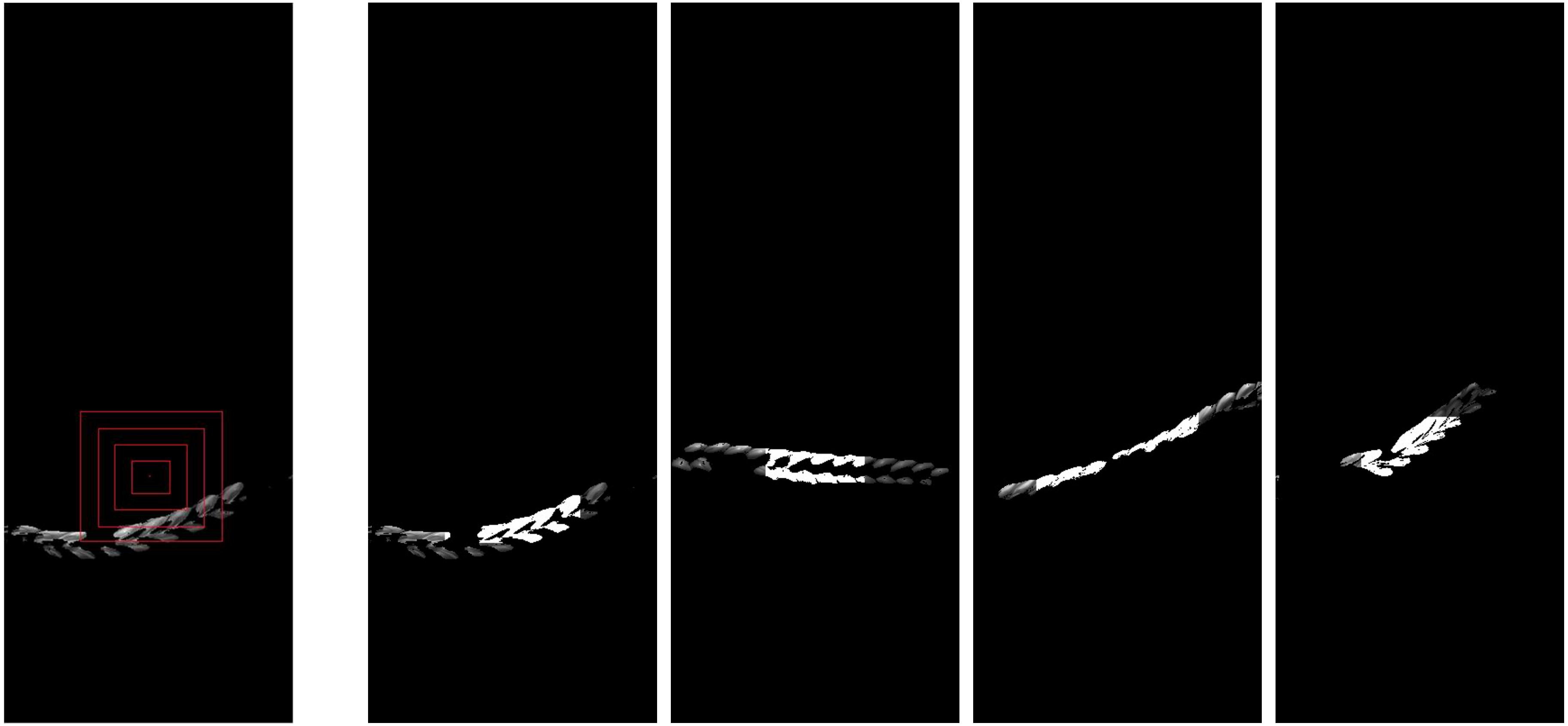


REGION OF INTEREST

- Now that we have a clean image, we can extract the specific information that we want from it.
- Our focus is on the middle grains(5-6).
- Starting from middle point of area where spikes are located, we expand its area until we cover 2000 pixels.
- This results in a collection of about 4,5 grains within the radius of the center grain from where we started.
- Each grain consist of around 20 X 20 pixels.



PICK GRAINS FROM MIDDLE OF SPIKE

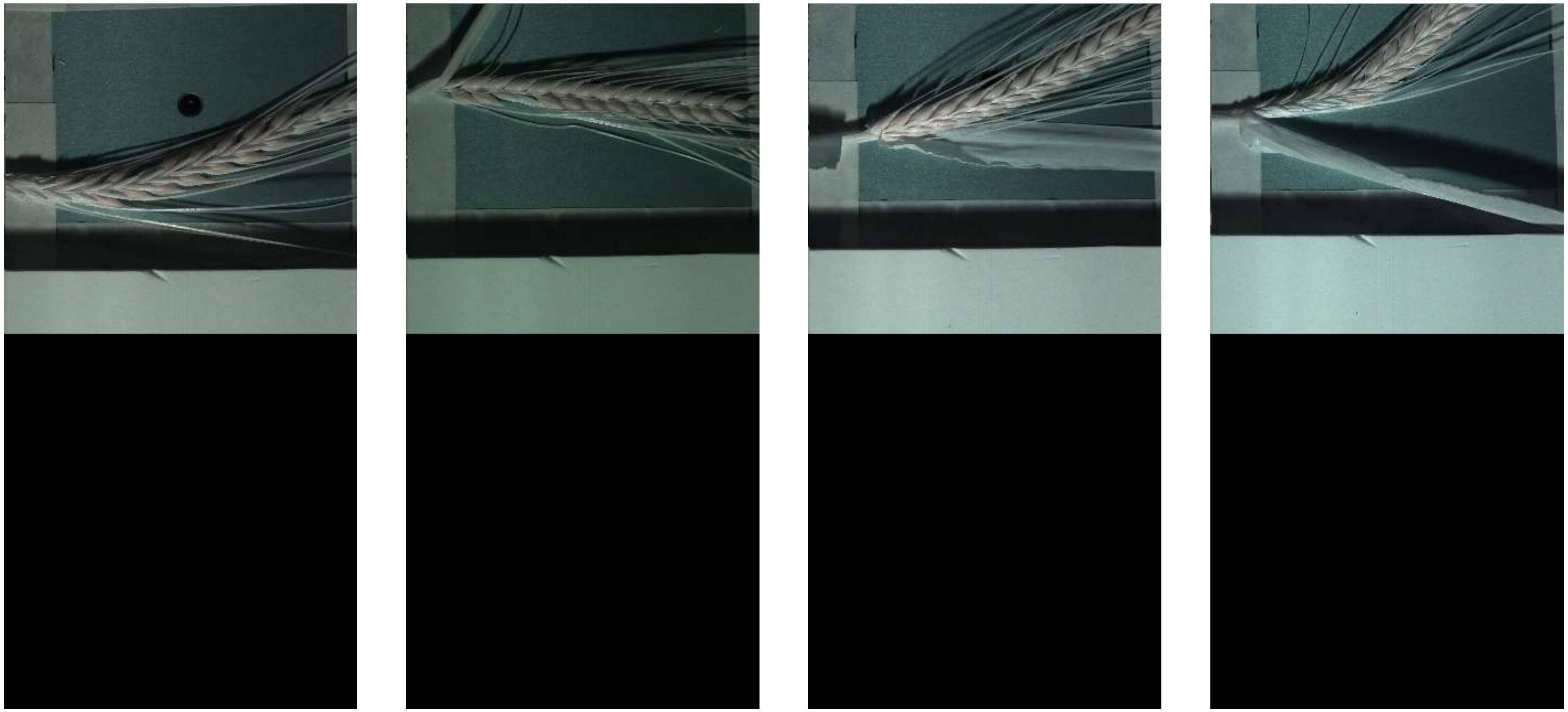


SEGMENTATION - VNIR IMAGES

VNIR images have a dimension of 1600 x 3200.

Three channels: 159, 152, 120.

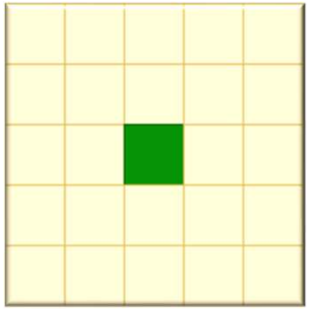
Only area where spikes are located are kept for processing



RESIZE VNIR

1600 x 3200

320 x 640

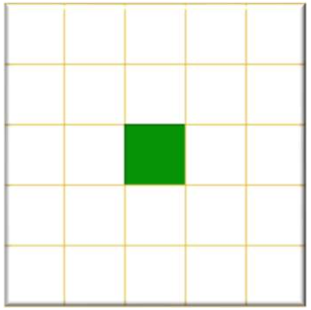


Mean

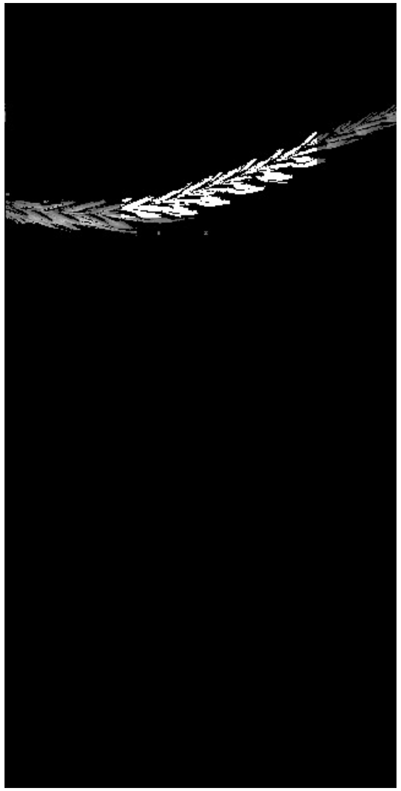
- High computation
- More than 30 min for single image (i5, 8GB ram)

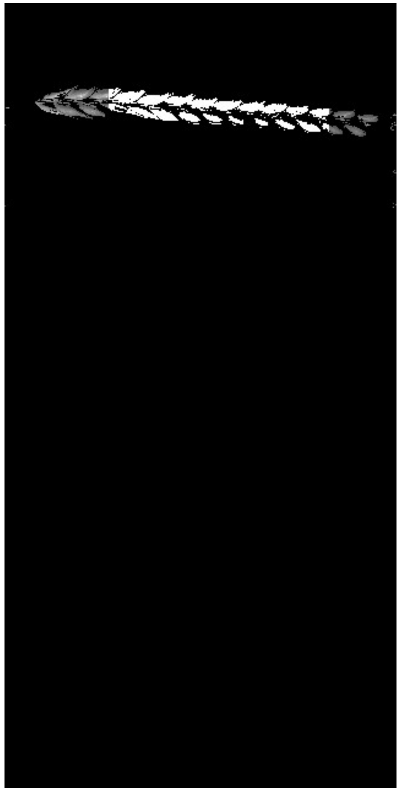
Median

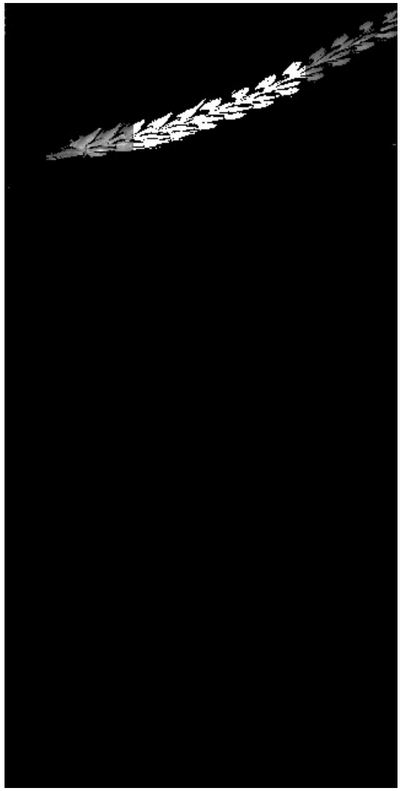
- Take less amount of time
- Around 3 min

