Paint analysis using multi-spectral data

James M. Curran[1], Wen Zhijian[1], Morgan Carpenter[2], Patrick Buzzini[2]

1. The University of Auckland, New Zealand
2. Sam Houston State University, Huntsville, TX, USA

Disclosure

This project is supported by Award No. 15PNIJ-21-GG-04163-RESS awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect those of the Department of Justice.

A case:

- A vehicle is involved in a hit-and-run accident.
- Paint chips are recovered from the victim's car at the point of impact.
- A suspect vehicle is located, showing **damage to its paintwork** consistent with the collision.

A Forensic Question:
Can the paint on the victim's car be **matched to the suspect vehicle**?
Are there ways to **distinguish between paints of the same colour** to identify the source?

A Forensic Research Question

- If I have more than one type of instrument to quantify the evidence, is there *value to be gained* in using all of them?
- That is – do we *improve* things by using them all?

- Additionally – we know how to collect the data, but how do we use the data from multiple instruments?
- This has been labelled *multiblock* analysis, or more recently *data fusion*.

What do we mean by *add value* or *improve*?

Remember (annoying) people like me are ultimately interested in computing the likelihood ratio

$$\underbrace{\frac{\Pr(H_1 \mid E, I)}{\Pr(H_2 \mid E, I)}}_{\text{Posterior Odds}} = \underbrace{\frac{\Pr(E \mid H_2, I)}{\Pr(E \mid H_1, I)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H_1, I)}{\Pr(H_2, I)}}_{\text{Prior Odds}}$$

What do we mean by *add value* or *improve*?

If we take a two stage approach – i.e. there is a match step, then the *LR* will simplify to

$$\frac{\Pr(E \mid H_2, I)}{\Pr(E \mid H_1, I)} = \frac{1}{f}$$

If we don't then the *LR* might look something like

$$\frac{\Pr(E \mid H_2, I)}{\Pr(E \mid H_1, I)} = \frac{f\left(\bar{X} \mid \bar{Y}, C, H_1, I\right)}{f\left(\bar{Y} \mid, U, H_2, I\right)}$$

What do we mean by *add value* or *improve*?

- In the matching paradigm, we want to see if we can improve classification accuracy
  - That is we want to increase true positives, and true negatives
  - And to decrease false positives and false negatives

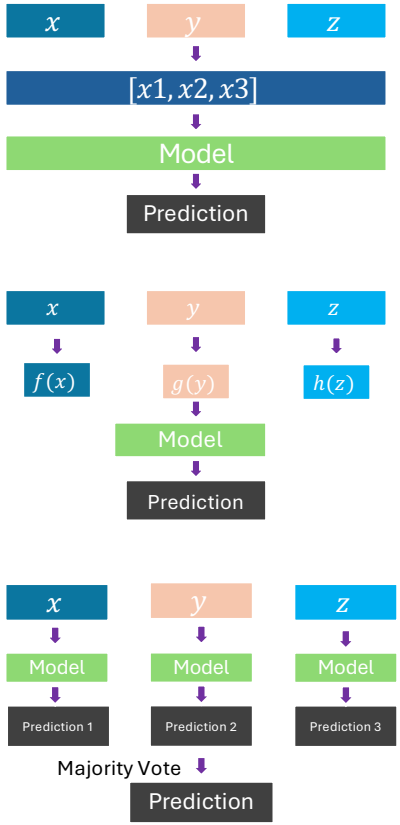
- In the continuous paradigm (no match step), it is little harder to define.
  - We might want to say we want to refine the posterior probabilities that a measurement comes from a particular class.
  - This is not the focus of today's talk

Data Fusion

Low-Level Fusion (Data-Level Fusion)
- Combines raw data from multiple sources.
- Maintains most information but may be noisy and computationally heavy.

Medium-Level Fusion (Feature-Level Fusion)
- Extracts features from each source first, then fuses them.
- Reduces data size, highlights relevant information.

High-Level Fusion (Decision-Level Fusion)
- Each source is analysed independently; results or decisions are combined.
- Easy to implement, robust to noise, but may lose detailed information.



