

**APPLICATION OF COMPUTER VISION TECHNIQUES  
TO DETECT PRINCIPAL ADULTERANTS IN  
TURMERIC POWDER**

**THESIS SUBMITTED BY  
DIPANKAR MANDAL**

***Doctor of Philosophy (Engineering)***

**DEPARTMENT OF PRINTING ENGINEERING  
FACULTY COUNCIL OF ENGINEERING AND TECHNOLOGY  
JADAVPUR UNIVERSITY  
KOLKATA 700032, INDIA**

**2024**

# JADAVPUR UNIVERSITY

KOLKATA-700032, INDIA

INDEX NO. 233/18/E

*1) Title of the Thesis*

***APPLICATION OF COMPUTER VISION  
TECHNIQUES TO DETECT PRINCIPAL  
ADULTERANTS IN TURMERIC POWDER***

*2) Name, Designation &  
Institution of the Supervisor*

***Dr. Arpitam Chatterjee***

Associate Professor,  
Department of Printing Engineering,  
Jadavpur University, Salt Lake Campus,  
Sector - III, Block-LB, Plot-8, Kolkata–  
700098, India  
Phone: 9836020770  
Fax: 03323357254

*3) List of Publications*

***Papers published in International Journals:***

- 1) **Mandal, D.**, Chatterjee, A., Tudu, B., “A deep neural network and random forests driven computer vision framework for identification and prediction of metanil yellow adulteration in turmeric powder” *Concurrency and Computation: Practice and Experience* (2021)
- 2) **Mandal, D.**, Chatterjee, A., Tudu, B., “A color channel based on multiple Random Forest coupled with a computer vision technique for the detection and prediction of Sudan dye-I adulteration in turmeric powder” *Color Research and Application* (2021)

- 3) **Mandal, D.**, Chatterjee, A., and Tudu, B.,  
"Identifying the degree of cornstarch  
adulteration in turmeric powder using  
optimized CNN" Intelligent Decision  
Technologies, Article in Press, DOI  
10.3233/IDT-240656

**Papers published in International**

**Conference proceeding**

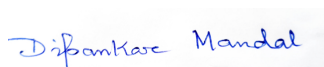
- 4) **Mandal, D.**, Chatterjee, A., Tudu, B.,  
“Identification of commercially available  
turmeric powder using color projection  
features” Second International Conference  
on Research in Computational Intelligence  
and Communication Networks,2016
- 5) **Mandal, D.**, Chatterjee, A., Tudu, B.,A  
Machine Vision Based Approach Towards  
Identification of Adulterant in turmeric  
Powder” Third International Conference  
on Research in Computational Intelligence  
and Communication Networks,2017
- 6) **D. Mandal**, A. Chatterjee, and B. Tudu;  
Convolution neural network driven  
computer vision system for identification  
of metanil yellow adulteration in turmeric  
powder. E. Hassanien (eds.) Emerging  
Technologies in Data Mining and  
Information Security. - 978-981-15-9926-2
- 7) Nandi, D., Datta, S., Prasad, S., Ghosh, S.,  
Mukherjee, A., Maity, S., **Mandal, D.**,  
Chatterjee A., An Approach of Flour  
Adulteration Detection in Turmeric  
Powder using Computer Vision, ICSTA –  
2023..

## Statement Originality

*I, Dipankar Mandal, registered on the 18<sup>th</sup> of April, 2018 do hereby declare that this thesis entitled “Application of computer vision techniques to detect principal adulterants in turmeric powder” contains a literature survey and original research work done by the undersigned candidate as part of Doctoral studies.*

*All information in this thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.*

*I also declare that I have checked this thesis as per the “Policy on Anti Plagiarism, Jadavpur University, 2019”, and the level of similarity as checked by iThenticate software is 4%.*



Dipankar Mandal

Index No. 238/18/E

Date: 21.3.2024



**Associate Professor**

Department of Printing Engineering  
Jadavpur University, Salt-lake Campus  
Kolkata - 700 098

**Dr. Arpitam Chatterjee**

Associate Professor

Department of Printing Engineering  
Jadavpur University, Salt Lake Campus,  
Kolkata-700098, India.

# CERTIFICATE FROM THE SUPERVISOR

Date: 21.3.2024

*This is to certify that the thesis entitled “Application of computer vision techniques to detect principal adulterants in turmeric powder” submitted by Dipankar Mandal, who got his name registered on 18<sup>th</sup> of April, 2018 for the award of Ph.D. (Engg.) degree of Jadavpur University is based upon his own work under the supervision of Dr. Arpitam Chatterjee and neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.*



**Associate Professor**

Department of Printing Engineering  
Jadavpur University, Salt-lake Campus  
Kolkata - 700 098

***Dr. Arpitam Chatterjee***

*Associate Professor*

*Department of Printing Engineering*

*Jadavpur University, Salt Lake Campus Sector-III, Block-LB, Plot-8,*

*Kolkata-700106. India*

*Dedicated to my*

*All my Family members & well-wishers*

## Acknowledgments

The pleasure after completing any job would remain unfinished by not including the people who make it happen and whose continuous supervision, suggestion, and support lead to success. Throughout this time, I appreciate their precious suggestions, invaluable constructive criticism, and cordial counsel. They provided me with honest and enlightening opinions on various topics about my doctoral research, for which I genuinely thank them.

I am privileged to express a profound sense of gratitude and indebtedness to Dr. Arpitam Chatterjee, Department of Printing Engineering, Jadavpur University, Kolkata, for very cooperative and precious guidance at every stage of the thesis program. The finishing of this thesis would be impossible without his persistent supervision.

I am grateful to the Head of the Printing Engineering Department, Jadavpur University, Prof. Soumen Basak for his support and for providing the laboratory facilities during this journey.

I would also like to convey my sincere gratitude to the PRC and RAC members for their beneficial advice throughout the Ph.D. program.

I must convey profound appreciation and gratitude to Prof. Bipan Tudu, Department of Instrumentation & Electronics Engineering, Jadavpur University for his immense support and valuable suggestions at every step of this research program.

I would like to convey my gratitude and thanks to all the faculty members, and technical and non-technical staff of the Department of Printing Engineering, Jadavpur University for their help.

Finally, I would like to express my gratitude to all my family members for their constant motivation toward my research work. I would like to show my respect to all my well-wishers.



---

Dipankar Mandal

21<sup>st</sup> March, 2024

## Abstract

Spices are essential to our daily lives, and a vast variety of spices are used extensively in our nation. The main subject of this study is turmeric (*Curcuma longa*), a perennial herb that is widely used in South and Southeast Asia, especially India, as a main spice. Three to five curcuminoids, the main bioactive substances in turmeric known for their medicinal qualities, are present. India is the world's biggest producer and exporter of turmeric, with a 1700 crore-a-year revenue. But because of its great commercial demand, dishonest traders can adulterate it. Turmeric is adulterated primarily for two reasons: to make it seem better and to make it have more volume. Illegal colors like Sudan dye and metanil yellow are applied to enhance luster. In cases where there is not enough turmeric, additional ingredients such as corn starch are added to increase volume. As a result, the turmeric business has to give quality control top priority. Therefore, to maintain the high standards of quality for this spice, methods for identifying these adulterants must be developed. The present methods available in the market are conventional. They require more time for sample preparation and as well as measurement procedure is time consuming. The common example of these technique are gas chromatography, HPLC, NIR, Spectroscopy, etc. These methods are invasive in nature .So rapid and non-invasive detection of adulterant mixed with turmeric is necessary .Machine vision is a new technique in this field to detect the presence of adulterants turmeric powder. In this work with the help of machine vision system coupled with deep neural network and convolution neural network three adulterants were detected as mentioned below.

In the first study of the thesis work metanil yellow adulteration was detected using machine vision. Here five varieties of turmeric powder were examined in this study; one was made in-house, and the other four were packaged organic turmeric powders from four well-known companies that claim to be 100% pure. MET of the Sigma Aldrich brand was procured from research lab suppliers. One part was labelled "pure" since it had not been subjected to mixing



of MET. The remaining three portions were marked as "adulterated" after being adulterated with three different percentages of MET. The weight-on-weight (w/w) percentages of MET were 5%, 10%, and 15%. The successful completion of classification and prediction tasks from the sub-images derived from the acquired photos was made possible by the deployment of Random Forest (RF) and Deep Neural Network (DNN) models. The findings show that the suggested methodology predicts the degree of MET adulteration with an accuracy surpassing 94% and achieves classification accuracy over 98% for differentiating between pure and adulterated turmeric powder.

In the next study Sudan Dye-I was detected using machine vision. Five different pure turmeric powder varieties were selected for testing. Of the five varieties, four were advertised as 100% pure organic turmeric powders and were purchased from the neighbourhood grocery store, while one pure variation was home made. The Sudan dye of Sigma Aldrich brand was procured from a lab chemical supplier. Each of the 5 variants of *pure* turmeric was divided into 4 equal portions that resulted total of 20 portions. Among the 20 portions, 1 portion of each variant was kept unadulterated while the rest 3 portions were adulterated with 3 different percentages of concentration (w/w) that are 5%, 10%, and 15%. Such mixing resulted in 5 *pure* and 15 *adulterated* samples comprising of 5 adulterated samples for each concentration level.

Finally corn starch was detected using machine vision. 5 variants of *pure* turmeric powders were taken for experimentation. Among those 5 variants, 4 were branded organic turmeric powders that claim 100% purity and were procured from the local supermarket and 1 pure variant was prepared in-house. The starch was added with the turmeric powder in four different percentages by volume, i.e. 10%, 20%, 30%, and 40%. Here convolution neural network (CNN) is used for both the detection as well as prediction of degree of adulteration starch in turmeric powder. Here accuracy achieved in tune of 98%. This method has shown remarkable improvement of CNN model to address the limitations of conventional methods.

# Contents

	<b>Page No.</b>
<b>CHAPTER 1: Introduction and scope of the thesis</b>	<b>1</b>
1.1. Introduction	1
1.2. A brief overview of turmeric	2
1.2.1. Medicinal values of turmeric	2
1.2.2. Main constituents of turmeric:	4
1.3. Adulterations in turmeric powder	5
1.3.1. Conventional methodologies for adulterant detection in Spices	6
1.3.1.1. Gas chromatography-mass spectrometry (GS-MS)	7
1.3.1.2. High-pressure liquid chromatography (HPLC)	7
1.3.1.3. UV-Vis spectroscopy	8
1.4. Literature survey on turmeric adulteration detection using conventional methods	9
1.5. Computer vision system	15
1.5.1. Organization of computer vision system	15
1.5.1.1. Image Acquisition	17
1.5.1.2. Operations and Analysis on Image	18
1.5.2. Feature Extraction	19
1.5.3. Decision making	20
1.5.4. Benefits of computer vision	20
1.5.5. Applications of Computer Vision	21
1.6. Literature review on applications of computer vision in food quality assessment	21
1.7. Problem description	26
1.8. Objective and scope of the thesis	26
1.9. Thesis structure	27
References	28

<b>CHAPTER 2: Imaging chamber development and performance evaluation methods</b>	<b>41</b>
2.1. Introduction	41
2.2. Principle of Human vision system	42
2.3. Imaging chamber development	43
2.4. Performance evaluation methods	46
2.5. Model validation	47
2.5.1. Validation of generalization potential	47
2.5.2. k-fold cross validation	48
2.5.3. Confusion matrix	49
2.5.3.1. Metrics derived from confusion matrix	50
2.6. Evaluation of regression results	52
2.7. Implementation environment	54
2.8. Conclusion	54
References	55
<b>CHAPTER 3: Detection of Metanil Yellow Adulteration in Turmeric Powder</b>	<b>57</b>
3.1. Introduction	57
3.2. Materials and methods	59
3.2.1. Experimental in-house database preparation	59
3.2.2. Frequency domain feature extraction	62
3.2.3. Data analysis models	66
3.2.3.1. Random forest (RF) classifier and regressor	66
3.2.3.2. Deep neural network (DNN) classifier	68
3.3. Results and discussions	72
3.4. Conclusion	84
References	84

<b>CHAPTER 4: Detection of Sudan dye-I adulteration in turmeric powder</b>	<b>89</b>
4.1. Introduction	89
4.2. Materials and methods	91
4.2.1. Experimental in-house database preparation	91
4.2.2. Color feature extraction	93
4.2.3. Data analysis model	98
4.3. Results and discussions	102
4.4. Conclusion	109
References	110
<b>CHAPTER 5: Detection of starch adulteration in turmeric powder</b>	<b>114</b>
5.1. Introduction	114
5.2. Materials and methods	116
5.2.1. Experimental in-house database preparation	116
5.2.2. Development of CNN model	118
5.2.3. Hyperparameter optimization of the CNN model	119
5.3. Results and discussions	123
5.4. Conclusion	127
References	128
<b>CHAPTER 6: Concluding remarks</b>	<b>132</b>
6.1. Introduction	132
6.2. Major findings	132
6.3. Responses to the research questions	134
6.4. Future scope of work	135
6.5. Conclusion	136

## List of Figures

Fig. 1.1.	Medicinal activities of turmeric	3
Fig. 1.2.	Chemical structure of turmeric	5
Fig. 1.3.	Working principle of GC-MS.	7
Fig. 1.4.	Working principle of HPLC system	8
Fig. 1.5.	Working principle of UV-Vis spectroscopy	9
Fig. 1.6.	Overview of Computer Vision System	16
Fig. 2.1.	Human optical system	43
Fig. 2.2.	The schematic of the operating principle of the imaging chamber	45
Fig. 2.3.	The actual imaging chamber with labeling to the different components	45
Fig. 2.4.	The actual imaging chamber in imaging mode after loading the sample	46
Fig. 2.5.	Illustration of (a) undesired and (b) desired loss plots along with (c) undesired and (d) desired accuracy plots for training and validation data. (These are randomly generated plots for explanation not plots of any studies conducted in this thesis work)	47
Fig. 2.6.	A 10-fold cross-validation where (shaded cells are training folds and white cells are testing folds)	48
Fig. 2.7.	Confusion matrix of binary classification	49
Fig. 2.8.	ROC plot	52
Fig. 2.9.	Residual plot with random data	54

Fig. 3.1.	Schematic representation of dataset preparation strategy	61
Fig. 3.2.	Samples of images from the experimental image database	61
Fig. 3.3.	Image analysis using histogram of HSV channels.	62
Fig. 3.4.	FFT spectra of pure and adulterated turmeric images.	64
Fig. 3.5.	Example of annular rings on FFT spectra and plots of extracted features with different numbers of annular rings.	65
Fig. 3.6.	Oob errors plotting of RF model with different number of trees.	68
Fig. 3.7.	General architecture of DNN.	72
Fig. 3.8.	Model validation plots of training and validation set using (A) accuracy and (B) loss.	73
Fig. 3.9.	Confusion matrix of different classifiers on unknown testing set	76
Fig. 3.10.	AUC-ROC plots of different classifiers.	79
Fig. 3.11.	Residual plots for different regression models on (a) test data and total dataset.	83
Fig. 4.1.	Samples of images from the image dataset	92
Fig. 4.2.	Color channel-based distinguishability analysis for pure and adulterated samples	94
Fig. 4.3.	PCA clustering of different classes for different color channels a) H, b) V, c) Y and d) Cr.	96
Fig. 4.4.	The centroids of different classes using K-means clustering for color channels a) H, b) V, c) Y and d) Cr	97
Fig. 4.5.	Ensample mechanism with multiple RF	99

