tinyML® Foundation

Enabling Ultra-low Power Machine Learning at the Edge

tinyML Summit April 22 - 24, 2024

TINY ML

www.tinyML.org

tinyML driven open-source portable spectrophotometer for rapid and non-destructive quality assessment of fruits and vegetables

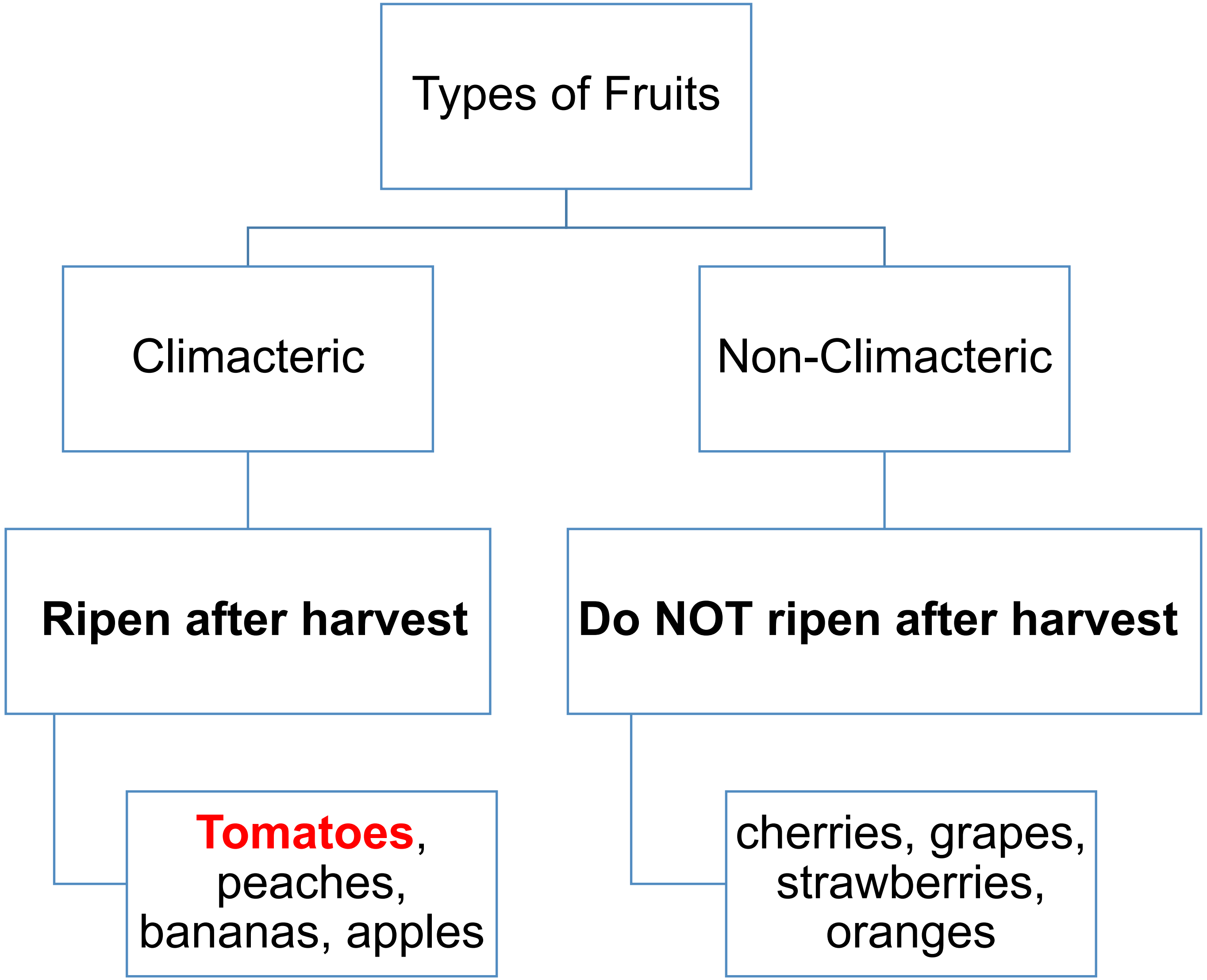
by

ARUN SHARMA
Assistant Professor
Food Engineering Department
National Institute of Food Technology Entrepreneurship and Management (NIFTEM)
(An Institute of National Importance (INI) declared through NIFTEM Act, 2021)
under Ministry of Food Processing Industries (MoFPI), Government of India
Plot No. 97, Sector 56, HSIIDC Industrial Estate, Kundli (131028), Sonepat, Haryana, India
Phone:+91-130-2281239, Mobile:+91-9416656492
Email: arunsharma1712@gmail.com

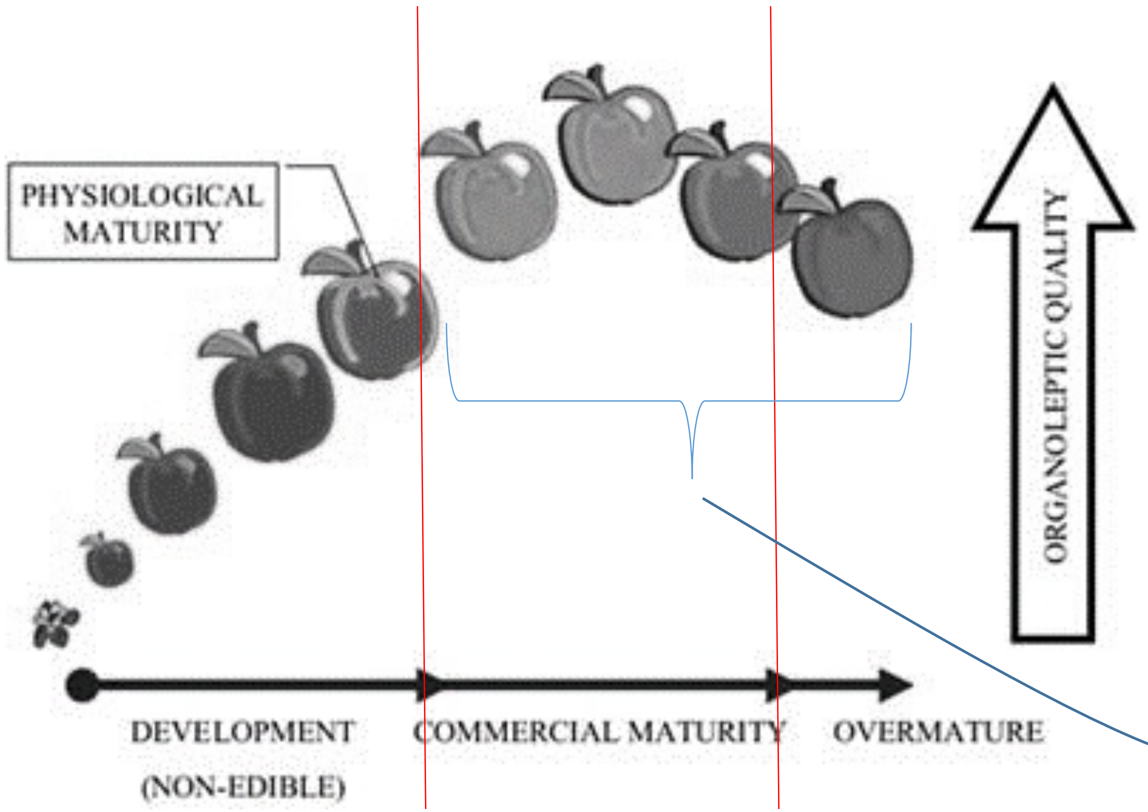
April, 2024

INTRODUCTION
(Domain Specific)

- Often consumers face uncertainty when purchasing fruits and vegetables as outer surface appearance can be deceptive.
- What appears fresh on the surface may not hold true on the inside, leading to customer dissatisfaction.



PHYSIOLOGICAL MATURITY AND HORTICULTURAL MATURITY



- Postharvest ripening
- Changes in physicochemical attributes such as colour, firmness, weight loss, total soluble solids (TSS), acidity, pH etc.

Image Source: https://sites.google.com/site/maturitypostharvest97/maturity

Fruit **ripening** is **genetically programmed** development overlapping with fruit maturity and senescence[1] which can be **affected** by fruit genotype, environment and pre- and post-harvest handling and storage conditions[2,3]

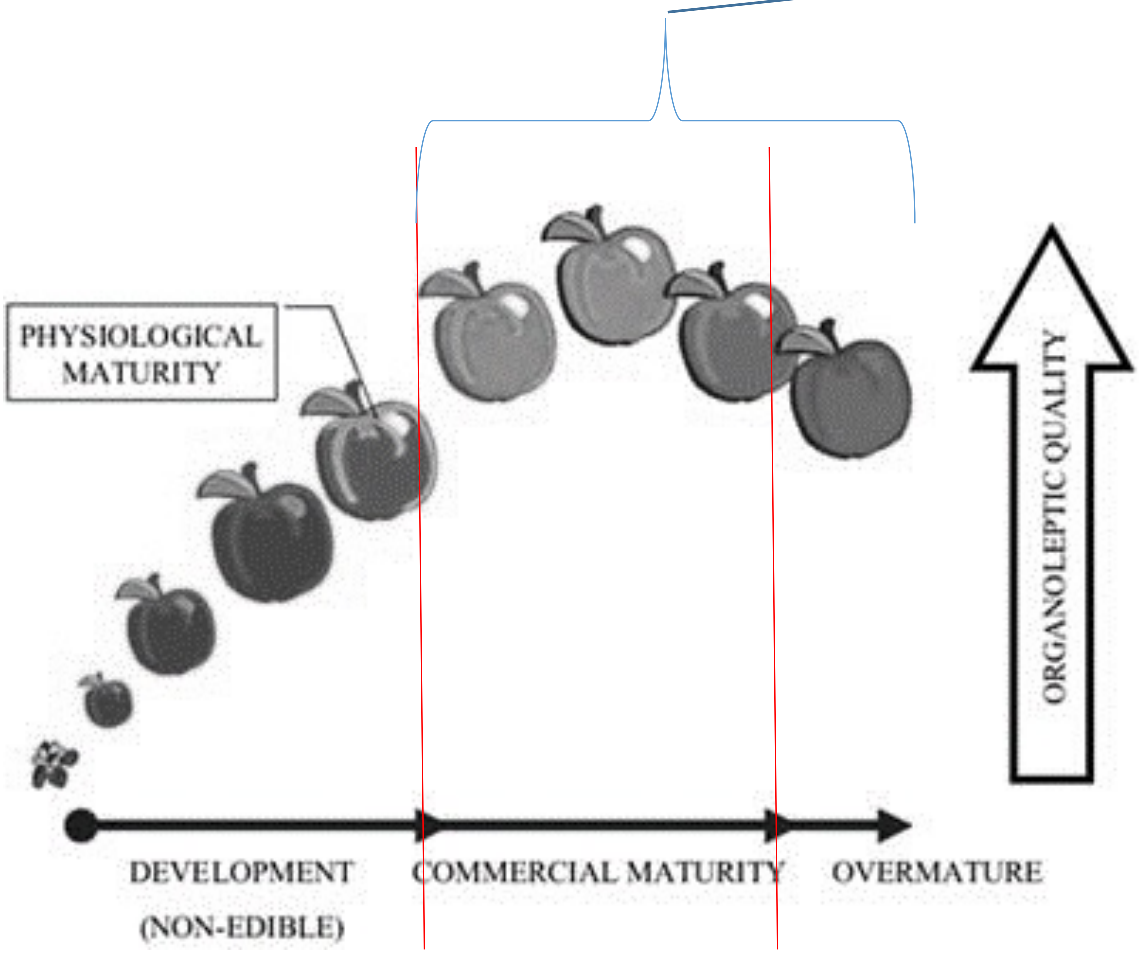


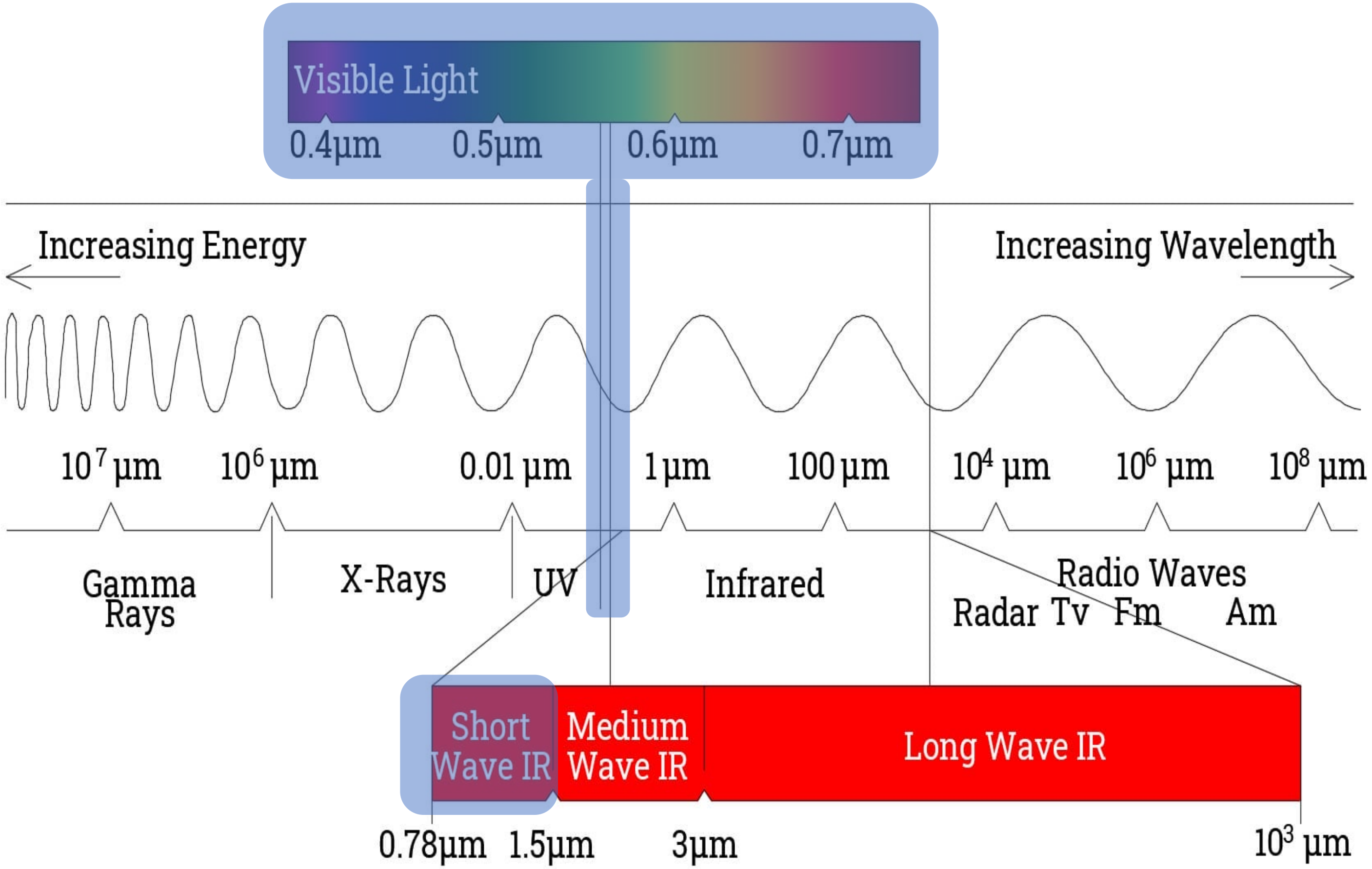
Image Source: https://sites.google.com/site/maturitypostharvest97/maturity

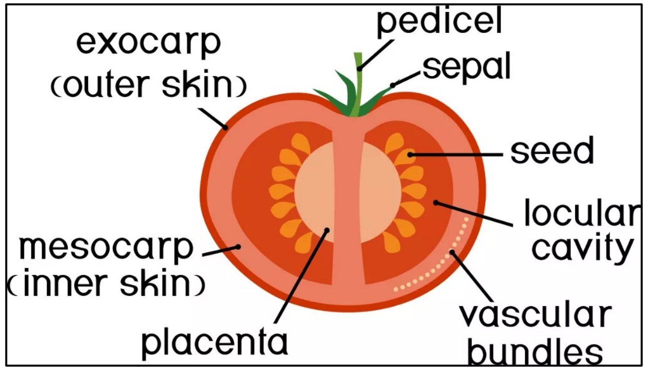
Fruit Quality Assessment

(Changes in physicochemical Parameters as indicators of fruit ripening stage)

Destructive / invasive methods	Non-destructive/ Non-invasive methods
• Mechanical and Chemical based laboratory methods • Laborious, time-consuming and require costly equipment and physical damage	• **Spectroscopy** (UV-Vis-NIR), ultrasonic, RFID, electronic nose, machine vision • rapid, non-destructive and economical methods

Electromagnetic Spectrum for Identification and Quantification of physicochemical attributes in intact fruits



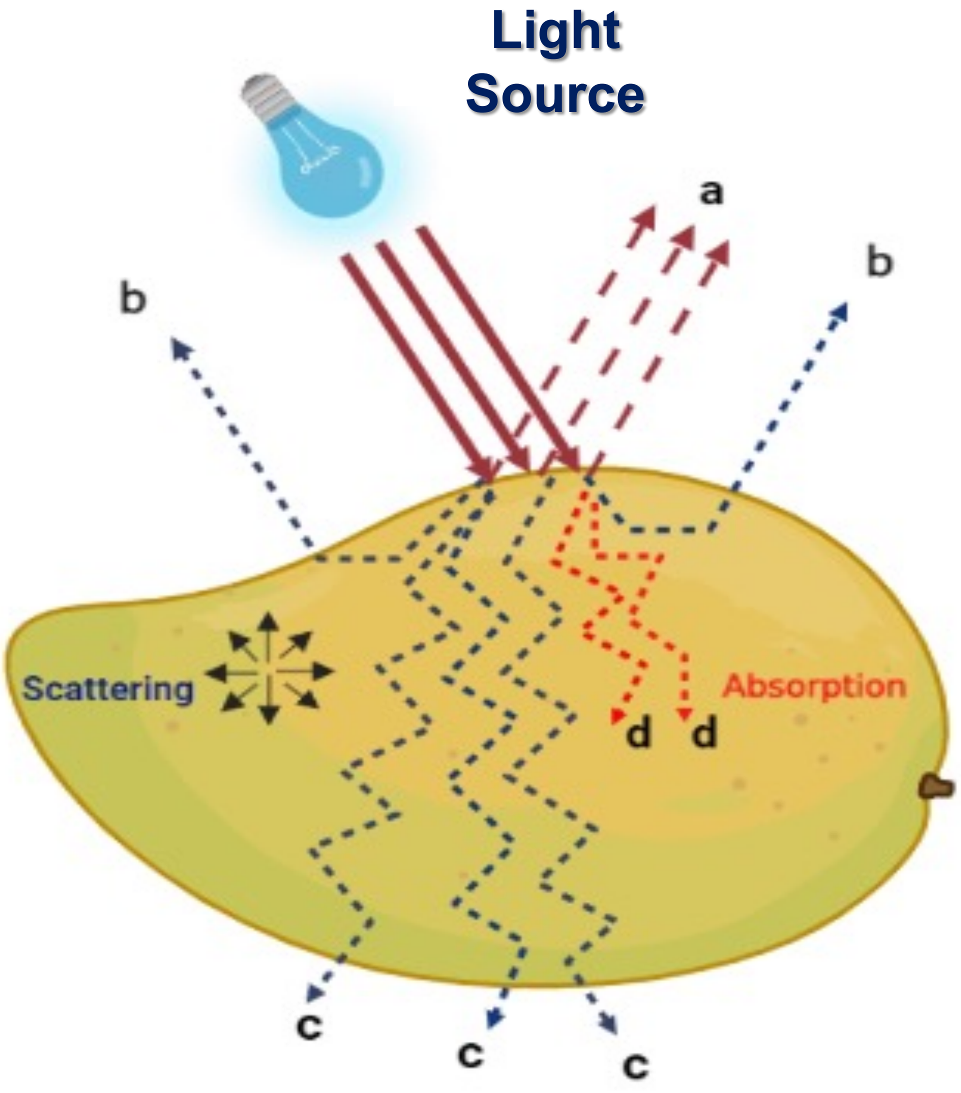


- **Visible region**: Colour and fruit surface/textural characteristics (**exocarp** properties)
- **SWNIR region**: Deeper penetration into fruit (**mesocarp** composition such as macronutrients, sugars etc.)[8,9]
- SWNIR absorption: Overtones and combination bands of molecular vibrational modes, thus beer-lambert law not applicable
- Chemometrics calibration models: Establishing relationships between VIS-SWNIR spectra and fruit attributes for quantitative and qualitative fruit composition assessment by analysing reflected, absorbed, or transmitted radiations at different wavelengths[10]