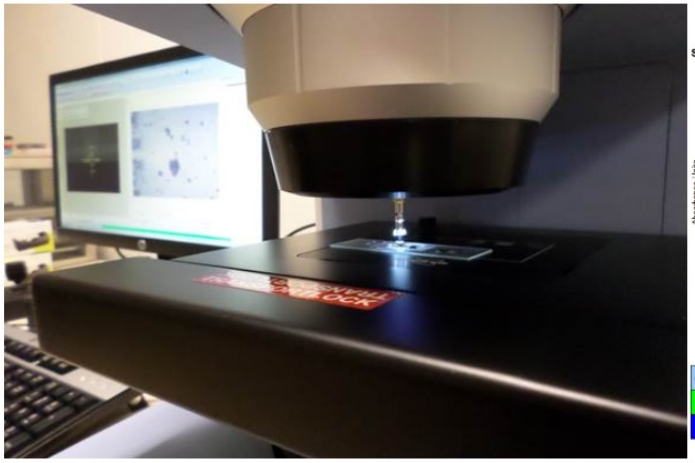
Data

- Five different suppliers
- Five different colour
- Seven samples per spray paint

Suppliers	Colours	Name	Vendor
Ace	Black	12 oz. Premium Gloss Black Paint + Primer Enamel Spray	Home Depot
Ace	Blue	12 oz. Premium Gloss Navy Paint + Primer Enamel Spray	Home Depot
Ace	Red	12 oz. Premium Gloss Banner Red Paint + Primer Enamel Spray	Home Depot
Ace	Silver	12 oz. Premium Gloss Chrome Aluminum Paint + Primer Enamel Spray	Home Depot
Ace	White	12 oz. Premium Gloss White Paint + Primer Enamel Spray	Home Depot
Behr	Black	12 oz. Black Gloss Interior/Exterior Spray Paint	Home Depot
Behr	Blue	12 oz. M510-7 Inked Gloss Interior/Exterior Spray Paint	Home Depot
Behr	Gray	12 oz. N460-3 Lunar Surface Gloss Interior/Exterior Spray Paint	Home Depot
Behr	Red	12 oz. P150-7 Flirt Alert Gloss Interior/Exterior Spray Paint	Home Depot
Behr	White	12 oz. 52 White Gloss Interior/Exterior Spray Paint	Home Depot
Colorshot	Blue	10 oz. Gloss Stamped Passport Navy Blue Spray Paint	Home Depot
Colorshot	Black	10 oz. Gloss Little Black Dress Black Spray Paint	Home Depot
Colorshot	Red	10 oz. Gloss Stiletto Red Spray Paint	Home Depot
Colorshot	Silver	9 oz. Metallic Silver Lining Silver Spray Paint	Home Depot
Colorshot	White	10 oz. Gloss Marshmallow White Spray Paint	Home Depot
Krylon	Black	12-oz Krylon COLORmaxx Gloss Black Spray Paint	Lowes
Krylon	Blue	12-oz Krylon COLORmaxx Gloss Navy Blue Spray Paint	Lowes
Krylon	Red	12-oz Krylon COLORmaxx Gloss Banner Red Spray Paint	Lowes
Krylon	White	12-oz Krylon COLORmaxx Gloss White Spray Paint	Lowes
Krylon	Silver	12-oz Krylon Glitter Blast Gloss Silver Flash Glitter Spray Paint	Lowes
Rust-Oleum	Silver	11 oz. Metallic Silver Protective Spray Paint	Home Depot
Rust-Oleum Painter's Touch 2X	Black	12 oz. Gloss Black General Purpose Spray Paint	Home Depot
R	ter's Touch	12 oz. Gloss Navy Blue General Purpose Spray Paint	Home Depot
R	ter's Touch	12 oz. Gloss Apple Red General Purpose Spray Paint	Home Depot
R	ter's Touch	12 oz. Gloss White General Purpose Spray Paint	Home Depot

Instruments

- UV–Vis Micro spectrophotometry
- Scanning Electron Microscopy / Energy Dispersive Spectroscopy (SEM/EDS)
- Fourier Transform Infrared Spectroscopy (FTIR)
- Raman Spectroscopy (532 nm)
- Raman Spectroscopy (785 nm)