Fig. 4.6.	The OOB error plots for RF model of color channels a) H, b) V, c) Y and d) Cr. Red line shows the chosen number of trees chosen for the model.	102
Fig. 4.7.	Confusion matrix for different RF models on the test set	105
Fig. 4.8.	Residual plots for RF prediction models for a) H, b) V, c) Y, and d) Cr color channels. The red line at 0 is for reference	106
Fig. 4.9.	Residual plot for a) different RF models and b) final prediction output.	107
Fig. 5.1.	Samples of images from the image dataset	117
Fig. 5.2.	BGA process representation	121
Fig. 5.3.	Convergence plot for BGA optimization	122
Fig. 5.4.	CNN model validation plots for (a) accuracy and (b) loss between training and validation set	123
Fig. 5.5.	The confusion matrix of CNN classification	125
Fig. 5.6.	ROC-AUC plot for classification results in different classes	126

## List of Tables

Table 1.1.	Some commonly used adulterants in turmeric	6
Table 1.2.	Major works reported in turmeric adulteration detection using conventional methods	9
Table 1.3.	Major works reported in food quality assessment using CV based detection methods	22
Table 3.1.	RF algorithm (for both classification and prediction)	67
Table 3.2.	DNN layer parameters	71
Table 3.3.	Parameter settings of DNN and RF	74
Table 3.4.	10-fold cross-validation results (mean of 10 runs)	75
Table 3.5.	Performance evaluation of different classifiers	78
Table 3.6.	AUC-ROC and log-loss values of different classifiers	80
Table 3.7.	MAE, RMSEP, and $R^2$ value comparison between different regression models	81
Table 3.8.	Performance comparison against $R^2$ and RMSEP	82
Table 4.1.	Pseudocode of RF	100
Table 4.2.	Parameter settings for different RF models	101
Table 4.3.	Results of 10-fold cross-validation for RF models	103
Table 4.4.	Classification results using metrics of confusion matrix	104
Table 4.5.	Prediction performance by different RF models	106
Table 4.6.	Performance comparison between different classifiers and prediction models	108
Table 4.7.	Prediction performance comparison in terms of coefficient of determination ( $R^2$ )	109



Table 5.1.	Initial model parameter settings	119
Table 5.2.	Binary code representations of optimization parameters	120
Table 5.3.	Optimized parameter settings	122
Table 5.4.	Result of 10 fold cross validation	124
Table 5.5.	Comparative evaluation of classification performance	125
Table 5.6.	Comparison on processing time	127
Table 6.1.	Performance comparison between the proposed methods in this thesis	133

## **List of Abbreviations**

NIR	Near-Infrared
DNA	Deoxyribo Nucleic Acid
HPLC	High-Pressure Liquid Chromatography
GS-MS	Gas Chromatography-Mass Spectrometry
UV-VIS	Ultraviolet-Visible
SERS	Surface-Enhanced Raman Spectroscopy
FT-IR	Fourier Transform Infrared
MHz	Megahertz
RF	Radio Frequency
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
PLS	Partial Least Square Discrimination
LS	Least Square
RB	Radial Basis
SVR	Support Vector Regression
NN	Neural Network
PLSDA	Partial Least Square
SIMCA	Soft Independent Modeling of Class Analogies
GA	Genetic Algorithm
kNN	K Nearest Neighbor

CNN	Convolution Neural Network
Vis	Visible
RMSEP	Root Mean Square Error Of Prediction
Mm	Millimeter
WHO	World Health Organization
FSA	The Food Standards Agency
ML	Machine Learning
RQ	Research Question
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
RF	Random Forest
MET	Metanil Yellow
FT-IR	Fourier Transform Infrared
FTIR-PLS	Fourier Transform Infrared Partial Least Square
RP-HPLC	Reversed-Phase High-Performance Liquid Chromatography
UPLC-DAD	Ultra Performance Liquid Chromatography-Diode Array Detector
2D-HPTLC	Two-Dimensional High-Performance Thin-Layer Chromatography
w/w	Weight to Weight
USB	Universal Serial Bus
LED	Light Emitting Diode
JEPG	Joint Photographic Experts Group

RGB	Red Green Blue
HSV	Hue, Saturation, Value
FT	Fourier Transform
AMF	Annular Mean Frequency
FFT	Fast Fourier Transform
DT	Decision Tree
OOB	Out of Bag
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
LR	Linear Regression
AUC-ROC	Area Under the Receiver Operating Characteristic Curve.
IARC	International Agency for Research on Cancer
MMPSD	Micro Matrix Power Spectral Density
MPixel	Mega Pixel
DC	Direct Current
gm	Gram
LinR	Linear Regression
LogR	Logistic Regression
FT Raman	Fourier Transform Raman

FC	Fully Connected
SGD	Stochastic Gradient Descent
OPLS-DA	Orthogonal Partial Least Squares-Discriminant Analysis
LIBS	Laser-Induced Breakdown Spectroscopy

# CHAPTER 1

## Introduction and scope of the thesis

### 1.1. Introduction

Spices are an essential part of our daily life as they add rich flavor and vibrant color, and provide nourishment and immunity to our food. They are not only popular in our country but have gained global acceptance. Spices play a significant role not just in our kitchens but also in medicine and various other fields. India is known as the "spice bowl of the world" as it produces and consumes various spices in huge quantities [1.1]. Among different spices, *Curcuma longa* L. ( $C_{21}H_{20}O_6$ ) or turmeric is one of the most popular ones which is used across different Asian cuisines and has many important therapeutic benefits [1.2]. In addition to its medicinal applications, turmeric has been often utilized as a key component in many cosmetic products. China, India, Pakistan, and other Asian nations like Burma, Bangladesh, Taiwan, Sri Lanka, Indonesia, Myanmar, etc. are the major producers of turmeric [1.3]. However, there is a tendency for widespread adulteration of turmeric hence, the product quality is to be verified through different safety norms and regulations set by the Food Adulteration Act of India (PFA, 2008) [1.4] to avoid any unwanted adulterations.

## **Chapter 1: Introduction and scope of the thesis**

---

This chapter provides a brief introduction about turmeric its usages, benefits, and risks of its adulteration. The different conventional methods with some of the major applications of those methods in detecting turmeric adulterations have been briefed in this chapter. The problems of the conventional methods have then been discussed in the light of reported literature followed by the possibilities of machine vision as a potential solution to those. A comprehensive survey of applications of machine vision in adulteration detection focusing the spice adulteration detection has also been presented in this chapter. The chapter presents the scope and research questions of this thesis. Finally, the chapter outlines have been presented.

### **1.2. A brief overview of turmeric**

#### **1.2.1. Medicinal values of turmeric**

Turmeric has notable medical use. It has been used in traditional medical systems to take care of a variety of diseases [1.5]. Its biological effects include anti-inflammatory, hypocholesterolemic, chromogenic, cytotoxic, spasmolytic, anti-rheumatic, anti-bacterial, antiviral, anti-diabetic, and anti-hepatotoxic properties. Additionally, turmeric is said to have anti-cancerous properties [1.6]. The therapeutic benefits of turmeric are attributed to its curcuminoids, turmeric oil, and total Turmeric contains the yellow curcumin pigment and methylated curcumin, both of which have antioxidant properties. A few of them are stated below and consolidated in Fig. 1.1.

## Chapter 1: Introduction and scope of the thesis

---



**Fig. 1.1:** Medicinal activities of turmeric. [1.7]

- *Household aid for chronic cough*

For chronic cough, colds, and throat issues, turmeric powder is a valuable home remedy [1.8].

- *Prevention and treatment of cancers*

Curcumin is a potentially effective chemical and therapeutic substance for the prevention and treatment of cervical cancer [1.9]. It has been observed that curcumin treatment reduces the growth of cervical cancer cells by altering molecular pathways tied to the human papillomavirus (HPV) in the cancer cells that are associated with it.

- *Treatment of Asthma*

Consuming curcumin capsules revealed that the tablets are beneficial for treating bronchial asthma [1.10] without causing any clinically noteworthy side effects.

- *Prevention of type2diabetes*

Turmeric can help those who are susceptible to diabetes prevent type 2 diabetes [1.11] from developing in them.



## Chapter 1: Introduction and scope of the thesis

---

- *Safeguard for liver diseases*

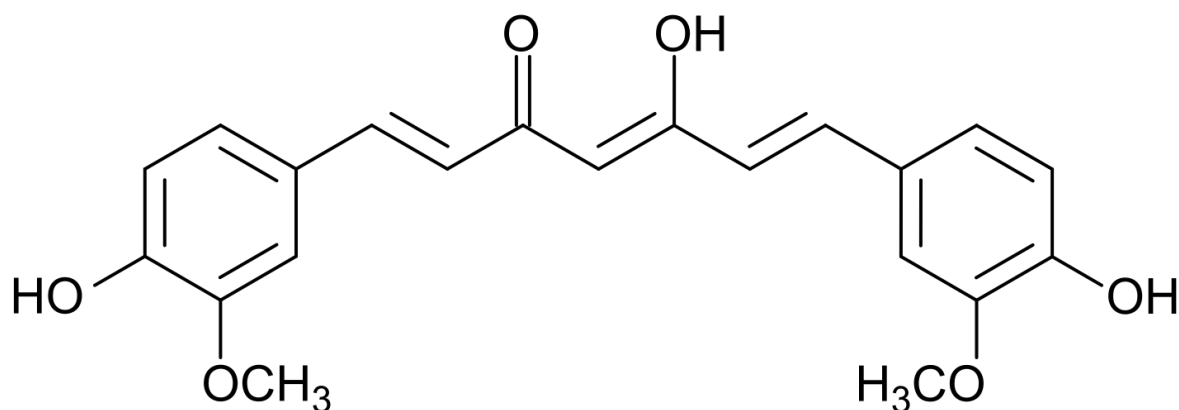
Curcumin may help prevent and treat cirrhosis [1.12] in rats because it can lower the expression of transforming growth beta (TGF- $\beta$ ). Curcumin is also a potentially effective fibrinolytic and antifibrotic medication for the treatment of chronic liver disorders [1.13].

- *Relief from rheumatoid arthritis*

Rheumatoid arthritis patients reported minimal negative effects from curcumin medication [1.14], which also improved their condition by reducing joint swelling and pain.

### 1.2.2. Main constituents of turmeric:

Turmeric constituents consist of three curcuminoids [1.15]. The main constituent is curcumin which is *diferuloylmethane*; which is accountable for its vibrant yellow color, and other constituents are *desmethoxycurcumin*, and *bisdemethoxycurcumin*. In addition to these main three constituents, it also consists of volatile oils [1.16] such as tumerone, atlantes, and zingier, and also other constituents which are sugars. Turmeric has more than 200 components that have been isolated. The approximate composition value of dried turmeric leaves is 6.0% crude protein, 0.5% crude fat, 34.5% total fiber, 43.3% carbs, 9.4% crude ash, and 6.3% moisture. This content is expressed in grams per 100 grams of dried turmeric leaves. Steam distillation of rhizomes can provide a variety of essential oils [1.17] (5-8%), including sesquiterpenes (53%), zingiberene (25%), and many more. The main ingredient that gives turmeric its yellow color is curcumin. It has also been proven to be the primary component responsible for the majority of the therapeutic benefits. Six oxygen atoms make up curcumin: two are assigned to the carbonyl group, two to the methyl group (C=O), and one is located in the middle of the molecule, joining the two carbonyl molecules. The chemical structure of turmeric has been shown in Fig. 1.2.



**Fig. 1.2:** Chemical structure of turmeric. [1.18]

### 1.3. Adulterations in turmeric powder

Adulteration is the addition of unwanted ingredients to food products. This can cause health hazards and decrease the nutritional value of food. These are frequently added to increase profits because they can lower the cost of the product or improve its look. However, there are instances of natural adulterations for example at the time of preparation of rice there may be the mixing of sand particles with rice grain. But apart from such natural cases majority of the adulteration is intentional and driven by economic benefit.

For turmeric powder adulteration is a proven challenge. There are different adulterants commonly used for turmeric adulterations either to improve its appearance in terms of its yellow color which is also termed as chemical adulteration or to increase the volume by materials having lower food and economic value termed as volume adulteration [1.19]. A list of different commonly used adulterants has been provided in Table 1.1.

## Chapter 1: Introduction and scope of the thesis

---

**Table 1.1:** Some commonly used adulterants in turmeric

Name of the adulterants	Type of adulteration
Metanil Yellow	Chemical adulteration
Sudan Dye I	Chemical adulteration
Lead chromate	Chemical adulteration
Egg-yellow color	Chemical adulteration
Tartrazine	Chemical adulteration
Pistachio hull waste	Volume adulteration
Wheat flour	Volume adulteration
Corn starch	Volume adulteration
Dry bread	Volume adulteration
Maize flour	Volume adulteration

### 1.3.1. Conventional methodologies for adulterant detection in Spices

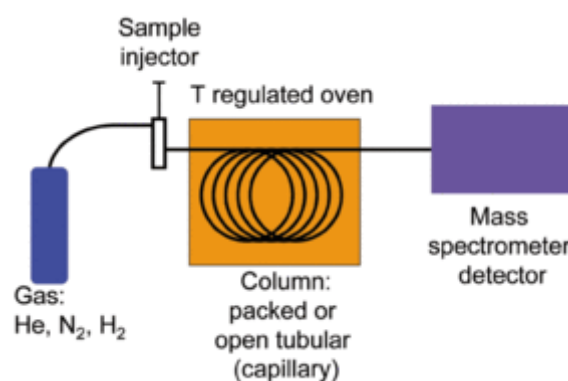
Conventionally the existence of chemical constituents including adulterants can be detected using different chemical, spectral, and instrumental analysis methods including high-pressure liquid chromatography (HPLC) [1.20], thin layer chromatography (TLC) [1.21], gas chromatography-mass spectrometry (GS-MS) [1.22], ultraviolet-visible (UV-VIS) spectroscopy [1.23], surface-enhanced Raman spectroscopy (SERS) [1.24], Fourier transforms infrared (FT-IR) spectroscopy [1.25] and similar devices. A brief discussion of the working principle of these methods has been provided in this section.

## Chapter 1: Introduction and scope of the thesis

---

### 1.3.1.1. Gas chromatography-mass spectrometry (GS-MS)

GC-MS is an advanced analytic method that combines the features of mass spectrometry and gas chromatography to identify various components inside the material being examined. This adaptable instrument can be used to identify and separate unidentified organic compounds, as well as to measure mass gases and run a permanent chromatograph. The mass spectrometer and gas chromatographer are the two components of the GC-MS as shown in Fig. 1.3. GC-MS is utilized to identify and quantify food additives such as preservatives, artificial colors, and flavor enhancers [1.26]. This helps in ensuring that food products comply with regulatory standards regarding permissible additive levels.