Spectrum image Source: Internet; Tomato image: Kazakova Maryia/Shutterstock

Different Modes of Light Phenomenon

1. Specular reflection (a)
2. Diffused reflection / Interactance (b)
3. Transmittance (c)
4. Absorption (d)



Literature Review

Significant literature reviews on Vis-NIR spectroscopy for quality estimation of fruits and vegetables

Sr. No.	Theme of review	Reference
1	Overview of NIR spectroscopy, basic concepts, equipment, chemometric analysis	(Nicolaï et al., 2007)
2	Applications of visible and near infrared (Vis-NIR) spectroscopy for quality and variety estimation of various fruit	(Wang et al., 2015)
3	Instrumentation and calibration procedures for NIR spectral data acquisition systems	(Xie et al., 2016)
4	Data processing techniques for fruits quality inspection through spectral data	(Srivastava & Sadistap, 2018b)
5	Bibliometric analysis of spectroscopy in food quality estimation	(Aleixandre-Tudó et al., 2019)
6	Off-line and in-line monitoring of postharvest produce quality using Vis-NIR spectroscopy	(Cortés et al., 2019)
7	Multivariate calibration of spectroscopic sensors	(Saeys et al., 2019)
8	Principles, theory and modelling of light transfer in horticulture produce	(Lu et al., 2020)
9	Instrumentation, chemometric analysis and applications of Vis-NIR spectroscopy	(Walsh et al., 2020)
10	Comparison of Vis-NIR spectroscopy calibration models	(Arruda de Brito et al., 2022)

Literature Review

Physiochemical Parameter	Equipment	Calibration Models	Year & Reference
Colour, TA, dry matter	F-750, SCiO and H-110 F spectrophometer	PCA, PLS	2022 [8]
pH	Flame-T-Vis-NIR spectrometer (Ocean Optics, USA)	PLS	2021 [9]
Carotenoid concentration	Portable Vis-NIR spectrometer F-750 Felix instruments	PLS, ANN	2020 [10]
Lycopene	Portable Vis-NIR spectrometer F-750 Felix instruments		2020 [11]
Colour, firmness, TSS, Acidity	F-750, SCiO and H-110 F spectrophometer	PCA	2020 [12]
Electric conductivity, TSS, colour	MicroNIR 1700 Viavi Solutions	PLS	2019 [13]
SSC, lycopene	Spectrometer developed by authors	PCA, PLS	2019 [14]

Literature Review

Physiochemical Parameter	Equipment	Calibration Model	Year & Reference
Firmness	Portable Vis/SWNIR spectrometer (Modelo Inc., USA)	PCR	2018 [15]
TA, pH, TSS	AvaSpec-ULS 2048, Netherlands	PLS	2018 [16]
TA, pH, TSS, colour	FieldSpec Pro FR spectrometer, Analytical Spectral Devices Inc, USA	PLS-DA	2017 [17]
SSc, lycopene and polyphenols	FieldSpec Pro FR spectrometer, Analytical Spectral Devices Inc, USA	PLS	2014 [18]
Firmness, TSS, Colour	Spectromete NIR portable, PHASIR 0917, Germany	PCR	2012 [19]
TSS, TA	Bruker Tensor 27 FTIR spectrometer, France	PLS	2011 [20]
Firmness, TSS, TA	USB4000, Ocean Optics Inc, USA	Multivariate regression	2010 [21]

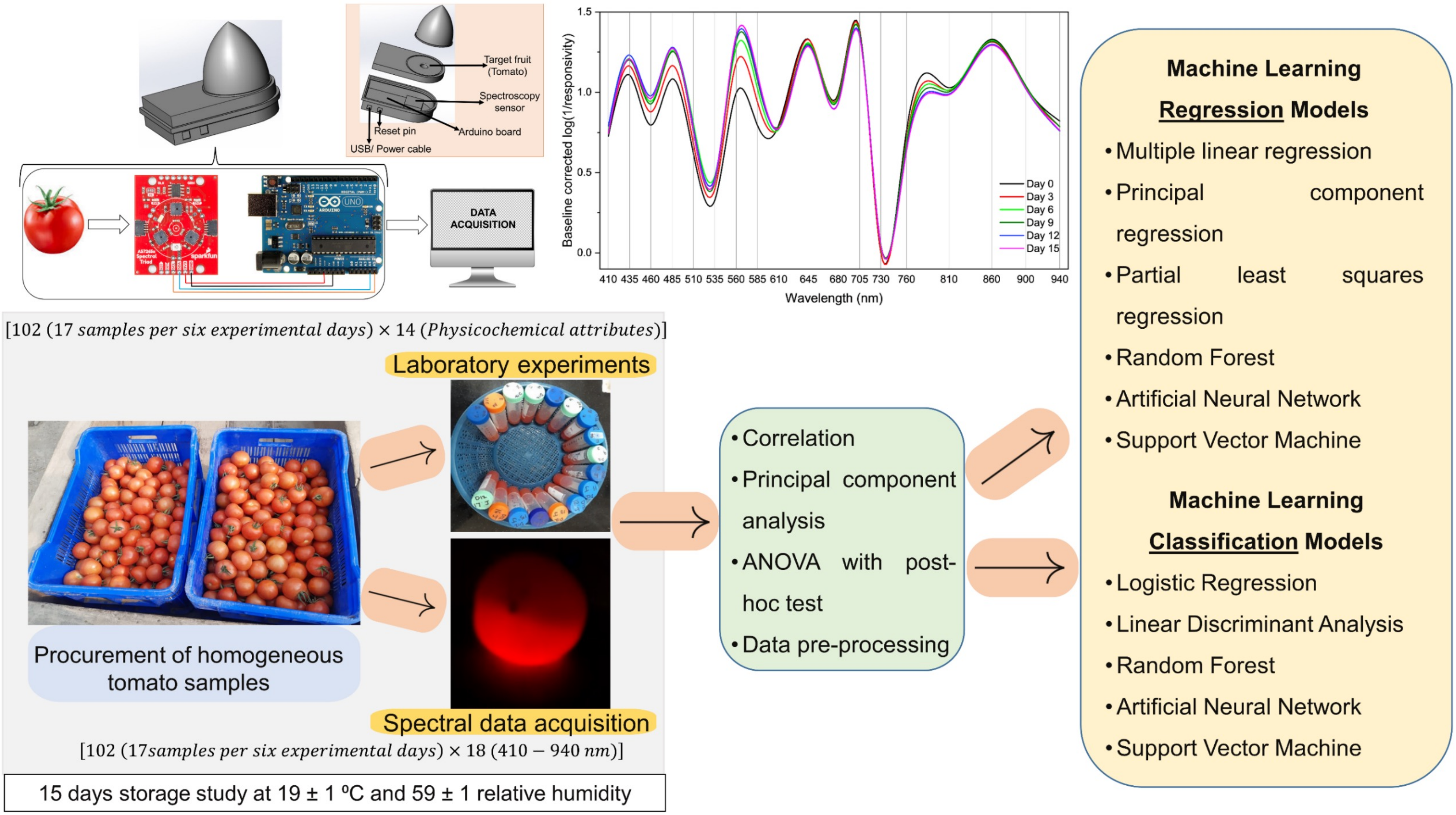
RESEARCH GAP

- Based on the literature review, it has been observed that:
  - Despite plethora of research papers from past one and half decade, it is important to foresee that the technology is still struggling to find its space in commercial post-harvest practices.
  - Most of research in the field of portable Vis-SWNIR fruit spectroscopy are 'derivative' in nature as they primarily showcase the application of existing commercially available spectroscopic devices coupled with traditional multivariate analysis techniques subjected to different fruits.
  - The chemometric calibration models developed on the spectral data obtained from such commercial equipment **often remain theoretical** due to lack of integration mechanisms to incorporate developed models back into the commercial instruments, which are often proprietary devices. This gap poses a significant hurdle in transfer of innovations and advancements for practical implementation of this technology.

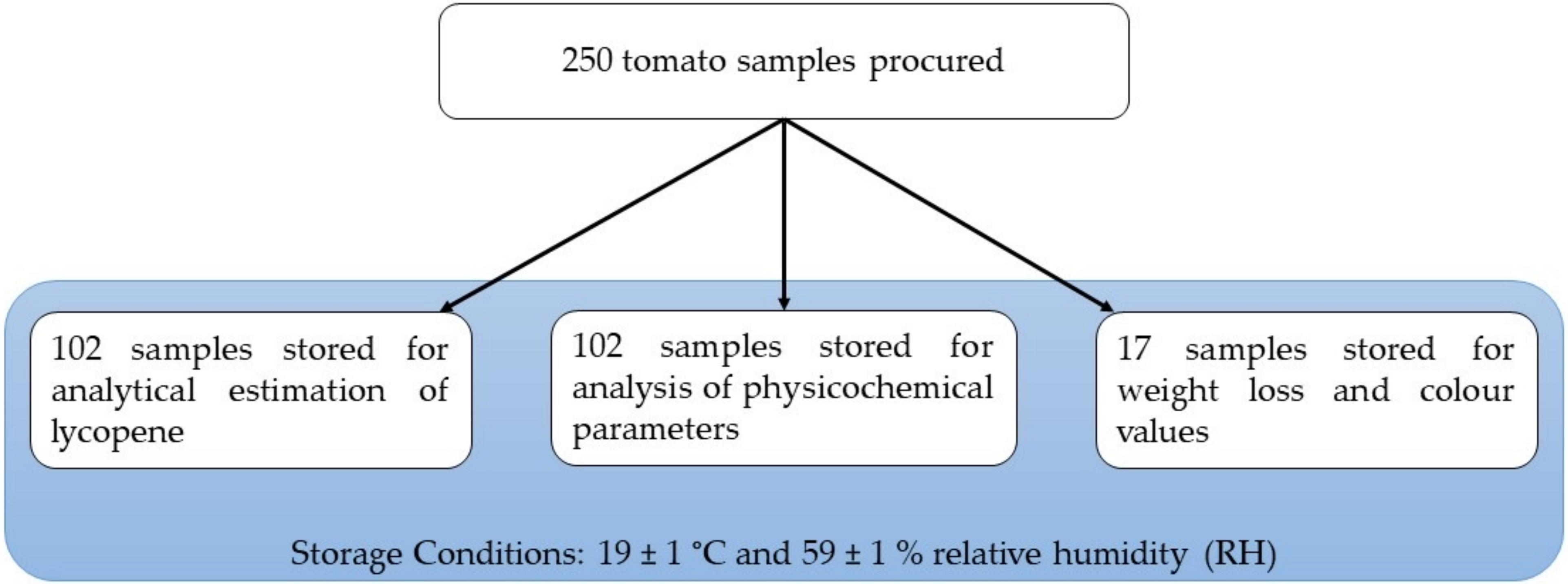
The way out

- The obvious and potential solution to this research gap is to develop intelligent devices based on **open source platforms** so that the developed machine learning calibration models could get seamless integration with existing ones and organically enhance instrument capabilities.
- The **proposed spectrometer** assembly is **system integration** of multi-spectral sensor chipset with open-source microcontroller housed in **uniquely designed cabinet** for spectral data acquisition in reflectance mode.
- Target fruit for present study: Tomatoes
  - Tomatoes (*Solanum lycopersicum*) are most suitable to study the ripening process of climacteric fruits[4]. Highest production among global vegetable crops[5]. Rich source of bioactive compounds, carotenoids, including lycopene[6]. Its consumption leads to reduced risk and incidence of cancer, and cardiovascular diseases[7].

DESIGN OF EXPERIMENT



Procurement of raw material



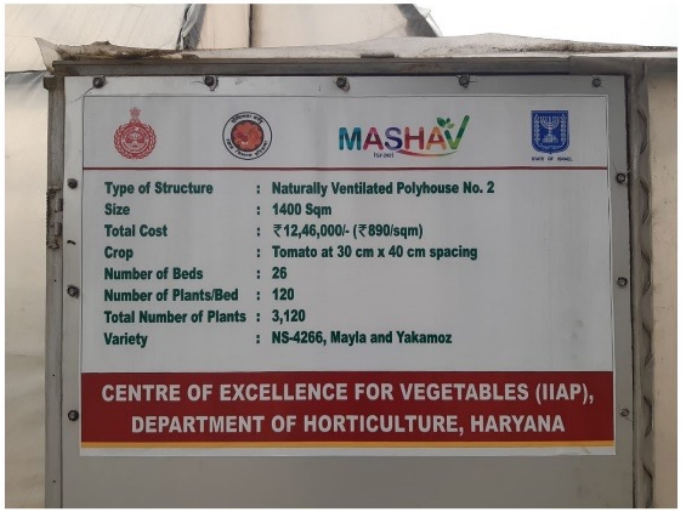
Sample Requirement/procurement plan w.r.t. 17 samples per experimental day for two different tests (a), (b) i.e. 34 samples per experimental day	
a) Total sample required for non-destructive and some destructive tests on specified dates	102
b) Total sample required for lycopene content on specified dates	102
Total number of homogeneous sample procurement for storage study (i)+(ii)	102+102+26 (extra) = 230
Number of tomato samples in 1 Kg	8
Tomatoes procured	30 Kg (Approx.)