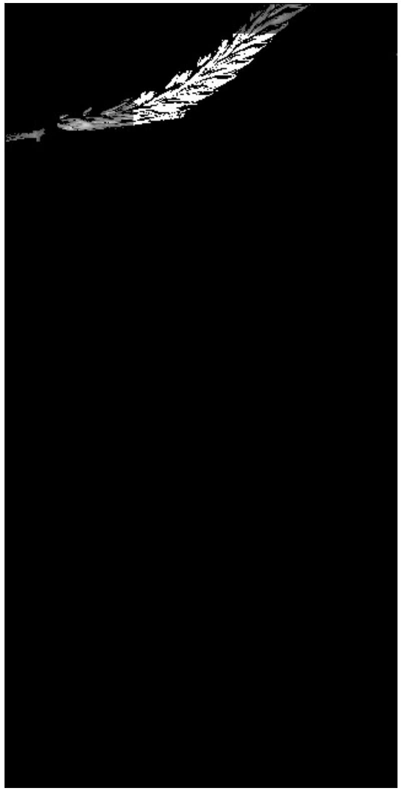


SEGMENTATION – VNIR IMAGES



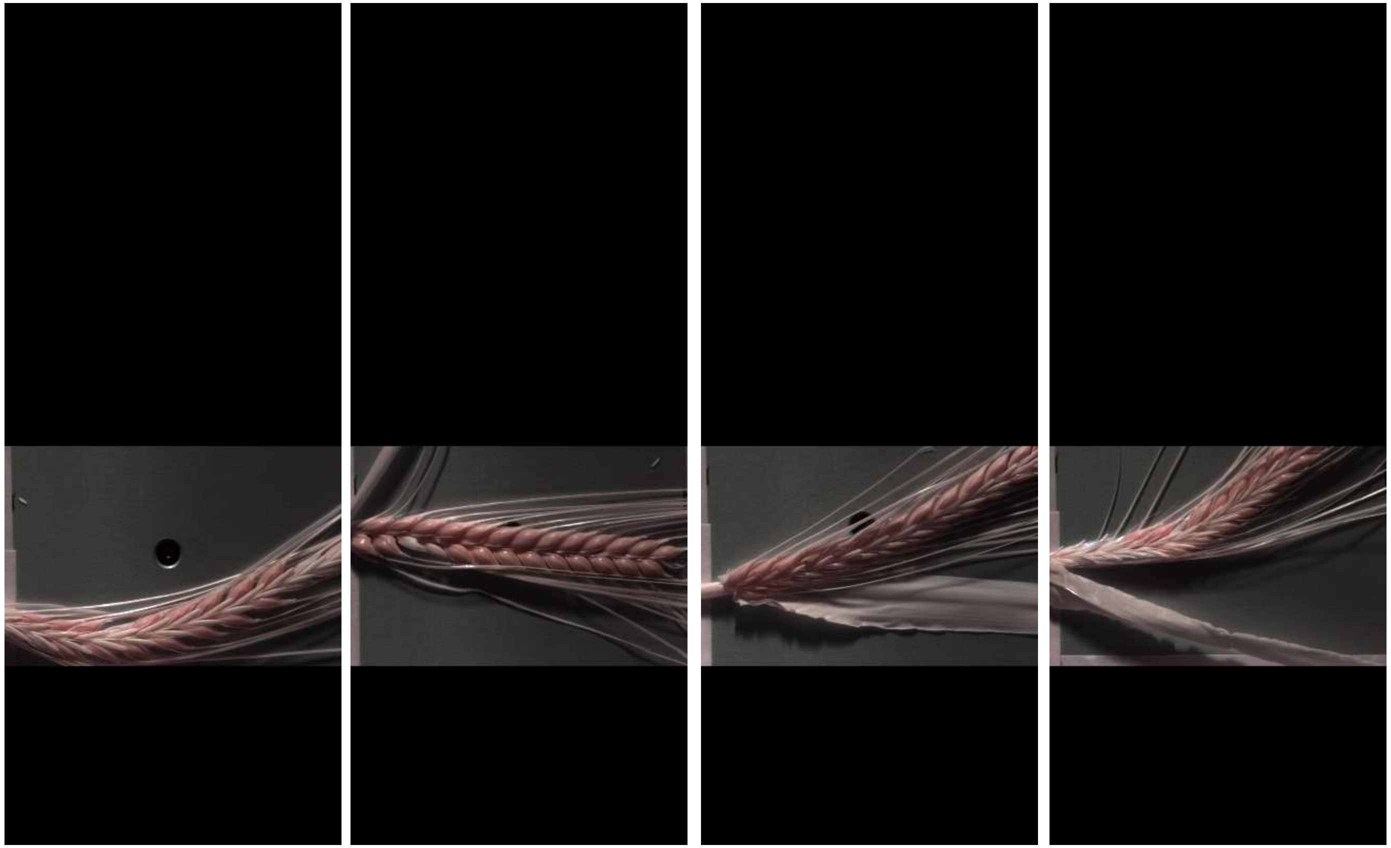






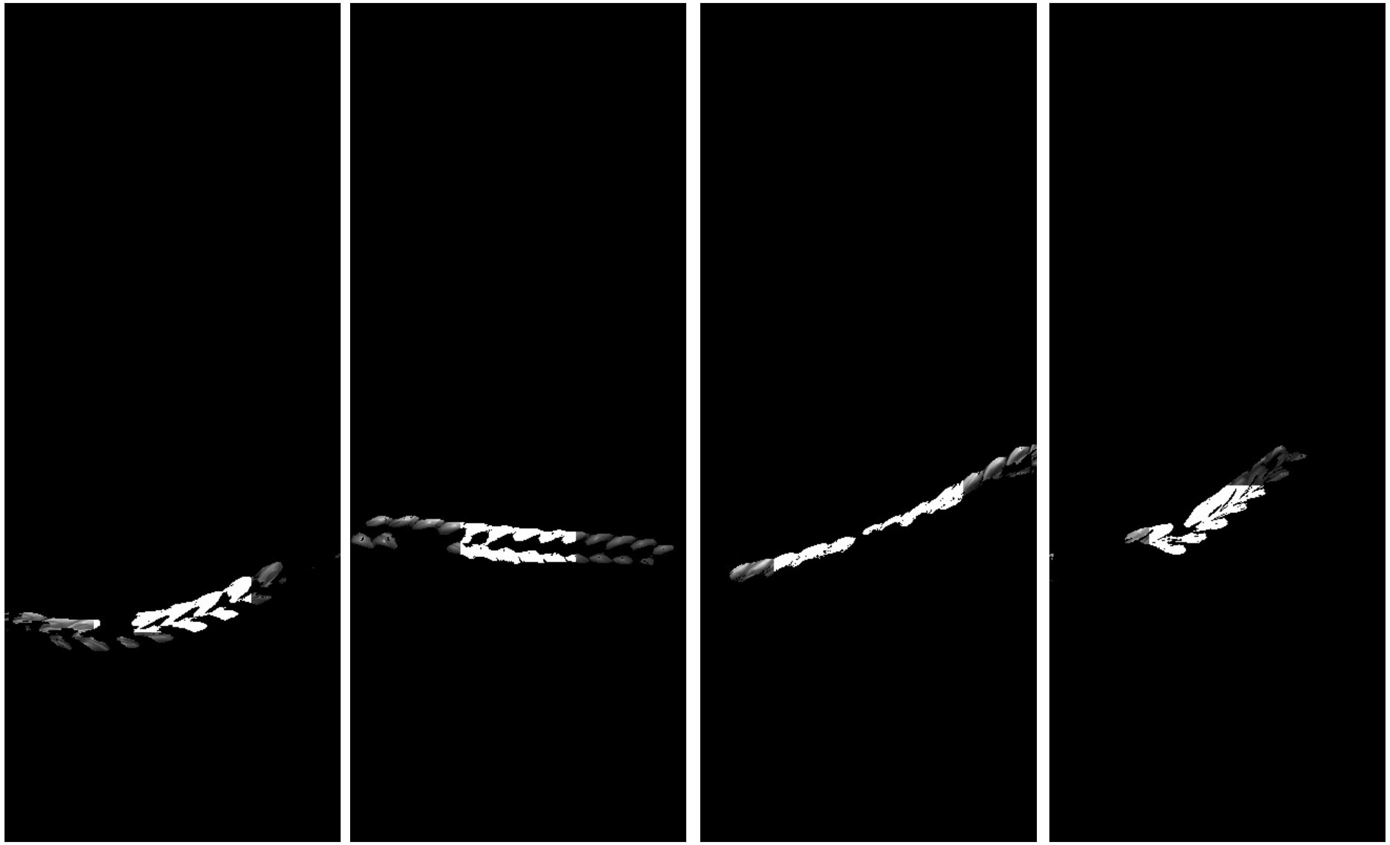
SEGMENTATION

- 97% of images are properly segmented
- Five Images were damaged
- 14 images improperly segmented



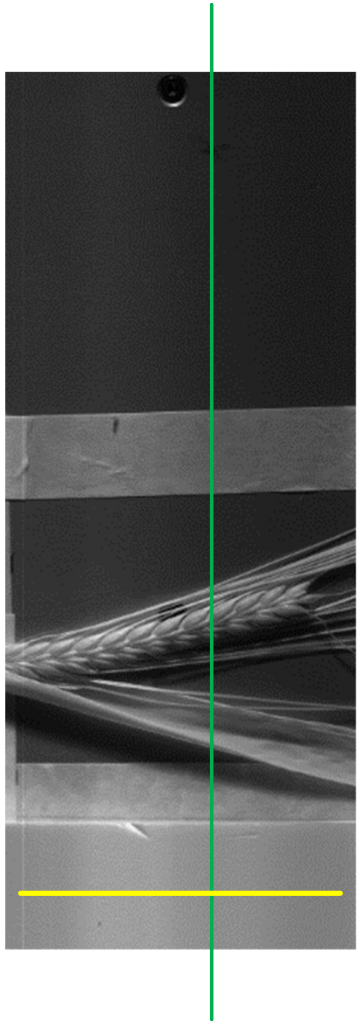
SEGMENTATION

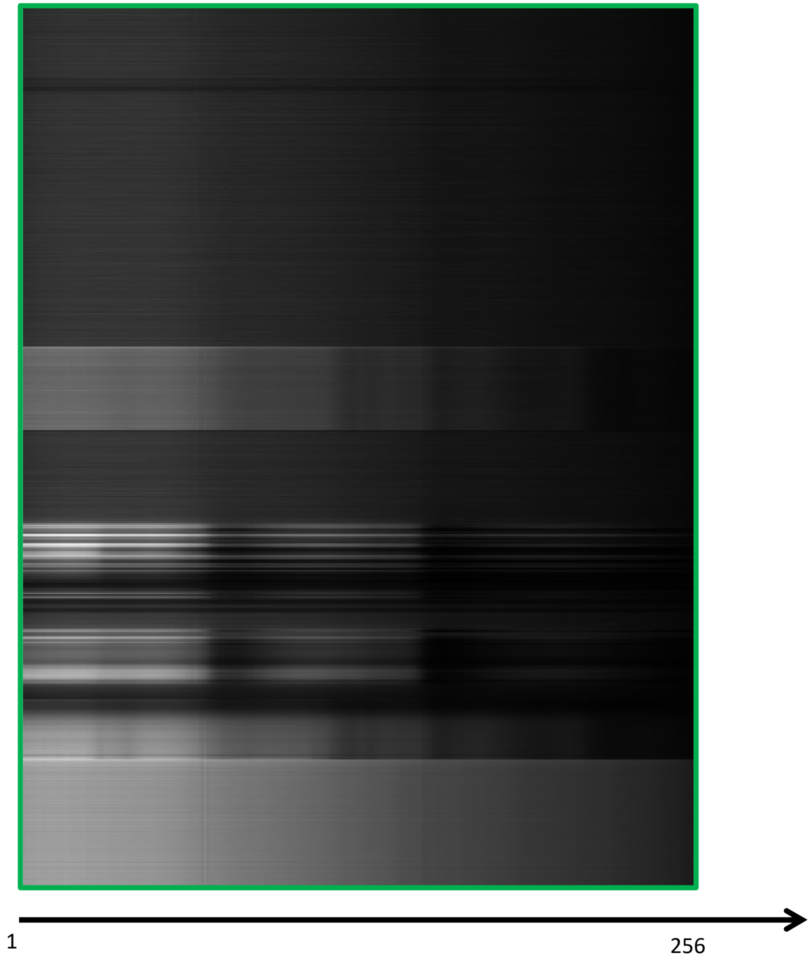
- 97% of images are properly segmented
- Five Images were damaged
- 14 images improperly segmented



DAYLIGHT INTENSITY

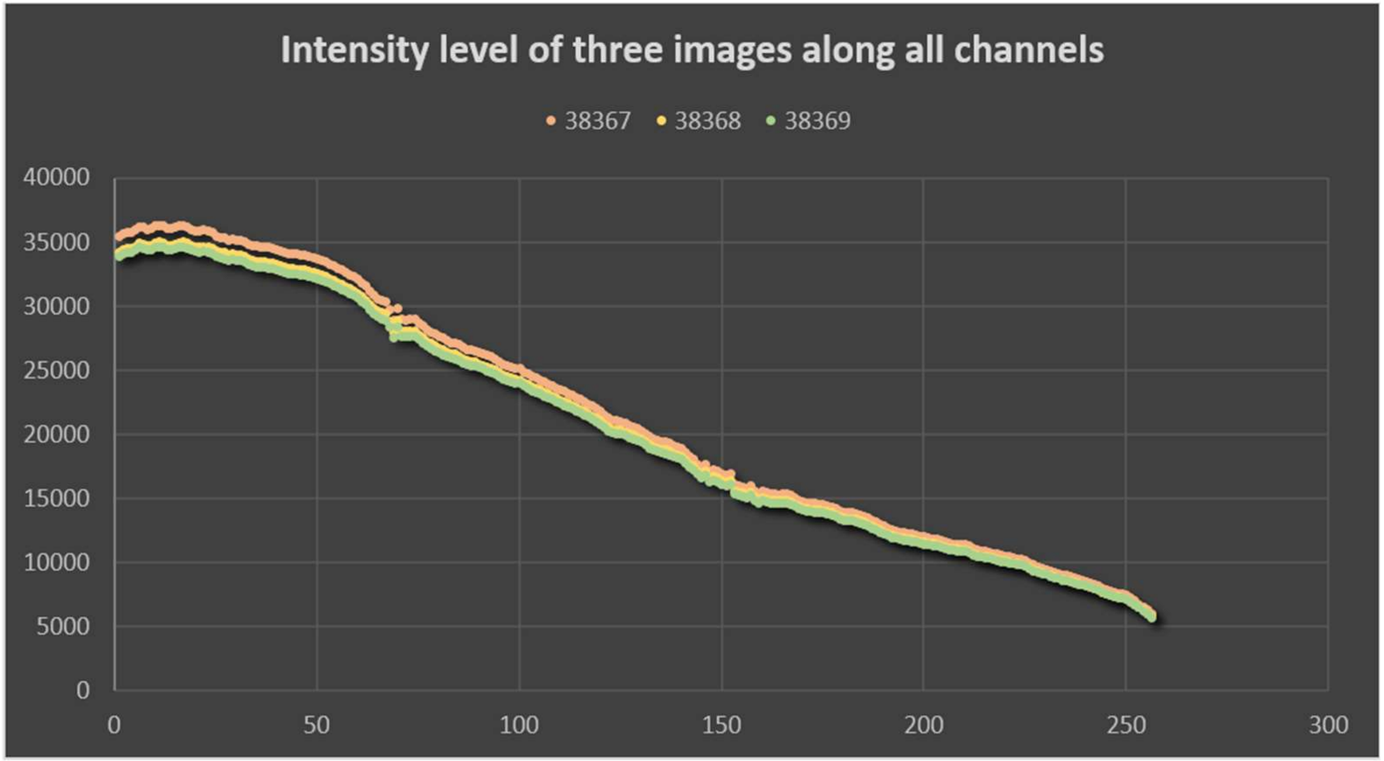
- Mean value of yellow highlighted area is taken as reference for daylight intensity value