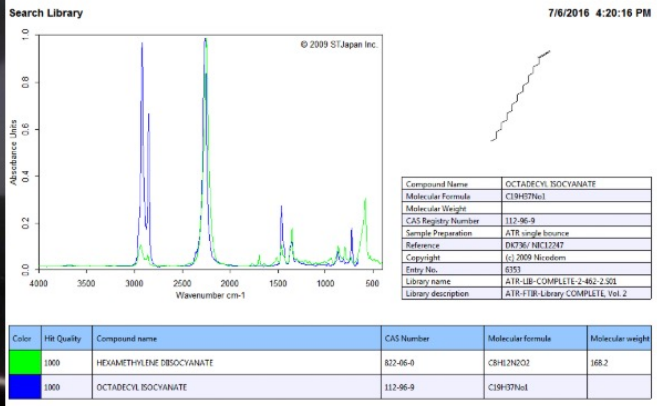
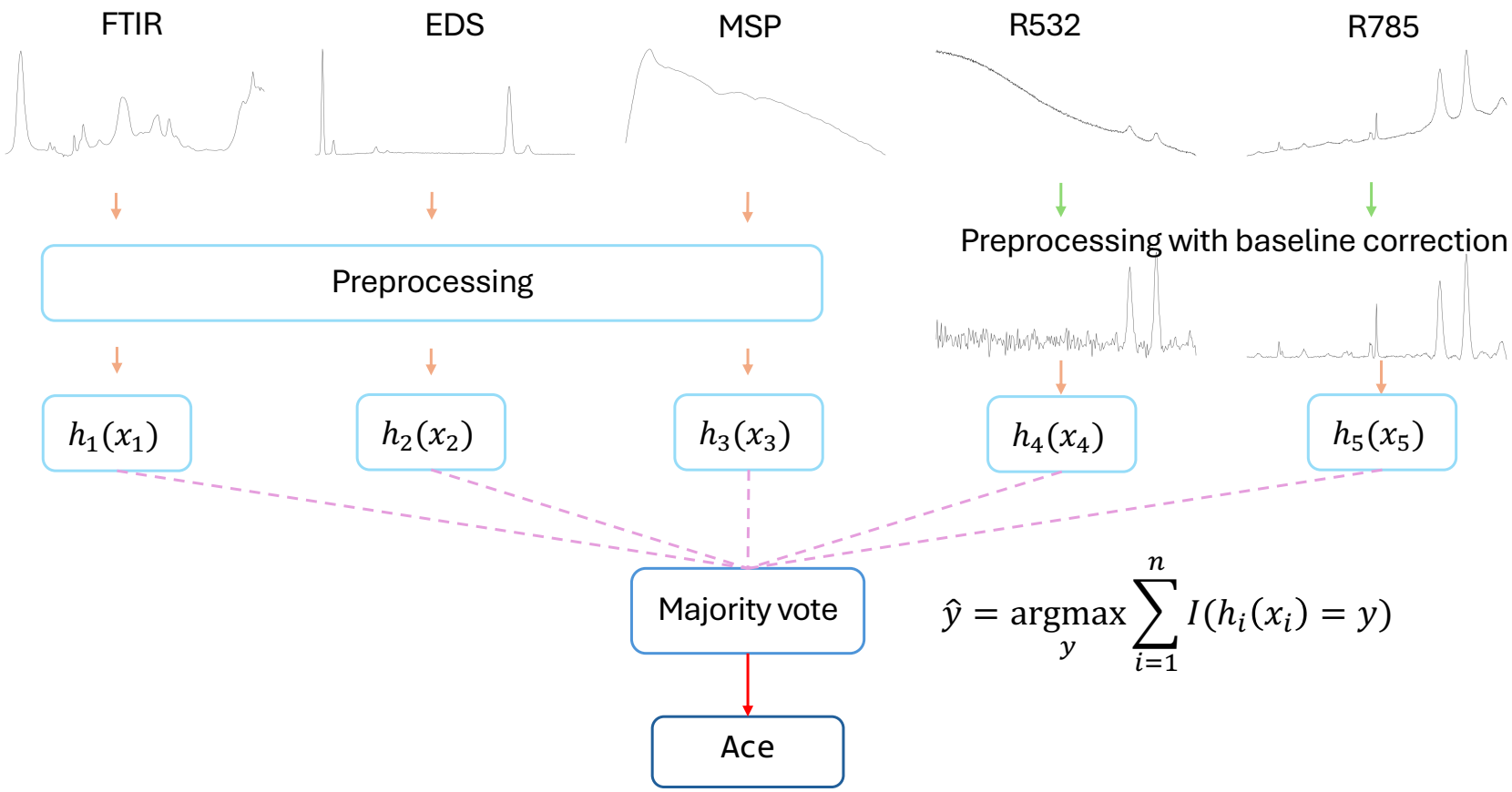
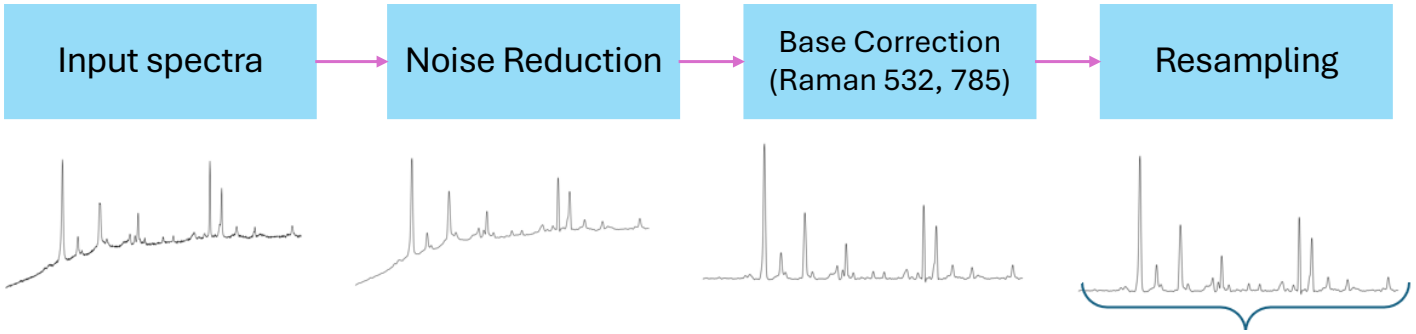