Procurement of raw material

Tomato Variety: NS-4266

View of ployhouse

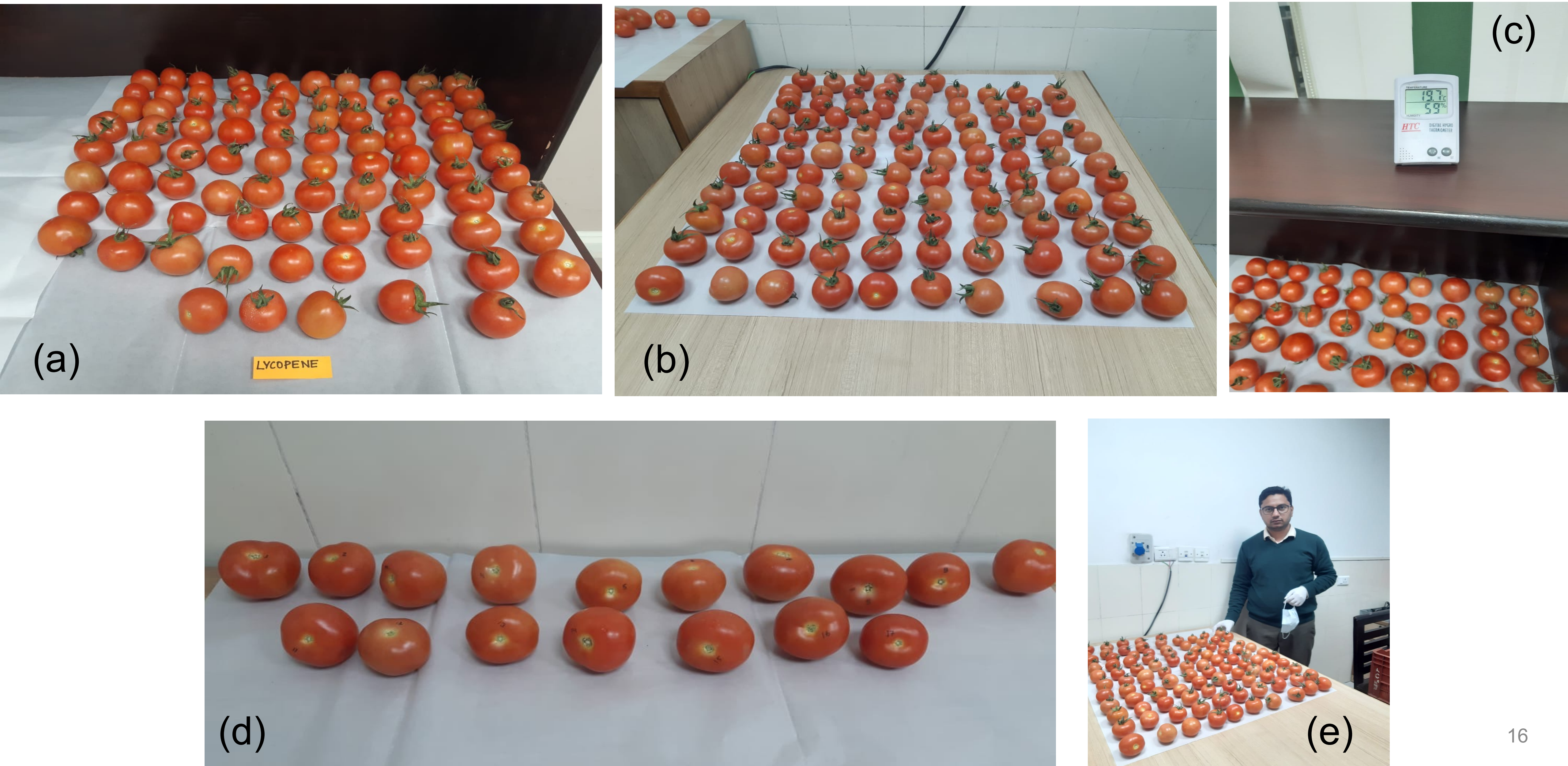
Harvesting



Outside view of ployhouse and details of polyhouse

Samples procured for experiment

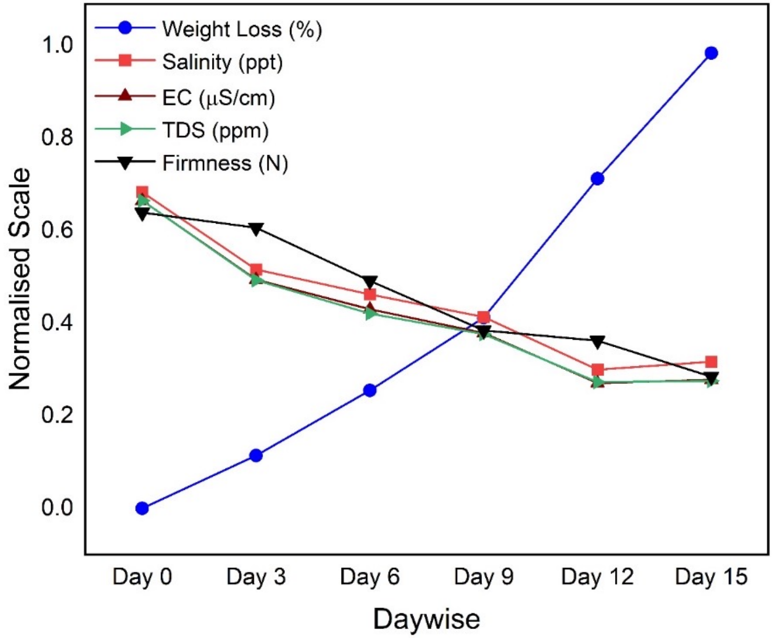
Storage study of raw material

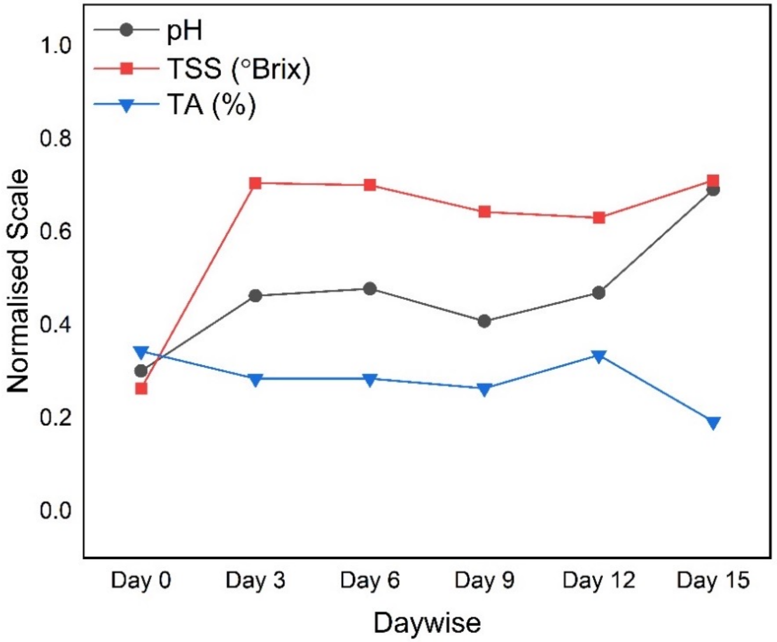


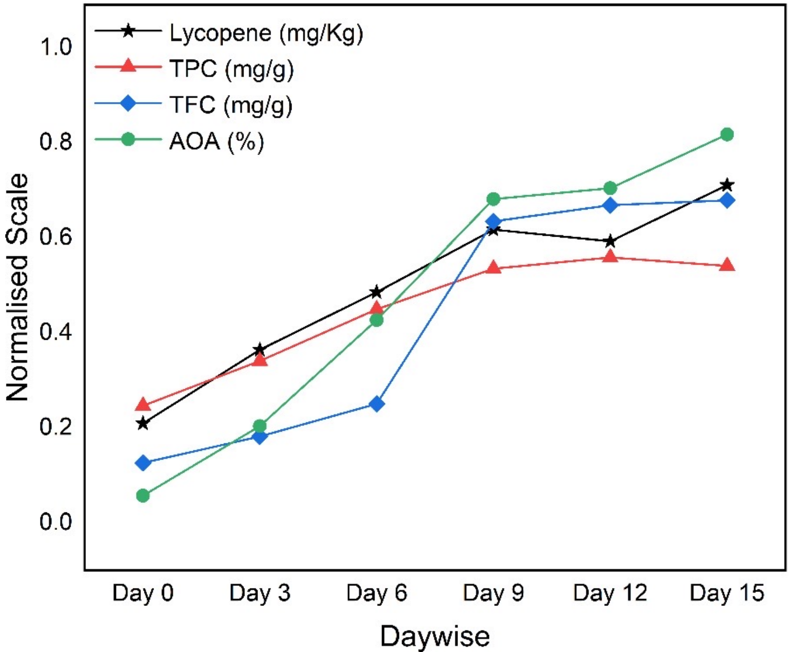
Data Acquisition

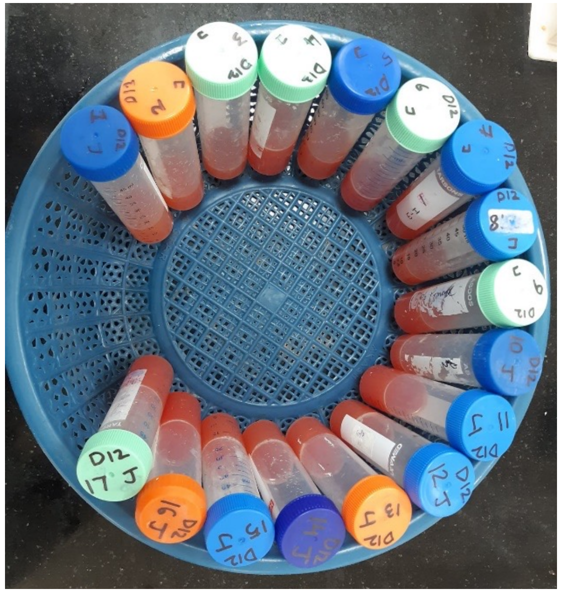
Parameters	Estimation technique/equipment
Weight loss	Analytical balance (BSA 2245-CW, Sartorius) [22]
Colour (L, a, b)	Konica Minolta Chroma meter CR-400 [23]
Firmness	Texture Analyser (Stable Micro system, U.K.) [22]
Total soluble solids (TSS)	Refractometer (Atago, Tokyo, Japan) [24]
Titrable acidity (TA)	Titration method [25]
pH	pH meter (Oakton Eco Tester) [25]
Electrical conductivity (EC)	Microprocessor-based digital conductivity meter (LT-51, Labtronics) [13]
Salinity	Microprocessor-based digital conductivity meter (LT-51, Labtronics) [13]
Total dissolved solids (TDS)	Microprocessor-based digital conductivity meter (LT-51, Labtronics) [13]
Lycopene content	AOAC method using HPLC [26]
Total phenolic content (TPC)	Folin Ciocalteu (FC) reagent method [27]
Total flavonoid content (TFC)	UV Spectroscopy method [22]
Antioxidant activity (AOA)	DPPH Assay [28]
Spectral Data	Proposed spectrometer assembly

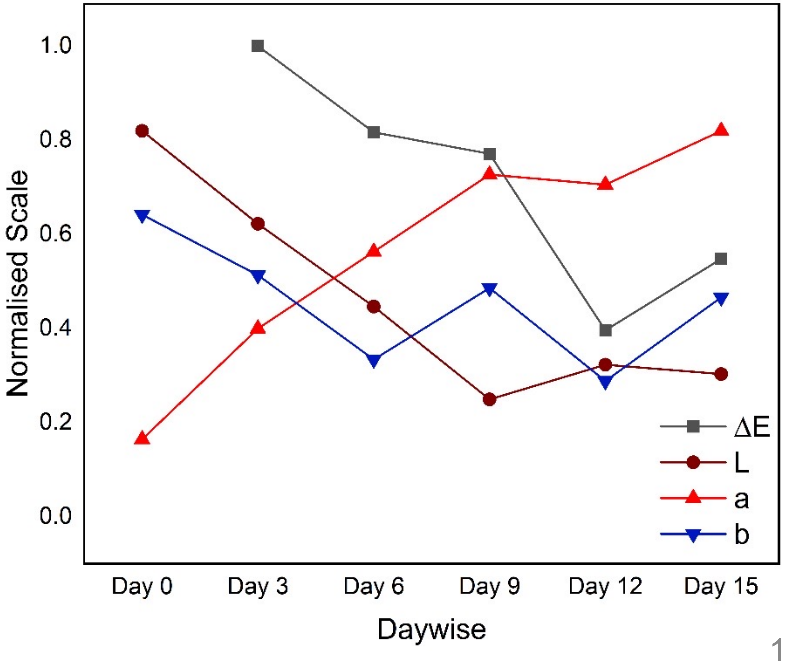
Postharvest storage study of tomatoes to assess variations in physicochemical parameters











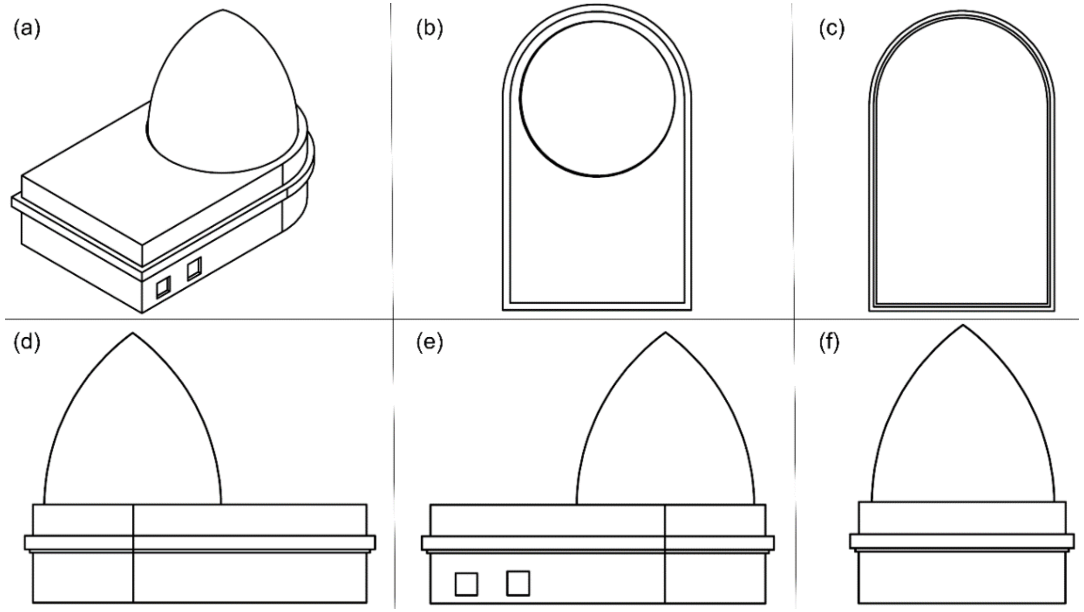
Attributes	Range (Day 0)	(Day 15)	Distribution (Shapiro–Wilk test)	Homogeneity of variance (Levens test)	Parametric or non-parametric test
Salinity	3.50 ± .32 PPT	2.53 ± .43 PPT	Normal ($p \le 0.231$)	Uniform ($p \le 0.054$)	ANOVA ($p \le 0.000$)
EC	3.50 ± .32 μS/cm	2.53 ± .43 μS/cm	Normal ($p \le 0.523$)	Uniform ($p \le 0.810$)	ANOVA ($p \le 0.000$)
TDS	2.26 ± .21 PPM	1.64 ± .27 PPM	Normal ($p \le 0.434$)	Uniform ($p \le 0.770$)	ANOVA ($p \le 0.000$)
Firmness	563.35 ± .87.01 N	388.76 ± 71.82 N	Normal ($p \le 0.334$)	Uniform ($p \le 0.250$)	ANOVA ($p \le 0.000$)
pH	4.08 ± .146	4.45 ± .123	Non-normal ($p \le 0.033$)	Non-uniform ($p \le 0.027$)	Kruskal–Wallis test ($p \le 0.000$)
TSS	1.58–3.20° ± 0.42°Brix	2.86–4.13 ± 0.37°Brix	Non-normal ($p \le 0.001$)	Non-uniform ($p \le 0.000$)	Kruskal–Wallis test ($p \le 0.002$)
TA	0.682 ± .200	0.470 ± .187	Non- normal ($p \le 0.000$)	Uniform ($p \le 0.386$)	Kruskal–Wallis test ($p \le 0.139$)
Lycopene	33.76 ± 8.76 mg/kg	72.14 ± 9.91 mg/kg	Normal ($p \le 0.555$)	Uniform ($p \le 0.514$)	ANOVA ($p \le .001$)
TPC	15.97 ± 4.41 mg/g	22.21 ± 2.85 mg/g	Normal ($p \le 0.579$)	Non-uniform ($p \le 0.008$)	Kruskal–Wallis test ($p \le 0.001$)
TFC	372.00 ± 76.85 mg/g	682.28 ± 132.72 mg/g	Non-normal ($p \le 0.001$)	Non-uniform ($p \le 0.001$)	Kruskal–Wallis test ($p \le 0.001$)
AOA	43 ± 3.84%	86.48 ± 8.84%	Non-normal ($p \le 0.001$)	Non-uniform ($p \le 0.012$)	Kruskal–Wallis test ($p \le 0.001$)
L	37.92 ± 0.89	37.43 ± 1.06	Non-normal ($p \le 0.007$)	Uniform ($p \le 0.104$)	Kruskal–Wallis test ($p \le 0.000$)
a	6.52 ± 1.469	15.45 ± 1.07	Non-normal ($p \le 0.001$)	Non-uniform ($p \le 0.008$)	Kruskal–Wallis test ($p \le 0.001$)
b	10.28 ± 0.76	9.48 ± 0.79	Normal ($p \le 0.246$)	Uniform ($p \le 0.732$)	ANOVA ($p \le 0.000$)

Abbreviations: TDS, total dissolved solids; EC, electrical conductivity; TSS, total soluble solids; TA, titratable acidity, *L*, *a*, *b*, colour values on Hunter scale; TPC, total phenolic content; TFC, total flavonoid content; AOA, antioxidant activity; ANOVA, analysis of variance.

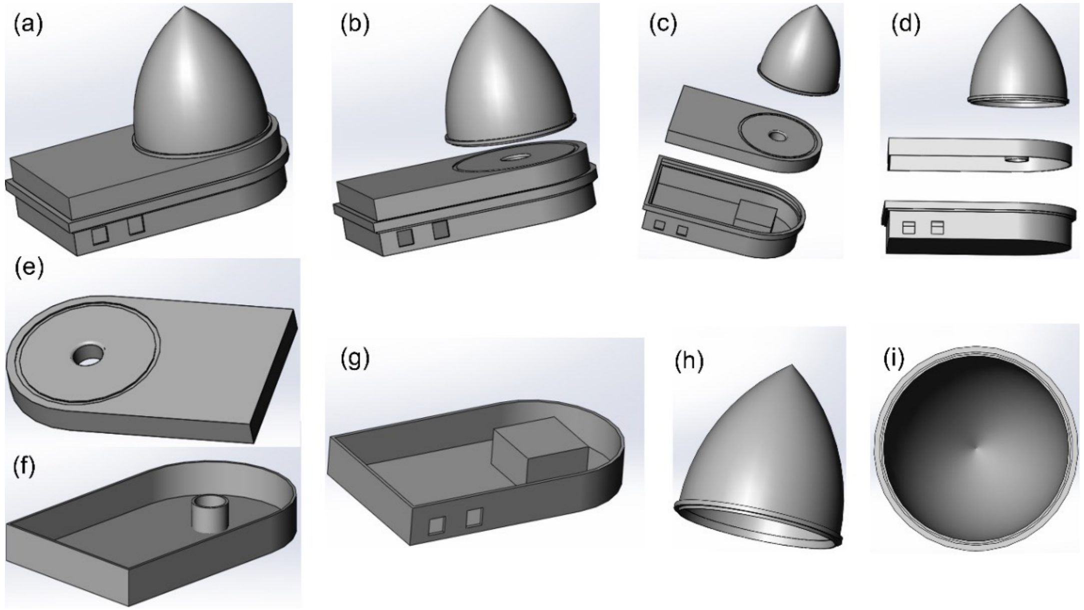
Design and development of sensor cabinet for spectral data acquisition

The design of the spectrophotometer cabinet is divided into three main parts:

1) Lower part design
2) Upper part design
3) Dome-shaped cover



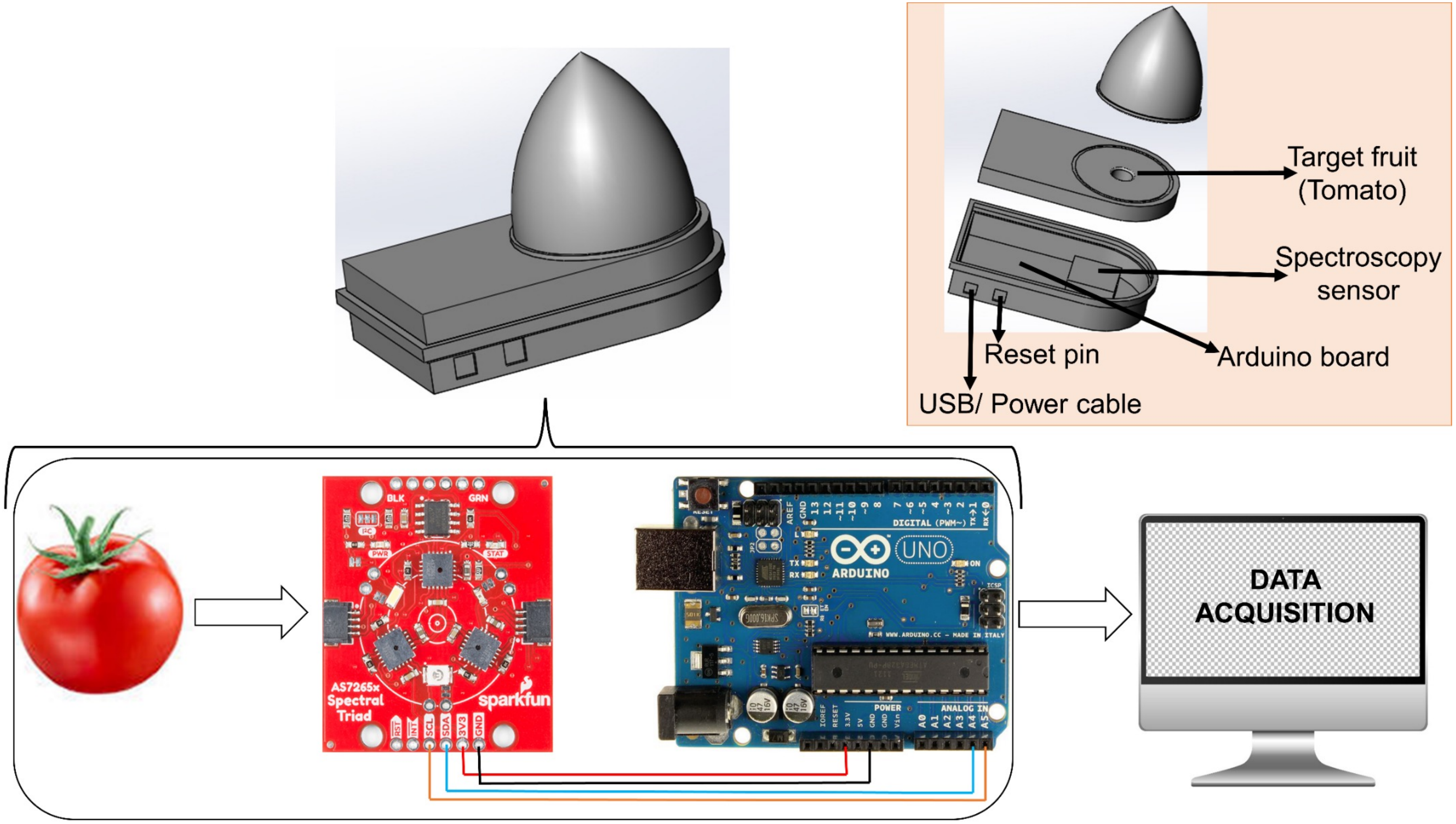
(a) Isometric view (b) Top view (c) Bottom view (d) Left-side view (e) Right-side view (f) Front view



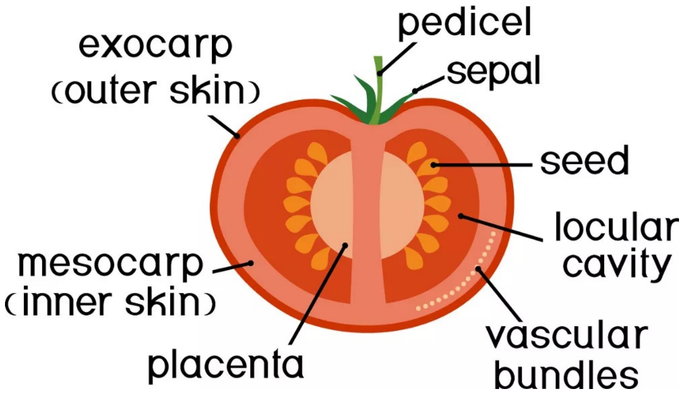
Perspective view of designed assembly: a, b, c and d, view of upper part: e and f, view of lower part: g, view of dome shaped cover: h and i