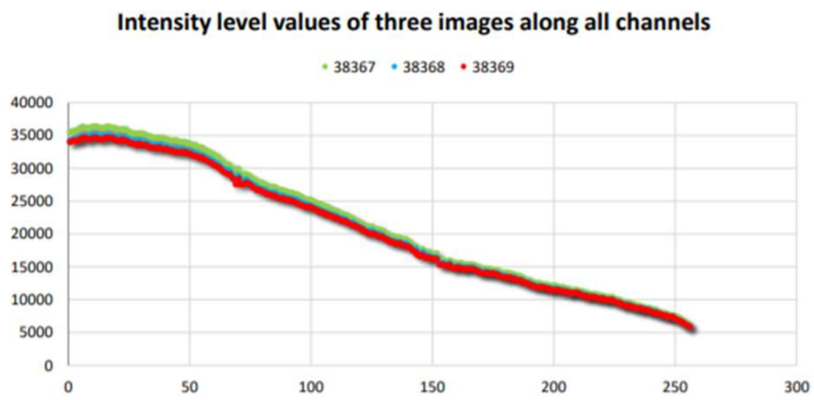
DAYLIGHT INTENSITY

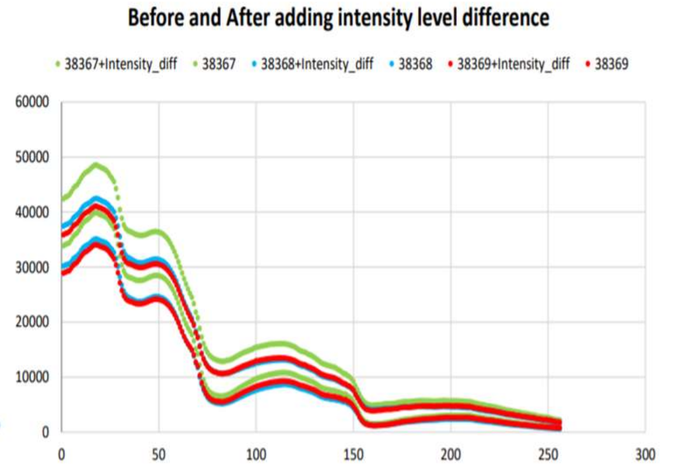
- Mean value of each channel is taken as reference for daylight intensity value

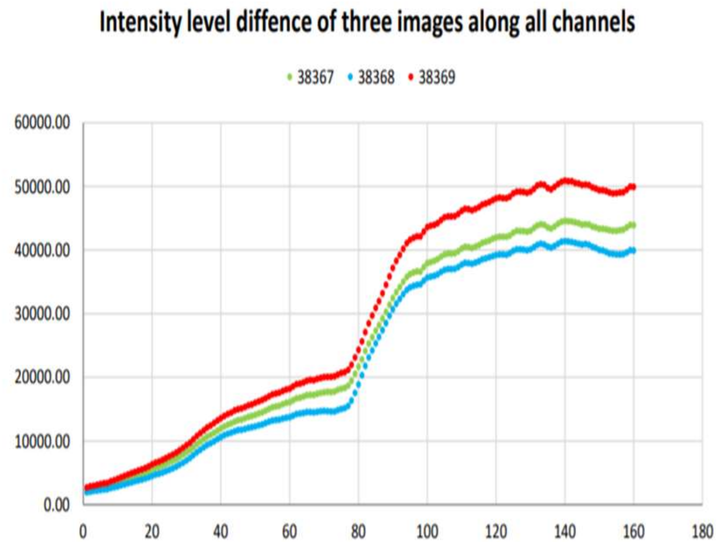


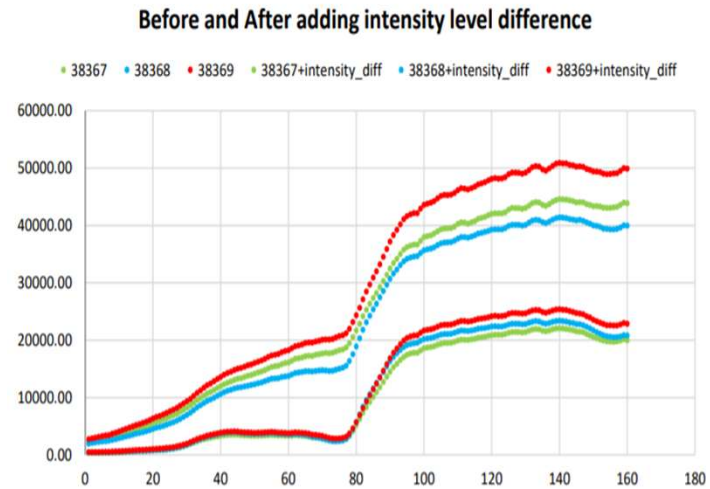
NORMALIZATION

- Difference of intensity level of each image added on respective image though all channels



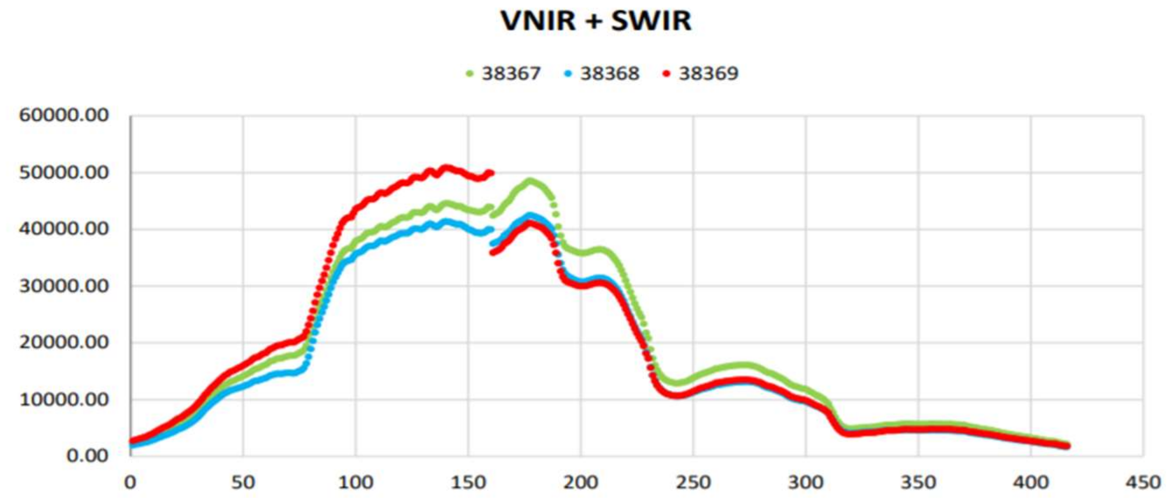


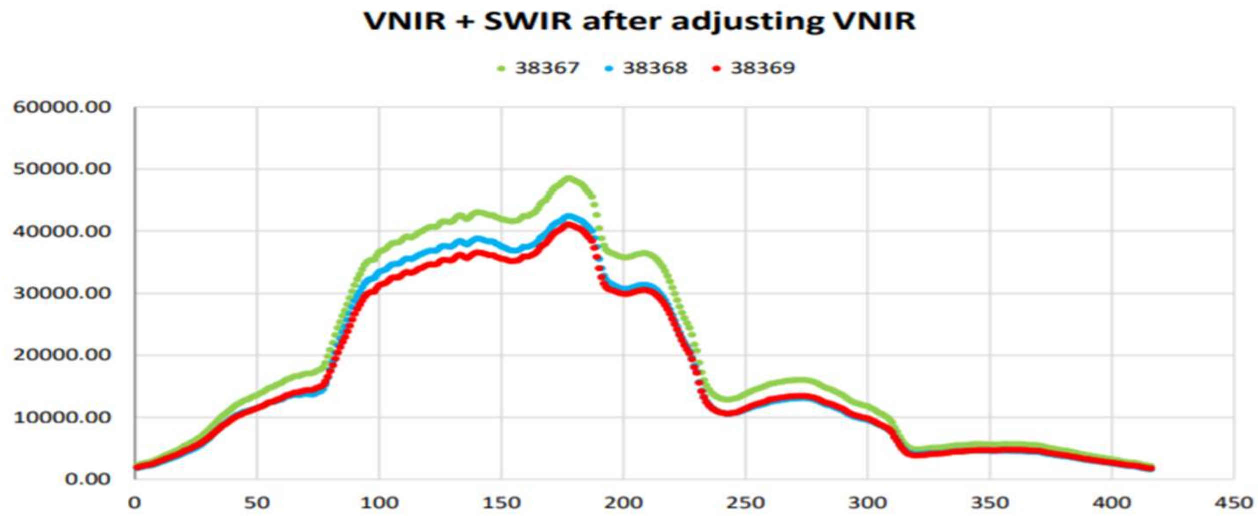




COMBINING IMAGES

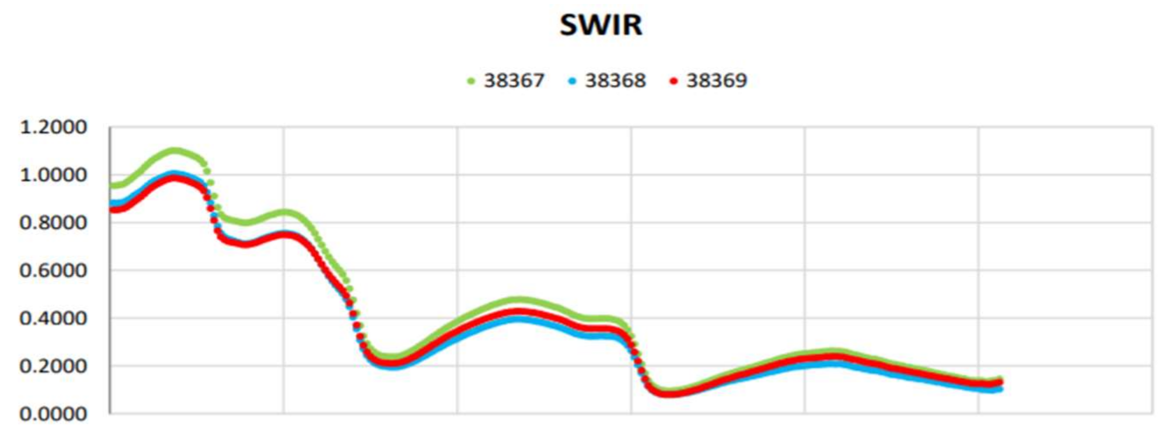
- There is a sharp difference between the channels 160 and 161.
- We keep SWIR intact and adjust VNIR to get a smooth curve by matching the values of 160 and 161 channels.

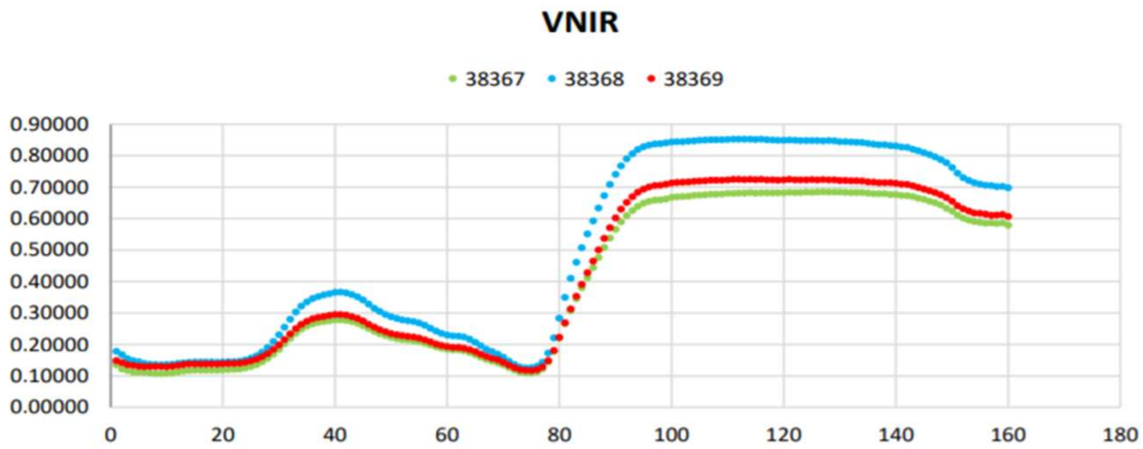




NORMALIZATION

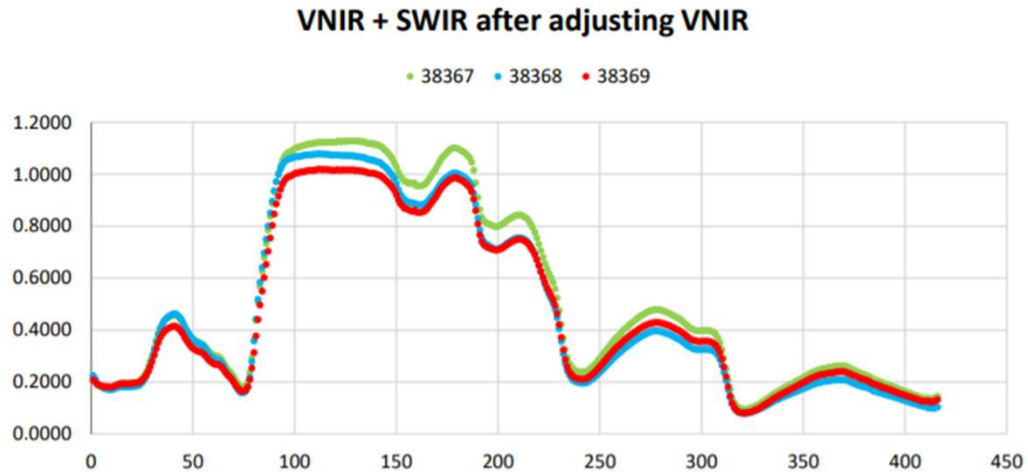
- Second, we normalized the dataset by dividing the daylight intensity values for each image elementwise across all channels.