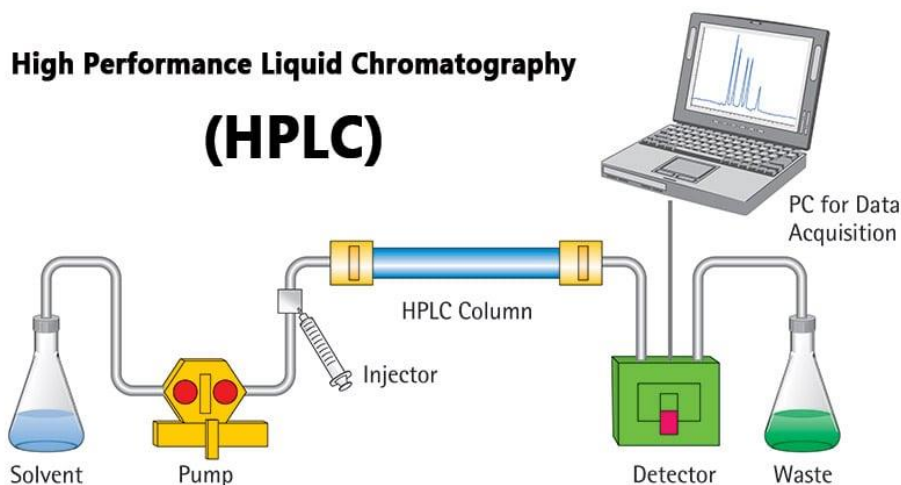
**Fig. 1.3:** Working principle of GC-MS. [1.27]

### 1.3.1.2. High-pressure liquid chromatography (HPLC)

HPLC involves the passage of a liquid sample through a packed column under high pressure, typically at or above 400 atmospheres. The column is packed with a stationary phase, and separation is based on the interactions between the sample components and the stationary phase. A detector records the elution time of each compound, producing a chromatogram as shown in Fig.1.4. HPLC is utilized for the determination of food additives, preservatives, vitamins, flavors, and other compounds in food and beverage products [1.28]. HPLC is used

## Chapter 1: Introduction and scope of the thesis

for the separation and quantification of biomolecular like amino acids, peptides, nucleic acids, and proteins.

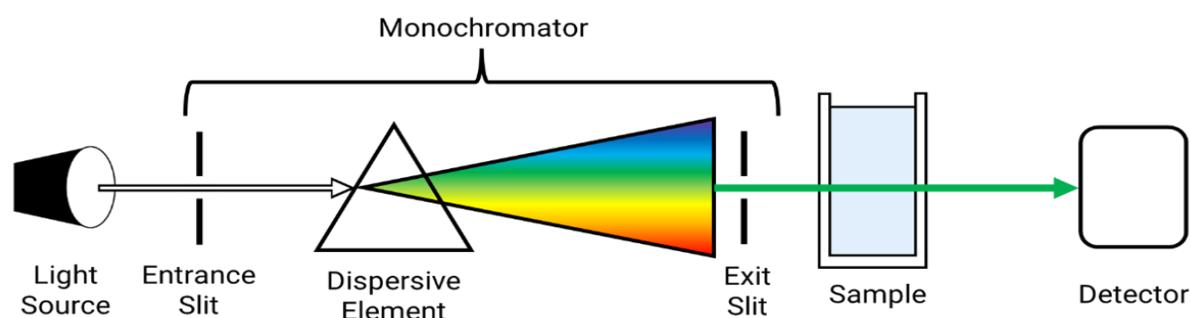


**Fig. 1.4:** Working principle of HPLC system. [1.29]

### 1.3.1.3. *UV-Vis spectroscopy*

The basis for UV-Vis spectroscopy's operation is Beer-Lambert's law. Analytical assessments using UV-Vis spectroscopy are typically completed between 200 and 800nm. The sample is placed in a cuvette and is exposed to monochromatic light as a radiation source. The sample's molecules absorb the light. This energy is converted into electrical impulses via a photon-sensitive sensor as shown in Fig.1.5 and the necessary signal processing is done beforehand. Using multivariable methods and UV-Vis spectroscopy, metanil dye I and IV were found in paprika powder samples [1.30]. The diode array assembly served as the detector. The detector had a 300–600 nm range and a 1 nm spectral resolution. It was said that the adulterant detection was satisfactory.

## Chapter 1: Introduction and scope of the thesis



**Fig. 1.5:** Working principle of UV-Vis spectroscopy. [1.31]

### 1.4. Literature survey on turmeric adulteration detection using conventional methods

This section provides some of the major works reported in the field of turmeric adulteration detection using conventional spectral and analytical methods in Table 1.2.

**Table 1.2:** Major works reported in turmeric adulteration detection using conventional methods

Applications	Working Principle	Data Analysis technique	Resulted performance	Ref.
Detection of metanil yellow in turmeric powder	Spectral Analysis	Principal Component Analysis, Partial least square regression	The coefficient of determination and r.m.s of calibration were 0.96-0.99 and 0.44-0.91 respectively	[1.32]
Determination of Sudan red and metanil yellow in turmeric powder	NIR Spectroscopy	Partial least square regression, soft independent model of class analogy	For Sudan Red $R^2 = 0.90$ and RMSE = 0.059 and for metanil yellow $R^2 = 0.91$ and RMSE = 0.058.	[1.33]
Determination of	Fluorescence		LOD value	[1.34]

## Chapter 1: Introduction and scope of the thesis

metanil yellow	europium doped carbon dots		0.390 $\mu\text{g/mL}$ with a linear range of 1.0–15.0 $\mu\text{g/mL}$	
Determination of additives and chemical contaminations in turmeric	FT-IR spectroscopy	Partial least square regression	$R^2$ for sudan red.97, root mean square error of prediction 1.3%	[1.35]
Authentication of turmeric powder	IR spectroscopy	PLS-DA, SICMA	The precision is 82% and 92% for SICMA and PLS-DA respectively	[1.36]
Determination of sudan dye-I in spices	UV-vis spectroscopy	PLS-DA, KNN, SICMA	Classification results were 99.3%, 96.3% and 90.4% respectively for PLS-DA, KNN and SIMCA methods	[1.37]
Detection of Pb adulteration	Portable X-ray fluorescence analysis (pXRF) and colorimetric test		Could detect when the Pb concentrations is more than 5-70 mg/kg in dried turmeric roots and 1000 mg/kg for powders	[1.38]
Detection of egg-yellow color and starch adulteration	2.4GHz cavity perturbation technique (CPT)	Different prediction models like LR, SVM, etc. along with hardware and software optimization	The models' accuracy resulted in the range of 85 – 99%	[1.39]

## Chapter 1: Introduction and scope of the thesis

Detection of metanil yellow	Characterization of Nano nickel cobalt oxide modified graphite electrode (NiCo <sub>2</sub> O <sub>4</sub> @GP) using X-ray diffraction (XRD) and scanning electron microscopy (SEM)	Differential pulse voltammetry (DPV) and cyclic voltammetry (CV)	A linear range of 5 to 1000 $\mu$ M was resulted under optimal experimental conditions	[1.40]
Detection of wheat flour, pistachio hull waste and dry bread adulteration	NIR spectroscopy in the range of 12,000–4000 $\text{cm}^{-1}$	PCA and PLSR	Calibration for validation was in the range of 0.265–0.306 while the prediction accuracy was in the range of 97.9–98.6%.	[1.41]
Detection of maize floor adulteration	Fluorescence quenching method using front-face synchronous fluorescence spectroscopy (FFSFS) and fluorescence	PLSR	$R^2$ was more than 0.95 with RMSE less than 0.06 and 9% LOD	[1.42]



## Chapter 1: Introduction and scope of the thesis

	titration			
Detection of tartrazine colored rice flour adulteration	Spectral mages of turmeric with different degree of adulteration and a cloud based implementation	Different CNN models	About 95% accuracy was achieved	[1.43]
Simultaneous detection of metanil yellow, Sudan I, and Sudan red G along with curcumin, demethoxycurcumin, and bisdemethoxycurcumin	Liquid chromatography-tandem mass spectrometry (LC-MS/MS) method		The LOQ and LOD were reported as 1-2 and 0.1-0.2 ng/mL, respectively.	[1.44]
Purity analysis using detection of C. longa DNA	Loop-mediated isothermal amplification (LAMP)		Resulted 10 times improved sensitivity was achieved than the reported PCR based results	[1.45]
Detection of metanil yellow, corn starch powder and Sudan dye –IV adulteration	Handheld NIR spectroscopy	PCADA and SIMCA	About 93% accuracy has been achieved	[1.46]
Detection of lead chromate (PbCrO <sub>4</sub> ) adulteration	Powder X-ray diffraction method with varying X-ray		LOD of 0.5% has been reported	[1.47]

## Chapter 1: Introduction and scope of the thesis

	tube voltage, current and scanning parameters		
Detection of lead chromate adulteration	Fourier transform-Raman (FT-Raman) spectroscopy	PLSR	0.5% LOD was reported [1.48] at based on the 1750-200 $\text{cm}^{-1}$ range of spectroscopy
Detection of Sudan Red and white turmeric adulteration	Raman imaging and FT-IR spectroscopy	PLSR	$R^2$ value of 0.97 for Sudan Red and 0.95 for white turmeric were reported [1.49]
Detecton and separation of metanil yellow, curcumin, bisdemethoxycurcumin and demethoxycurcumin	Gradient reverse phase high pressure liquid chromatography (RP-HPLC)		$R^2$ value of 0.99 and less than 2% relative standard deviation value were achieved [1.50]
Metanil yellow adulteration detection	Handheld NIR	PLSR	0.33% LOD and 1.10% LOQ were resulted [1.51]
Detection of tartrazine colored rice flour	Different spectral bands and peaks of multispectral imaging	PCA and Bhattacharya distance	$R^2$ value in the tune of 0.98 – 0.99 was resulted [1.52]
Detection of metanil yellow adulteration	Variation in parameters like capacitance,		Resulted $R^2$ was more than 0.9 with repeatability error about [1.53]

## Chapter 1: Introduction and scope of the thesis

	impedance, conductance, and current–voltage characteristics of electrical impedance spectroscopy	2 – 4%	
Food Adulteration Act of India (PFA, 2008) Determination of sudan dye-I in chili and curry containing foodstuffs	HPLC-PDA	Calibration curve	Recovery was in the range of 51-86% in sauce and 89-100% in powdered spice [1.54]
Extraction and separation of sudan dye	In line MMPSD		Recoveries in the range of 94.5–99.1%, [1.55]
Determination of banned dyes in spices	LC-Mass spectrometry		Recoveries in the range of 75.7-92.3% [1.56]

All the listed works have conveyed excellent performance. However there are some inherent limitations of most of the spectral and analytical methods which include expensive devices, non-portability (except NIR based devices), and skilled manpower for operation, precise sample preparation and IoT-based system integration. Looking into those limitations the scope of computer vision was explored as it has proven advantages of simple and low cost implementation across diverse engineering applications.

## **Chapter 1: Introduction and scope of the thesis**

---

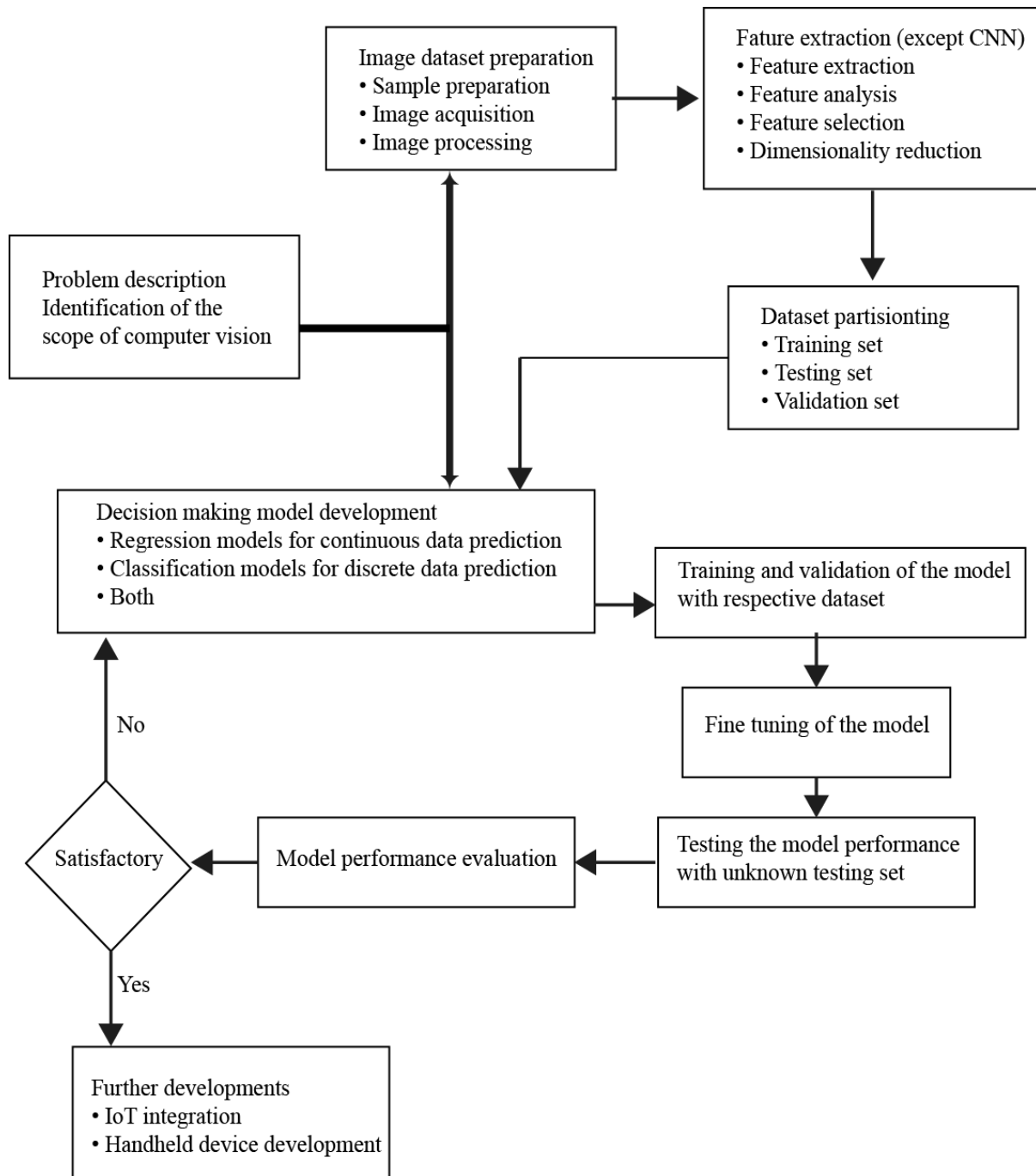
### **1.5. Computer vision system**

Computer vision [1.57], also known as machine vision, mimics the characteristics of human perception. It includes two components; measurements of features and pattern recognition based on those features. Typically, a machine vision system includes uniform lighting equipment for constant illumination of the object, a high-resolution camera or scanner to confine the image of that object, and a processor or controller for further processing of the image data [1.58]. To provide image-based automatic examination for quality control, process control, and robot guidance, it is necessary to have knowledge of and skills in computer vision. It is mostly utilized for optical gauging, quality control, sorting, part assembly inspection, presence and absence, and industrial process management [1.59].

#### **1.5.1. Organization of computer vision system**

The main operations of a computer vision system comprises of image acquisition, object detection, object recognition and decision making. The overview of the CV system can be shown as Fig. 1.6. and the operating principle of these operations is described below.

## Chapter 1: Introduction and scope of the thesis



**Fig. 1.6:** Overview of Computer Vision System

## Chapter 1: Introduction and scope of the thesis

---

### 1.5.1.1. Image Acquisition

Image acquisition is the capturing of image from various sources. This is the first step in computer vision. Depending on application and requirements, there are various types of image capturing devices like scanner, digital camera, medical imaging devices etc. Some of the major parameters governing the image acquisition are as follows.