Schematic Diagram for Spectral Data Acquisition

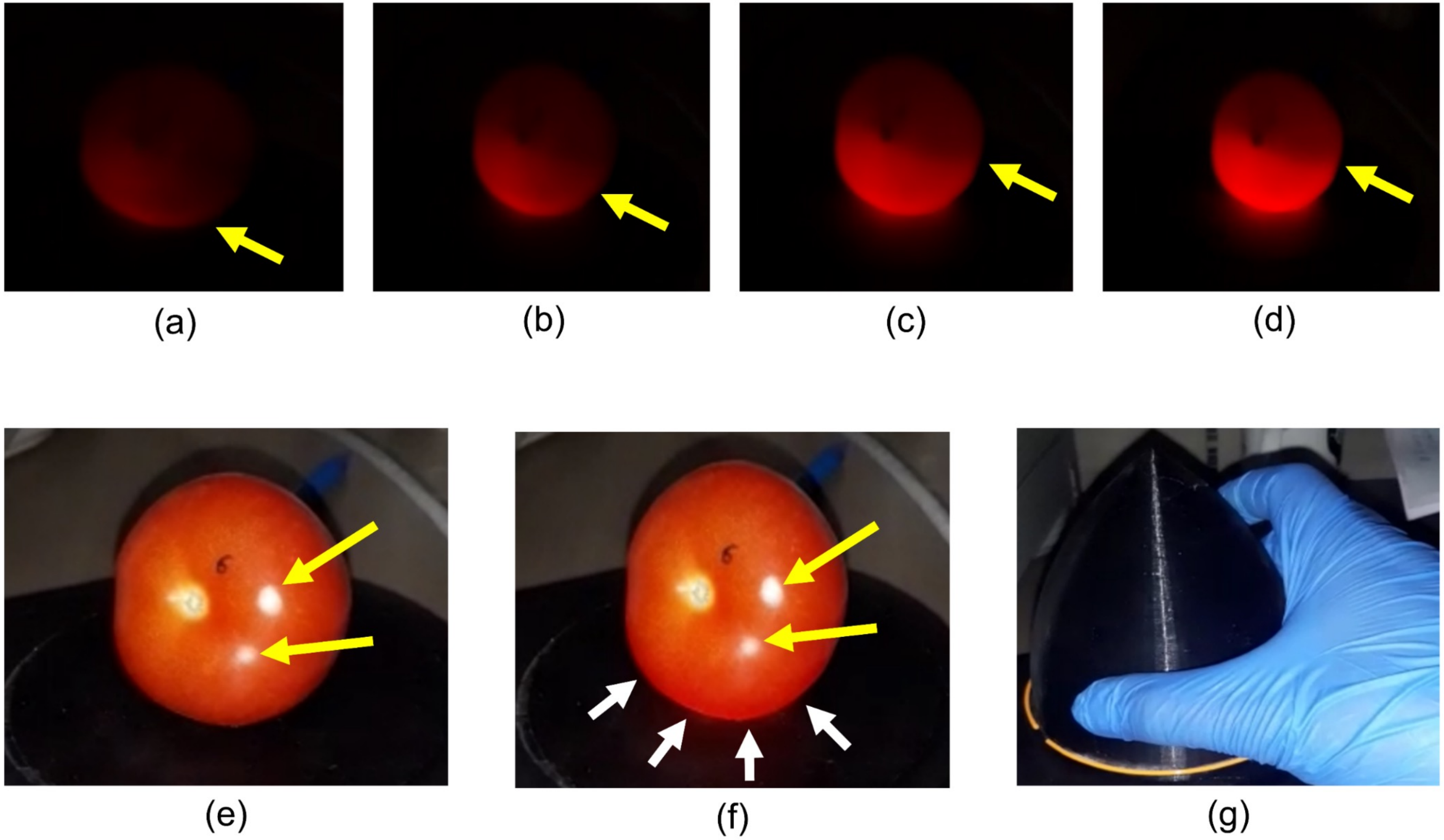
The spectral data at 18 different wavelengths from 410 – 940 nm with 20 nm full width at half maximum (FWHM) was acquired by system integration of commercially available AS7265x 18-channel multispectral sensor chipset with Arduino Uno, an open-source microcontroller board based on the Microchip ATmega328P microcontroller assembly, housed in an ergonomically designed cabinet



(Image source: Internet; all images used are copyright free)

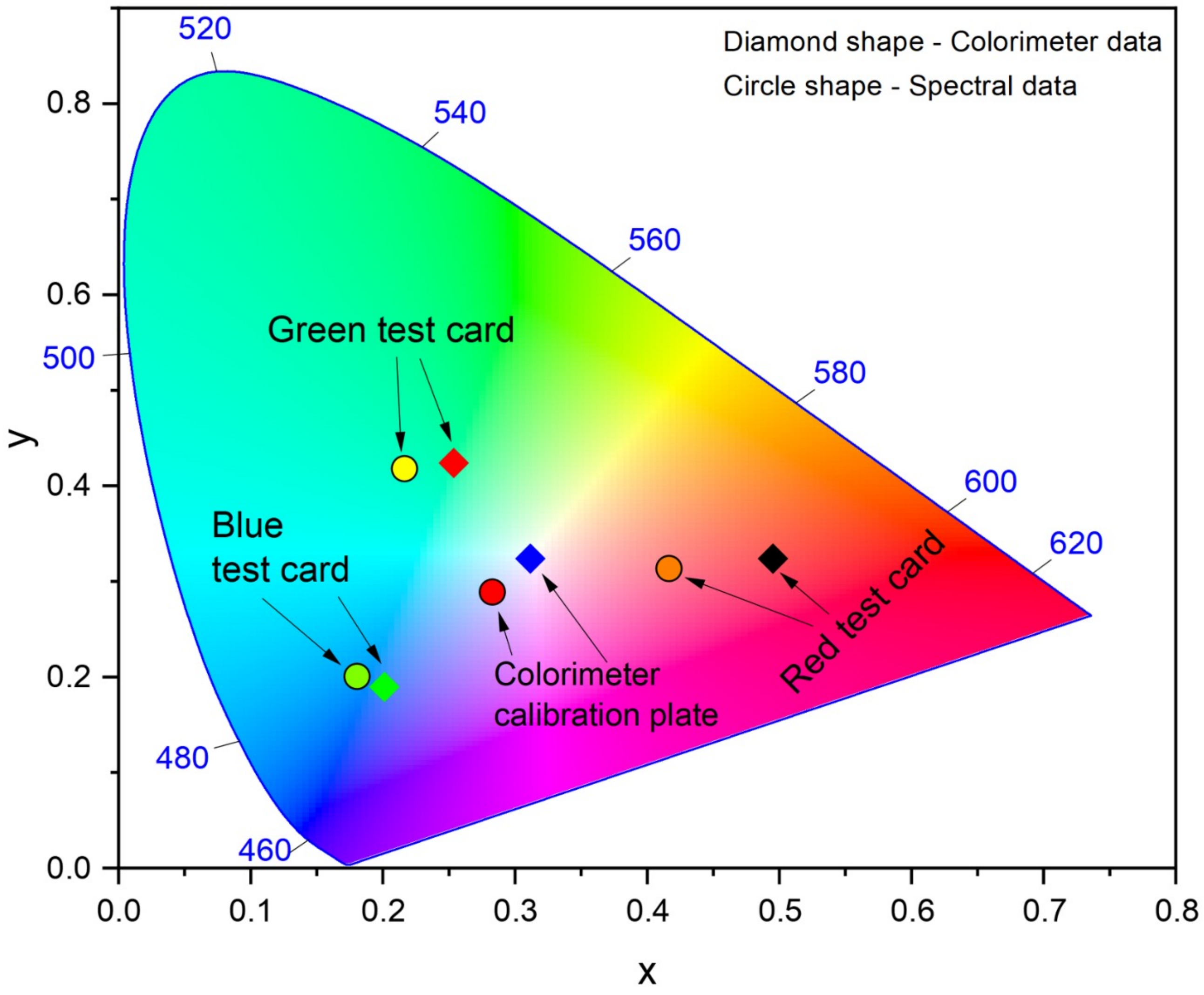


(Image: Kazakova Maryia/Shutterstock)



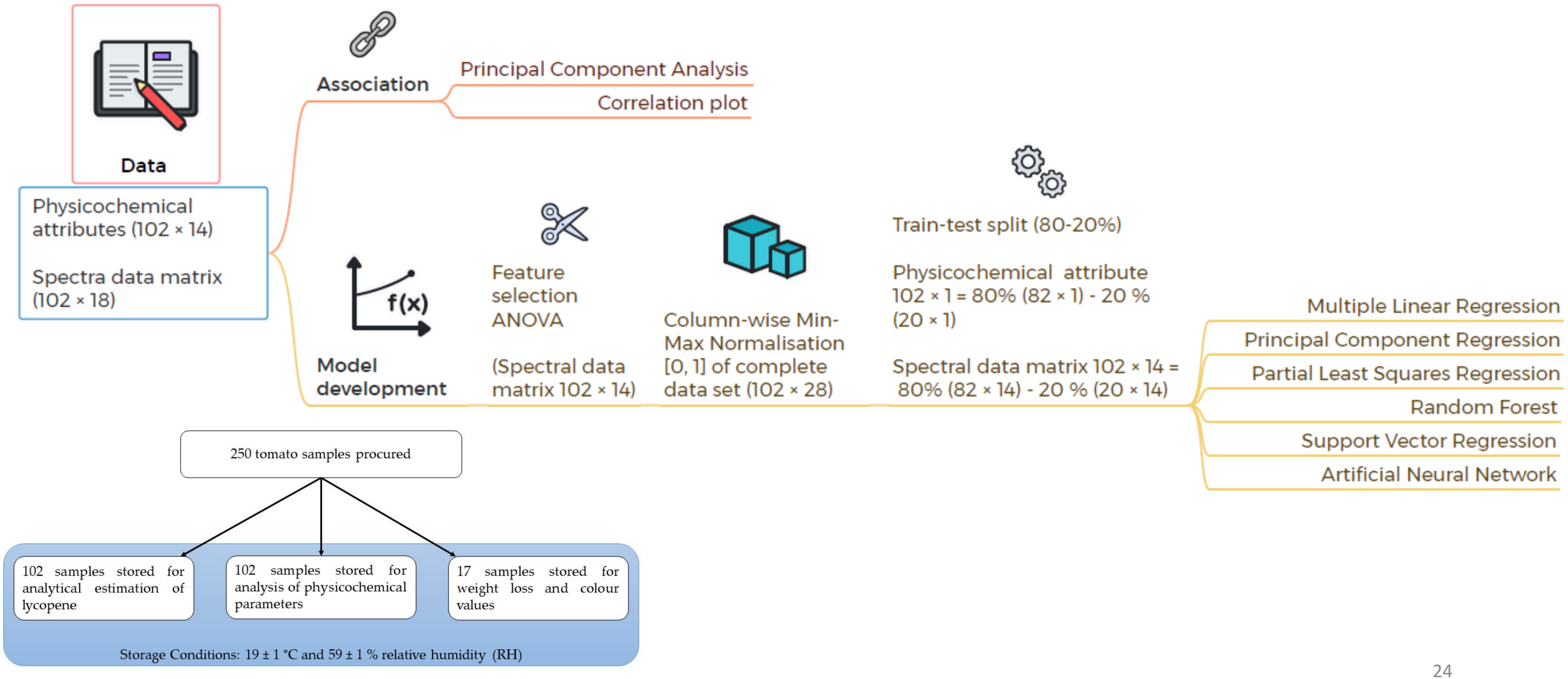
Spectral data acquisition screenshot images captured in slow motion; (a) to (d) represents gradual penetration and diffused reflectance of Vis-SWNIR radiations with no external light source; (e) and (f) represents specular reflectance of external light source used to capture images; (g) target fruit fully covered with dome shaped cover.

Validation of spectral data acquired through proposed spectrophotometer assembly with colorimeter (CR400 Minolta, Japan) on chromaticity plot (CIE 1931)

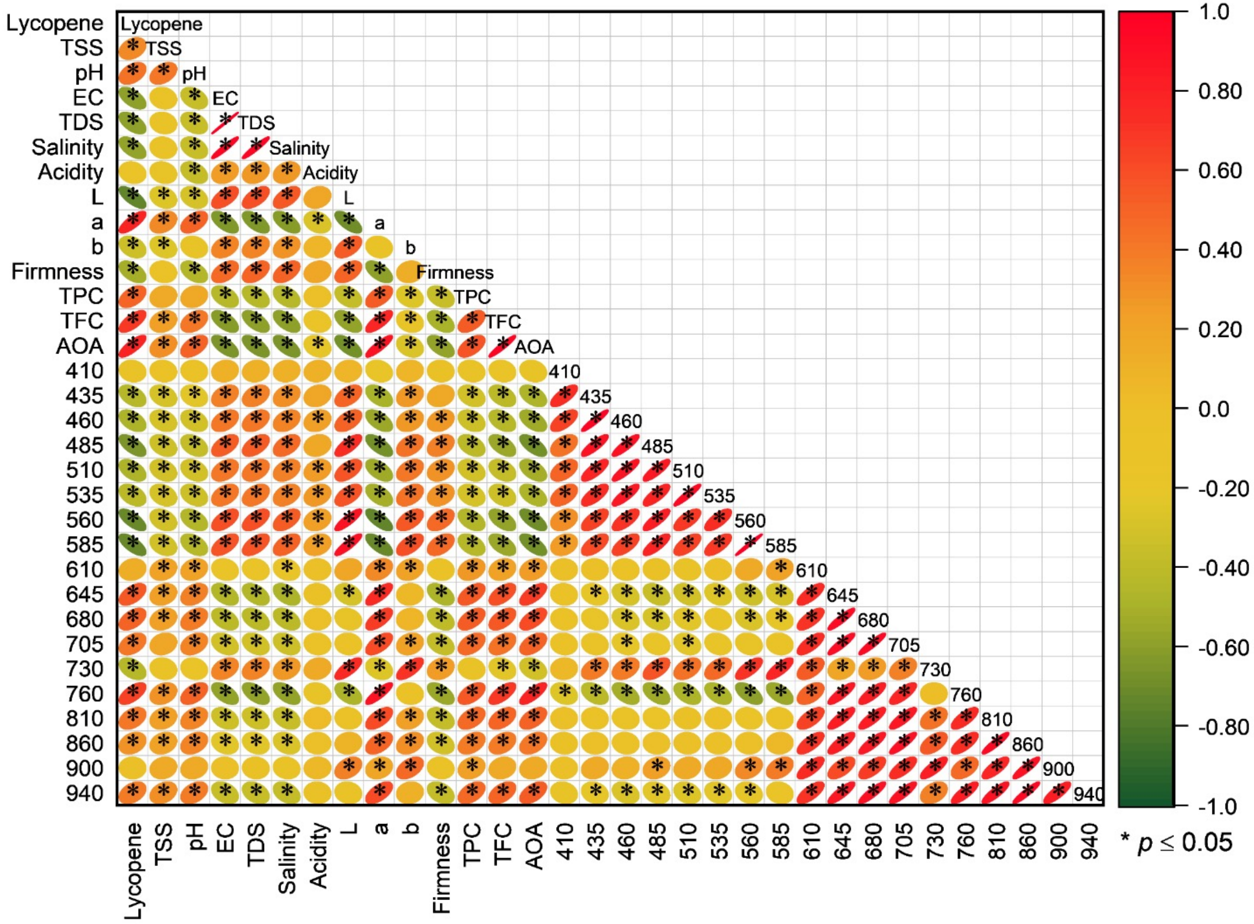


- Comparison of Colorimeter and Spectral Coordinates Colorimeter (x, y) vs. Spectral (x, y) on white calibration plate, Red (255, 0, 0), Green (0, 255, 0) and Blue (0, 0, 255) test cards
  - White: Colorimeter (0.3205, 0.3363), Spectral (0.2827, 0.2829)
  - Red: Colorimeter (0.5086, 0.3363), Spectral (0.4165, 0.3136)
  - Green: Colorimeter (0.2611, 0.4400), Spectral (0.2163, 0.4179)
  - Blue: Colorimeter (0.20719, 0.19720), Spectral (0.18039, 0.2006)
- Euclidean Distance (√((x2 - x1)^2 + (y2 - y1)^2)) between Colorimeter and Spectral Coordinates: <u>White: 0.0602, Red: 0.1136, Green: 0.0512, Blue: 0.0272</u>
- Lower Euclidean distance indicates similarity between spectral and colorimeter data
- Largest Euclidean distance for red colour attributed to non-linear curved boundary of chromaticity graph

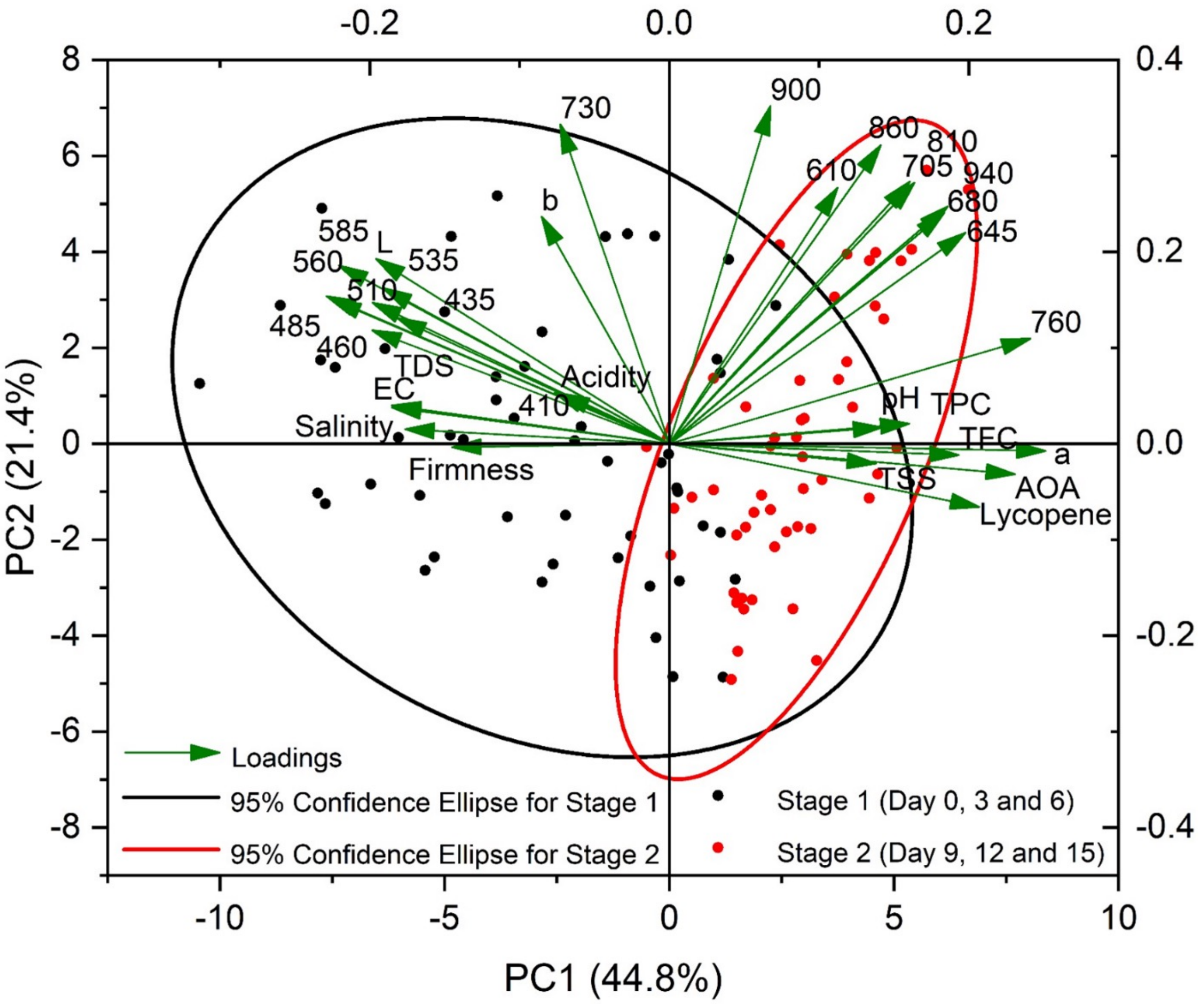
Development of machine learning prediction models for estimation of physicochemical attributes using spectral data



Quantitative and Qualitative analysis of association between physicochemical and spectral variables