COMBINING IMAGES

- Similar to previous approach, the images were combined by keeping SWIR intact and adjusting VNIR values.



DATASET

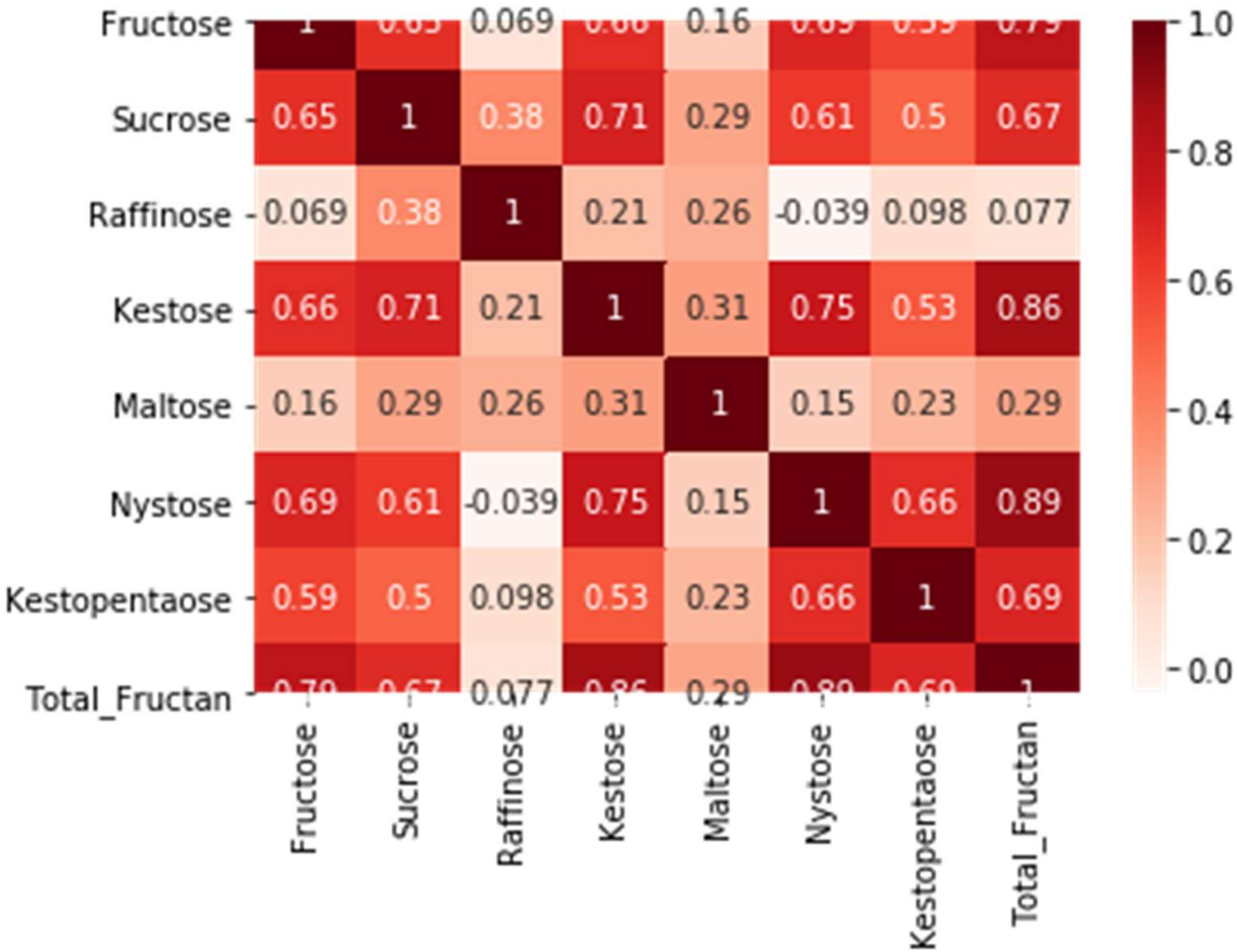
	Img_name	Band_1	Band_2	Band_3	Band_4	...	Band_413	Band_414	Band_415	Band_416
0	38367	2272.05	2419.64	2530.86	2612.02	...	2340.47	2268.82	2193.45	2151.58
1	38368	1822.57	1917.46	2003.75	2079.78	...	1804.88	1721.17	1650.61	1630.31
2	38369	1931.79	2051.13	2138.02	2228.5	...	1931.9	1871.97	1819.16	1791.44
3	38370	2080.95	2201.4	2295.21	2375.41	...	2479.15	2409.34	2317.89	2264.44
4	38371	1931.68	2033.02	2117.09	2179.04	...	2805.46	2697.94	2626.43	2521.02
...	...	...	...	...	...	...	...	...	...	...
455	38866	2126.79	2285.02	2377.65	2469.73	...	2146.35	2030.96	1953.62	1904.95
456	38867	2109.3	2248.25	2359.03	2449.73	...	1968.03	1907.35	1815.73	1779.05
457	38868	2291.33	2430.71	2570.37	2662.76	...	2472.47	2387.72	2280.07	2238.45
458	38869	2077.55	2176.74	2277.05	2366.14	...	2053.77	1940.01	1905.01	1823.89
459	38870	1811.26	1924.09	2030.43	2103.25	...	1278.11	1218.63	1211.74	1186.9

DATASET

	Img_name	Band_1	Band_2	Band_3	Band_4	...	Band_413	Band_414	Band_415	Band_416
0	38367	0.221308	0.199913	0.192579	0.183369	...	0.135571	0.138096	0.139288	0.144988
1	38368	0.223737	0.210706	0.194744	0.18722	...	0.0987096	0.097514	0.101424	0.102856
2	38369	0.207885	0.19952	0.190341	0.186545	...	0.122287	0.125059	0.126839	0.131405
3	38370	0.186489	0.170946	0.15947	0.152563	...	0.130311	0.129185	0.132639	0.133727
4	38371	0.173454	0.158179	0.148503	0.137108	...	0.104787	0.105215	0.106914	0.108773
...	...	...	...	...	...	...	...	...	...	...
455	38866	0.21357	0.194173	0.178069	0.171785	...	0.137663	0.137719	0.140751	0.144272
456	38867	0.271764	0.251519	0.242375	0.23356	...	0.131109	0.131851	0.132176	0.136252
457	38868	0.242911	0.221712	0.211098	0.209131	...	0.144037	0.146799	0.148005	0.15427
458	38869	0.274681	0.253926	0.236602	0.23139	...	0.132767	0.131657	0.133418	0.136164
459	38870	0.194395	0.171286	0.160191	0.149746	...	0.157732	0.157595	0.160872	0.163162

DATASET AGAINST SUGAR VALUES

CORRELATION?



SUGAR CONTENTS

Mean value of each channel for all the images build our dataset

MODELS

SPLIT THE DATASET BASED ON GENOTYPE

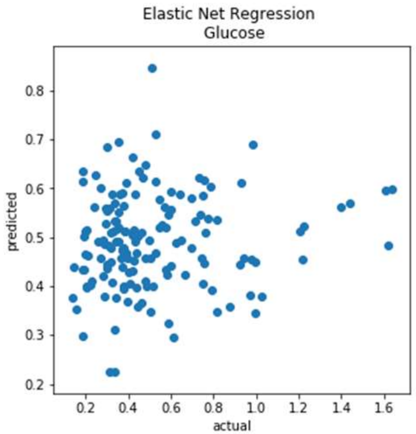
ALL ALGORITHMS TRIED

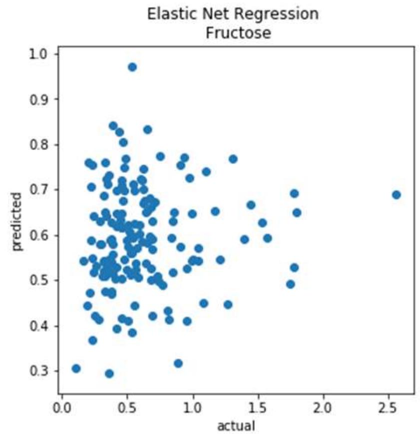
- Ridge Regression
- Lasso
- Elastic Net
- SVR
- Random Forest Regression
- PLS Regression
- Stepwise OLS regression
- Deep Neural Networks

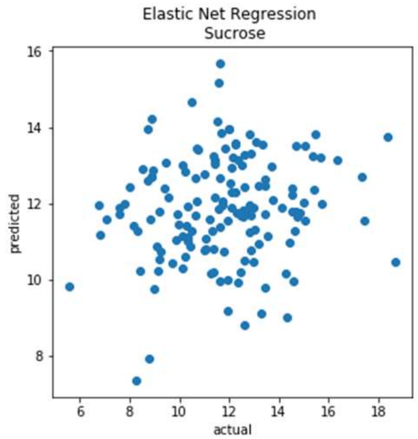
RIDGE REGRESSION

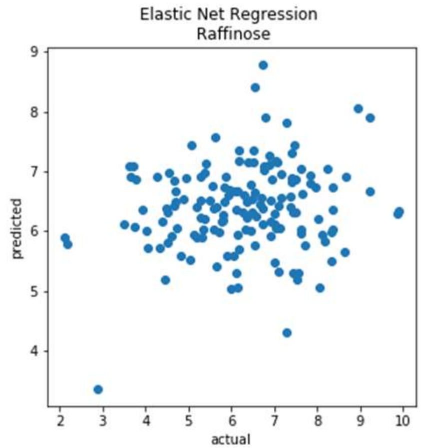
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.511074901	-3.88061	0.442623544
Fructose	0.626878935	-3.58233	0.658037805
Sucrose	4.516844231	-4.40465	30.37255988
Raffinose	2.893977891	-5.39806	13.55439056
Kestose	1.917278963	-7.69109	5.886601115
Maltose	0.578016671	-8.63727	0.5331861
Nystose	1.677731598	-7.77544	4.987306438
Kestopentaose	0.532623374	-8.84136	0.455078583
Total_Fructan	8.090380188	-6.3398	111.5696898