*Lighting:* Most CV systems consist of hardware components like lighting modules and sensor devices. CV systems create images by analyzing the reflected, scattered or transmitted light from the scene. The lighting techniques involve characteristics of light source and placement with respect to the camera/acquisition device. The color of illumination is also an important feature as the object's appearance and recorded intensities considerably vary according to that.

*Lens:* The lens or lens-mirror assemblies are used for proper projection of reflected/transmitted light on the image sensors. The lenses vary in optical quality and price. The quality and resolution of the captured images depends on the lens arrangements. A fixed lens as part of a standalone vision system typically uses autofocus, which could be a mechanically adjusted lens that can automatically focus on the part [1.60]. Autofocus lenses usually have a fixed field of view at a given distance. The image is presented to the sensor in the form of light.

*Image sensor:* The sensors in a CV system convert this light into a digital image which is then sent to the processor for analysis. Some of the commonly used image sensors are charge-coupled device (CCD) [1.61] and complementary metal oxide-semiconductors (CMOS) [1.62]. These have their own sensitivity characteristics against different spectrums of lights and associated advantages and disadvantages.

## Chapter 1: Introduction and scope of the thesis

---

*Vision processing:* Vision processing consists of algorithms that review the image and extract required information, run the necessary inspection, and make a decision. Finally, communication is typically accomplished by either a discrete I/O signal or data sent over a serial connection to a device that is logging information on the data.

*Communications:* The communication between the sensor and the output device is done by either a discrete I/O signal or data sent over a serial connection to a device that is logging information. Discrete I/O points may be connected to a programmable logic controller which will use that information to control a work cell or an indicator such as a solenoid which might be used to trigger a particular mechanism. Data communication by a serial connection can be done by RS232 serial output [1.63]. Some systems employ a higher-level industrial protocol like Ethernet/IP, which may be connected to a device like a monitor or other operator interface to provide an operator interface specific to the application for convenient process monitoring and control.

### ***1.5.1.2. Operations and Analysis on Image***

Different operations are to be performed on captured images. Some of the major operations commonly performed under this head have been briefed below.

*Denoising:* Most of the captured images consist of noises. Noises can be of different types and some are inherent due to hardware and lighting conditions. This step attempts to remove the noises that mostly cause blurring, contrast distortions, and other sorts of distortions that can cause distortions in features [1.64]. However, there are cases where the noises are not removed to simulate the real-world scenario where the distortions are inevitable.

## Chapter 1: Introduction and scope of the thesis

---

*Segmentation:* In this step the image is analyzed and different segments are segregated. The goal of this step is to isolate the region of interest (ROI) from the captured scene [1.65]. However, this also may not be needed for cases where only the ROI is captured strategically.

### 1.5.2. Feature Extraction

The segmented images are subjected to the feature extraction where some of the inherent characteristics that are mathematically expressed being extracted from the images. Very simple examples of such features are mean, median and such statistical values of the images which are calculated from the pixel values.

There are three types of features which are most commonly extracted; color feature, texture feature and shape feature. In case of color features a common practice is to convert mages from its native RGB color space to other color spaces which provide much better interpretation of colors. Texture features are calculated mostly based on the correlational measures for example gray level co-occurrence matrix (GLCM) [1.66], local binary pattern (LBP) [1.67], etc. The shape features are particularly important in case of object detection and identification sort of problems. It can vary from a simple shape feature like chain code [1.68] to mathematically complicated methods that can work on 3-D shapes. It is important to note that in most of the cases a single of those three feature classes may not be sufficient for robust performance hence depending on the applications sets of features are extracted that can contain different types of feature with proper normalizations.

In many cases the features are extracted in a transformed domain. Frequency domain transforms such as discrete Fourier transform (FT) [1.69], discrete cosine transform (DCT) [1.70] and wavelet transform (WT) [1.71] are the most popular such transforms in image



## Chapter 1: Introduction and scope of the thesis

---

processing and CV. These transforms provide better interpretation of the images, particularly in the case of local or region-based operations.

### 1.5.3. *Decision making*

This is the final step where based on the features the decision of the CV is declared. Decisions broadly are of two types; classification and regression.

*Classification:* As the name says these tasks include identification of the class of the captured image among different classes being considered in the problem. For example, if there are pictures of different spices a classification problem's goal is to identify the spice from the picture where each type of spice is a class. So here the results are non-continuous and discrete in nature.

*Regression:* In this case, the decision is continuous i.e. some value prediction which can be any value on a defined scale depending on the problem. For example, if from the picture it was a task to predict what the concentration of adulteration in a spice is then it would be a regression problem as the concentration is a continuous number.

There are tasks where both of these tasks are to be performed and in general, the decision is often called prediction which covers both classification and regression results in general.

### 1.5.4. **Benefits of computer vision**

Where human vision is the best for qualitative interpretation of a complex, unstructured scene, CV excels at quantitative measurement of a structured scene because of its speed, accuracy, and repeatability. For example, on a production line, a CV system can inspect hundreds, or even thousands, of parts per minute [1.72]. A CV system built around the right camera resolution and optics can easily inspect object details too small to be seen by the

## **Chapter 1: Introduction and scope of the thesis**

---

human eye. In removing physical contact between a test system and the parts being tested, CV can be also helpful in preventing damage to machine parts. It can be instrumental in reducing maintenance costs and implementing preventive maintenance. CV can also bring additional safety prevent human contamination of clean rooms and protect human workers from hazardous environments [1.73].

### **1.5.5. Applications of Computer Vision**

The CV systems are a non-invasive techniques for the detection and analysis of different parameters. Nowadays without using conventional laborious methods machine vision systems are also being used. The application areas are widespread across food quality evaluation, biomedical image analysis, industrial automation, image processing, financial services and robotics. The sorting of vegetables and fruits, and the location of defects e.g., the location of dark contaminants and insects in cereals can be detected using CV. In the case of biomedical image analysis tumor detection, measurement of size and shape of internal organs, X-ray inspection and blood cell count can also be done using CV. In the case of the field of robotics and automation, not only obstacles can be avoided by automobiles through recognition and interpretation of objects in a scene but even collision can be avoided. Although there are huge reported literatures on application of CV in diverse engineering applications here a literature review on application of CV in food quality detection has been presented since this is the focus of this thesis.

### **1.6. Literature review on applications of computer vision in food quality assessment**

The focus of the literature review presented n Table 1.3 remains on the application of CV in food quality assessment where the term food includes food ingredients as well.

## Chapter 1: Introduction and scope of the thesis

**Table 1.3:** Major works reported in food quality assessment using CV based detection methods

Application area	Working Principle	Resulted performance	Ref.
Sugarcane quality	Color features of the images and different pre-trained CNN models	The performance resulted in 93-98% accuracy across different CNN models	[1.74]
Apple bruises detection	NIR images were further subjected to different image processing algorithms like segmentation, Hough transform, etc.	With more than 95% accuracy early bruises in apples were detected	[1.75]
Quality assessment of mango fruits	In-house image database were prepared both spectral and soft X-ray imaging followed by data analysis using multilayer perceptron neural network (MLP NN) and CNNs	For color images and X-ray images about 91.52% and 98.7% accuracy were achieved, respectively	[1.76]
Quality inspection and sorting of chili products	YOLO-V5 framework was developed	The resulted precision was about 0.90	[1.77]
Detection of fruit	The types of fruits was	About 95% accuracy in the	[1.78]

## Chapter 1: Introduction and scope of the thesis

freshness	identified using CNN and the freshness assessment as well as shelf life prediction was made using Gaussian Naïve Bayes (GNB) and random forest	classification of fruit and about 88% accuracy in shelf-life prediction were achieved	
Detection of black tea fermentation	Colored images were subjected to feature extraction where about 150 statistical features were extracted and subjected to k-NN, SRC and SVM	About 87.39% - 98.75% accuracies were achieved using different classifiers	[1.79]
Intramuscular fat content in pork cut	Semantic segmentation-based deep learning model	About 93.28% accuracy was achieved across the prediction of fat content across different cuts	[1.80]
Rice plant defect detection	Used modified- Xception network for classification with the features extracted by ResNet	Achieved about 99% accuracy on domestic and public dataset.	[1.81]
Apple quality inspection and grading	Used FCN-ResNet 18 for quality inspection and SSD-MobileNet for grading	Achieved 86.7% and 92.6% accuracy for grading and about 95% average accuracy for quality assessment	[1.82]
Wheat-grain	Hyper parameter	About 98% accuracy was	[1.83]

## Chapter 1: Introduction and scope of the thesis

classification	tuned CNN	archived	
Prediction of rice seed germination	Four varieties of rice from Tamilnadu were classified using pre-trained Alexnet, Resnet, inception v3.0 models	Accuracy for different networks varied between 89 – 99%	[1.84]
Grape leaf disease detection	Artificial bee colony based feature selection followed by SVM and KNN-based classification of diseases	About 90 - 93% correct prediction was achieved with 8SVM and 85 – 90% for KNN	[1.85]
Detection of Nitrosamine n red meat	Feature selection was conducted using p-value of Mann-Whitney U test and classified using PLS-DA algorithm	The results showed 100% sensitivity and 71.21% specificity with a $R^2$ value n the tune of 0.85	[1.86]
Grading and sorting of cashew	Color, texture, size and shape features were extracted upon taking Multisresolutional wavelet and Contourlet transform which were subjected to SVM and back	SVM and back propagation neural networks resulted more than 99% and 97% accuracy, respectively	[1.87]

## Chapter 1: Introduction and scope of the thesis

---

	propagation neural network for classification		
Detection of foreign objects in walnuts	Walnut images containing different natural and man-made foreign objects were classified using deep neural network	About 99% and 95% accuracy were achieved for the detection and classification of foreign objects	[1.88]
Identification of Enzyme growth in banana peel	Seventeen color features were extracted and a genetic programming model was used for identification of polyphenol oxidase and peroxidase enzymes	$R^2$ value of 0.98 and 0.97 were achieved for polyphenol oxidase and peroxidase enzymes detection, respectively	[1.89]
In-line sorting of mango	Different shape parameters along with Fourier transformed features were extracted from the mango images and used for classification	About 98% classification accuracy was achieved	[1.90]

## **Chapter 1: Introduction and scope of the thesis**

---

### **1.7. Problem description**

The literature review conveys that the detection of adulterations in spices is a challenge to human health. As turmeric is one of the most popular spices in India and other Asian countries it is often been adulterated. Hence, the detection of adulteration in turmeric also carries the utmost importance. The literature review reveals mostly different analytical and instrumental methods are used for adulteration detection. Despite appreciable accuracy such methods have drawbacks like expensive instruments, lengthy and precise sample-making processes, invasive and destructive testing methods, and portability issues that limit their realization in different locations in the supply chain for on-the-spot detection. The sample-making for obtaining the result is also time-consuming and skilled labor-intensive in some cases. The chemical reagents used for the detection purpose are costly, too. These motivate towards the development of noninvasive, on-destructive, and chemical-free approaches for food quality assessment.

### **1.8. Objective and scope of the thesis**

The literature review reveals that computer vision frameworks have shown potential for those limitations associated with conventional methods. It also conveys the promising potential of deep learning methodologies over the conventional machine learning (ML) approaches. Hence the scope of the thesis is to explore the potential of CV as an effective means for detecting the presence of adulteration and the degree of adulterations also. This drives us to find the answers to the following research questions (RQs) through this thesis work.

- RQ1: Can computer vision be a possible alternative to existing techniques?
- RQ2: Can a deep learning-driven computer vision system perform well in adulteration detection of turmeric powder?

## Chapter 1: Introduction and scope of the thesis

---

- RQ3: Can convolution neural network (CNN) perform competitively in adulteration detection?
- RQ4: Can a computer vision-based model address the limitations of existing techniques?

### 1.9. Thesis structure

The thesis comprises six chapters. The contents of the different chapters have been briefed below;

**Chapter 1, “Introduction and scope of the thesis”** the significance of turmeric and its adulteration scenario. Then the existing methodologies for adulteration detection have been discussed. A concise overview of computer vision technology and its different components has been presented here. It includes a detailed literature survey from different aspects. Then it frames the problem description and outlines the objectives of this thesis work. Finally, it frames research questions that have been answered through this thesis work.

**Chapter 2, “Imaging Chamber Development and Performance Evaluation Methods”** describes the development of an indigenous imaging chamber towards consistency of illumination. The different components and design aspects of the imaging chamber have been described in detail. The chapter also presents the important evaluation parameters and metrics that are commonly used in the CV fraternity to judge the performance of the CV models.

**Chapter 3 “Detection of Metanil Yellow adulteration in turmeric powder”** depicts the identification and prediction of the metanil yellow adulterant in turmeric powder with the machine vision. In this work, a new feature detection method has been presented. The chapter also presents the application of deep neural networks (DNN) and random forests (RF) for the detection of adulterated turmeric as well as the degree of adulteration.



## Chapter 1: Introduction and scope of the thesis

---

**Chapter 4 “Detection of Sudan dye-I adulteration in turmeric powder”** has been designed to find the Sudan dye-I adulteration in turmeric powder. Here a different color channel-based feature extraction has been proposed. For prediction purposes, a new framework of ensemble multiple RFs has been proposed.

**Chapter 5 “Detection of Starch adulteration in turmeric powder”** deals with the detection of starch in turmeric powder. Here the potential of CNN has been investigated where the CNN model has been developed and optimized using the soft computing method.

**Chapter 6 “Concluding Remarks”** finally summarizes the major findings, merits, and limitations of the presented methods, answers to the RQs, possible future extensions, and concluding remarks.

## References

- 1.1. Parthasarathy, V. A. (Ed.). (2008). *Organic spices*. New India Publishing
- 1.2. Rathore, S., Mukim, M., Sharma, P., Devi, S., Nagar, J. C., & Khalid, M. (2020). Curcumin: A review for health benefits. *Int. J. Res. Rev*, 7(1), 273-290.
- 1.3. Khodang, C., & Sharma, A. (2022). Post-harvest losses during the marketing of organic turmeric in the Kakching district of Manipur, India.
- 1.4. Rukundo, I. R., & Danao, M. G. C. (2020). Identifying turmeric powder by source and metanil yellow adulteration levels using near-infrared spectra and PCA-SIMCA modeling. *Journal of food protection*, 83(6), 968-974.
- 1.5. Debjit Bhowmik, C., Kumar, K. S., Chandira, M., & Jayakar, B. (2009). Turmeric: a herbal and traditional medicine. *Arch. Appl. Sci. Res*, 1(2), 86-108.

## Chapter 1: Introduction and scope of the thesis

---

- 1.6. Rathaur, P., Raja, W., Ramteke, P. W., & John, S. A. (2012). Turmeric: The golden spice of life. *International Journal of pharmaceutical sciences and research*, 3(7), 1987.
- 1.7. <https://images.app.goo.gl/4ABa8gL6HFnJxZrL8>
- 1.8. Jose, S. (2019). A Study to Assess the Effectiveness of Steam Inhalation with Tulsi Leaves and Turmeric Powder to Reduce Common Cold in 5 to 15 year Old School Going Children at Selected Rural Areas Coimbatore (Doctoral dissertation, PPG College of Nursing, Coimbatore).
- 1.9. Unlu, A., Nayir, E., Kalenderoglu, M. D., Kirca, O., & Ozdogan, M. (2016). Curcumin (Turmeric) and cancer. *J buon*, 21(5), 1050-1060.
- 1.10. Abidi, A., Gupta, S., Agarwal, M., Bhalla, H. L., & Saluja, M. (2014). Evaluation of efficacy of curcumin as an add-on therapy in patients of bronchial asthma. *Journal of clinical and diagnostic research: JCDR*, 8(8), HC19.
- 1.11. Pivari, F., Mingione, A., Brasacchio, C., & Soldati, L. (2019). Curcumin and type 2 diabetes mellitus: prevention and treatment. *Nutrients*, 11(8), 1837.
- 1.12. Bruck, R., Ashkenazi, M., Weiss, S., Goldiner, I., Shapiro, H., Aeed, H., ... & Pines, M. (2007). Prevention of liver cirrhosis in rats by curcumin. *Liver International*, 27(3), 373-383.
- 1.13. Elhaggagy, A. A., Alsaggaf, S., & Amin, H. A. (2014). The preventive and therapeutic role of curcumin in liver cirrhosis. *Life Science Journal*, 11(10), 328-38.