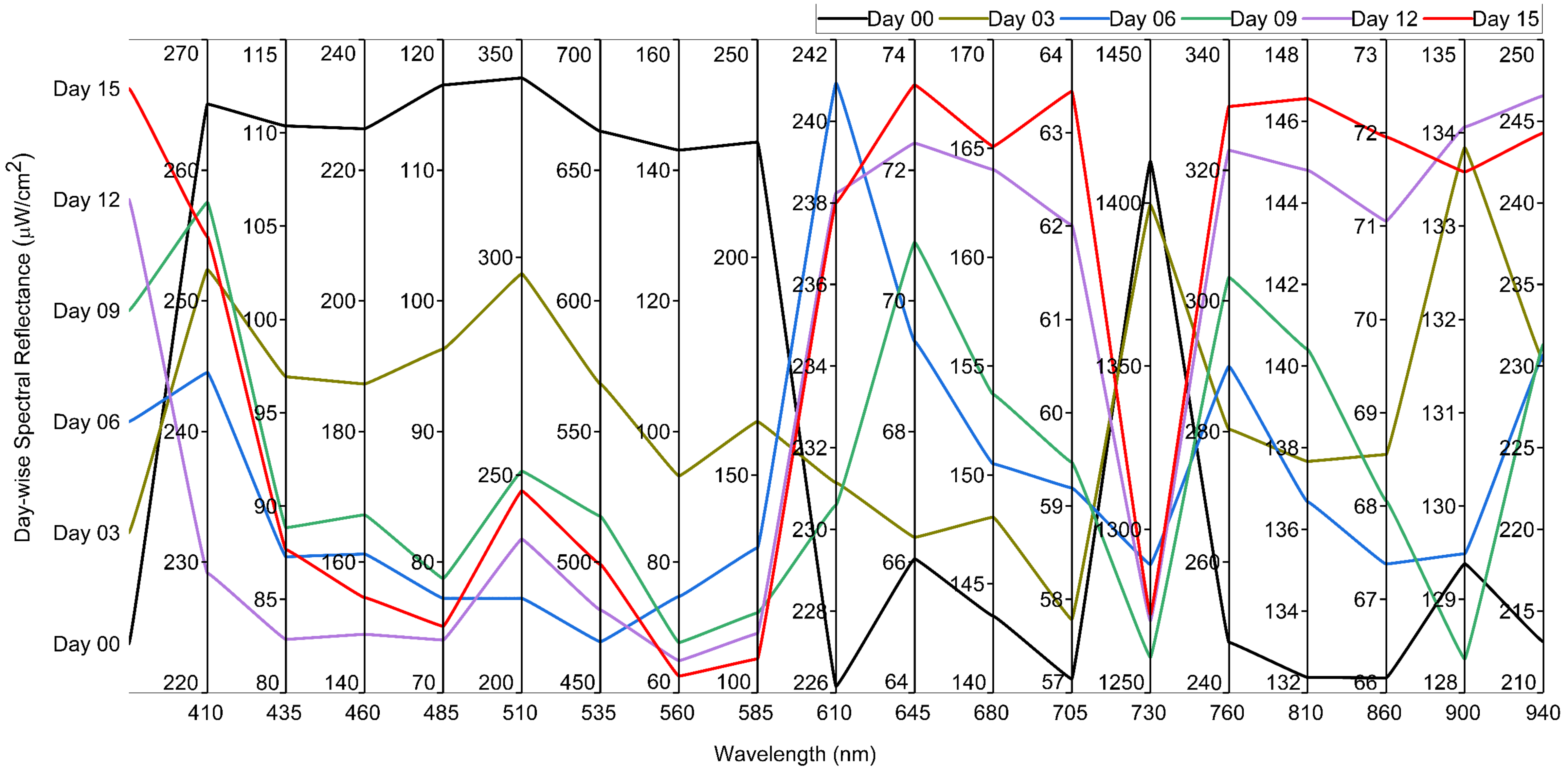


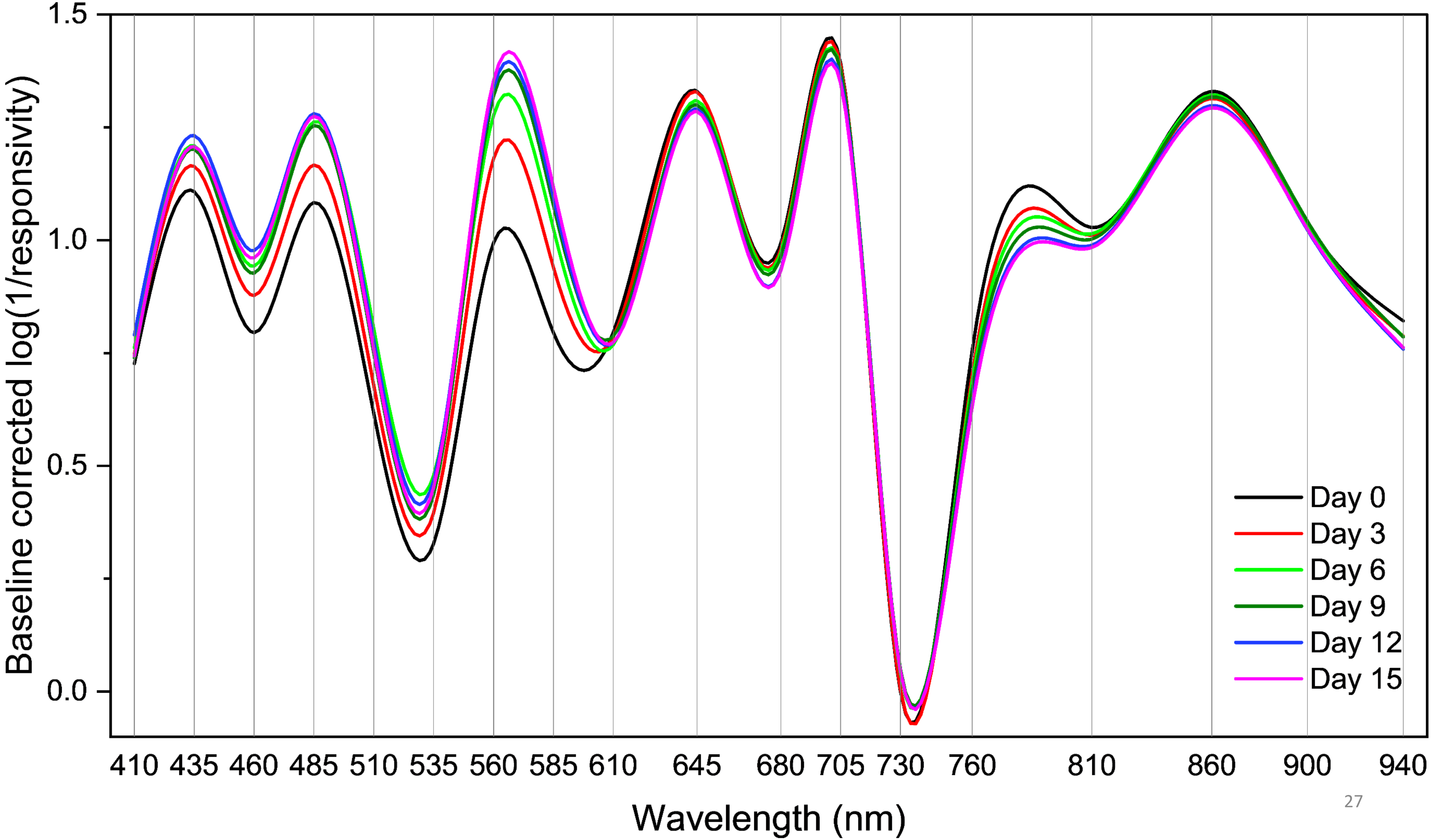
Correlation plot



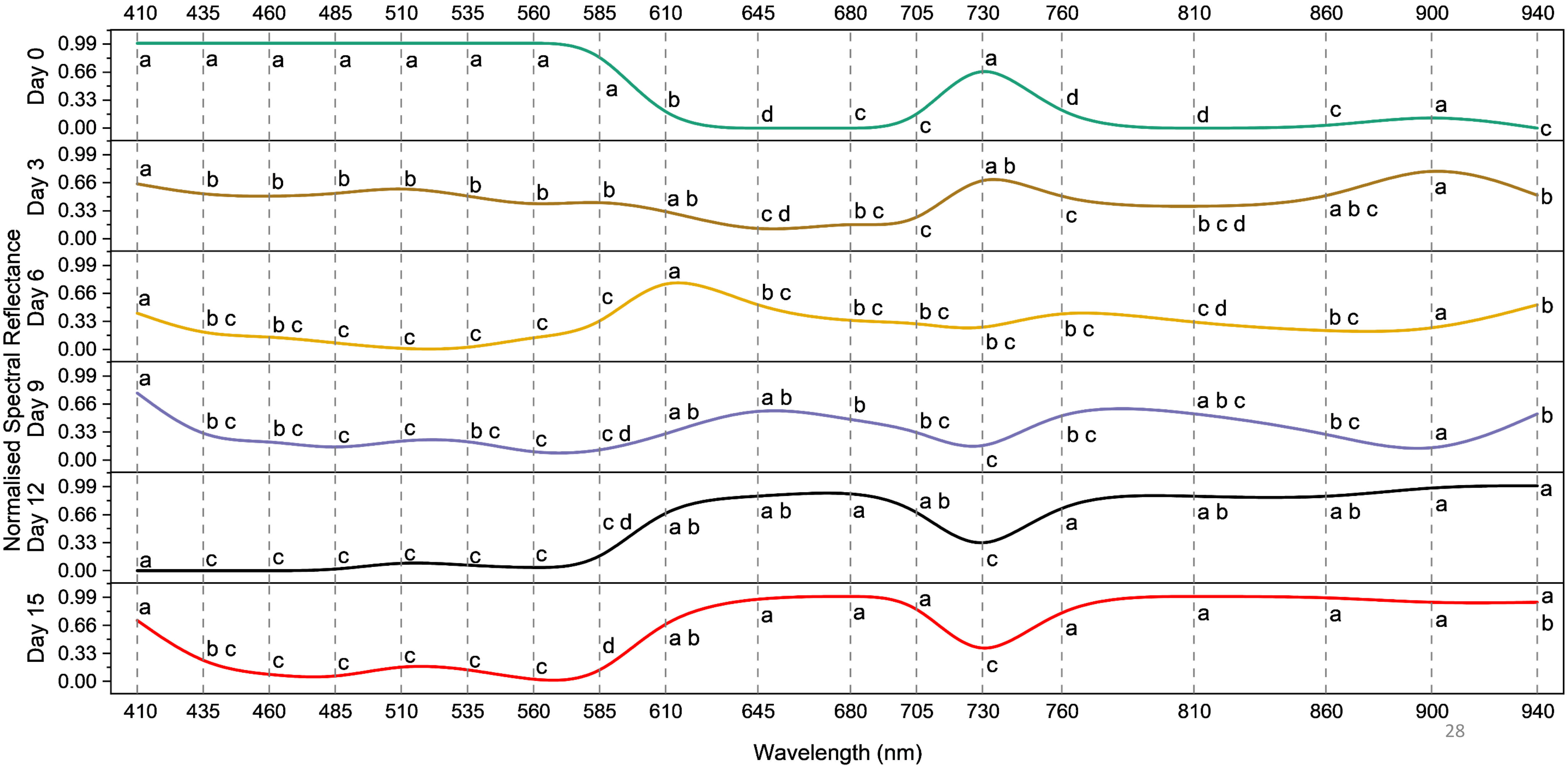
PCA Biplot

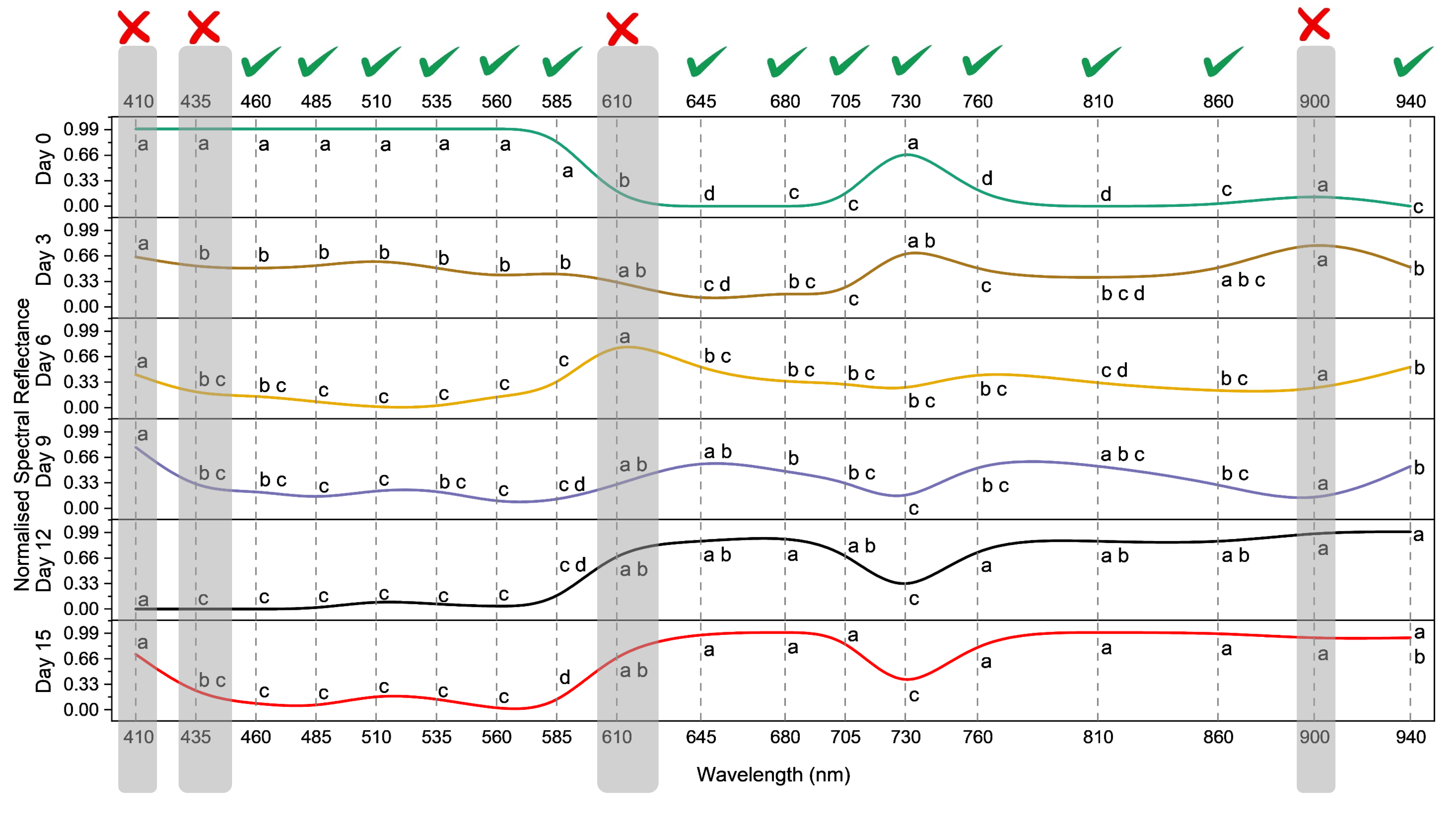
Parallel Index Plot




1.5
1.0
0.5
0.0
Baseline corrected log(1/responsivity)
410 435 460 485 510 535 560 585 610 645 680 705 730 760 810 860 900 940
Wavelength (nm)
Day 0
Day 3
Day 6
Day 9
Day 12
Day 15

Analysis of variance using Tukey's post-hoc test for identification of statistical variance in spectra with respect to storage duration





Normalised Spectral Reflectance
Day 0
Day 3
Day 6
Day 9
Day 12
Day 15
Wavelength (nm)
410 435 460 485 510 535 560 585 610 645 680 705 730 760 810 860 900 940

Multicollinearity diagnostics of spectral wavelengths

Wavelengths (nm)	Iterations							
	1st	2nd	3rd	4th	5th	6th	7th	8th
460	7.09	7.03	6.87	6.73	6.47	3.90	3.89	3.88
485	27.27	26.81	26.57	19.59	16.30			
510	33.93	33.21	33.21					
535	27.29	26.92	26.60	9.00	6.18	5.91	5.58	5.49
560	118.26	16.34	15.72	14.79	14.60	10.21	10.21	7.97
585	207.45							
645	19.71	18.89	17.96	15.65	12.25	10.31	9.90	8.24
680	23.82	22.12	21.85	21.20	11.16	11.01	8.99	8.94
705	29.28	28.86	25.54	25.37				
730	24.06	5.53	4.96	4.95	4.76	4.62	3.41	3.27
760	24.78	19.12	17.04	14.83	14.83	13.76	11.40	
810	15.47	14.64	8.87	8.79	7.99	7.89	6.80	6.35
860	62.70	62.43						
940	31.37	31.35	15.42	15.26	14.60	14.56		

Summary of criterion based selection approach to identify best subset of predictors for colour value ‘a’

Best Subsets Regression

Model Index	Predictors
1	\`560\`
2	\`560\` \`645\`
3	\`535\` \`560\` \`645\`
4	\`560\` \`645\` \`730\` \`810\`
5	\`560\` \`645\` \`680\` \`730\` \`810\`
6	\`535\` \`560\` \`645\` \`680\` \`730\` \`810\`
7	\`460\` \`535\` \`560\` \`645\` \`680\` \`730\` \`810\`

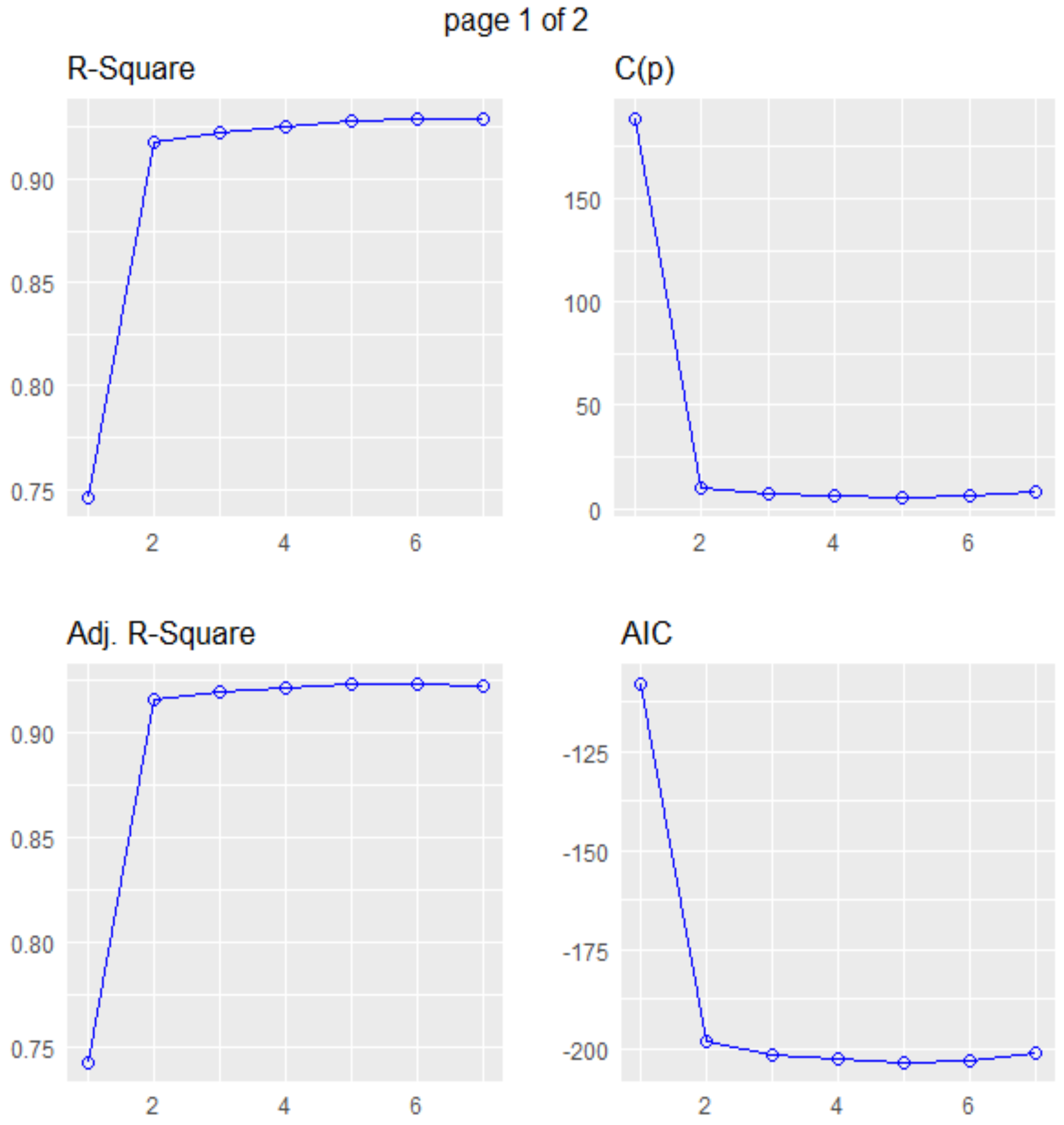
Subsets Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP
1	0.7457	0.7425	0.7327	188.5071	-107.7677	-344.2015	-100.5476	1.2289
2	0.9177	0.9156	0.9096	10.2956	-198.2328	-431.2423	-188.6059	0.4030
3	0.9227	0.9197	0.9135	7.0362	-201.3892	-434.0002	-189.3556	0.3834
4	0.9255	0.9216	0.9136	6.1100	-202.4050	-434.6178	-187.9646	0.3744
5	0.9281	0.9234	0.9145	5.2965	-203.4130	-435.0665	-186.5660	0.3657
6	0.9294	0.9237	0.9139	6.0171	-202.8184	-434.0376	-183.5646	0.3644
7	0.9294	0.9227	0.9117	8.0000	-200.8373	-431.8368	-179.1768	0.3693