Image from: https://www.advancedmicroanalytical.com/AMAServices.aspx?mode=tech&ID=15

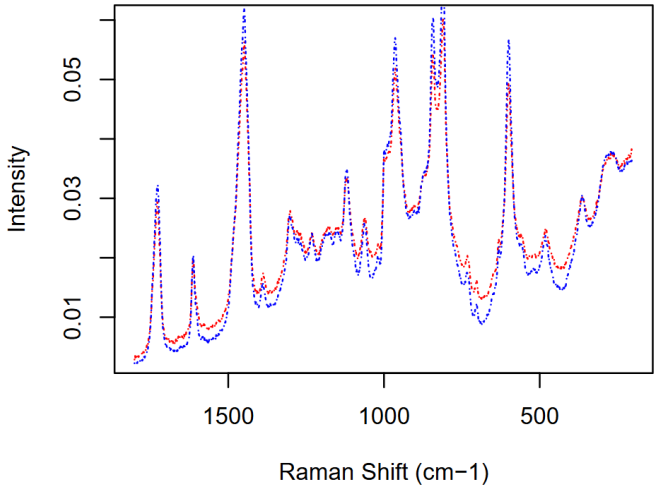
Method


FTIR
EDS
MSP
R532
R785
Preprocessing
Preprocessing with baseline correction
$h_1(x_1)$
$h_2(x_2)$
$h_3(x_3)$
$h_4(x_4)$
$h_5(x_5)$
Majority vote
Ace
$\hat{y} = \underset{y}{\operatorname{argmax}} \sum_{i=1}^{n} I(h_i(x_i) = y)$

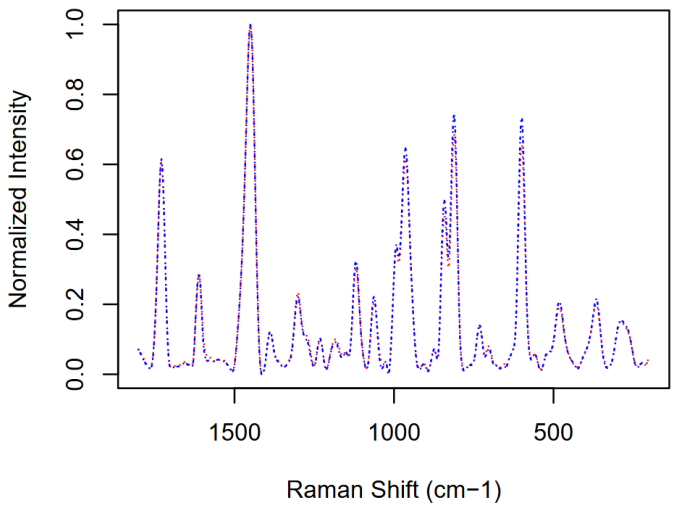
Spectrum preprocessing



Before preprocessing (RoI)



After preprocessing (RoI) $1433 \rightarrow 512$



Aim: Can we predict manufacturer within the same colour?