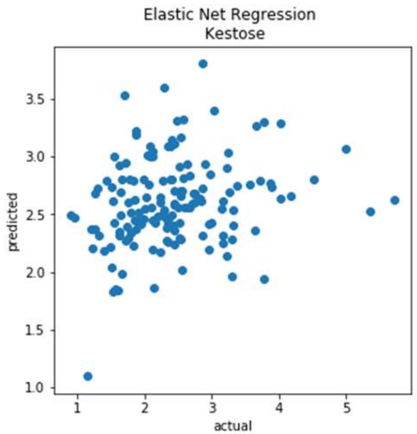
RIDGE REGRESSION

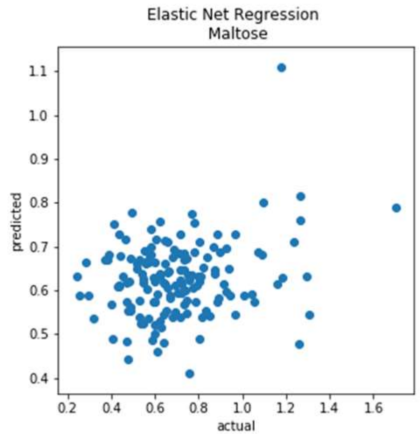


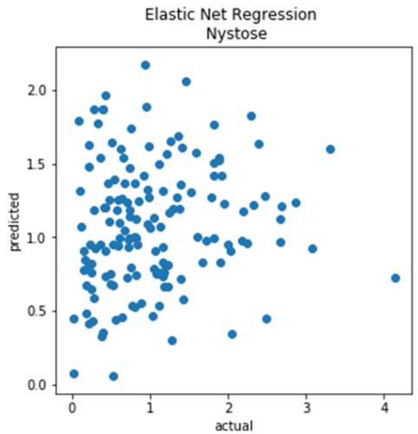


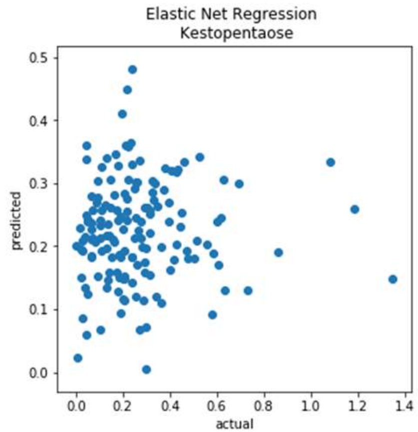








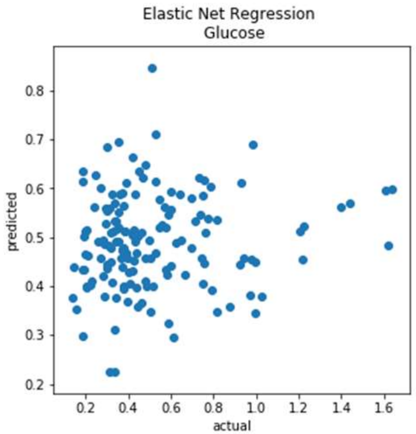


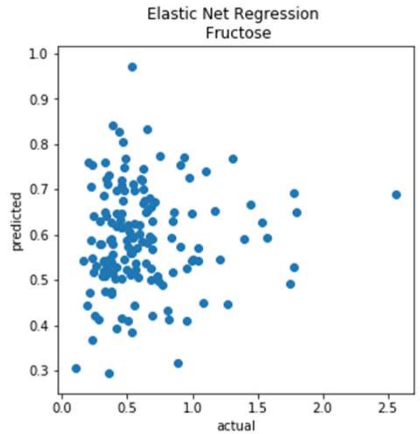


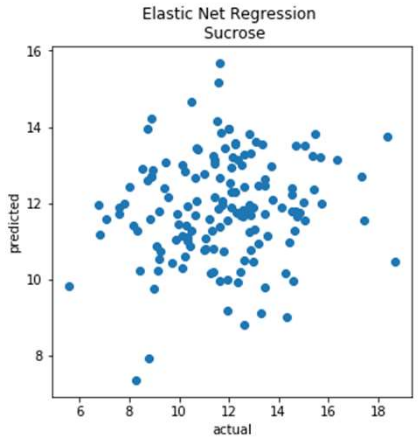
ELASTIC NET

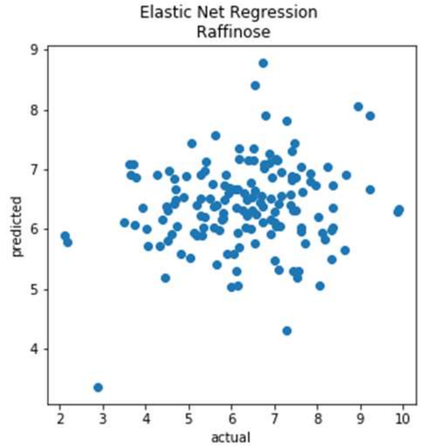
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.225395	-0.03672	0.09402
Fructose	0.272316	-0.04594	0.150201
Sucrose	2.072446	-0.17048	6.577736
Raffinose	1.223988	-0.10549	2.342003
Kestose	0.667266	-0.06532	0.721554
Maltose	0.179537	-0.05939	0.058611
Nystose	0.634522	-0.17936	0.670258
Kestopentaose	0.160034	-0.14356	0.05288
Total_Fructan	3.773329	-0.58032	24.02191

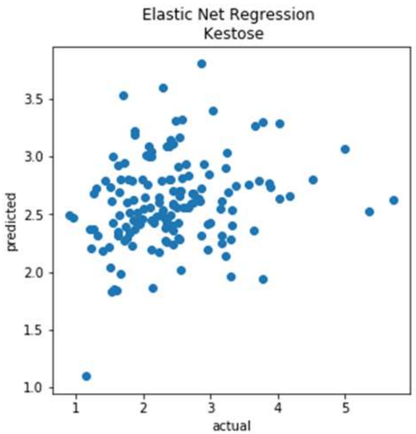
ELASTIC NET

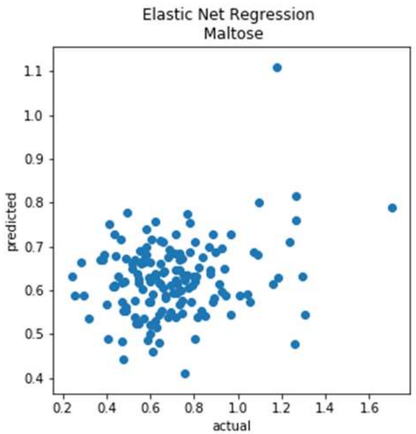

Elastic Net Regression
Glucose
predicted
actual

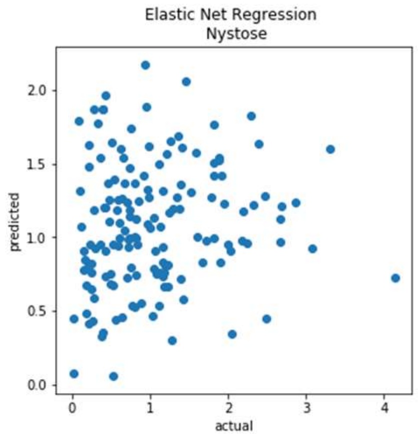

Elastic Net Regression
Fructose
predicted
actual

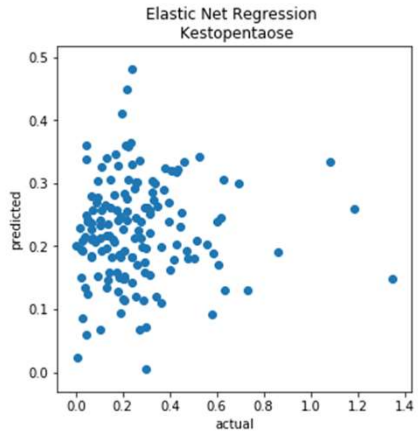

Elastic Net Regression
Sucrose
predicted
actual


Elastic Net Regression
Raffinose
predicted
actual


Elastic Net Regression
Kestose
predicted
actual


Elastic Net Regression
Maltose
predicted
actual

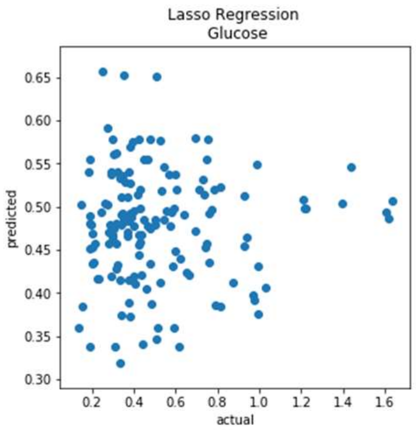

Elastic Net Regression
Nystose
predicted
actual

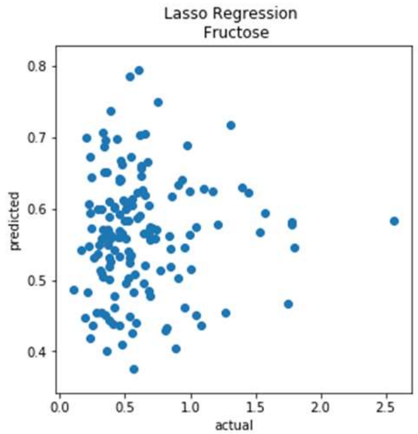

Elastic Net Regression
Kestopentaose
predicted
actual

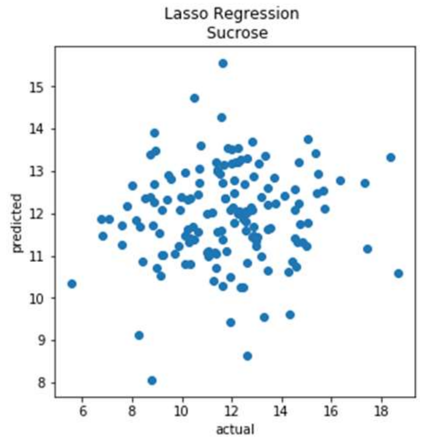
LASSO

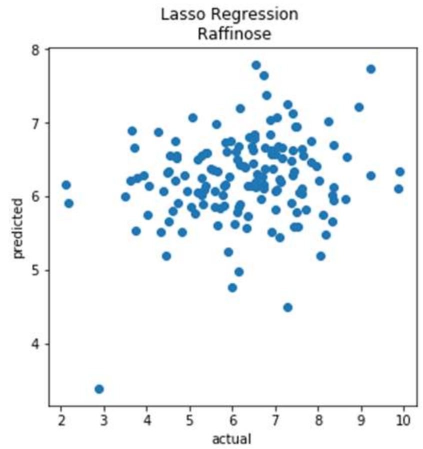
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.22319	-0.05337	0.09553
Fructose	0.264749	-0.05339	0.15127
Sucrose	2.049952	-0.15177	6.472586
Raffinose	1.161771	-0.00836	2.136232
Kestose	0.662869	-0.04797	0.709804
Maltose	0.17603	-0.04968	0.058074
Nystose	0.612399	-0.05749	0.600999
Kestopentaose	0.158774	-0.09359	0.050569
Total_Fructan	3.405795	-0.30661	19.86133