AIC: Akaike Information Criteria ; SBIC: Sawa's Bayesian Information Criteria ;
SBC: Schwarz Bayesian Criteria ; MSEP: Estimated error of prediction

MLR equation:

$$'a' = 0.50346 - 0.62139 \times (R_{560}) + 0.45872 \times (R_{645}) - 0.15258 \times (R_{730}) + 0.17484 \times (R_{810})$$



Summary of criterion based selection approach to identify best subset of predictors for lycopene

Best Subsets Regression

Model Index	Predictors
1	`560`
2	`645` `730`
3	`560` `645` `730`
4	`535` `560` `645` `730`
5	`535` `560` `645` `730` `810`
6	`535` `560` `645` `680` `730` `810`
7	`460` `535` `560` `645` `680` `730` `810`

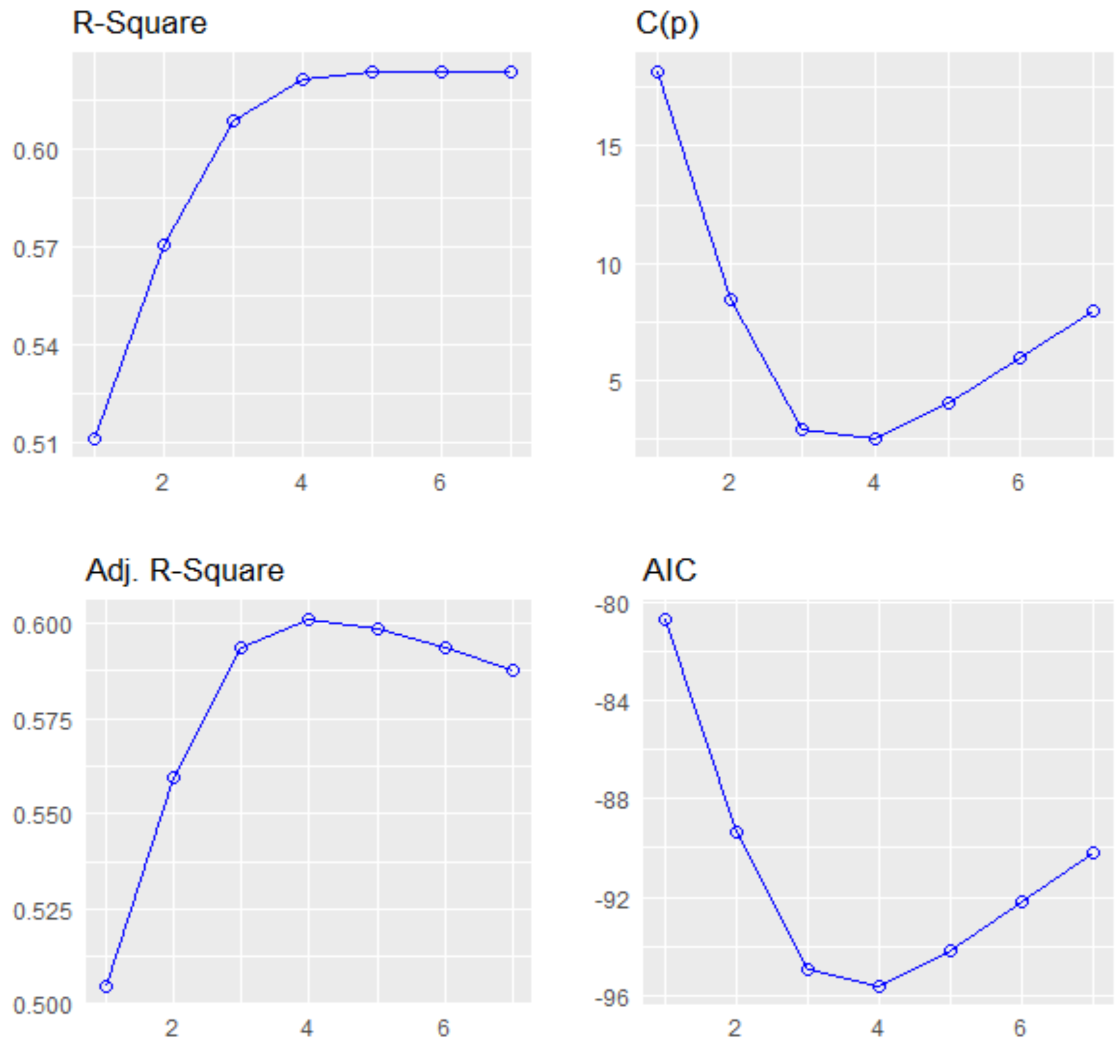
Subsets Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP
1	0.5108	0.5047	0.4806	18.1723	-80.6822	-314.0210	-73.4620	1.7099
2	0.5705	0.5596	0.539	8.4371	-89.3533	-322.2340	-79.7264	1.5205
3	0.6086	0.5935	0.5689	2.9552	-94.9615	-327.1516	-82.9279	1.4038
4	0.6210	0.6013	0.5704	2.5029	-95.6171	-327.3370	-81.1768	1.3769
5	0.6236	0.5988	0.5608	4.0062	-94.1655	-325.5986	-77.3185	1.3860
6	0.6236	0.5935	0.5503	6.0000	-92.1725	-323.3882	-72.9187	1.4046
7	0.6236	0.5880	0.5391	8.0000	-90.1725	-321.1720	-68.5120	1.4238

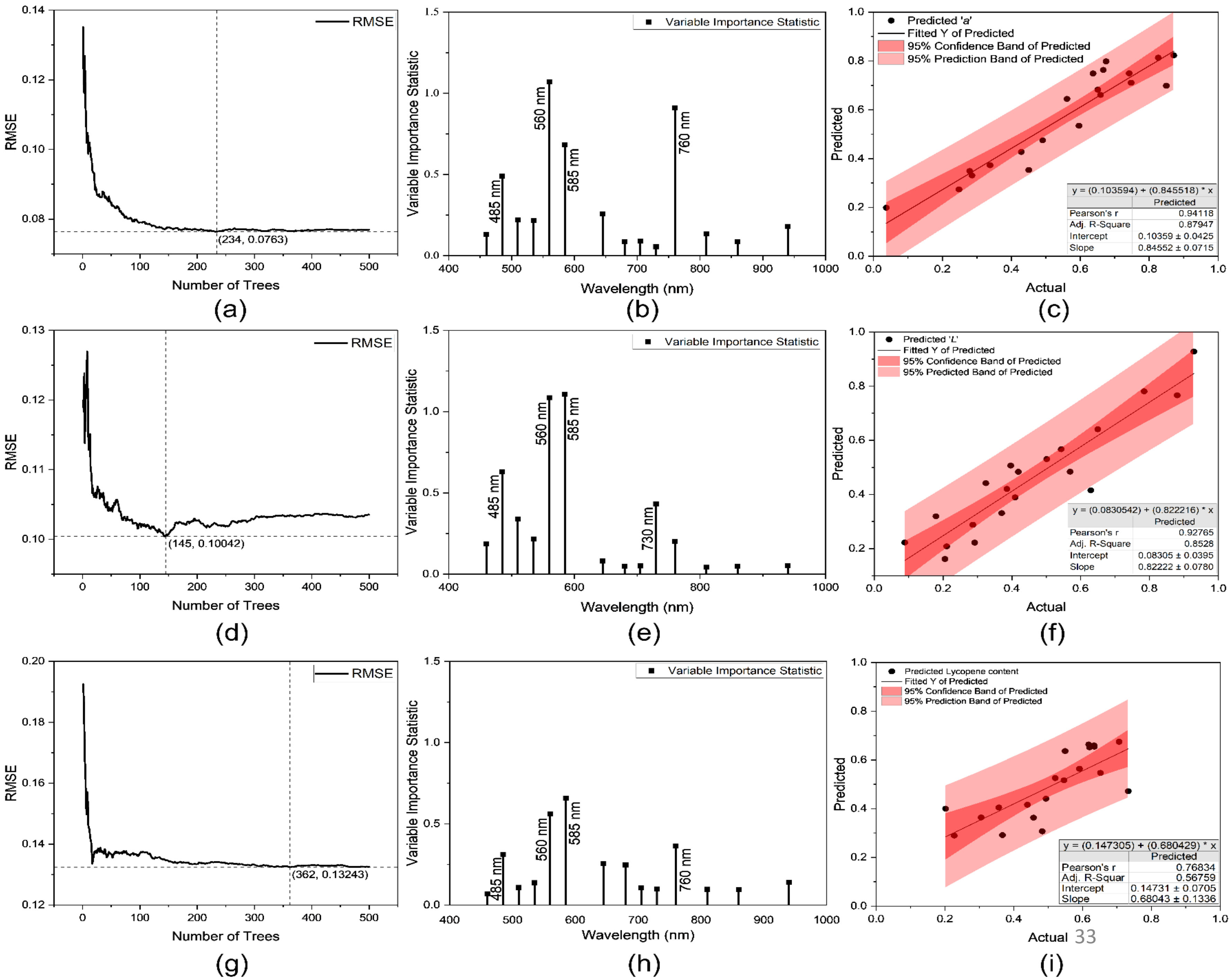
AIC: Akaike Information Criteria ; SBIC: Sawa's Bayesian Information Criteria ;
SBC: Schwarz Bayesian Criteria ; MSEP: Estimated error of prediction

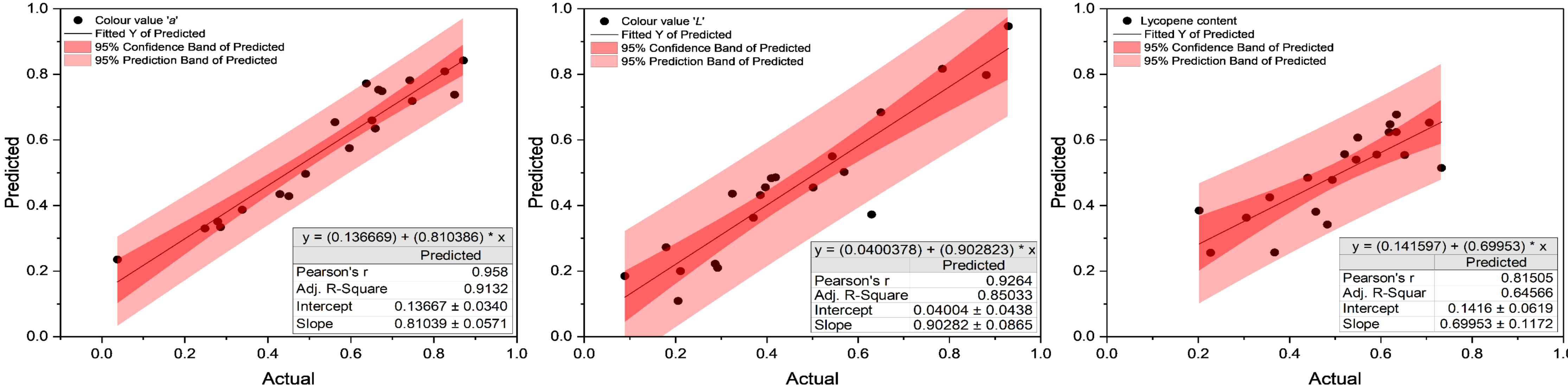
MLR equation:

$$\text{Lycopene} = 0.52382 - 0.30808 \times (R_{560}) + 0.43641 \times (R_{645}) - 0.31730 \times (R_{730})$$



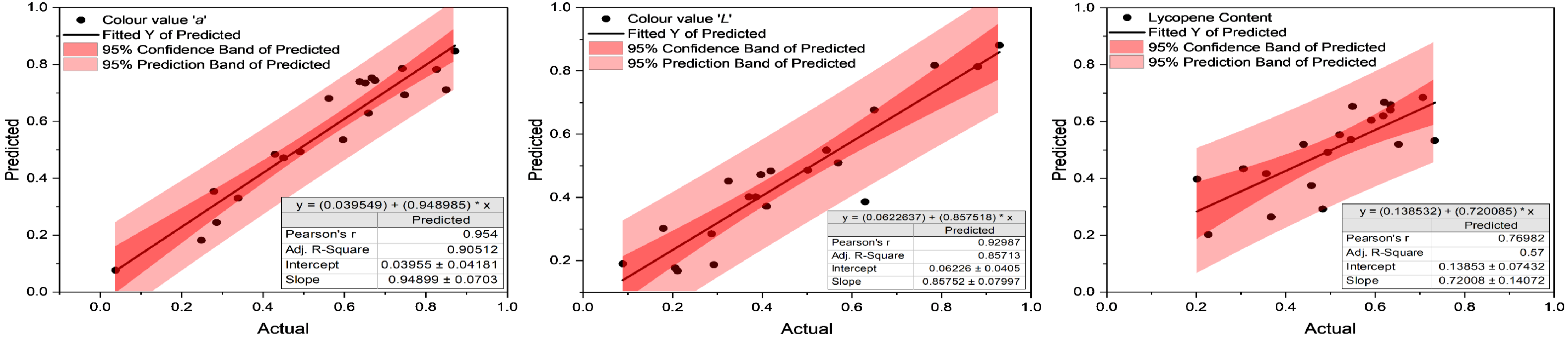
Number of trees in **Random Forest (RF)** versus RMSE (Train), variable importance statistic and linear fit line between actual values in test dataset and predicted values for '*a*' (a to c), '*L*' (d to f) and lycopene content (g to i).





(a) (b) (c)

Linear fit line between actual values and predicted values from **SVM** for ‘*a*’, ‘*L*’ and lycopene content at (a) to (c)



(a) (b) (c)

Linear fit line between actual values and predicted values from **ANN** for ‘*a*’, ‘*L*’ and lycopene content at (a) to (c)

Performance comparison of linear and non-linear models

	MLR			PCR			PLSR			ANN		RF	SVM	
Attributes	Best predictor subset (spectral wavelength - nm)	R^2 (CV)	RMSE (Test)	N_c	R^2 (CV)	RMSE (Test)	N_l	R^2 (CV)	RMSE (Test)	Size, Decay	RMSE (Test)	RMSE (Test)	N_s	RMSE (Test)
'a'	560, 645, 730, 810	0.9317^A	0.0608^e	3	0.9051^C	0.062^d	4	0.9255^B	0.0622^d	2, 0.001	0.0678^c	0.0774^a	63	0.0751^b
'L'	535, 560, 645, 730	0.8694^A	0.0869^a	3	0.8583^C	0.085^c	3	0.8657^B	0.0828^e	4, 0.01	0.0834^d	0.0852^c	66	0.0861^b
AOA	560, 645, 730	0.7476^A	0.1807^a	3	0.6673^C	0.155^c	3	0.6699^B	0.1559^b	1, 0.01	0.1497^d	0.1335^f	65	0.1479^e
'b'	460, 535, 730	0.6791^A	0.1651^a	3	0.5096^C	0.1645^b	5	0.5975^B	0.1348^f	1, 0.001	0.1527^d	0.1632^c	70	0.1518^e
Lycopene	560, 645, 730	0.6361^A	0.107^a	3	0.587^B	0.098^b	3	0.5808^C	0.0977^c	2, 0.01	0.098^b	0.0978^{bc}	72	0.087^d
TFC	535, 560, 645	0.4313^A	0.183^b	3	0.3712^C	0.1761^d	3	0.3738^B	0.1707^e	1, 0.01	0.1569^f	0.1876^a	67	0.1778^c
TPC	560, 810	0.4809^A	0.191^a	3	0.3712^B	0.1843^d	1	0.3425^C	0.1899^b	2, 0.01	0.1886^c	0.1794^e	77	0.1776^f
TDS	645, 730	0.471^A	0.183^b	3	0.3862^B	0.1762^f	3	0.3599^C	0.1765^e	4, 0.01	0.1787^d	0.1905^a	75	0.182^c
EC	730, 810	0.4151^A	0.1653^c	3	0.3456^B	0.1547^f	3	0.3174^C	0.1563^d	1, 0.01	0.1557^e	0.1677^b	71	0.1685^a
Salinity	560, 680	0.4295^A	0.1883^a	3	0.3366^B	0.1818^d	3	0.3137^C	0.1825^b	1, 0.01	0.1821^c	0.1647^f	73	0.1675^e
TSS	460, 560, 730	0.3623^A	0.1514^e	4	0.2986^B	0.1461^b	3	0.2909^C	0.1425^e	1, 0.001	0.144^c	0.1291^f	69	0.1436^d
Firmness	645, 730	0.3255^A	0.1628^f	3	0.235^B	0.1644^e	1	0.2052^C	0.1905^a	2, 0.1	0.1857^c	0.1894^b	77	0.1744^d
pH	560, 680	0.3229^A	0.1222^e	4	0.2175^C	0.1236^c	3	0.2248^B	0.1248^b	2, 0.001	0.1178^f	0.1225^d	73	0.1264^a
Acidity	460	0.2381^A	0.1511^c	1	0.0165^B	0.1557^b	1	0.00814^C	0.151^c	1, 0.01	0.1631^a	0.1498^d	71	0.1433^e

Means that do not share a letter are significantly different. Upper-case letter indicate statistical difference between linear models. Lower-case letter indicate statistical difference between linear and non-linear models. N_c = number of PCs in PCR, N_l = number of LVs in PLSR, N_s = number of support vectors in SVM.