Model/Classifiers

- KNN: K-Nearest Neighbours
- Naive Bayes
- Random Forest
- SVM: Support Vector Machine
- XGBoost: Extreme Gradient Boosting

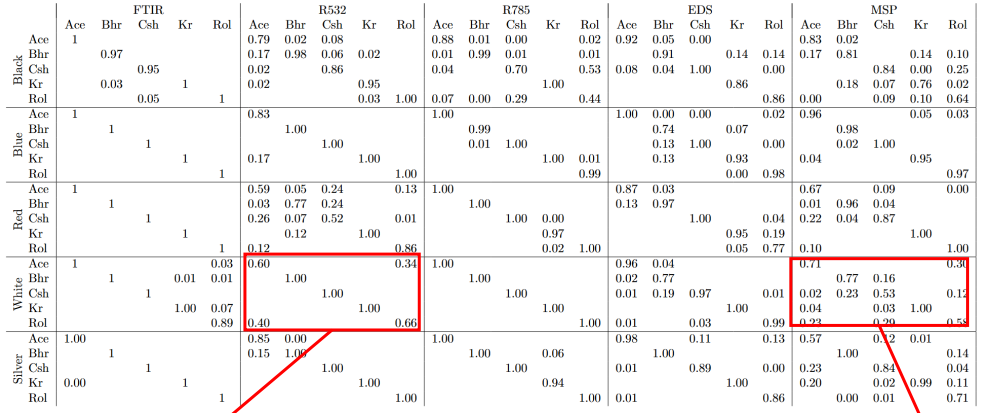
Training

- Randomly Split data into a training set (80%) and a testing set (20%)
- Employ a 3 fold-cross validation in the training set to optimise the hyperparameters
- Use testing set to evaluate the model performance
- Repeat the experiment 20 times to compute the average accuracy in the test set.

Training and Classification result:

Paint Colour	Classifier	MV	FTIR	R532	R785	EDS	MSP
Black	KNN	1.00	1.00	0.86	0.73	0.95	0.68
	Naive Bayes	0.99	0.97	0.83	0.73	0.87	0.65
	Random Forest	1.00	0.98	0.92	0.80	0.91	0.77
	SVM (Linear Kernel)	0.99	0.89	0.56	0.77	0.73	0.69
	XGBoost	0.98	0.85	0.73	0.68	0.81	0.76
Blue	KNN	1.00	1.00	0.96	0.96	0.95	0.98
	Naive Bayes	1.00	0.99	0.96	1.00	0.92	0.93
	Random Forest	1.00	1.00	0.97	1.00	0.93	0.97
	SVM (Linear Kernel)	0.95	0.80	0.71	0.94	0.87	0.78
	XGBoost	0.99	0.87	0.81	0.79	0.79	0.82
Red	KNN	1.00	1.00	0.67	1.00	0.96	0.93
	Naive Bayes	1.00	1.00	0.70	1.00	0.86	0.87
	Random Forest	1.00	1.00	0.75	1.00	0.91	0.90
	SVM (Linear Kernel)	0.96	0.80	0.42	1.00	0.99	0.55
	XGBoost	0.99	0.89	0.61	0.90	0.86	0.74
White	KNN	1.00	0.98	0.79	1.00	0.88	0.69
	Naive Bayes	1.00	0.94	0.83	1.00	0.90	0.68
	Random Forest	1.00	0.98	0.85	1.00	0.94	0.72
	SVM (Linear Kernel)	0.99	0.80	0.78	0.80	0.64	0.63
	XGBoost	0.98	0.87	0.72	0.89	0.71	0.65
Silver	KNN	0.99	0.96	0.82	0.98	0.89	0.82
	Naive Bayes	1.00	0.97	0.90	0.97	0.94	0.81
	Random Forest	1.00	1.00	0.97	0.99	0.95	0.82
	SVM (Linear Kernel)	0.95	0.68	0.74	0.72	0.82	0.50
	XGBoost	0.99	0.88	0.88	0.85	0.76	0.74