## Chapter 1: Introduction and scope of the thesis

---

- 1.14. Bagherniya, M., Darand, M., Askari, G., Guest, P. C., Sathyapalan, T., & Sahebkar, A. (2021). The clinical use of curcumin for the treatment of rheumatoid arthritis: A systematic review of clinical trials. *Studies on Biomarkers and New Targets in Aging Research in Iran: Focus on Turmeric and Curcumin*, 251-263.
- 1.15. Pfeiffer, E., Höhle, S., Solyom, A. M., & Metzler, M. (2003). Studies on the stability of turmeric constituents. *Journal of Food Engineering*, 56(2-3), 257-259.
- 1.16. Monton, C., Charoenchai, L., Suksaeree, J., & Sueree, L. (2016). Quantitation of curcuminoid contents, dissolution profile, and volatile oil content of turmeric capsules produced at some secondary government hospitals. *journal of food and drug analysis*, 24(3), 493-499.
- 1.17. Funk, J. L., Frye, J. B., Oyarzo, J. N., Zhang, H., & Timmermann, B. N. (2010). Anti-arthritic effects and toxicity of the essential oils of turmeric (*Curcuma longa* L.). *Journal of agricultural and food chemistry*, 58(2), 842-849.
- 1.18. <https://images.app.goo.gl/h/J1TXv5812dr7erMA9>
- 1.19. Fonovich, T. M. (2013). Sudan dyes: are they dangerous for human health?. *Drug and chemical toxicology*, 36(3), 343-352.
- 1.20. Sahil, K., Prashant, B., Akanksha, M., Premjeet, S., & Devashish, R. (2011). Gas chromatography-mass spectrometry: applications. *International Journal of Pharmaceutical and Biological Archives*, 2(6), 1544-1560
- 1.21. 19. Lehotay, S. J., & Hajšlová, J. (2002). Application of gas chromatography in food analysis. *TrAC Trends in Analytical Chemistry*, 21(9-10), 686-697
- 1.22. Bidlingmeyer, B. A. (1993). *Practical HPLC methodology and applications*. John Wiley & Sons.

## Chapter 1: Introduction and scope of the thesis

---

- 1.23. Shoeb, M., Islam, M. M., Reza, M. S., Nahar, N., & Islam, M. M. (2022). HPLC analysis of artificial preservatives, stimulants and sweeteners in carbonated beverages in Bangladesh. *Current Research on Biosciences and Biotechnology*, 3(2), 215-221.
- 1.24. Perkampus, H. H. (2013). *UV-VIS Spectroscopy and its Applications*. Springer Science & Business Media.
- 1.25. Bharathi, S. K. V., Sukitha, A., Moses, J. A., & Anandharamakrishnan, C. (2018). Instrument-based detection methods for adulteration in spice and spice products—A review. *Journal of Spices and Aromatic Crops*, 27(2), 106-118.
- 1.26. 24. Kar, S., Tudu, B., Bag, A. K., & Bandyopadhyay, R. (2018). Application of near-infrared spectroscopy for the detection of metanil yellow in turmeric powder. *Food analytical methods*, 11, 1291-1302
- 1.27. <https://images.app.goo.gl/CVb2uRvKt8vBUKyD7>
- 1.28. Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2022). Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: The role of preprocessing methods in analysis. *Vibrational Spectroscopy*, 120
- 1.29. <https://images.app.goo.gl/CvEx2BeGkxUSruKHA>
- 1.30. Abdel-Lateef, M. A., Albalawi, M. A., Al-Ghamdi, S. N., Mahdi, W. A., Alshehri, S., & El Hamd, M. A. (2023). Determination of metanil yellow dye in turmeric powder using a unique fluorescence Europium doped carbon dots. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 287, 122124
- 1.31. <https://images.app.goo.gl/ecs9jbm7unEXk4ez5>

## Chapter 1: Introduction and scope of the thesis

---

- 1.32. Dhakal, S., Schmidt, W. F., Kim, M., Tang, X., Peng, Y., & Chao, K. (2019). Detection of additives and chemical contaminants in turmeric powder using FT-IR spectroscopy. *Foods*, 8(5), 143.
- 1.33. Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2021). An evaluation of IR spectroscopy for authentication of adulterated turmeric powder using pattern recognition. *Food Chemistry*, 364, 130406
- 1.34. Di Anibal, C. V., Odena, M., Ruisánchez, I., & Callao, M. P. (2009). Determining the adulteration of spices with Sudan I-II-III-IV dyes by UV–visible spectroscopy and multivariate classification techniques. *Talanta*, 79(3), 887-892
- 1.35. Cornet, V., Govaert, Y., Moens, G., Van Loco, J., & Degroodt, J. M. (2006). Development of a fast analytical method for the determination of Sudan dyes in chili- and curry-containing foodstuffs by high-performance liquid chromatography–photodiode array detection. *Journal of agricultural and food chemistry*, 54(3), 639-644.
- 1.36. Rajabi, M., Sabzalain, S., Barfi, B., Arghavani-Beydokhti, S., & Asghari, A. (2015). In-line micro-matrix solid-phase dispersion extraction for simultaneous separation and extraction of Sudan dyes in different spices. *Journal of Chromatography A*, 1425, 42-50.
- 1.37. Botek, P., PouStka, J., & Hajslova, J. (2007). Determination of banned dyes in spices by liquid chromatography-mass spectrometry. *Czech Journal of Food Science*, 25(1), 17-24
- 1.38. Nollet, L. M. (2023). Non-Destructive Testing Methods of Spices. In *Analysis of Food Spices* (pp. 285-304). CRC Press.

## Chapter 1: Introduction and scope of the thesis

---

- 1.39. Kaur, T., Gamez, A., Olvera, J. L., Schaefer, B. C., & Corona-Chavez, A. (2023). I-TAINTED: Identification of turmeric adulteration using the cavity perturbation technique and technology optimized machine learning. IEEE Access.
- 1.40. Nag, S., Das, D., Naskar, H., Tudu, B., Bandyopadhyay, R., & Roy, R. B. (2022). Detection of metanil yellow adulteration in turmeric powder using nano nickel cobalt oxide modified graphite electrode. IEEE Sensors Journal, 22(13), 12515-12521.
- 1.41. Tamiji, Z., Habibi, Z., Pourjabbar, Z., Khoshayand, M. R., Sadeghi, N., & Hajimahmoodi, M. (2022). Detection and quantification of adulteration in turmeric by spectroscopy coupled with chemometrics. Journal of Consumer Protection and Food Safety, 17(3), 221-230.
- 1.42. Xie, J. Y., Tan, J., Tang, S. H., & Wang, Y. (2022). Fluorescence quenching by competitive absorption between solid foods: Rapid and non-destructive determination of maize flour adulterated in turmeric powder. Food Chemistry, 375, 131887.
- 1.43. Lanjewar, M. G., Morajkar, P. P., & Parab, J. (2022). Detection of tartrazine colored rice flour adulteration in turmeric from multi-spectral images on smartphone using convolutional neural network deployed on PaaS cloud. Multimedia Tools and Applications, 81(12), 16537-16562.
- 1.44. Kotha, R. R., Tareq, F. S., Byrdwell, C., & Luthria, D. L. (2021). Rapid and Sensitive Analytical Assessment of Curcuminoids and Three Common Turmeric Adulterants in a Single Run Using Liquid Chromatography and Tandem Mass Spectrometry. ACS Food Science & Technology, 1(11), 2174-2181.

## Chapter 1: Introduction and scope of the thesis

---

- 1.45. Sheu, S. C., Wu, Y. C., Lien, Y. Y., & Lee, M. S. (2021). Specific, sensitive and rapid *Curcuma longa* turmeric powder authentication in commercial food using loop-mediated isothermal nucleic acid amplification. *Saudi Journal of Biological Sciences*, 28(10), 5931-5936.
- 1.46. Ranjan, R., Kumar, N., Kiranmayee, A. H., & Panchariya, P. C. (2021, March). Application of handheld NIR spectroscopy for detection of adulteration in turmeric powder. In *2021 7th international conference on advanced computing and communication systems (ICACCS)* (Vol. 1, pp. 1238-1241). IEEE.
- 1.47. Paranthaman, R., Moses, J. A., & Anandharamakrishnan, C. (2021). Development of a method for qualitative detection of lead chromate adulteration in turmeric powder using X-ray powder diffraction. *Food Control*, 126, 107992.
- 1.48. Erasmus, S. W., van Hasselt, L., Ebbinge, L. M., & van Ruth, S. M. (2021). Real or fake yellow in the vibrant colour craze: Rapid detection of lead chromate in turmeric. *Food Control*, 121, 107714.
- 1.49. Chao, K., Dhakal, S., Schmidt, W. F., Qin, J., Kim, M., Peng, Y., & Huang, Q. (2020). Raman and IR spectroscopic modality for authentication of turmeric powder. *Food chemistry*, 320, 126567.
- 1.50. Sahu, P. K., Panda, J., Jogendra Kumar, Y. V. V., & Ranjitha, S. K. (2020). A robust RP-HPLC method for determination of turmeric adulteration. *Journal of liquid chromatography & related technologies*, 43(7-8), 247-254.

## Chapter 1: Introduction and scope of the thesis

---

- 1.51. Rukundo, I. R., Danao, M. G. C., Weller, C. L., Wehling, R. L., & Eskridge, K. M. (2020). Use of a handheld near infrared spectrometer and partial least squares regression to quantify metanil yellow adulteration in turmeric powder. *Journal of near infrared spectroscopy*, 28(2), 81-92.
- 1.52. Rukundo, I. R., Danao, M. G. C., Weller, C. L., Wehling, R. L., & Eskridge, K. M. (2020). Use of a handheld near infrared spectrometer and partial least squares regression to quantify metanil yellow adulteration in turmeric powder. *Journal of near infrared spectroscopy*, 28(2), 81-92.
- 1.53. Das, C., Chakraborty, S., Bera, N. K., Acharya, K., Chattopadhyay, D., Karmakar, A., & Chattopadhyay, S. (2019). Impedimetric approach for estimating the presence of metanil yellow in turmeric powder from tunable capacitance measurement. *Food analytical methods*, 12, 1017-1027.
- 1.54. Biyani, M., Biyani, R., Ushijima, H., Saito, M., Takamura, Y., Tamiya, E., & Biyani, M. (2018). Instant enumeration of total viable bacterial counts for food quality assurance using 'DEP-On-Go'sensor. *Analytical methods*, 10(14), 1585-1592.
- 1.55. Rajabi, M., Sabzalain, S., Barfi, B., Arghavani-Beydokhti, S., & Asghari, A. (2015). In-line micro-matrix solid-phase dispersion extraction for simultaneous separation and extraction of Sudan dyes in different spices. *Journal of Chromatography A*, 1425, 42-50.
- 1.56. Li, J., Ding, X. M., Liu, D. D., Guo, F., Chen, Y., Zhang, Y. B., & Liu, H. M. (2013). Simultaneous determination of eight illegal dyes in chili products by liquid chromatography–tandem mass spectrometry. *Journal of chromatography B*, 942, 46-52.



## Chapter 1: Introduction and scope of the thesis

---

- 1.57. Snyder, W. E., & Qi, H. (2004). Machine vision (Vol. 1). Cambridge University Press.
- 1.58. Burke, M. W. (2012). Image Acquisition: Handbook of machine vision engineering: Volume 1 (Vol. 1). Springer Science & Business Media.
- 1.59. Davies, E. R. (2009). The application of machine vision to food and agriculture: a review. *The Imaging Science Journal*, 57(4), 197-217.
- 1.60. Ren, H., & Wu, S. T. (2012). Introduction to adaptive lenses. John Wiley & Sons.
- 1.61. Mehta, S., Patel, A., & Mehta, J. (2015, April). CCD or CMOS Image sensor for photography. In *2015 international conference on communications and signal processing (ICCSP)* (pp. 0291-0294). IEEE.
- 1.62. Rajachidambaram, J. S. (2011). Evaluation of amorphous oxide semiconductors for thin film transistors (TFTs) and resistive random access memory (RRAM) applications.
- 1.63. Dawoud, D. S., & Dawoud, P. (2022). Serial communication protocols and standards. River Publishers.
- 1.64. Clark, I. A., Callaghan, M. F., Weiskopf, N., Maguire, E. A., & Mohammadi, S. (2021). Reducing susceptibility distortion related image blurring in diffusion MRI EPI data. *Frontiers in Neuroscience*, 15, 706473.
- 1.65. Haralick, R. M., & Shapiro, L. G. (1985). Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1), 100-132.

## Chapter 1: Introduction and scope of the thesis

---

- 1.66. Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International journal of scientific and research publications*, 3(5), 1-5.
- 1.67. Pietikäinen, M. (2010). Local binary patterns. *Scholarpedia*, 5(3), 9775.
- 1.68. Liu, Y. K., & Žalik, B. (2005). An efficient chain code with Huffman coding. *Pattern Recognition*, 38(4), 553-557.
- 1.69. Sundararajan, D. (2001). *The discrete Fourier transform: theory, algorithms and applications*. World Scientific.
- 1.70. Ahmed, N., Natarajan, T., & Rao, K. R. (1974). Discrete cosine transform. *IEEE transactions on Computers*, 100(1), 90-93.
- 1.71. Zhang, D., & Zhang, D. (2019). Wavelet transform. *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*, 35-44.
- 1.72. Parker, J. R. (2010). *Algorithms for image processing and computer vision*. John Wiley & Sons.
- 1.73. Jarraya, S. K. (2018). Computer vision-based fall detection methods using the kinect camera: A survey. *International Journal of Computer Science & Information Technology (IJCSIT)* Vol, 10.
- 1.74. Modi, R. U., Chandel, A. K., Chandel, N. S., Dubey, K., Subeesh, A., Singh, A. K., ... & Kancheti, M. (2023). State-of-the-art computer vision techniques for automated sugarcane lodging classification. *Field Crops Research*, 291, 108797.
- 1.75. Tian, M., Zhang, J., Yang, Z., Li, M., Li, J., & Zhao, L. (2024). Detection of early bruises on apples using near-infrared camera imaging technology combined with adaptive threshold segmentation algorithm. *Journal of Food Process Engineering*, 47(1), e14500.