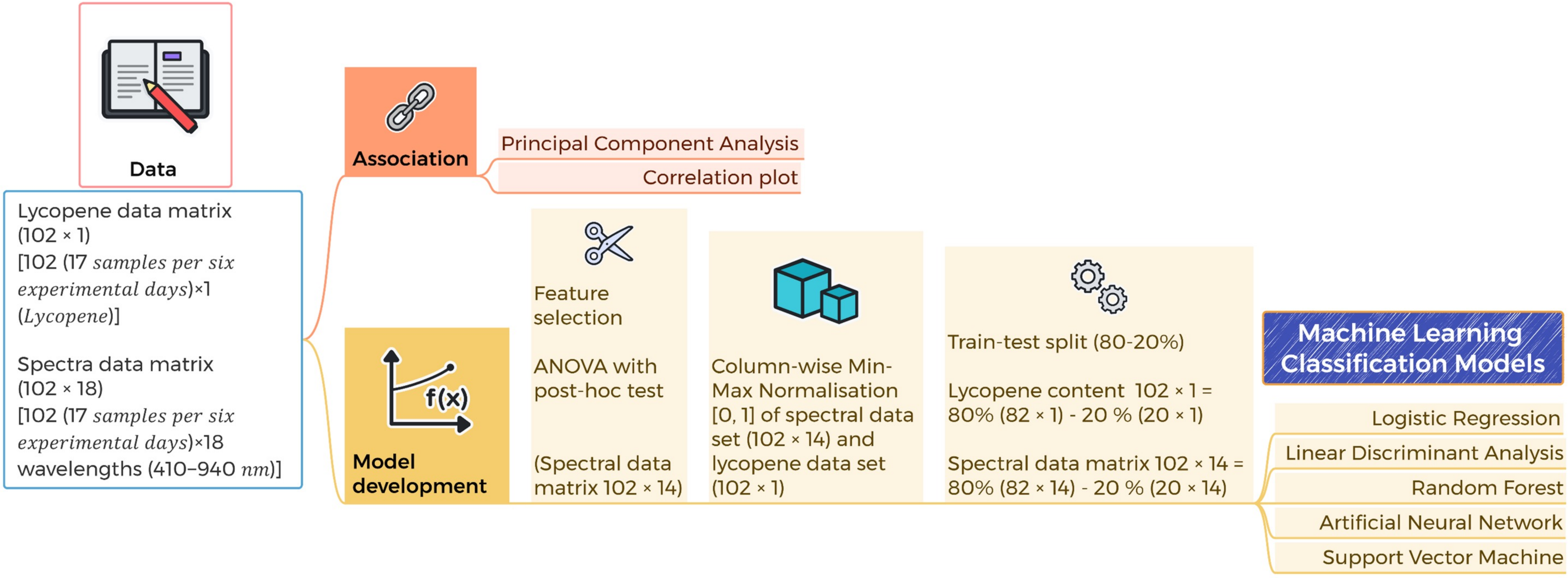
COMPARATIVE STUDY WITH LITERATURE

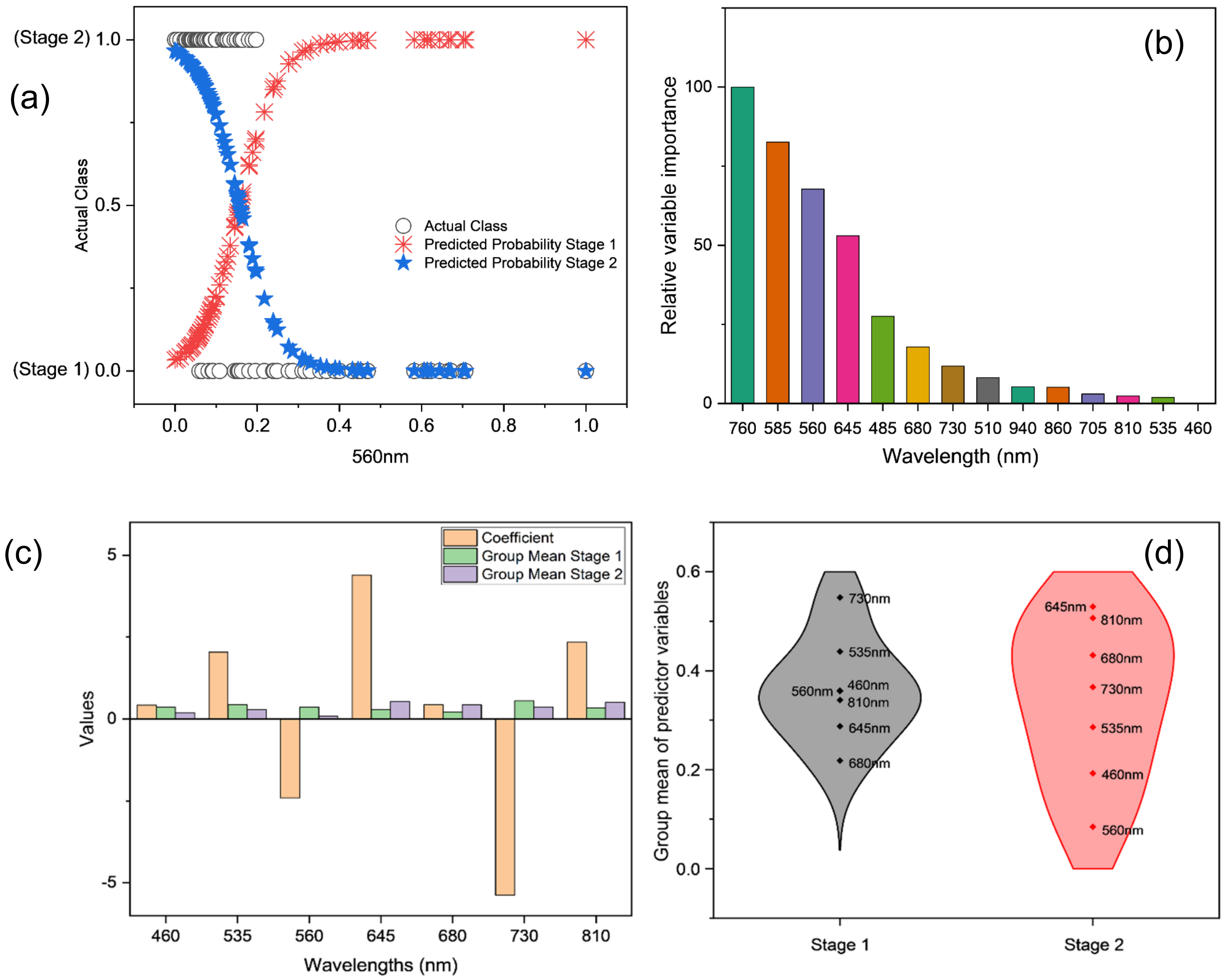
- Comparison with latest research on Vis-SWNIR spectroscopy-based prediction of physicochemical attributes of tomatoes from diverse regions (South America, Europe, East Asia, and Australia).

Study	Region	Attribute(s)	Spectrometer	R^2 (Calibration)	Model Type
Present Study	India	Colour (a, L, b), Lycopene	Proposed spectrometer assembly	0.93, 0.86, 0.68, 0.64	MLR
Brito et al. (2022)	Brazil	Colour, TA, DM	Portable F-750 (USA) (Interactance Mode)	0.85-0.94, 0.52-0.39	PLSR
Égei, Márton, et al. (2022)	Hungary	Lycopene	FieldSpec HandHeld 2TM	0.52-0.66	PLSR
Tilahun et al. (2018)	Korea	Colour (L, a, b)	Vis-SWNIR (Life & Tech, CO, Ltd, Yongin, Korea)	0.49, 0.92, 0.52	PLSR
Acharya et al. (2017)	Australia	DM, Colour (a)	Portable F-750 (USA)	0.85-0.96	PLSR

NOTE: In order to ensure **meaningful and accurate comparison** of results from different studies, it is crucial to consider **similar experimental design settings** such as fruit variety, cultivar and genotype, physiological maturity of fruit at harvest, storage conditions, spectroscopy instrumentation, mode of spectral data acquisition, source of illumination and type of detectors employed for spectra acquisition, data pre-processing and model development techniques such as cross-validation among other things.

Development of machine learning classification models for estimation of physicochemical attributes using spectral data





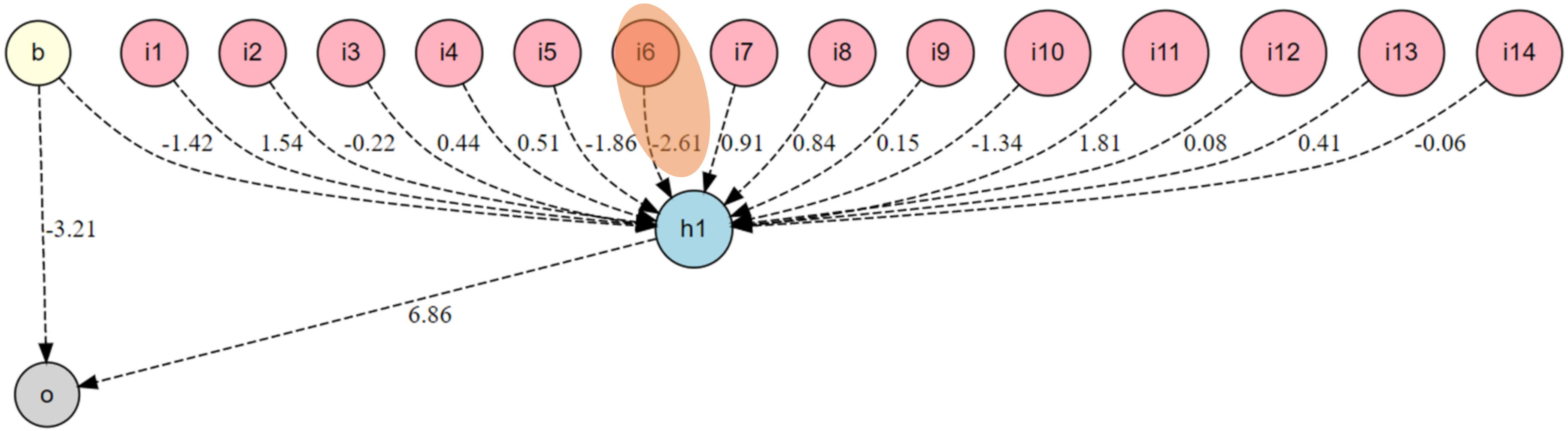
Development of machine learning classification models

The assessment of quantitative and qualitative association using correlation and PCA biplot revealed that wavelength range 500-600 nm was inversely but strongly associated with lycopene content.

The findings in this study align with existing literature, which has consistently linked lycopene content to wavelengths in the range of 500–750 nm.

(a): logistic regression (b): random forest (c and d): linear discriminant analysis

Artificial neural network (ANN) and comparison of confusion matrix



Predictor i6 (585 nm) had strongest influence on the classification outcome compared to the other predictor variables

Table 1 *A. Sharma et al. Vibrational Spectroscopy 130 (2024) 103628*

Confusion matrix and performance metrics of different classification models on test dataset.

	Probabilistic classifiers				Non-probabilistic classifiers					
	Logistic Regression		LDA		RF		ANN		SVM	
Accuracy on test dataset	80%		90%		80%		95%		90%	
Confusion Matrix	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2
Stage 1	7	1	8	0	6	0	9	0	8	0
Stage 2	3	9	2	10	4	10	1	10	2	10
95% Confidence Interval	(0.5634, 0.9427)		(0.683, 0.9877)		(0.5634, 0.9427)		(0.7513, 0.9987)		(0.683, 0.9877)	
F-1 score	0.818		0.909		0.833		0.952		0.909	
Sensitivity (Recall)	0.7		0.8		0.6		0.9		0.8	
Specificity	0.9		1		1		1		1	
Positive predicted Value (Precision)	0.875		1		1		1		1	
AUC for Stage 1	0.87		0.99		0.97		0.98		0.99	
AUC for Stage 2	0.87		0.99		0.97		0.98		0.99	
Log loss for Stage 1	6.508247		3.034263							
Log loss for Stage 2	2.728349		1.125162							

RESULTS AND DISCUSSION

- Close agreement between colorimeter and spectral data confirmed reliability and scientific validity of proposed spectrometer.
- Statistical and chemometric analysis revealed blue (380-440nm) and green (440-600nm) spectra varied with attributes associated with water content loss, while red (600-750nm) and SWNIR (750-1100nm) spectra varied with attributes associated with carotenoid content, such as lycopene, antioxidant activity, and colour value ‘a’.
- Non-linear models captured underlying relationships despite multi-collinearity among predictor variables while MLR showed outstanding performance with high R-squared and competitive RMSE (Test) results.
- MLR was most suitable for predicting colour value 'a' and correlated attributes like lycopene and antioxidant activity for specific tomato genotype and storage conditions.
- 500–750 nm wavelength range dominated the classification of tomatoes based on lycopene content
- ANN surpassed logistic regression, LDA, RF, and SVM, achieving a 95% test accuracy.

CONCLUSION

- Research aimed at non-destructive quality evaluation of raw tomatoes using a portable spectrophotometer integrating multi-spectral sensor chipset and open-source microcontroller.

- 15-day storage study collected voluminous data for 14 physicochemical attributes through lab estimation and non-destructive spectral data acquisition in reflectance mode.

- This study presented a unique spectroscopic approach for the non-invasive quality assessment through machine learning regression as well as classification of tomatoes.

- Notably, the overall range within which a particular compound exhibits its spectroscopic characteristics tends to remain consistent, however, specific precise classifier wavelength may vary due to different factors including instrumentation, fruit variety, physiological maturity of fruit at harvest, storage conditions, light source and sensors used for spectral acquisition, data pre-processing, and model development techniques like cross-validation, among others.

FUTURE WORK

- In order to ensure widespread applicability of this technology, it is recommended to replicate this study under different experimental settings on other climacteric fruits.

COMMERCIAL RELEVANCE OF WORK

- The proposed technology is going to be a win-win situation for both consumers and farmers.

- On one hand, it will empower consumers with data driven decisions and access to premium quality fruits and vegetables while on other hand premium quality horticulture produce command better prices, enabling farmers to increase their income.

- The real-time assessment capability of the instrument will also help farmers to identify fruits with potential issues before they deteriorate further. This will reduce post-harvest losses, ensuring more of the farmers' produce reaches the market in optimal condition, further increasing their income by efficient resource utilization.

PUBLICATIONS

Book Chapter published

- Krishna G., **Arun Sharma**, Neela Emanuel, Pramod K. Prabhakar, and **Ritesh Kumar**. "Sensors for Non-Destructive Quality Evaluation of Food." *Food Chemistry: The Role of Additives, Preservatives and Adulteration* (2021): 397-449.

Research Papers Published

- **Arun Sharma**, A.D. Tiwari, Monika Kumari, Nishant Kumar, Vikas Saxena, and **Ritesh Kumar**. "Artificial intelligence-based prediction of lycopene content in raw tomatoes using physicochemical attributes." *Phytochemical Analysis* (**2022**). [Impact Factor – 3.3]

- **Arun Sharma**, **Ritesh Kumar**, Nishant Kumar, Kuljinder Kaur, Vikas Saxena, and Priyadeep Ghosh. "Chemometrics driven portable Vis-SWNIR spectrophotometer for non-destructive quality evaluation of raw tomatoes." *Chemometrics and Intelligent Laboratory Systems* 242 (**2023**): 105001. [Impact Factor – 3.9]

- **Arun Sharma**, **Ritesh Kumar**, Nishant Kumar, and Vikas Saxena. "Machine learning driven portable Vis-SWNIR spectrophotometer for non-destructive classification of raw tomatoes based on lycopene content." *Vibrational Spectroscopy* (**2024**): 103628. [Impact Factor – 2.5]

INTELLECTUAL PROPERTY RIGHT

- An Intellectual Property Right in terms of 'Design Registration' titled 'Artificial Intelligence based Table-Top Spectrophotometer for Non-Destructive Quality Evaluation of Fruits and Vegetables' has been **granted** by Indian Patent Office on April 19, 2024.

INTELLECTUAL PROPERTY INDIA
PATENTS | DESIGNS | TRADE MARKS
GEOGRAPHICAL INDICATIONS

सत्यमेव जयते
GOVERNMENT OF INDIA

Controller General of Patents,Designs and Trademarks
Department of Industrial Policy and Promotion
Ministry of Commerce and Industry

Design Application Details

Application Number:	376671-001
Cbr Number:	200137
Cbr Date:	03/01/2023 15:03:00
Applicant Name:	**1. COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH** **2. NATIONAL INSTITUTE OF FOOD TECHNOLOGY ENTREPRENEURSHIP AND MANAGEMENT (NIFTEM)**

Design Application Status

Application Status:	Design Accepted and Published, Journal No is 16/2024 and Journal Date is 19/04/2024