Result



		FTIR					R532					R785					EDS					MSP				
		Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol
Black	Ace	1					0.79	0.02	0.08			0.88	0.01	0.00		0.02	0.92	0.05	0.00			0.83	0.02			
	Bhr		0.97				0.17	0.98	0.06	0.02		0.01	0.99	0.01		0.01		0.91		0.14	0.14	0.17	0.81		0.14	0.10
	Csh			0.95			0.02		0.86			0.04		0.70		0.53	0.08	0.04	1.00		0.00			0.84	0.00	0.25
	Kr		0.03		1		0.02			0.95					1.00					0.86			0.18	0.07	0.76	0.02
	Rol			0.05		1				0.03	1.00	0.07	0.00	0.29		0.44					0.86	0.00		0.09	0.10	0.64
Blue	Ace	1					0.83					1.00					1.00	0.00	0.00		0.02	0.96			0.05	0.03
	Bhr		1					1.00					0.99					0.74		0.07			0.98			
	Csh			1					1.00				0.01	1.00				0.13	1.00		0.00		0.02	1.00		
	Kr				1		0.17			1.00					1.00	0.01		0.13		0.93		0.04			0.95	
	Rol					1					1.00					0.99				0.00	0.98					0.97
Red	Ace	1					0.59	0.05	0.24		0.13	1.00					0.87	0.03				0.67		0.09		0.00
	Bhr		1				0.03	0.77	0.24				1.00				0.13	0.97				0.01	0.96	0.04		
	Csh			1			0.26	0.07	0.52		0.01			1.00	0.00				1.00		0.04	0.22	0.04	0.87		
	Kr				1			0.12		1.00					0.97					0.95	0.19				1.00	
	Rol					1	0.12				0.86				0.02	1.00				0.05	0.77	0.10				1.00
White	Ace	1				0.03	0.60				0.34	1.00					0.96	0.04				0.71				0.30
	Bhr		1		0.01	0.01		1.00					1.00				0.02	0.77					0.77	0.16		
	Csh			1					1.00					1.00			0.01	0.19	0.97		0.01	0.02	0.23	0.53		0.12
	Kr				1.00	0.07				1.00					1.00					1.00		0.04		0.03	1.00	
	Rol					0.89	0.40				0.66					1.00	0.01		0.03		0.99	0.23		0.29		0.58
Silver	Ace	1.00					0.85	0.00				1.00					0.98		0.11		0.13	0.57		0.12	0.01	
	Bhr		1				0.15	1.00					1.00		0.06			1.00			0.14		1.00			0.14
	Csh			1					1.00					1.00			0.01		0.89		0.00	0.23		0.84		0.04
	Kr	0.00			1					1.00					0.94					1.00		0.20		0.02	0.99	0.11
	Rol					1					1.00					1.00	0.01				0.86		0.00	0.01		0.71

R532

White	Ace	Bhr	Csh	Kr	Rol
Ace	0.60				0.34
Bhr		1.00			
Csh			1.00		
Kr				1.00	
Rol	0.40				0.66

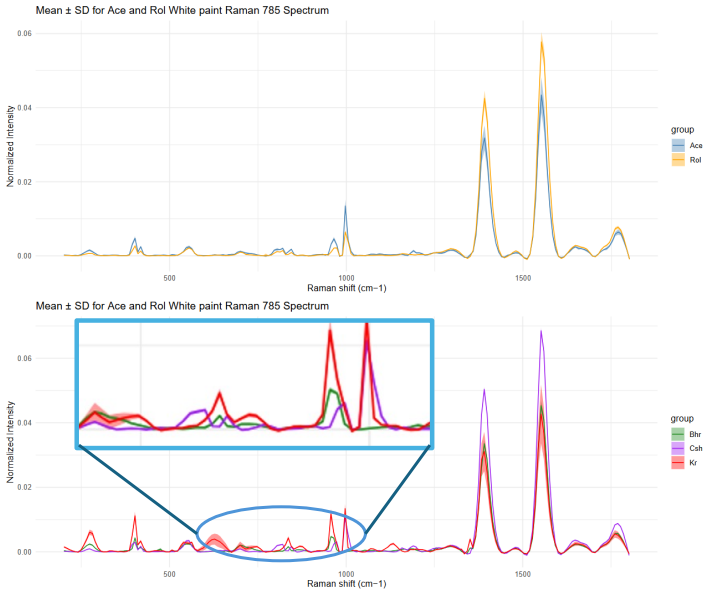
MSP

White	Ace	Bhr	Csh	Kr	Rol
Ace	0.71				0.30
Bhr		0.77	0.16		
Csh	0.02	0.23	0.53		0.12
Kr	0.04		0.03	1.00	
Rol	0.23		0.29		0.58