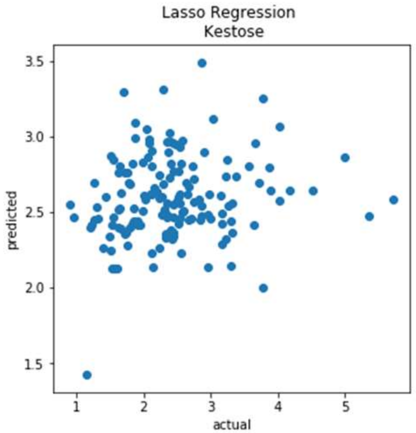
LASSO

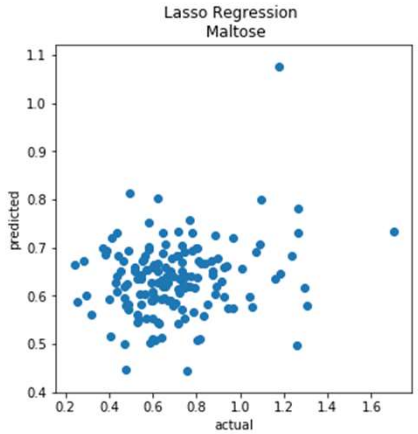

Lasso Regression
Glucose
predicted
actual

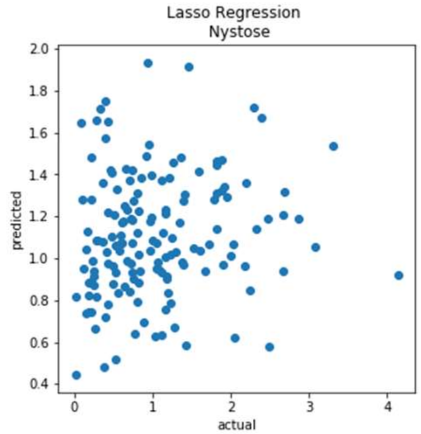

Lasso Regression
Fructose
predicted
actual

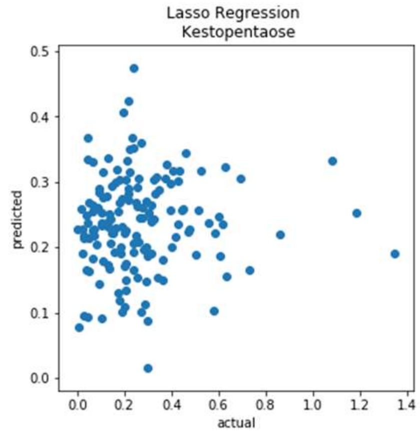

Lasso Regression
Sucrose
predicted
actual


Lasso Regression
Raffinose
predicted
actual

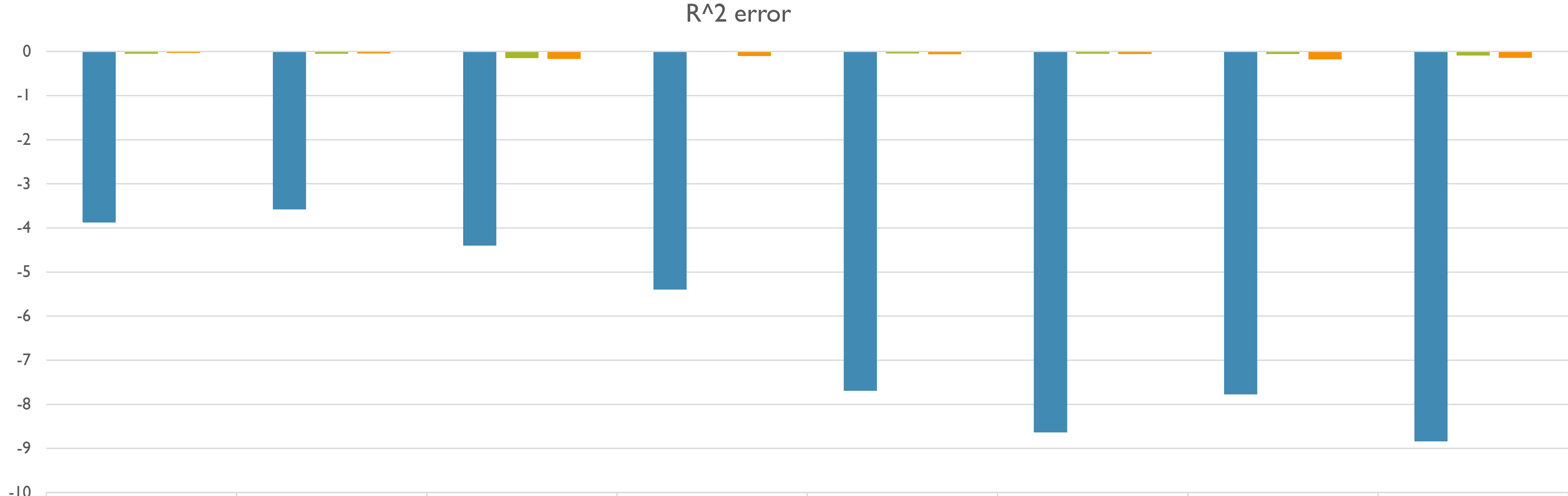

Lasso Regression
Kestose
predicted
actual


Lasso Regression
Maltose
predicted
actual


Lasso Regression
Nystose
predicted
actual


Lasso Regression
Kestopentaose
predicted
actual

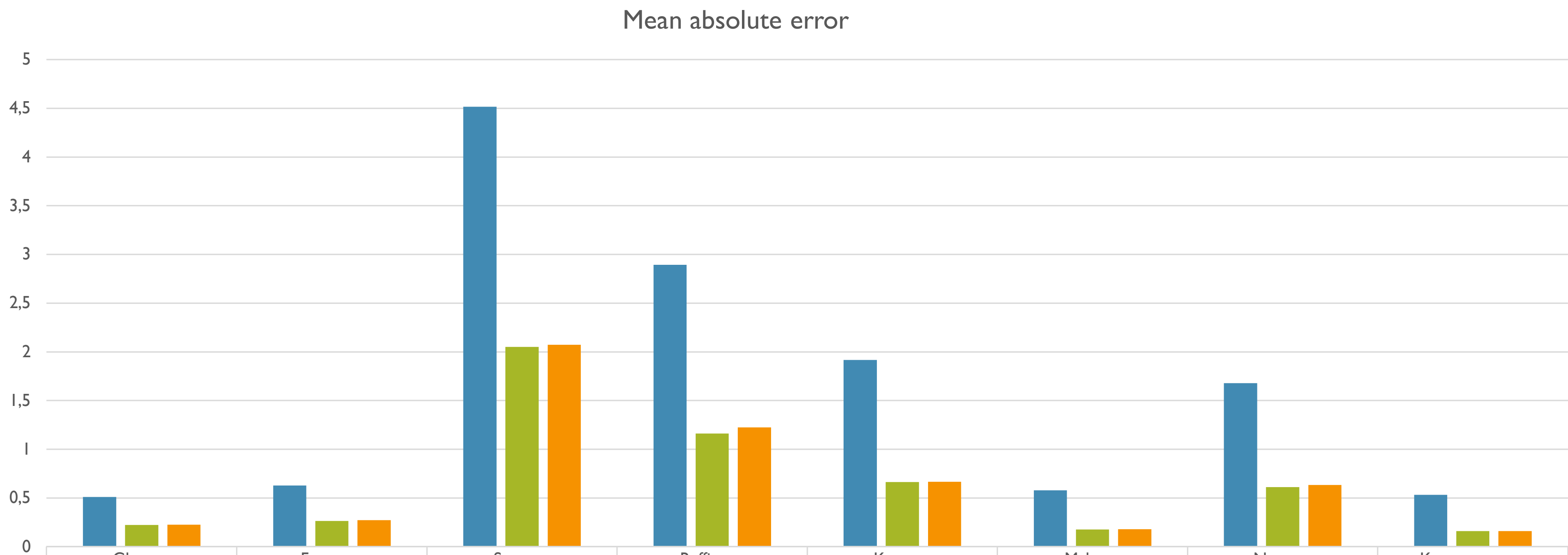
R^2 ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
Ridge Regression	-3,88061	-3,58233	-4,40465	-5,39806	-7,69109	-8,63727	-7,77544	-8,84136
Lasso	-0,05337	-0,05339	-0,15177	-0,00836	-0,04797	-0,04968	-0,05749	-0,09359
Elastic Net	-0,03672	-0,04594	-0,17048	-0,10549	-0,06532	-0,05939	-0,17936	-0,14356

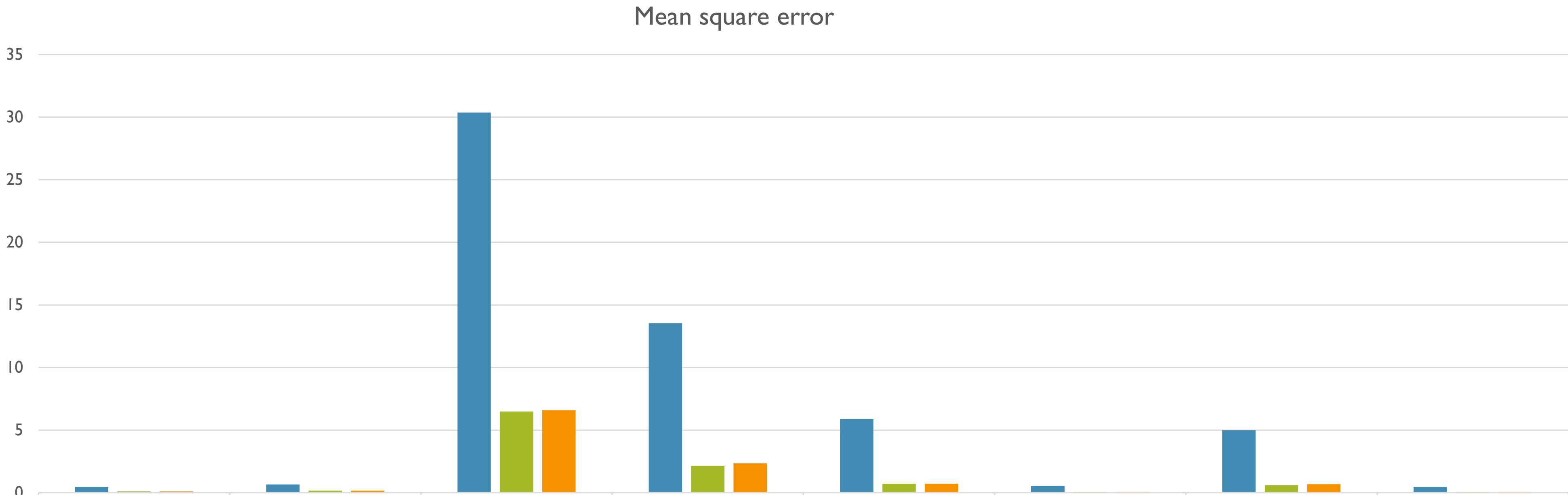
■ Ridge Regression ■ Lasso ■ Elastic Net

MEAN ABSOLUTE ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
Ridge Regression	0,511074901	0,626878935	4,516844231	2,893977891	1,917278963	0,578016671	1,677731598	0,532623374
Lasso	0,22319	0,264749	2,049952	1,161771	0,662869	0,17603	0,612399	0,158774
Elastic Net	0,225395	0,272316	2,072446	1,223988	0,667266	0,179537	0,634522	0,160034

MEAN SQUARE ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
Ridge Regression	0,442623544	0,658037805	30,37255988	13,55439056	5,886601115	0,5331861	4,987306438	0,455078583
Lasso	0,09553	0,15127	6,472586	2,136232	0,709804	0,058074	0,600999	0,050569
Elastic Net	0,09402	0,150201	6,577736	2,342003	0,721554	0,058611	0,670258	0,05288

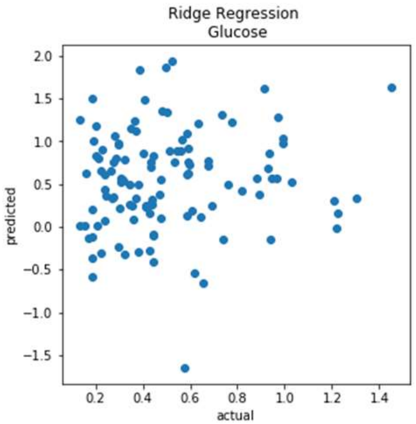
MODELS

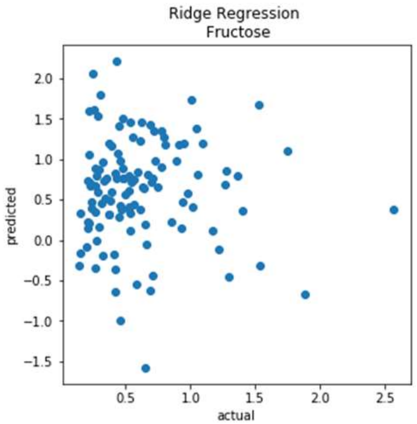
BASED ON RANDOMLY SPLITTING DATASET INTO TRAIN AND TEST (TAKING MEANS OF 10 REPLICATES)

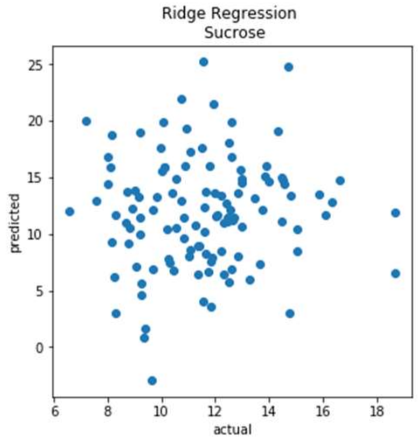
RIDGE REGRESSION

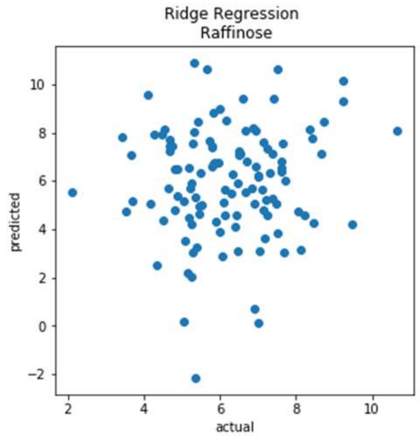
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.441288	-3.0415	0.324161
Fructose	0.579718	-3.48335	0.562616
Sucrose	3.550131	-2.64905	20.32791
Raffinose	2.058246	-2.58193	6.888539
Kestose	1.366227	-2.94252	2.942811
Maltose	0.392421	-2.48984	0.25688
Nystose	1.366642	-2.97526	2.933128
Kestopentaose	0.391586	-3.02246	0.249752
Total_Fructan	6.465502	-2.93823	67.35755

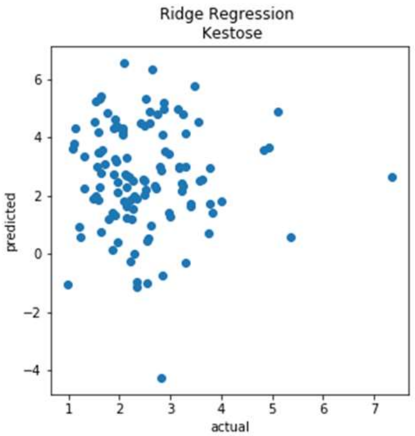
RIDGE REGRESSION

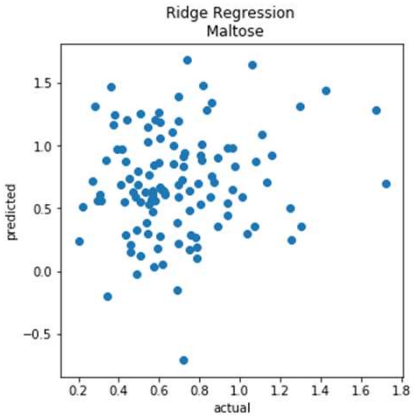


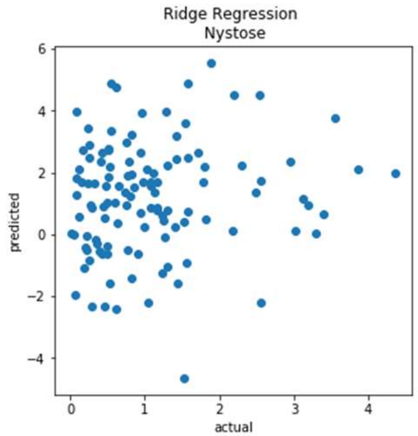


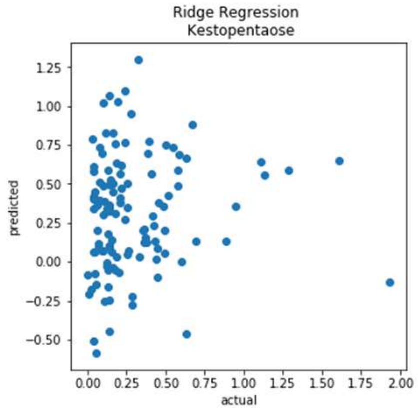








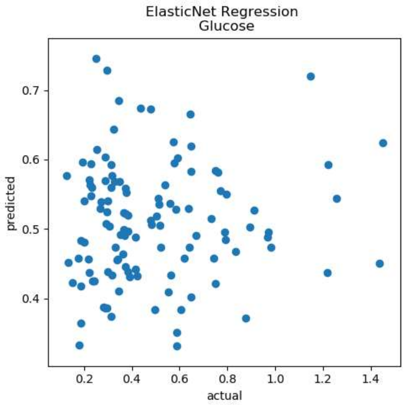


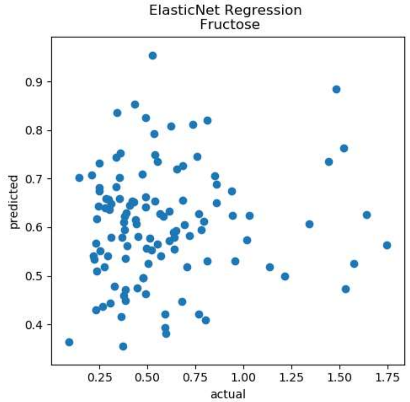


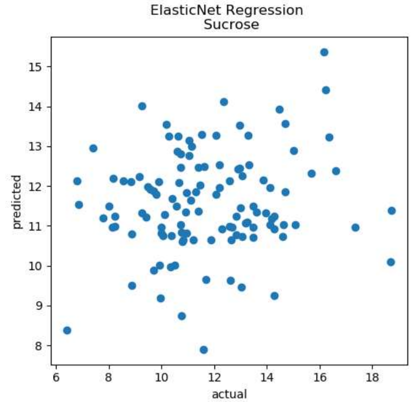
ELASTIC NET

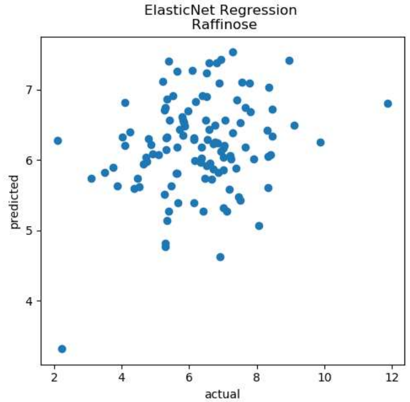
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.221439	0.056633	0.083809
Fructose	0.277698	0.04544	0.141775
Sucrose	1.960413	-0.08924	6.061645
Raffinose	1.106896	-0.01355	1.974434
Kestose	0.668826	0.007923	0.783563
Maltose	0.185754	0.132907	0.064505
Nystose	0.706149	0.02043	0.789722
Kestopentaose	0.196809	0.025075	0.073493
Total_Fructan	3.525995	-0.14442	21.76777