Back

Progressive Colour Change with Storage Time

0th day

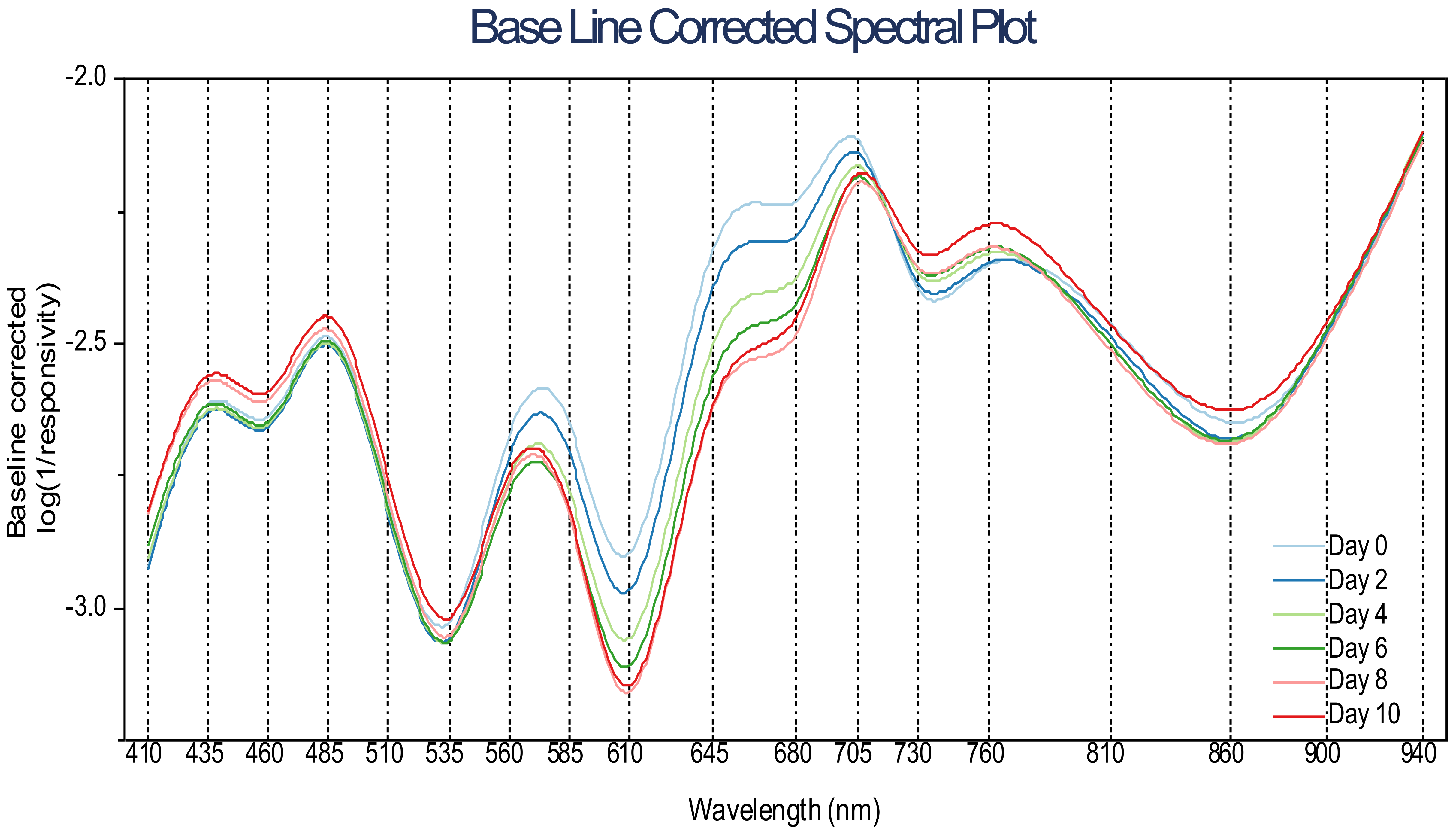
2nd day

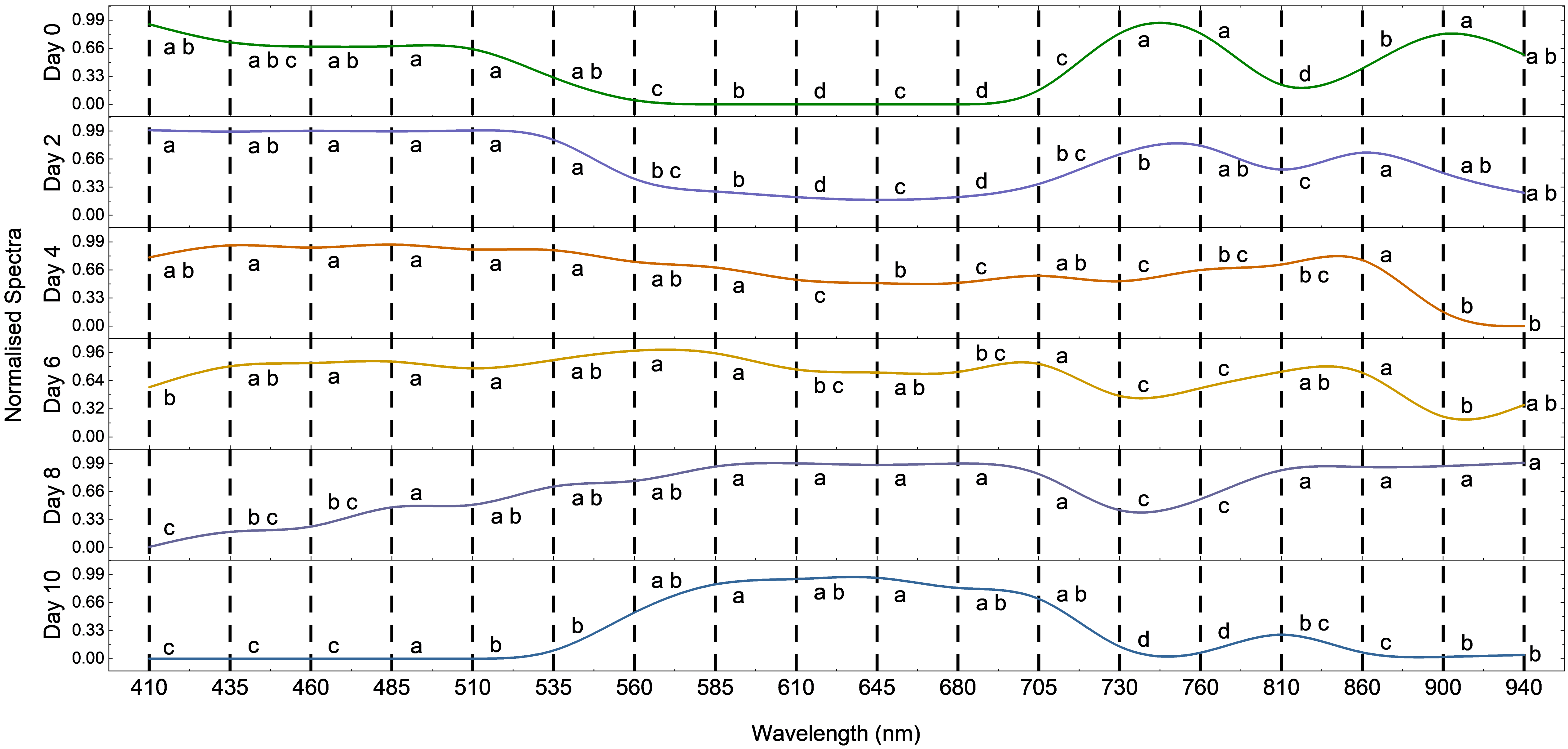
4th day

6th day

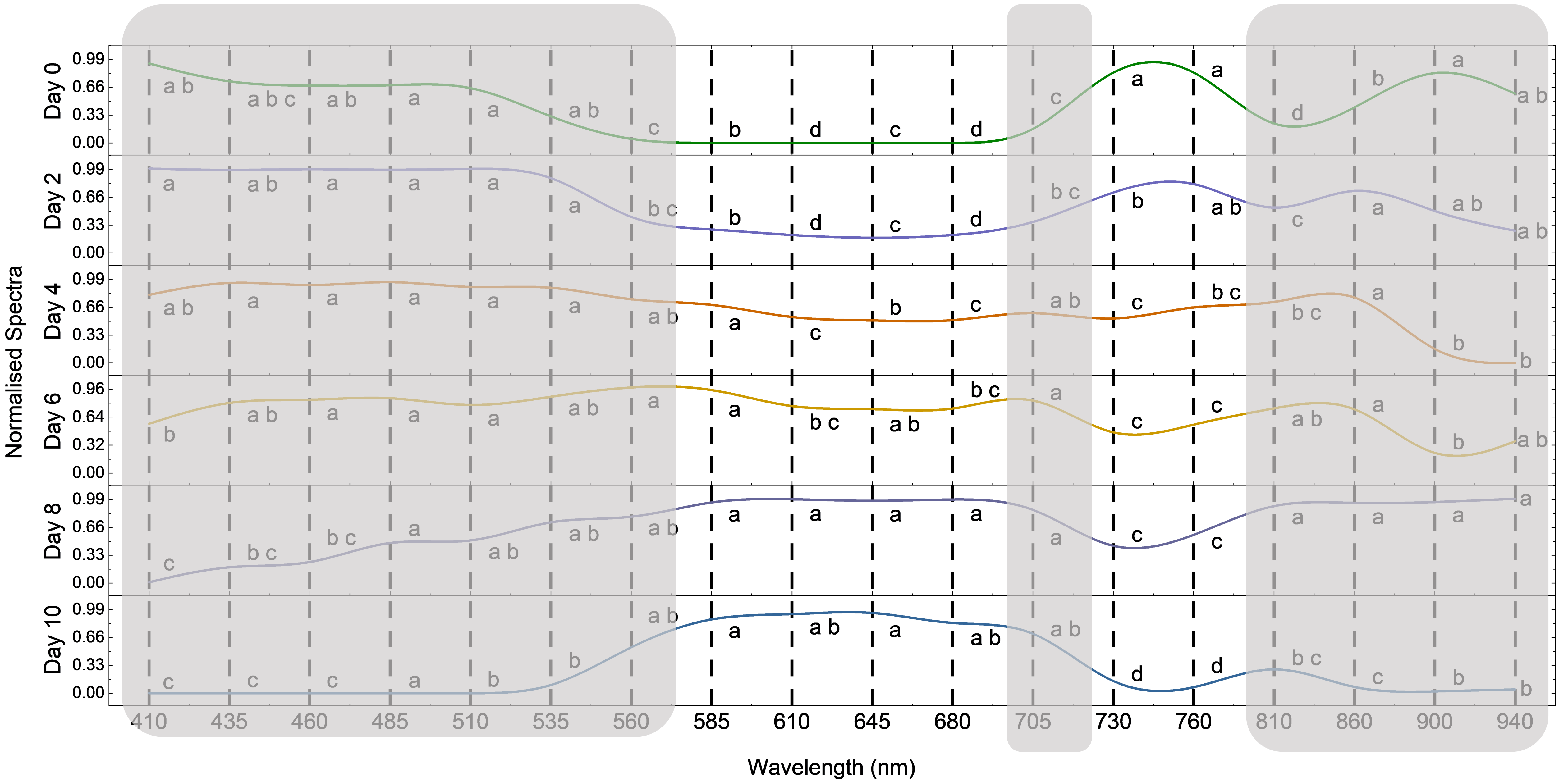
8th day

10th day

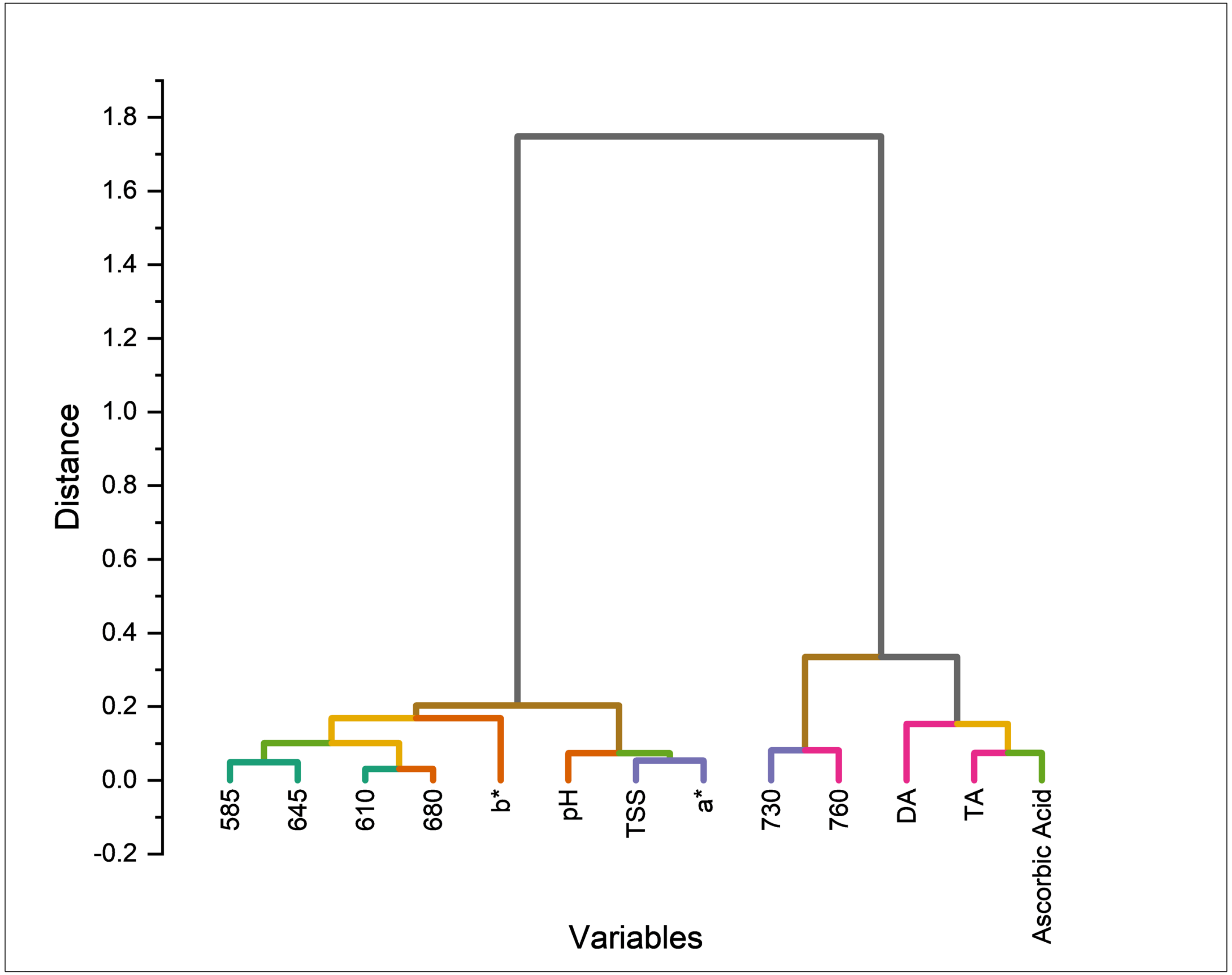

Base Line Corrected Spectral Plot
Baseline corrected
log(1/responsivity)
-2.0
-2.5
-3.0
410
435
460
485
510
535
560
585
610
645
680
705
730
760
810
860
900
940
Wavelength (nm)
Day 0
Day 2
Day 4
Day 6
Day 8
Day 10



Normalised Spectra
Day 0
Day 2
Day 4
Day 6
Day 8
Day 10
0.99
0.66
0.33
0.00
0.96
0.64
0.32
410
435
460
485
510
535
560
585
610
645
680
705
730
760
810
860
900
940
Wavelength (nm)

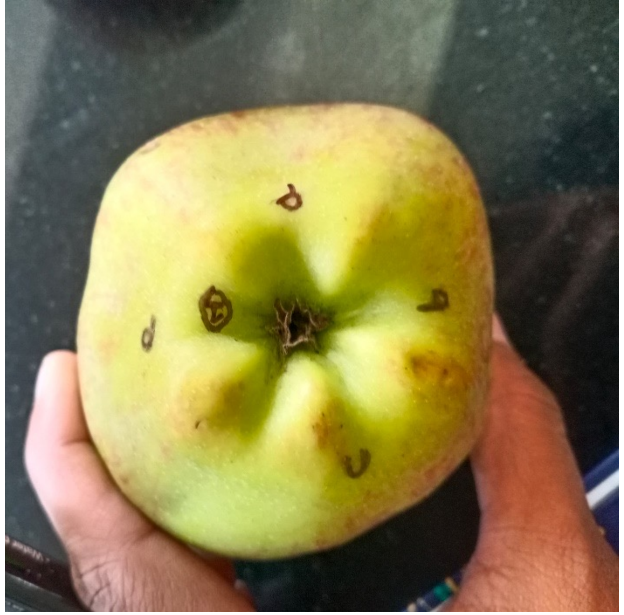

Normalised Spectra
Day 0
Day 2
Day 4
Day 6
Day 8
Day 10
0.99
0.66
0.33
0.00
0.96
0.64
0.32
410
435
460
485
510
535
560
585
610
645
680
705
730
760
810
860
900
940
Wavelength (nm)

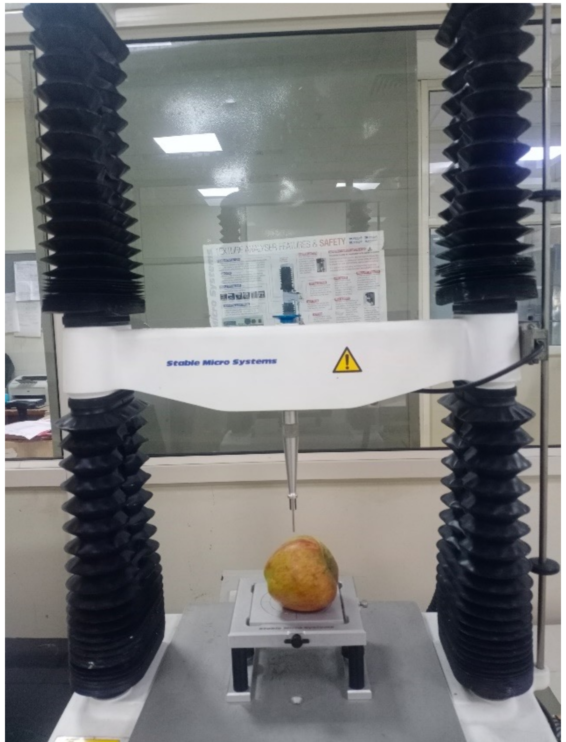
Hierarchical Cluster Analysis (HCA)




Physicochemical Analysis

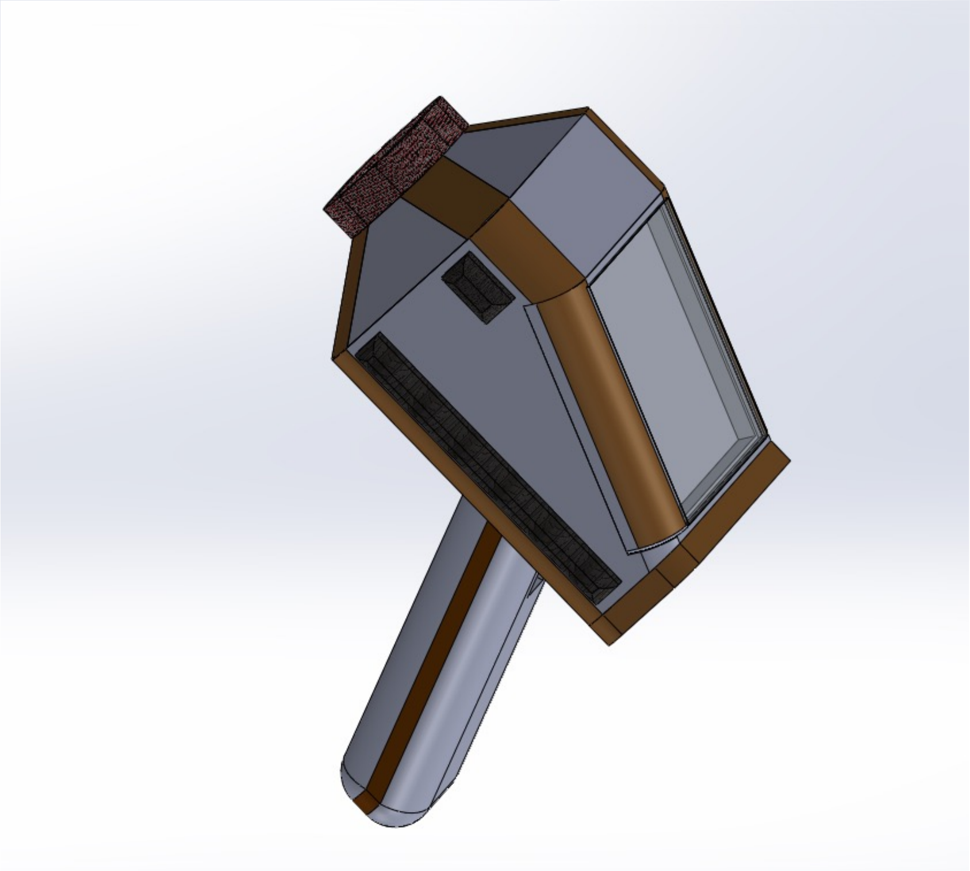

Non-Destructive analysis

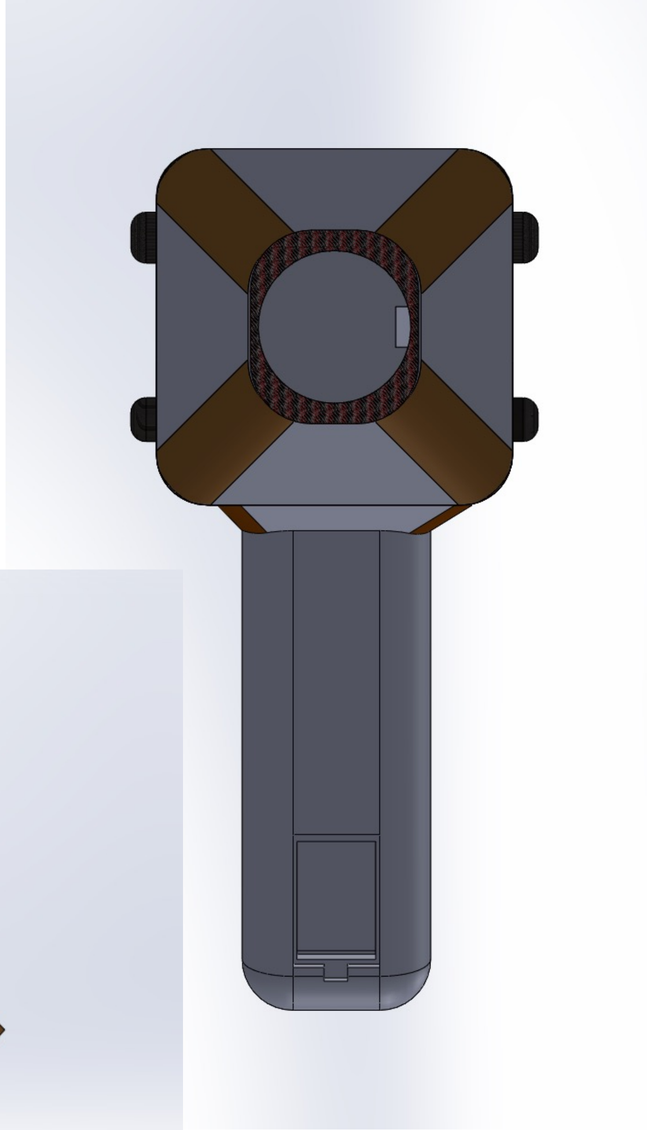



Stable Micro Systems


L*= 93.03
a*= -0.40
b*= 4.50
CAL
DEL UNDO
ESC
COLOR
ENTER
READY
T


Mean (1)
Last 03/04/24 15:51:34 ID=130
ΔA:0.95
03/04/24 15:51:40





ACKNOWLEDGEMENTS

Thankful acknowledgement to tinyML Foundation for the opportunity to present my work at tinyML Summit-2024.

I thankfully acknowledge following organisations for providing infrastructural and technical support for conducting the study:

- Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India
- Council of Scientific and Industrial Research–Central Scientific Instruments Organisation (CSIR–CSIO), Sector 30, Chandigarh, India
- National Institute of Food Technology Entrepreneurship and Management (NIFTEM), Kundli, Sonipat, Haryana, India

Thankful acknowledgement to audience for attending the presentation.

Open for collaborations and investors/start-ups

Thank You

ARUN SHARMA

Email: arunsharma1712@gmail.com

WhatsApp & Mobile: +91-9416656492