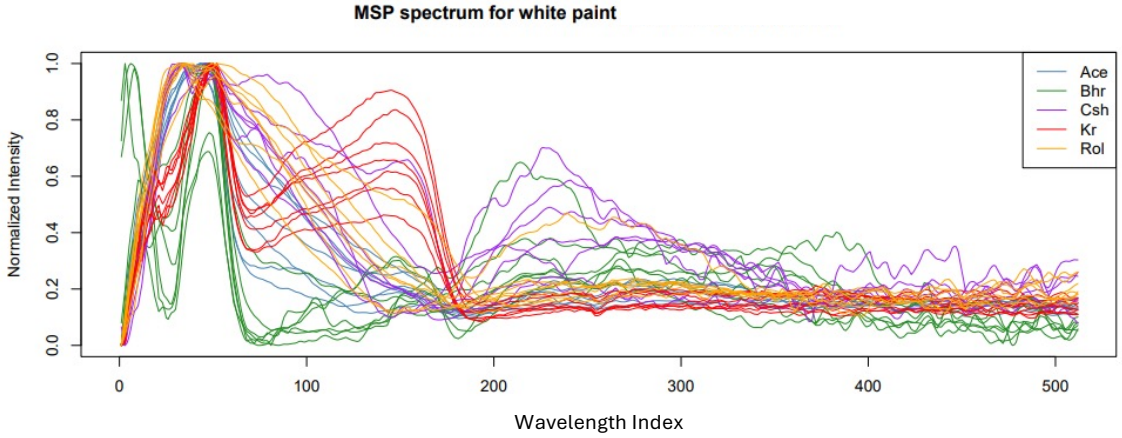
Diagnostic result for Raman 532 and compared with Raman 785

Raman 532

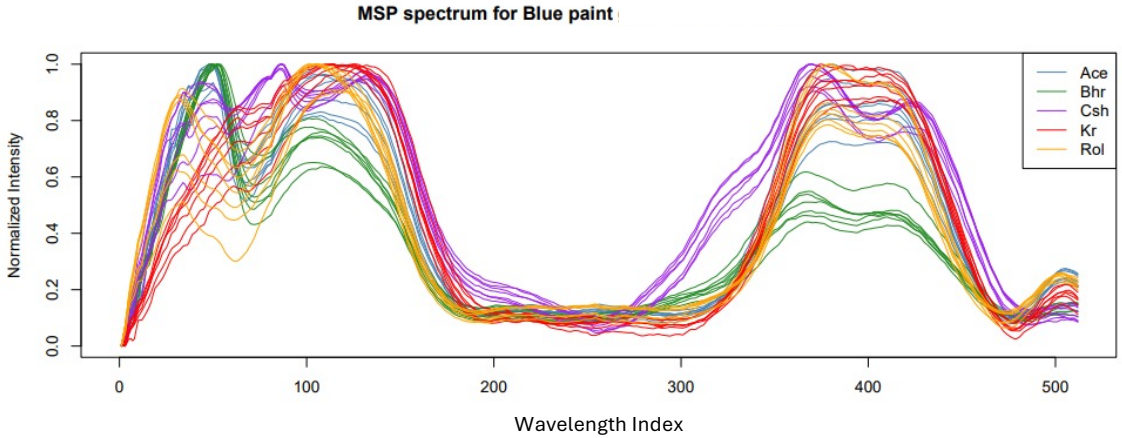
Raman 785



Diagnostic result for MSP



White	MSP Ace	Bhr	Csh	Kr	Rol
Ace	0.709				0.301
Bhr		0.767	0.155		
Csh	0.015	0.233	0.529		0.121
Kr	0.044		0.029	1	
Rol	0.233		0.286		0.578



White	MSP Ace	Bhr	Csh	Kr	Rol
Ace	0.963			0.053	0.03
Bhr		0.977			
Csh		0.023	1		
Kr	0.037			0.947	
Rol					0.97

Limitation and future work

Limitations

- **Limited dataset:**
  - Only seven observation per samples. It may reduces model generalisability.
- **Restricted colour range:**
  - Focusing only on common colours introduces bias and limits performance on uncommon paints.
- **Narrow supplier coverage:**
  - Only popular brands were included. It may reduces the applicability in real casework.