## Chapter 1: Introduction and scope of the thesis

---

- 1.76. Ashok, V., Bharathi, R. K., & Sheela, N. (2023). An Automatic Non-Destructive External and Internal Quality Evaluation of Mango Fruits based on Color and X-ray Imaging using Computer Vision Techniques. *Inteligencia Artificial*, 26(72), 223-243.
- 1.77. Abhijit, Akhil, S., Kumar, V. A., Jose, B. K., & Abubeker, K. M. (2023, June). Computer Vision Assisted Bird-Eye Chilli Classification Framework Using YOLO V5 Object Detection Model. In *International Conference on Power Engineering and Intelligent Systems (PEIS)* (pp. 217-226). Singapore: Springer Nature Singapore.
- 1.78. Reka, S. S., Bagelika, A., Venugopal, P., Ravi, V., & Devarajan, H. (2024). Deep Learning-Based Classification of Rotten Fruits and Identification of Shelf Life. *Computers, Materials & Continua*, 78(1).
- 1.79. Bhargava, A., Bansal, A., Goyal, V., & Shukla, A. (2023). Machine learning & computer vision-based optimum black tea fermentation detection. *Multimedia Tools and Applications*, 82(28), 43335-43347.
- 1.80. Liu, H., Zhan, W., Du, Z., Xiong, M., Han, T., Wang, P., ... & Sun, Y. (2023). Prediction of the intramuscular fat content of pork cuts by improved U2-Net model and clustering algorithm. *Food Bioscience*, 53, 102848.
- 1.81. Meena, R., Joshi, S., & Raghuwanshi, S. Xception model for disease detection in rice plant. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-18.
- 1.82. Garcés Cadena, A. A., Menéndez Granizo, O. A., Córdova, E. P., & Prado Romo, A. J. (2023). Clasificación de calidad de manzana para monitoreo de cosechabilidad utilizando visión por computador y algoritmos de aprendizaje profundo. *Ingeniare. Revista chilena de ingeniería*, 31, 0-0.

## Chapter 1: Introduction and scope of the thesis

---

- 1.83. Lingwal, S., Bhatia, K. K., & Tomer, M. S. (2021). Image-based wheat grain classification using convolutional neural network. *Multimedia Tools and Applications*, 80(28), 35441-35465.
- 1.84. Durai, S., Mahesh, C., Sujithra, T., & Shyamalakumari, C. (2022, January). Germination Prediction System for Rice seed using CNN Pre-trained models. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-9). IEEE.
- 1.85. Andrushia, A. D., & Patricia, A. T. (2020). Artificial bee colony optimization (ABC) for grape leaves disease detection. *Evolving Systems*, 11(1), 105-117.
- 1.86. Arora, M., & Mangipudi, P. (2021). A Meta-heuristic Approach for Design of Image Processing Based Model for Nitrosamine Identification in Red Meat Image. *Recent Patents on Engineering*, 15(3), 326-337.
- 1.87. Shyna, A., & George, R. M. (2017, April). Machine vision based real time cashew grading and sorting system using SVM and back propagation neural network. In *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)* (pp. 1-5). IEEE.
- 1.88. Rong, D., Xie, L., & Ying, Y. (2019). Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture*, 162, 1001-1010.
- 1.89. Nadafzadeh, M., Mehdizadeh, S. A., & Soltanikazemi, M. (2018). Development of computer vision system to predict peroxidase and polyphenol oxidase enzymes to evaluate the process of banana peel browning using genetic programming modeling. *Scientia horticultrae*, 231, 201-209.

## **Chapter 1: Introduction and scope of the thesis**

---

- 1.90. Ibrahim, M. F., Ahmad Sa'ad, F. S., Zakaria, A., & Md Shakaff, A. Y. (2016). In-line sorting of harumanis mango based on external quality using visible imaging. *Sensors*, 16(11), 1753.

# CHAPTER 2

## Imaging chamber development and performance evaluation methods

### 2.1. Introduction

Illumination plays a pivotal role in computer vision-based systems [2.1] as the appearance as well as features of the images significantly depends on the illumination under which image acquisition has been performed. To reduce the inconsistency due to illumination variation this chapter describes the imaging chamber which has been developed in-house. The goal of the use of an imaging chamber is to use the same imaging conditions throughout the different proposals that have been presented in this thesis. This in turn can bring more justified comparison among the proposals.

This chapter includes another important aspect called performance evaluation methods. Conventionally in the ML fraternity number of metrics are used that convey the performance of the ML models from different aspects. This chapter includes the mathematical development and interpretation of metrics that have been used across the different proposals presented in this thesis.

## **Chapter 2: Imaging chamber development and performance evaluation methods**

---

### **2.2. Principle of Human vision system**

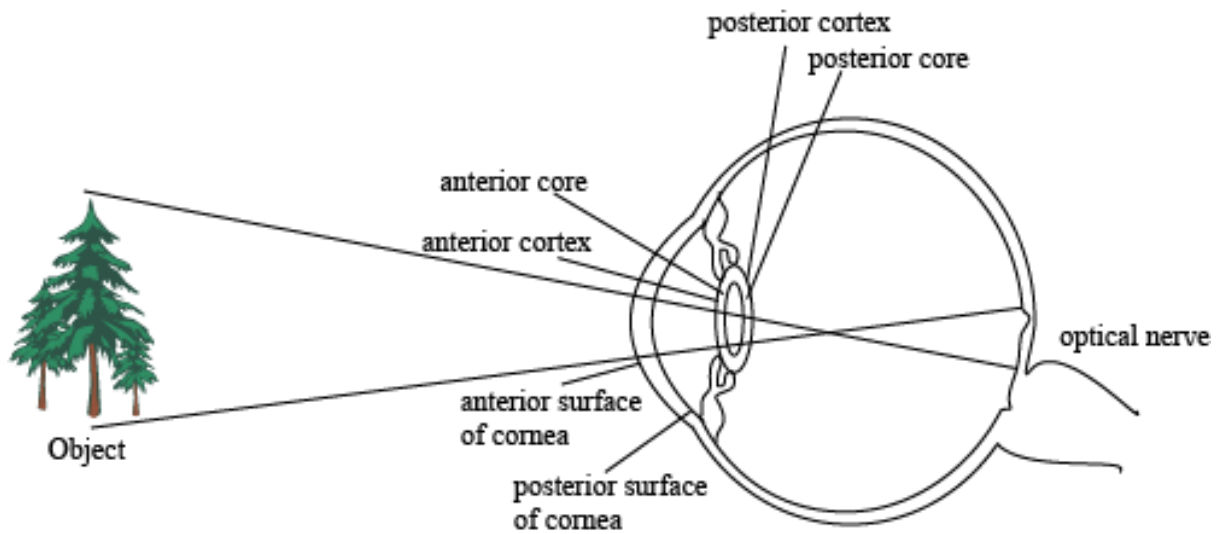
The ability to absorb and comprehend visual information from the surroundings is made possible by the intricate and sophisticated sensory system that is vision system. To convert light into images, which the brain subsequently interprets, requires cooperation between the eyes, optic nerves, and different brain regions. The physiological foundation of visual perception, or the capacity to recognize and interpret light, is the visual system. The system creates an image and a mental representation of the surrounding environment by detecting, transducing, and interpreting information about light within the visual range. The optical system, which includes the cornea and lens, and the neural system, which includes the retina and visual cortex, are the functional divisions of the visual system, which is connected to the eye. The formation of monocular images, the neural mechanisms underlying stereopsis and the assessment of distances to and between objects, motion perception, pattern recognition, accurate motor coordination under visual guidance, and color vision are just a few of the complex tasks that the visual system accomplishes based on the image-forming functionality of the eye. When combined, these help with higher level activities like identifying objects. Visual perception is the neuropsychological aspect of visual information processing; its dysfunction is referred to as visual impairment, and its total absence is referred to as blindness. The pupillary light reflex and circadian photo entrainment are two other non-image generating visual functions of the visual system that operate independently of visual perception. Major components of the human vision system consist of eye, optic nerve, macula and fovea, vitreous humor and visual cortex. The different parts of eye are described below.

Cornea-The clear front layer of the eye which aids in focusing inward light. Pupil- The opening in the center of the iris that controls the amount of light incoming into the eye. Lens- A translucent, flexible device which helps to focus light onto the retina. Retina-The light-

## Chapter 2: Imaging chamber development and performance evaluation methods

---

sensitive layer at the back of the eye which is made up of rods and cone-shaped photoreceptors that translate light into electrical signals. Rods- Photoreceptor, which are sensitive to low light levels. Cones-Photoreceptors, which are in charge of peripheral and night vision, are sensitive to color vision. The basic mechanism behind human vision system is depicted below



**Fig. 2.1:** Human optical system (courtesy: <https://entokey.com/the-human-eye-as-an-optical-system/>)

### 2.3. Imaging chamber development

The imaging chamber was developed in-house. The chassis of the chamber was made of wood and the dimensions (length  $\times$  breadth  $\times$  height) were  $0.5\text{ ft} \times 0.5\text{ ft} \times 1\text{ ft}$ . One of the walls of the chamber remained hinged so that it could be opened for loading and unloading samples. Fixed distance between the camera lens and sample under imaging was ensured throughout the experimentations in this thesis. The four corners of the chamber were mounted with LED strips. The strips were cut according to the diameter and each foot of the strip had 25 LEDs. The color temperature of the strip was 6500K which simulates the natural daylight. The strips were mounted using adhesive tapes which were in-built into the backside of the strips. The



## **Chapter 2: Imaging chamber development and performance evaluation methods**

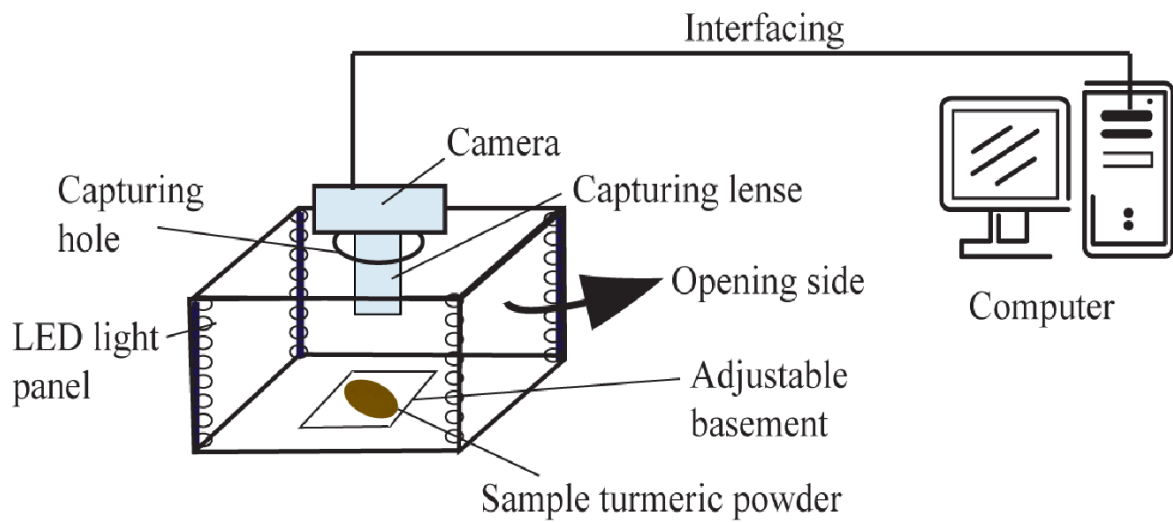
---

LEDs were of 2835 type and the driving current and wattage for each LED were 60mA and 0.2Watts. Each LED could give about 20 Lumen brightness hence when all the 4 sides of the panel were illuminated the brightness of the closed panel was about 2000 Lumen. The color rendering index (CRI) [2.2] of the LEDs was 80 (average of R1 – R8). The LED strips were connected to a 12V 150W DC adaptor which was attached to the chassis. The on/off switch was also provided at the outside wall. No external light sources like flashlights or mobile phones or cameras were used during the imaging process.

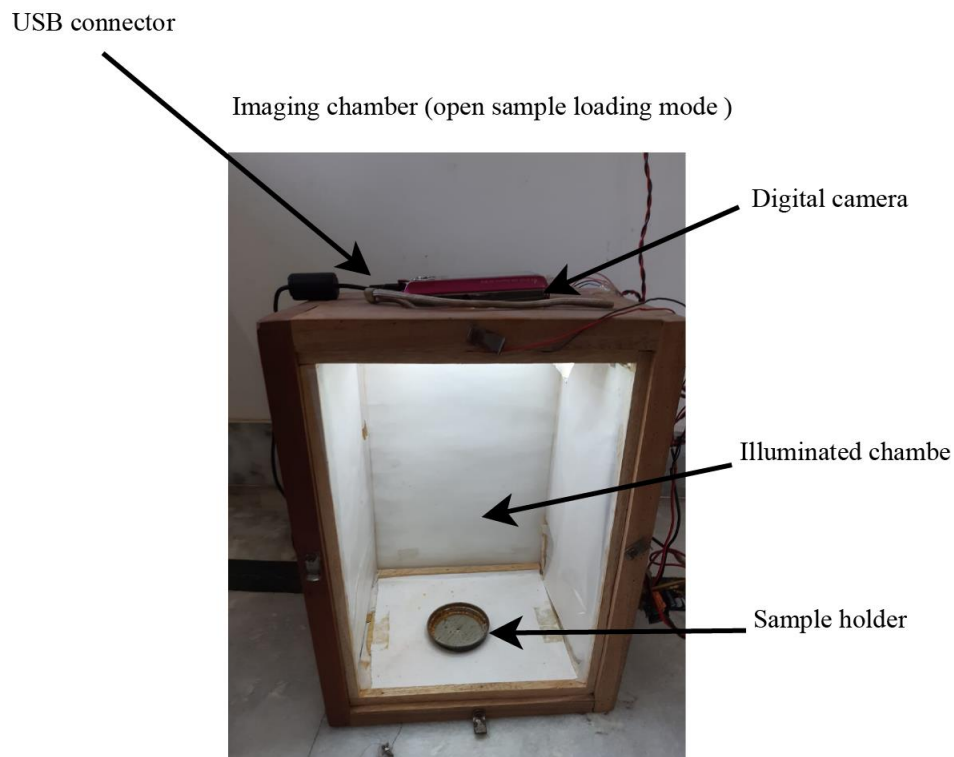
To obtain improved smoothening and distribution of light inside the chamber diffusers were used. In our chamber translucent diffuser paper was used. The thickness of the paper was about 0.06mm. The diffuser sheets were cut into the required dimension and pasted at the beads of the inner walls.

At the roof of the chassis, an opening was made to place the lens of the camera. The opening was made such that in the case of a digital camera the focusing of the lens could be adjusted through the lens movement. However, for mobile phone cameras which were also used in this thesis to simulate more real-life scenarios such lens movements are not present. As shown in Fig. 2.2 the digital camera was connected to the PC for image transfer using a USB 2.0 connector and Matlab programming platform. The loading mode and imaging mode of the actual imaging chamber with labels of different parts/components have been shown in Fig. 2.3. and the closed imaging mode has been shown in Fig. 2.4.