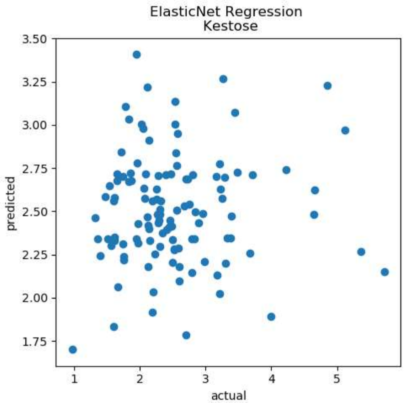
ELASTIC NET

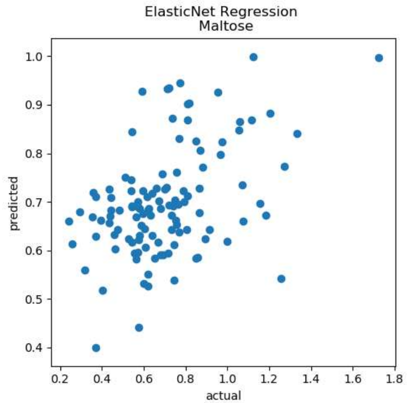

ElasticNet Regression
Glucose
predicted
actual

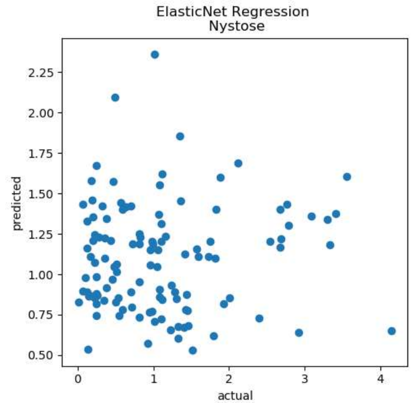

ElasticNet Regression
Fructose
predicted
actual

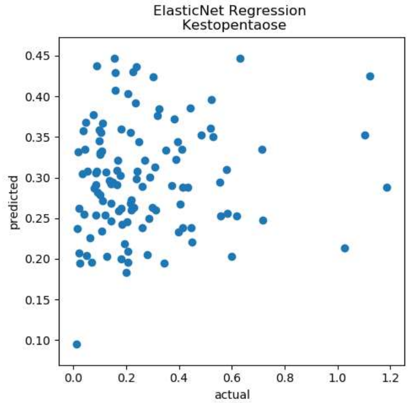

ElasticNet Regression
Sucrose
predicted
actual


ElasticNet Regression
Raffinose
predicted
actual


ElasticNet Regression
Kestose
predicted
actual


ElasticNet Regression
Maltose
predicted
actual

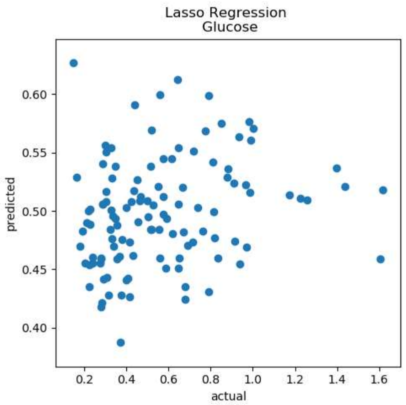

ElasticNet Regression
Nystose
predicted
actual

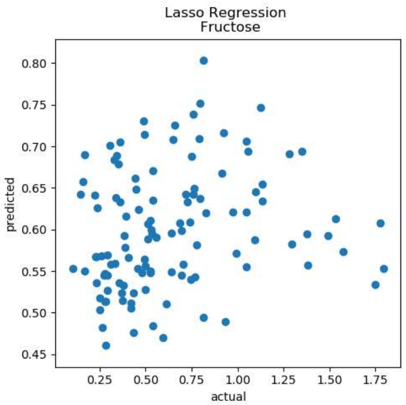

ElasticNet Regression
Kestopentaose
predicted
actual

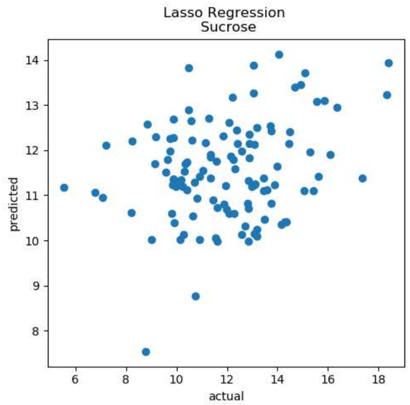
LASSO

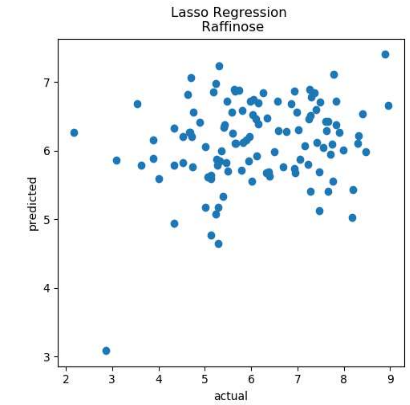
Sugar Contents	Mean absolute error	R^2 error	Mean square error
Glucose	0.220978	-0.01924	0.080412
Fructose	0.269796	-0.01317	0.126867
Sucrose	1.882335	0.014208	5.482845
Raffinose	1.104471	0.026226	1.918576
Kestose	0.652554	0.025089	0.73997
Maltose	0.187364	0.119368	0.070103
Nystose	0.678549	0.017393	0.732928
Kestopentaose	0.183541	-0.00063	0.058472
Total_Fructan	3.137438	-0.01234	17.19574

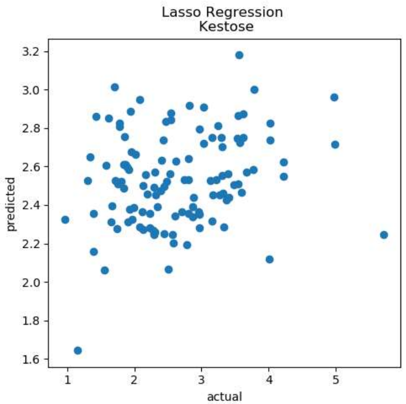
LASSO

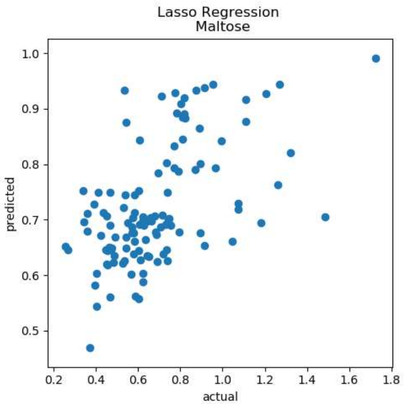

Lasso Regression
Glucose
predicted
actual

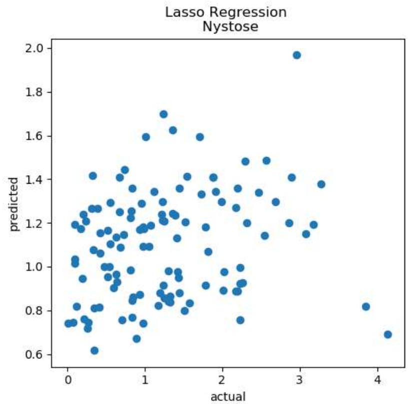

Lasso Regression
Fructose
predicted
actual

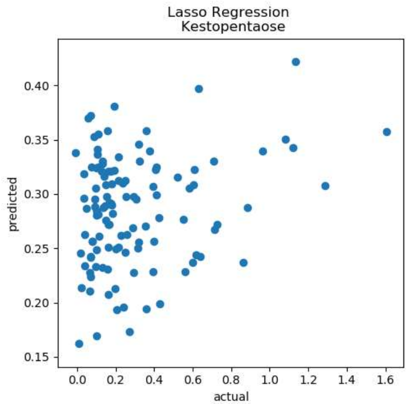

Lasso Regression
Sucrose
predicted
actual


Lasso Regression
Raffinose
predicted
actual

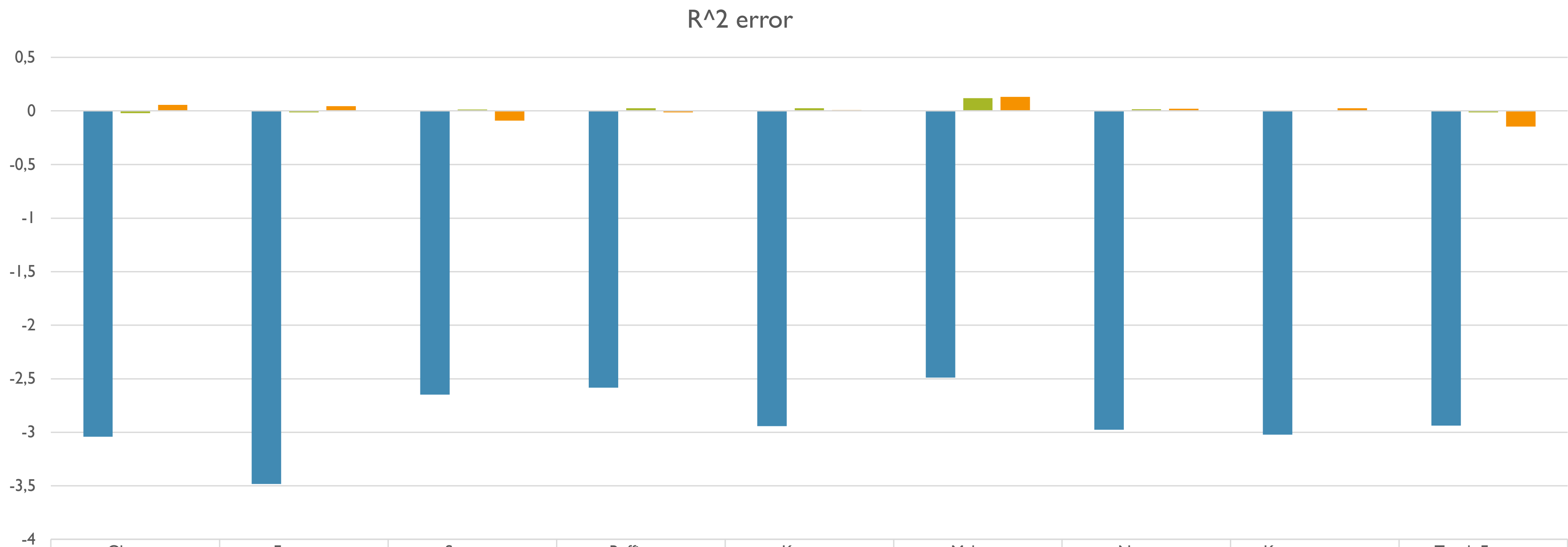

Lasso Regression
Kestose
predicted
actual


Lasso Regression
Maltose
predicted
actual


Lasso Regression
Nystose
predicted
actual


Lasso Regression
Kestopentaose
predicted
actual

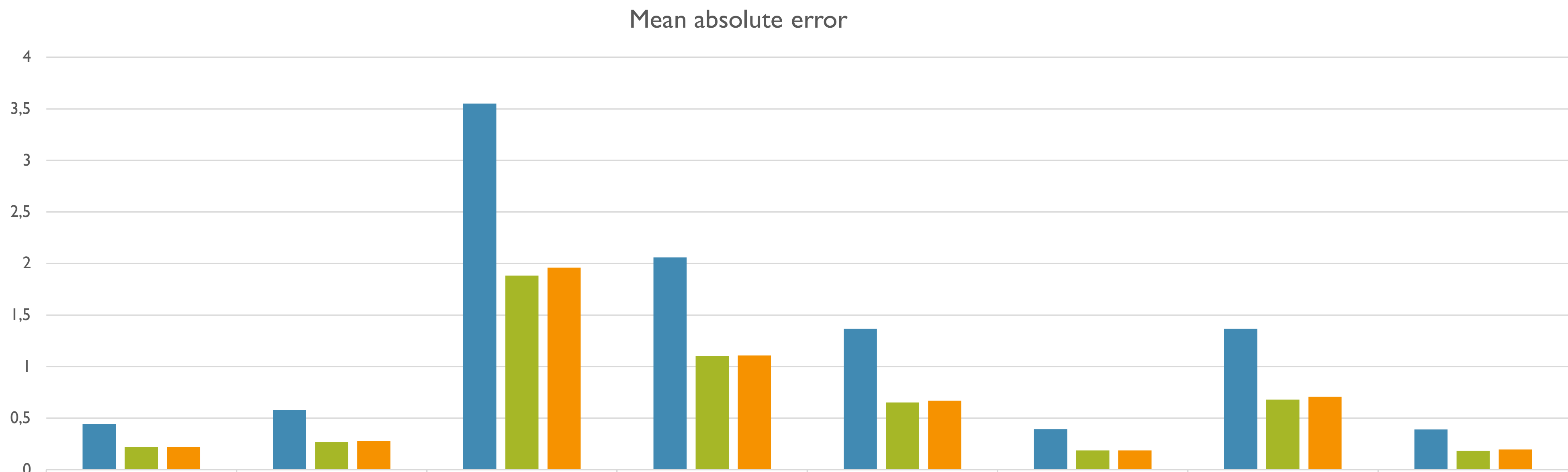
R^2 ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose	Total_Fructan
Ridge Regression	-3,0415	-3,48335	-2,64905	-2,58193	-2,94252	-2,48984	-2,97526	-3,02246	-2,93823
Lasso	-0,01924	-0,01317	0,014208	0,026226	0,025089	0,119368	0,017393	-0,00063	-0,01234
Elastic Net	0,056633	0,04544	-0,08924	-0,01355	0,007923	0,132907	0,02043	0,025075	-0,14442