Future Work

- **Dataset expansion:**
  - Include more paints across brands and manufacturing batches to capture broader variability.
- **Colour diversity:**
  - Incorporate rare and mixed colours.
- **Collaborative efforts:**
  - Work with manufacturers, law enforcement, and regulatory bodies to access diverse and up-to-date samples.

Thank you
Questions?

Model and hyper parameter fine-tunning

Classier	Hyper parameters and tuning range
Random Forrest	$m_t \in \{1, 2, \ldots, 10\}$
Naive Bayes	$l \in \{0, 0.5, 1\}$, $a \in \{0, 1.67, 3.33, 5\}$, $u \in \{0, 1\}$
KNN	$k \in \{1, 2, \ldots, 10\}$
SVM	$c \in \{0.1, 0.2, \ldots, 0.5\}$
XGBoost	$n \in \{100, 200\}$, $m \in \{1, 2, 3\}$

m_t: number of features randomly selected at each split
l: Laplace smoothing parameter
α: bandwidth adjustment factor
u: whether use kernel density estimation (0 = No, 1 = Yes)
k: Number of Neighbour
c: cost parameter (trade-off between model complexity and classification errors)
n: number of boosting rounds
m: maximum tree depth

Result

		FTIR					R532					R785					EDS					MSP				
		Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol	Ace	Bhr	Csh	Kr	Rol
Black	Ace	1					0.79	0.02	0.08			0.88	0.01	0.00		0.02	0.92	0.05	0.00			0.83	0.02			
	Bhr		0.97				0.17	0.98	0.06	0.02		0.01	0.99	0.01		0.01		0.91		0.14	0.14	0.17	0.81		0.14	0.10
	Csh			0.95			0.02		0.86			0.04		0.70		0.53	0.08	0.04	1.00		0.00			0.84	0.00	0.25
	Kr		0.03		1		0.02			0.95					1.00					0.86			0.18	0.07	0.76	0.02
	Rol			0.05		1				0.03	1.00	0.07	0.00	0.29		0.44					0.86	0.00		0.09	0.10	0.64
Blue	Ace	1					0.83					1.00					1.00	0.00	0.00		0.02	0.96			0.05	0.03
	Bhr		1					1.00					0.99					0.74		0.07			0.98			
	Csh			1					1.00				0.01	1.00				0.13	1.00		0.00		0.02	1.00		
	Kr				1		0.17			1.00					1.00	0.01		0.13		0.93		0.04			0.95	
	Rol					1					1.00					0.99				0.00	0.98					0.97
Red	Ace	1					0.59	0.05	0.24		0.13	1.00					0.87	0.03				0.67		0.09		0.00
	Bhr		1				0.03	0.77	0.24				1.00				0.13	0.97				0.01	0.96	0.04		
	Csh			1			0.26	0.07	0.52		0.01			1.00	0.00				1.00		0.04	0.22	0.04	0.87		
	Kr				1			0.12		1.00					0.97					0.95	0.19				1.00	
	Rol					1	0.12				0.86				0.02	1.00				0.05	0.77	0.10				1.00
White	Ace	1				0.03	0.60				0.34	1.00					0.96	0.04				0.71				0.30
	Bhr		1		0.01	0.01		1.00					1.00				0.02	0.77					0.77	0.16		
	Csh			1					1.00					1.00			0.01	0.19	0.97		0.01	0.02	0.23	0.53		0.12
	Kr				1.00	0.07				1.00					1.00					1.00		0.04		0.03	1.00	
	Rol					0.89	0.40				0.66					1.00	0.01		0.03		0.99	0.23		0.29		0.58
Silver	Ace	1.00					0.85	0.00				1.00					0.98		0.11		0.13	0.57		0.12	0.01	
	Bhr		1				0.15	1.00					1.00		0.06			1.00					1.00			0.14
	Csh			1					1.00					1.00			0.01		0.89		0.00	0.23		0.84		0.04
	Kr	0.00			1					1.00					0.94					1.00		0.20		0.02	0.99	0.11
	Rol					1					1.00					1.00	0.01				0.86		0.00	0.01		0.71