## Chapter 2: Imaging chamber development and performance evaluation methods



**Fig. 2.2:** The schematic of the operating principle of the imaging chamber



**Fig. 2.3:** The actual imaging chamber with labeling to the different components

## Chapter 2: Imaging chamber development and performance evaluation methods

---



**Fig. 2.4:** The actual imaging chamber in imaging mode after loading the sample

### 2.4. Performance evaluation methods

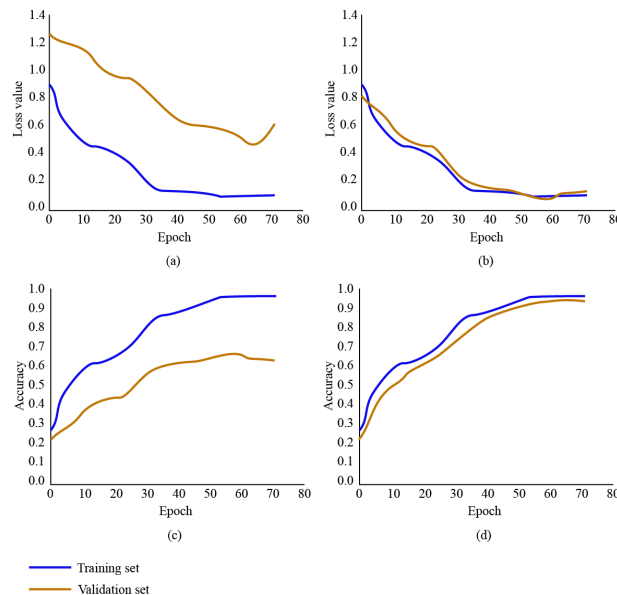
Different evaluation methods and metrics are commonly used in the ML fraternity to verify and validate the potential of the models. The methods and metrics that have been used in this thesis consistently in all the proposals have been briefed in this section for a better understanding and interpretation of the results. It can be noted here that in the supervised models of ML (as been used in this thesis) the entire experimental data is divided into three parts; training, validation, and testing. The training data is used to train the supervised model, validation data is used to validate the generalization of the model and testing is the unknown data to the model used for verification of the model's performance for unknown data.

## Chapter 2: Imaging chamber development and performance evaluation methods

### 2.5. Model validation

#### 2.5.1. Validation of generalization potential

Model validation [2.3] is an important step that ensures the avoidance of two undesired situations called under- and over-fitting. In under-fitting the model performs poorly for both training and validation data. In other words, the model is still capable to learn and perform better. Over-fitting is a more critical one where the model performs substantially with training data but not with validation data. Commonly these are referred to as generalization problems which limit the model for real-life implementation. To verify the generalization potential the common practice is to use the accuracy and loss plots where the accuracy and loss values both for training and validation sets in different epochs are plotted. The desired plot will result similarity in the plot nature between training and validation. Also, a small gap at the end between the plots is desired to ensure a small bias high variance nature of the model. Some examples of the plots for under- and over-fitting [2.4] cases have been shown in Fig. 2.5.



**Fig. 2.5:** Illustration of (a) undesired and (b) desired loss plots along with (c) undesired and (d) desired accuracy plots for training and validation data. (These are randomly generated plots for explanation not plots of any studies conducted in this thesis work).

## Chapter 2: Imaging chamber development and performance evaluation methods

### 2.5.2. *k*-fold cross validation

This is used to verify the consistency potential of the model. In this case, the entire dataset is split into  $k$  number of random splits called folds (in our thesis  $k=10$ ) [2.5]. Out of these  $k$  folds  $k-1$  folds are used for training the model and the rest 1 fold is used for testing. This is repeated for  $k$  to ensure each of the folds is subjected to training and testing data at least once. The selection of folds is also cyclic in manner e.g. in the case of 10-fold cross validation if in the first run folds 2 – 10 are subjected as training set and 1<sup>st</sup> fold is testing then in the next run fold 1 and 3 – 10 will be training set and 2<sup>nd</sup> fold will be testing. Figure 2.6 presents a 10-fold cross-validation graphically where the gray shaded cells represent training folds and the white cell is the testing cell. This concept is shown schematically. This is also sometimes used for parameter optimization of the model where the hyperparameter values for the fold where the model resulted in the highest accuracy and lowest loss are taken for experimentations.

Split1	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split2	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split3	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split4	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split5	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split6	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split7	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split8	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split9	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
Split10	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10

**Fig. 2.6:** A 10-fold cross-validation where shaded cells are training folds and white cells are testing folds. The tunable parameters of ML or optimization algorithms.

## Chapter 2: Imaging chamber development and performance evaluation methods

### 2.5.3. Confusion matrix

Confusion matrix [2.6] is one of the most popular metrics used across the ML fraternity to verify the performance potential of models, particularly, for classification tasks. For a binary classification task where there are two classes called positive and negative it can be represented as the matrix shown in Fig. 2.7. As it can be seen results are segregated into 4 attributes; namely true positive (TP), false positive (FP), true negative (TN) and false negative (FN). TP are the classified samples that have been correctly classified as positive while FP are the samples that are negative but have been identified as positive. Similarly, TN is the cases that have been correctly identified as negative but FN are positive but identified as negative. Although, the concept of a confusion matrix is developed on a binary classification basis it can be extended to multiple classes considering the one-vs-all (OVA) [2.7] strategy where the class to be identified is taken as positive and all rest are negative.

		Actual classes	
Predicted classes		Positive	Negative
	Positive	TP	FP
	Negative	FN	TN

**Fig. 2.7:** Confusion matrix of binary classification.

It can be noted that the confusion matrix can be shown both in terms of numbers of TP, FP, FN, and TN or also as a percentage. Since percentage provides a better picture of the performance in this thesis that has been followed for all the proposed methods.

## Chapter 2: Imaging chamber development and performance evaluation methods

---

### 2.5.3.1. Metrics derived from confusion matrix

The following metrics are derived from confusion matrix parameters which convey the potential of the models from different aspects.

- *Precision*: It reflects the cases which have been predicted as positive by the model are truly positive in reality [2.8]. This is also called ‘confidence’ and mathematically expressed as Eq. 2.1.

$$Precision = \frac{TP}{TP + FP} \quad (2.1)$$

*Recall*: It reflects how many of the real positive cases have been predicted as positive by the model [2.8]. This is also called ‘sensitivity’ and mathematically expressed as Eq. 2.2.

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

- *F-measure*: It is the harmonic mean of precision and recall as shown in Eq. 2.3 [2.8]. It is particularly useful to measure the consistency and balance of the model, which means whether the model can perform well across all the classes in the problem or the performance is biased towards some classes and poor for other classes.

$$F\_measure = \frac{2 \times recall \times precision}{(recall + precision)} \quad (2.3)$$

- *Accuracy*: It is a simple measure of how many samples have been correctly classified in individual classes and this can be extended to the entire testing dataset by summing the diagonal values and dividing by a total number of samples under consideration. The class-wise and overall accuracy can be expressed as Eq. 2.4 and 2.5, respectively.

$$Accuracy_k = \frac{TP_k + TN_k}{TP_k + TN_k + FP_k + FN_k}; k \text{ is the class index} \quad (2.4)$$

## Chapter 2: Imaging chamber development and performance evaluation methods

---

$$Accuracy_{overall} = \frac{\sum_k (TP_k + TN_k)}{Total\ number\ of\ samples} \quad (2.5)$$

- *Receiver Operating Characteristics (ROC)*: This is a comparison between true positive rate (TPR) and false positive rate (FPR) where TPR is ratio of number of predicted positive cases to the actual positive cases and FPR is similarly ratio of predicted negative cases to actual negative cases. TPR and FPR are also referred as sensitivity and specificity, respectively [2.9]. Equation 2.6 and 2.7 represent these parameters.

$$TPR = \frac{TP}{TP + FN} \quad (2.6)$$

$$FPR = \frac{TN}{TN + FP} \quad (2.7)$$

- *ROC-AUC*: The area under the curve (AUC) of ROC plot conveys the probability of accurate prediction for the model [2.10]. It is conventionally plotted against a naïve classifier with a 45° straight line which has a 50% probability of correct prediction, means it has equal chances to predict positive and negative for any sample. An example of ROC-AUC plot is shown in Fig. 2.7 where this naïve classifier is shown in dotted line. The classifier 1 in Fig. 2.8 has low AUC than classifier 2, hence classifier 2 is considered to be a better model.



## Chapter 2: Imaging chamber development and performance evaluation methods

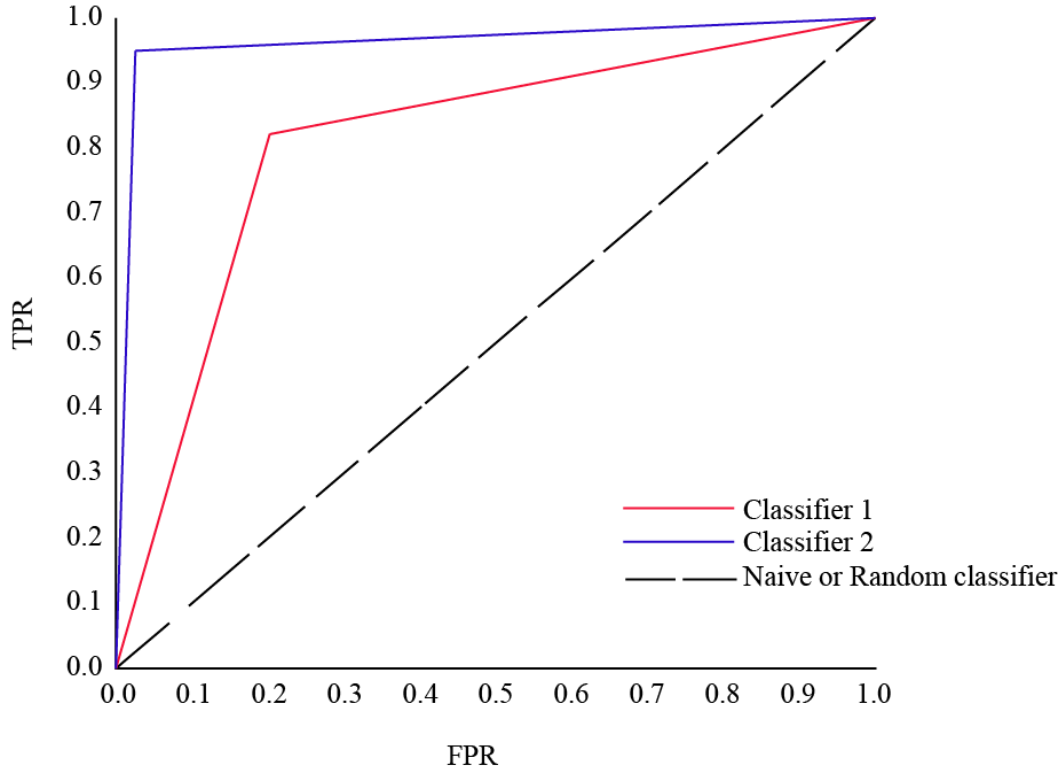


Fig. 2.8: ROC plot.

### 2.6. Evaluation of regression results

The previously described metrics are mostly used in the case of classification tasks. For regression-based predictions, there are separate metrics as class-wise metrics like confusion matrix cannot work there. Following are some of the most commonly used metrics for regression-based prediction tasks [2.11].

- *Root means squared error (RMSE)*: It is the measure of the difference between actual and prediction values and can be calculated as Eq. 2.8.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (2.8)$$

where;  $y$  and  $\hat{y}$  are actual and predicted values, respectively and  $j$  is the index of samples. A lower value of RMSE is desired for a good prediction model.

## Chapter 2: Imaging chamber development and performance evaluation methods

---

- *Mean absolute error (MAE)*: MAE is also popularly used as a loss function to assess the potential of the model and is expressed as Eq. 2.9.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (2.9)$$

where;  $y$  and  $\hat{y}$  are actual and predicted values, respectively and  $j$  is the index of samples. Being a loss function this metric also reflects better potential with lower values.

- *Coefficient of determination*: This measure is popularly referred to as  $R^2$  value. This is represented as Eq. 2.10.

$$R^2 = 1 - \frac{RSS}{TSS} \quad (2.10)$$

where;  $RSS$  stands for the residual sum of squares and is calculated as  $\sum_{j=1}^n (y_j - \hat{y}_j)^2$  and

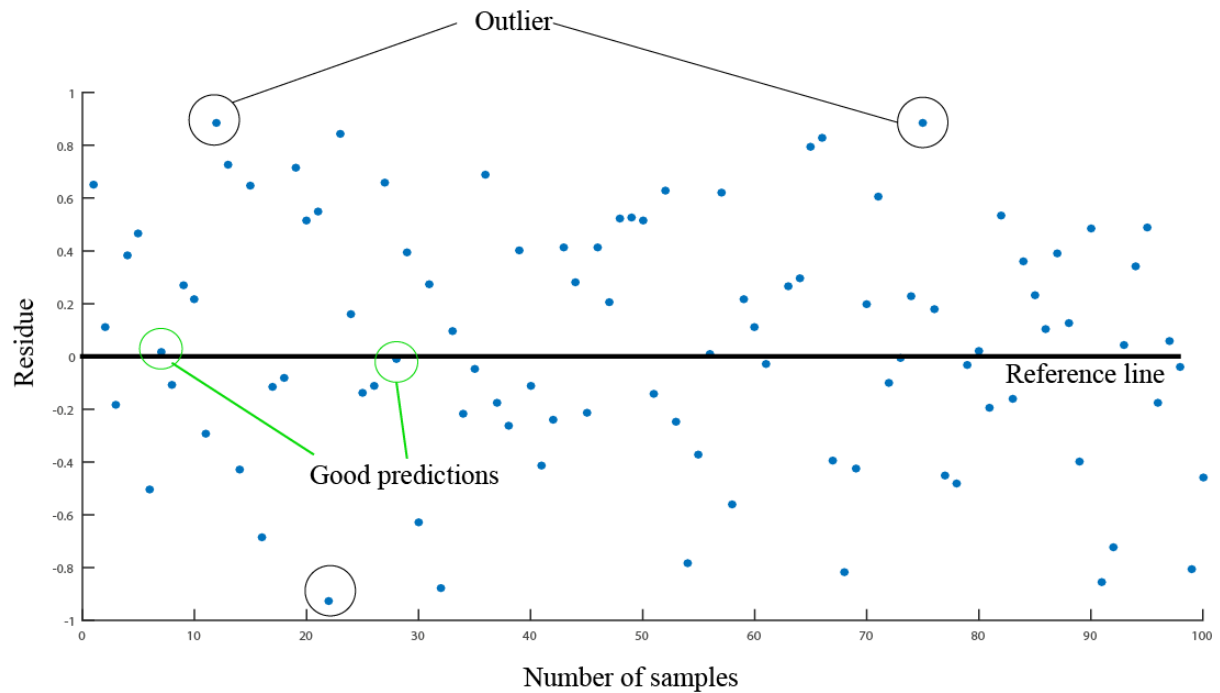
$TSS$  stands for a total sum of squares and is calculated as  $\sum_{j=1}^n (y_j - \bar{y})^2$  considering  $\bar{y}$  as

the mean of the samples. A higher value i.e. close to 1 represents good prediction potential of the model.

- *Residual plot*: This is a graphical representation of the predicted results [2.12]. In this plots the signed difference between actual and predicted values are plotted as scatter plot and a reference line is put at the value 0. This reference line helps to see the degree of deviation resulting from the predicted values. The more the points are closer and on the reference line better the prediction potential. This plot also helps to identify the number of outliers resulting from the prediction model. Examples of residual plots have been shown in Fig. 2.9.

## Chapter 2: Imaging chamber development and performance evaluation methods

---



**Fig. 2.9:** Residual plot with random data.

### 2.7. Implementation environment

All the methods proposed in this thesis were implemented in Windows PC having a corei3 processor with 8GB RAM. The image cropping and processing were performed in Matlab. The machine learning algorithms were implemented in Python using Spyder IDE and standard libraries.