Ridge Regression Lasso Elastic Net

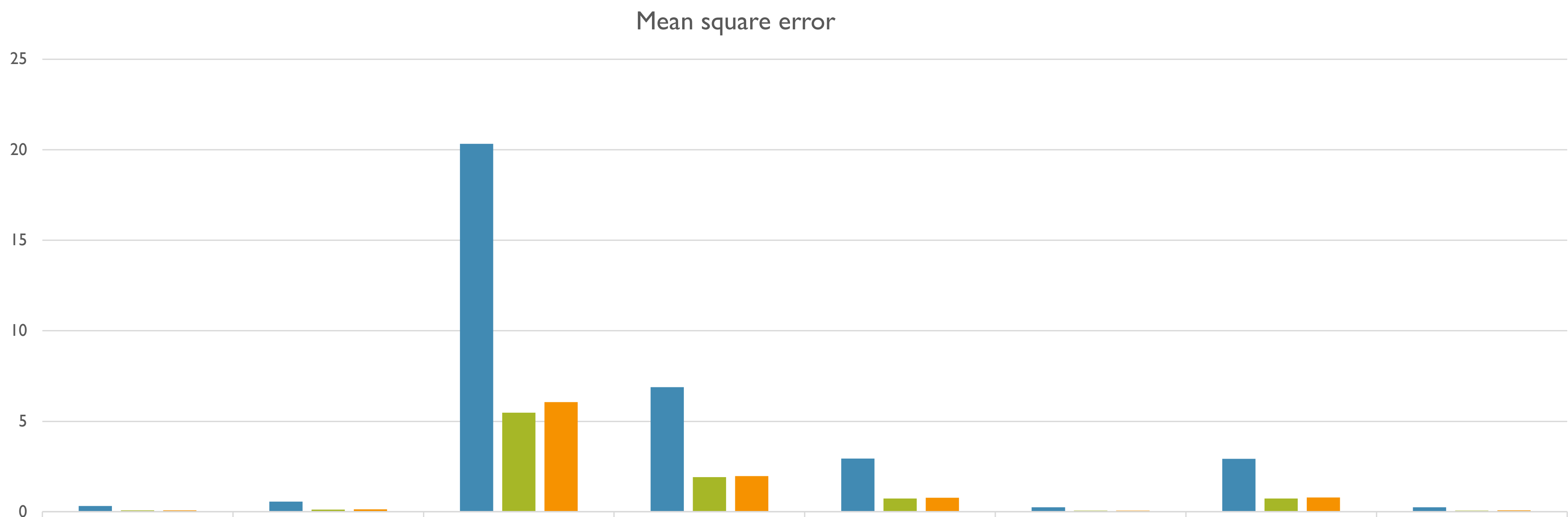
MEAN ABSOLUTE ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
Ridge Regression	0,441288	0,579718	3,550131	2,058246	1,366227	0,392421	1,366642	0,391586
Lasso	0,220978	0,269796	1,882335	1,104471	0,652554	0,187364	0,678549	0,183541
Elastic Net	0,221439	0,277698	1,960413	1,106896	0,668826	0,185754	0,706149	0,196809

Ridge Regression Lasso Elastic Net

MEAN SQUARE ERROR



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
Ridge Regression	0,324161	0,562616	20,32791	6,888539	2,942811	0,25688	2,933128	0,249752
Lasso	0,080412	0,126867	5,482845	1,918576	0,73997	0,070103	0,732928	0,058472
Elastic Net	0,083809	0,141775	6,061645	1,974434	0,783563	0,064505	0,789722	0,073493

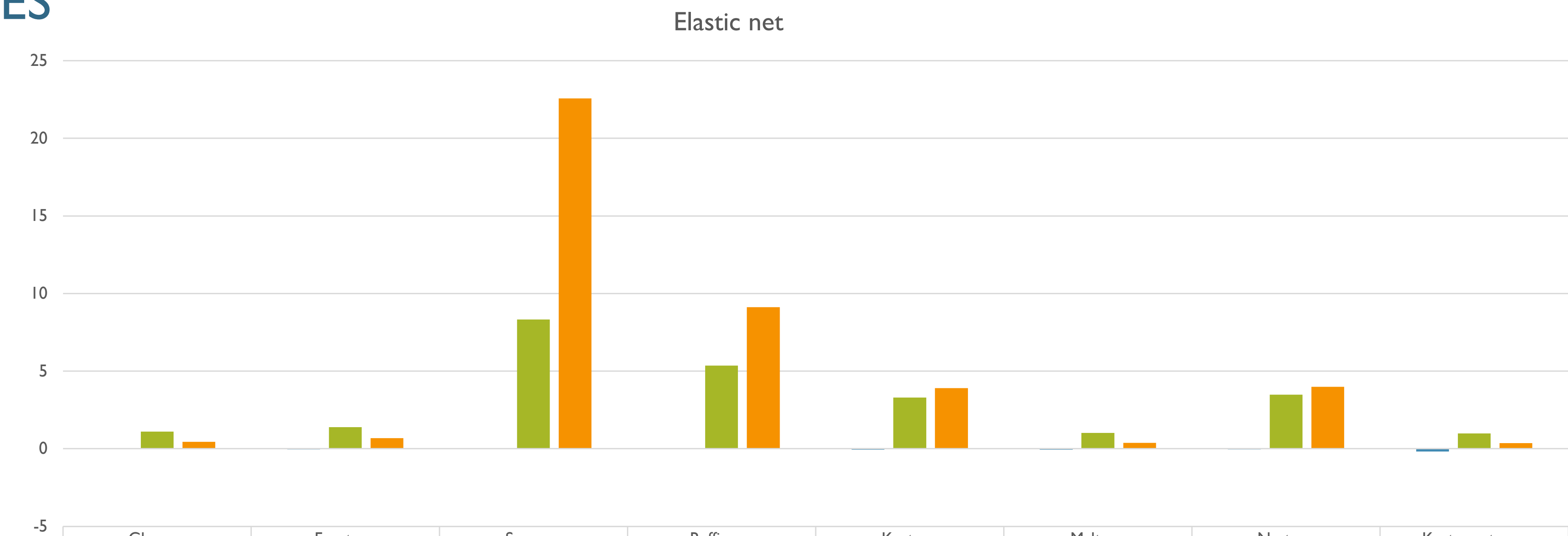
■ Ridge Regression ■ Lasso ■ Elastic Net

REMOVING NOISY IMAGES

- **Taking Sucrose in to consideration**
- **Relative Error > 0.4**

List of 28 images having relative error larger than 0.4			
38371	38571	38629	38805
38390	38583	38699	38806
38405	38588	38727	38824
38473	38589	38728	38828
38475	38608	38750	38858
38530	38622	38791	38861
38568	38627	38793	38865

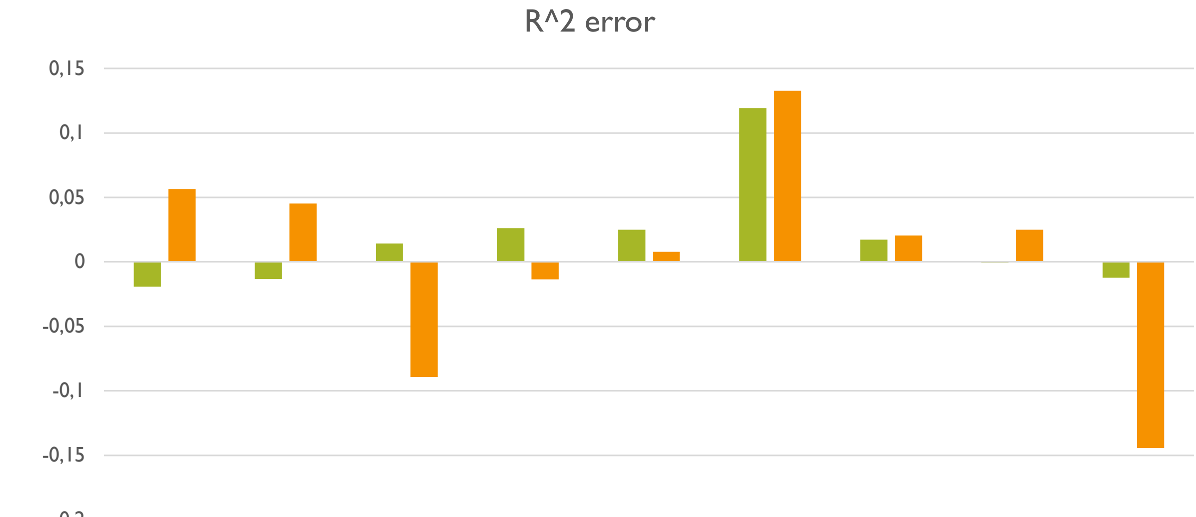
REMOVING NOISY IMAGES



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose
r2_score_error	-0,022733179	-0,040082646	-0,026646006	-0,018498523	-0,063862966	-0,056756575	-0,033618863	-0,176293146
mean_absolute_error	1,111945009	1,390164268	8,321508529	5,365557718	3,29616158	1,026738812	3,482611302	0,991954344
mean_squared_error	0,439913399	0,684157559	22,56793305	9,116560128	3,909442054	0,384637954	3,991911993	0,358894982

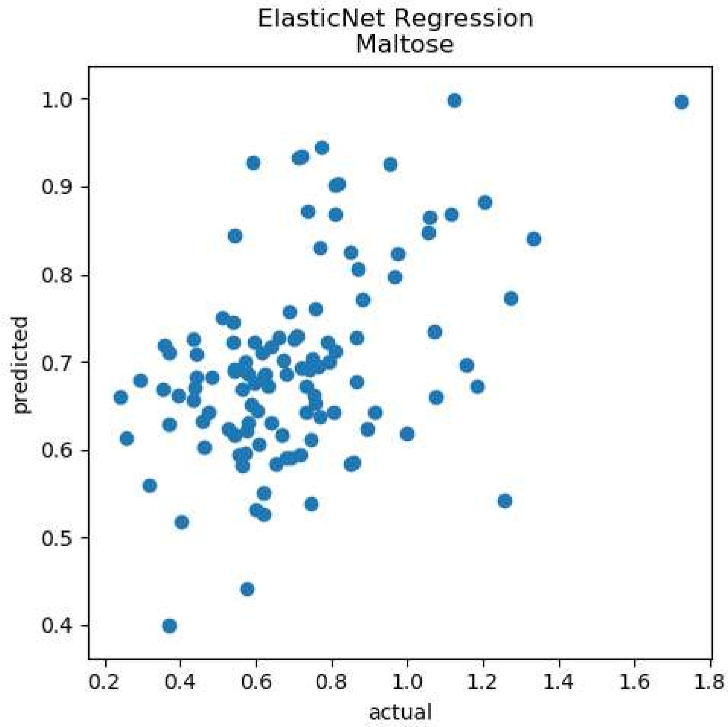
r2_score_error   mean_absolute_error   mean_squared_error

ELASTIC NET



	Glucose	Fructose	Sucrose	Raffinose	Kestose	Maltose	Nystose	Kestopentaose	Total_Fructan
Lasso	-0,01924	-0,01317	0,014208	0,026226	0,025089	0,119368	0,017393	-0,00063	-0,01234
Elastic Net	0,056633	0,04544	-0,08924	-0,01355	0,007923	0,132907	0,02043	0,025075	-0,14442

Lasso Elastic Net



SUMMARY

- The segmentation that was performed on the data given was satisfactory(97% of the images were properly segmented).
- We tried various normalization techniques and performed experiments on numerous regression models.
- No single regression model gave us the intended results.
- Example , Elastic Net performed good for Maltose, although not as expected.

THANK YOU!