### 2.8. Conclusion

This chapter presents the detail of the indigenous imaging chamber that has been used throughout the works carried under this thesis. The chapter also elucidates the major quality evaluation methods and metrics which have been used for all the proposals. The objective of this was to provide a preamble of the evaluation method which can be helpful to evaluate the proposals presented in this thesis.

## Chapter 2: Imaging chamber development and performance evaluation methods

---

### References

- 2.1 Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature.
- 2.2 Li, C., Ronnier Luo, M., Li, C., & Cui, G. (2012). The CRI-CAM02UCS colour rendering
- 2.3 Sornette, D., Davis, A. B., Ide, K., Vixie, K. R., Pisarenko, V., & Kamm, J. R. (2007). Algorithm for model validation: Theory and applications. *Proceedings of the National Academy of Sciences*, 104(16), 6562-6567.
- 2.4 Pothuganti, S. (2018). Review on over-fitting and under-fitting problems in Machine Learning and solutions. *Int. J. Adv. Res. Electr. Electron. Instrum. Eng*, 7, 3692-3695.
- 2.5 Arlot, S., & Lerasle, M. (2016). Choice of V for V-fold cross-validation in least-squares density estimation. *Journal of Machine Learning Research*, 17(208), 1-50.
- 2.6 Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-label confusion matrix. *IEEE Access*, 10, 19083-19095.
- 2.7 Gupta, R., Alok, A., & Ananthakrishnan, S. (2019). One-vs-all models for asynchronous training: An empirical analysis. *arXiv preprint arXiv:1906.08858*.
- 2.8 Beauxis-Aussalet, E., & Hardman, L. (2014, November). Simplifying the visualization of confusion matrix. In *26th Benelux conference on artificial intelligence (BNAIC)*.
- 2.9 Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*, 58(10), 982-990.

## **Chapter 2: Imaging chamber development and performance evaluation methods**

---

- 2.10 Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4), 561-577.
- 2.11 Oz, I., Bhatti, M. K., Popov, K., & Brorsson, M. (2019). Regression-based prediction for task-based program performance. *Journal of Circuits, Systems and Computers*, 28(04), 1950060.
- 2.12 Larsen, W. A., & McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics*, 14(3), 781-790.

# CHAPTER 3

## Detection of Metanil Yellow Adulteration in Turmeric Powder

### 3.1. Introduction

This chapter presents a computer vision framework to identify the metanil yellow (MET) adulteration in powdered turmeric. Among different adulterants, *MET* is one of the most commonly used ones. It is a synthetic compound ( $C_{18}H_{14}N_3NaO_3S$ ) and gives a bright yellow appearance to the turmeric powder. Its similarity to the turmeric powder appearance is one of the main reasons for its common usage as an adulterant to turmeric powder [3.1]. Its consumption can cause severe health hazards to gastrointestinal function, liver, and even to the extent of cancer [3.2]. Thus detecting the presence of *MET* in powdered turmeric is of immense importance but a very difficult task due to its appearance-wise similarity to the turmeric powder [3.3].

The works presented in this chapter has been partially published in author's following publication Mandal, D., Chatterjee, A., & Tudu, B. (2022). A deep neural network and random forests driven computer vision framework for identification and prediction of metanil yellow adulteration in turmeric powder. *Concurrency and Computation: Practice and Experience*, 34(1), e6500.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

Chemical and spectral analysis using analytical instruments [3.4] are the most widely reported techniques for MET adulteration in turmeric some of the major such works have been briefed here. The application of a gradient reverse-phase high-pressure liquid chromatographic (RP-HPLC) method is reported in. The application of NIR-spectroscopy has shown the coefficient of determination ( $R^2$ ) value up to 0.99[3.5]. The application of FT-Raman and FT-IR spectroscopy reported in has shown a correlation coefficient of 0.93 and 0.95 for FT-IR and FT-Raman, respectively [3.6]. The application of HPLC using fused core-column for the detection of different curcuminoids has shown more than 99% prediction accuracy.

Fourier transform infrared spectroscopy-partial least square (FTIR-PLS) has also been applied to the detection and analysis of total and individual curcuminoids which resulted in more than 98% prediction capability in terms of  $R^2$  values Application of reverse phase HPLC has also shown more than 99% prediction capability for purity determination of turmeric dietary supplements [3.7]. The equivalent degree of prediction capability has also been reported for geographical discrimination in curcuminoid content in turmeric using a rapid UPLC-DAD-validated analytical method [3.8]. Two-dimensional high-performance thin-layer chromatography (2D-HPTLC) [3.9] has also been employed to detect extraneous colors in turmeric powder. In[3.10] Savitzky-Golay first-derivative-based pre-processing of the NIR handheld spectrometer response was subjected to extract the first principal component (PC1) of principal component analysis (PCA) and that was used as a feature to classify adulterated turmeric powder using soft independent modeling class analogy (SIMCA) model. The amount of *MET* has been predicted using a handheld NIR device and partial least square regression and obtained more than 96% prediction accuracy in terms of  $R^2$

## **Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder**

---

values[3.11].In a DNA barcoding-based application of adulteration[3.12] detection in turmeric has been reported.

This chapter includes in-house database preparation with different degrees of adulteration and imaging of the samples using the indigenous imaging chamber described in Chapter 2. The working principle is based on machine vision which has emerging applications in various fields including food quality surveillance [3.13], geoinformation systems [3.14], medical applications [3.15], and security surveillance systems [3.16]. The chapter also presents new frequency domain image features. The application of random forest (RF) and deep neural networks (DNN) as classifiers and predictors have also been presented in the chapter. The chapter concludes with the obtained results and related discussions.

### **3.2. Materials and methods**

#### ***3.2.1. Experimental in-house database preparation***

Here five varieties of turmeric powder were examined in this study; one was made in-house, and the other four were packaged organic turmeric powders from four well-known companies that claim to be 100% pure. To make the in-house version of turmeric powder, with raw turmeric roots obtained from nearby markets. After 15 minutes of boiling in water, the roots were dried in an oven set at 70°C. To make powder, the dried roots were placed in a mixer blender .After grounding each of the five turmeric variations was divided into four equal portions. MET of the Sigma Aldrich brand was procured from research lab suppliers. One part was labeled "pure" since it had not been subjected to the mixing of MET. The remaining three portions were marked as "adulterated" after being adulterated with three different percentages of MET. The weight-on-weight (w/w) percentages of MET were 5%, 10%, and



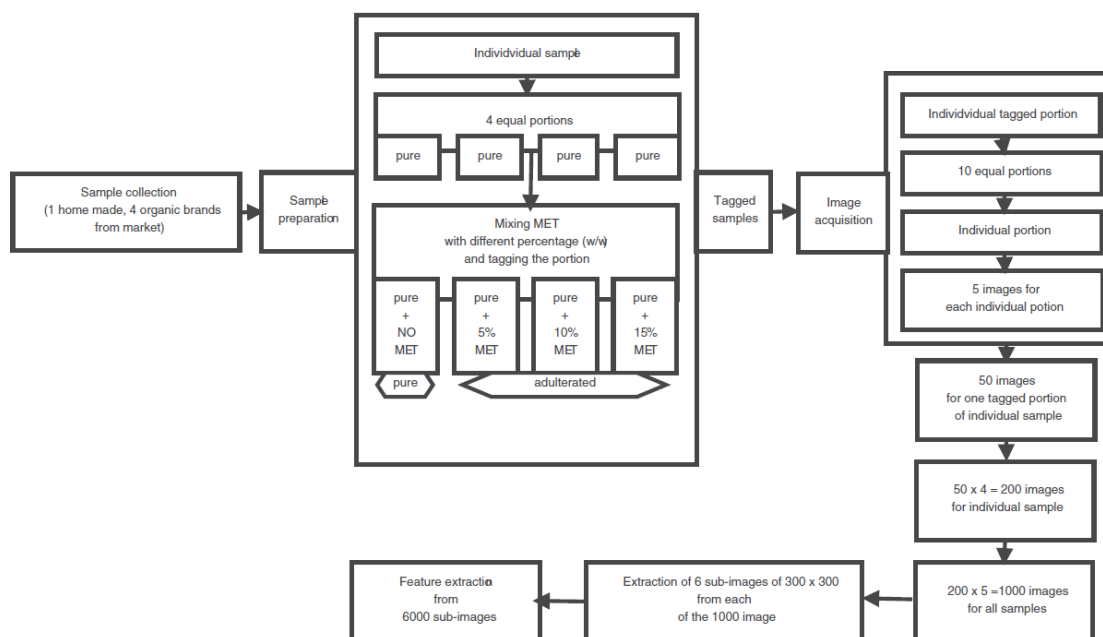
### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

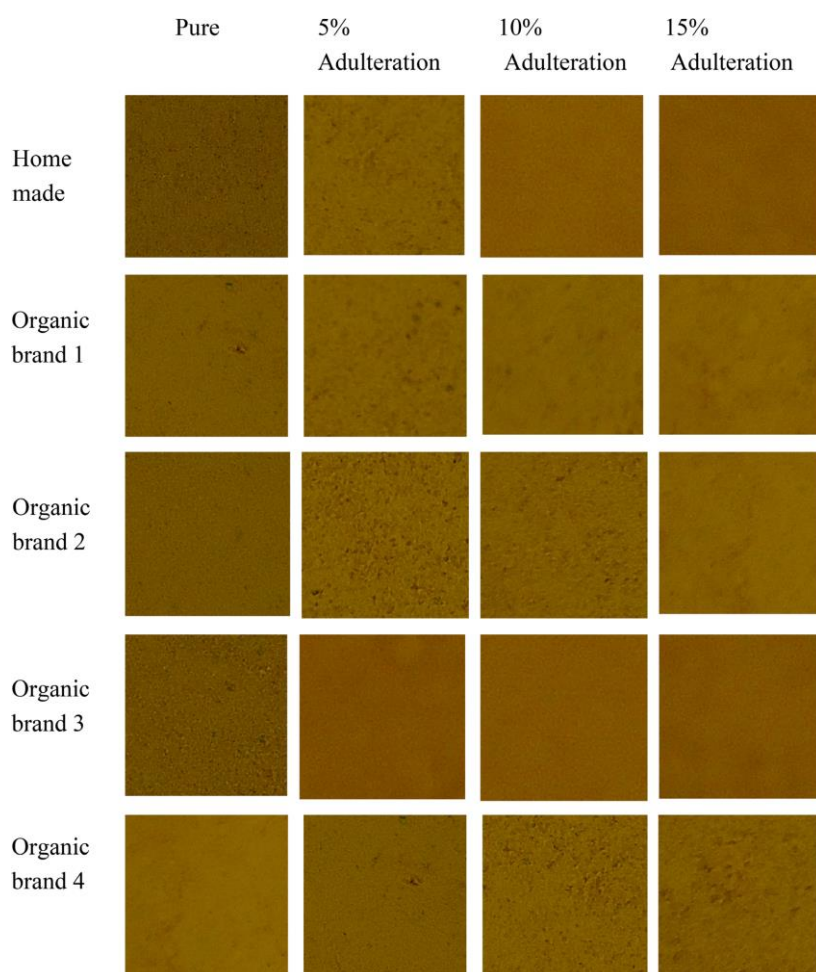
15%. The produced samples were kept dry and protected from the elements until the imaging procedure was applied. They were stored in airtight containers.

Each of the 4 portions (1 *pure* and 3 *adulterated*) for each of the 5 turmeric samples (1 homemade and 4 branded) was divided into 10 equal portions for imaging. The individual classes for each portion were subjected 5 times in the imaging chamber. That resulted in 50 images for each portion and a total of 200 images for each of the 5 variants. This way the final image database was comprised of 1000 ( $300 \times 300$ ) images with 250 *pure* labeled and 750 *adulterated* labeled samples having 250 images with each of 5%, 10%, and 15% *MET* adulteration. Further 6 sub-images from each of the 1000 images were extracted randomly to obtain the final image database with 6000 images. The image database preparation process has been presented in Fig. 3.1. The images were captured in JPEG format since it is the most widely used image format despite its lossy nature. Also achieving promising results with JPEG images can vouch for the generalization possibilities of the presented model in a mobile phone app-based development where the images will be captured by a mobile phone camera in JPEG format. Some of the samples from the captured images of the in-house database are shown in Fig. 3.2.

## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder



**Fig. 3.1:** Schematic representation of dataset preparation strategy



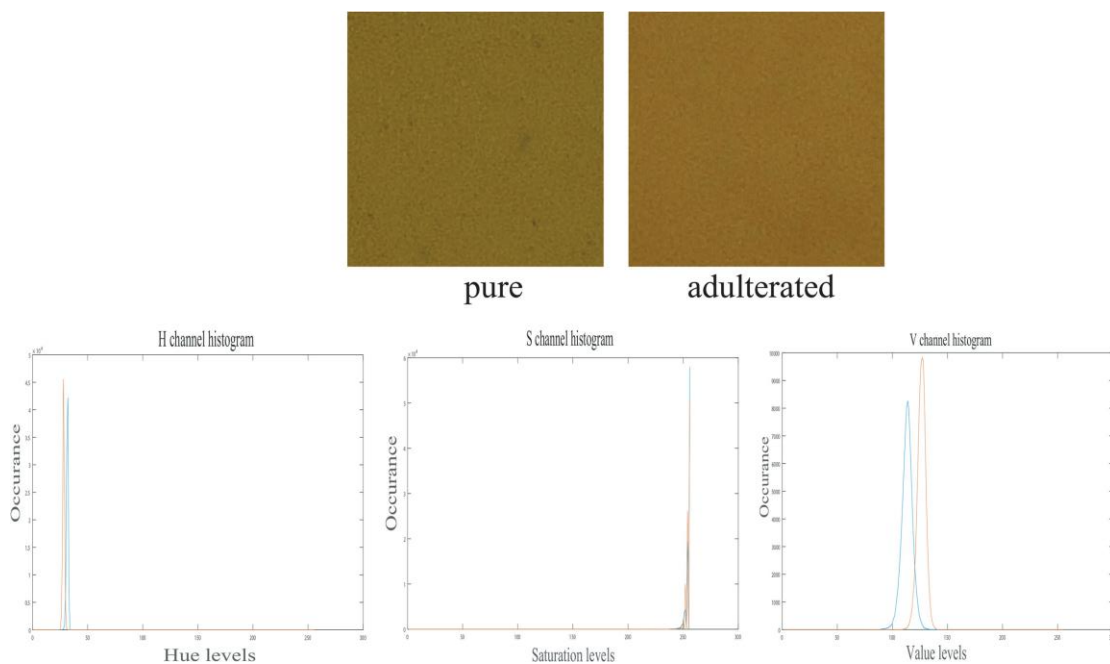
**Fig. 3.2:** Samples of images from the experimental image database

## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

### 3.2.2. Frequency domain feature extraction

The visual appearance of the pure and contaminated samples cannot be easily distinguished in their native red, green, and blue (RGB) color scheme, as can be observed in Fig.3.2. To provide more information than a device-dependent native image format like RGB, the collected photos were first transformed to hue, saturation, and value (HSV) color space. Fig. 3.3 shows the characteristics of H, S, and V channel histograms for pure and adulterated test images.

As it can be seen the shift in the peaks of the histogram is not very noticeable in the H and S channels. This illustrates how the color and saturation of turmeric powder are largely unaffected by the addition of MET. Given this, adulteration is hard to spot by straightforward visual inspection. On the other hand, the shift in the V channel is easily visible, and also the shift is towards the brighter side of the histogram. This conveys the fact that the addition of *MET* provides a brighter appearance of pure turmeric which often misleads buyers.



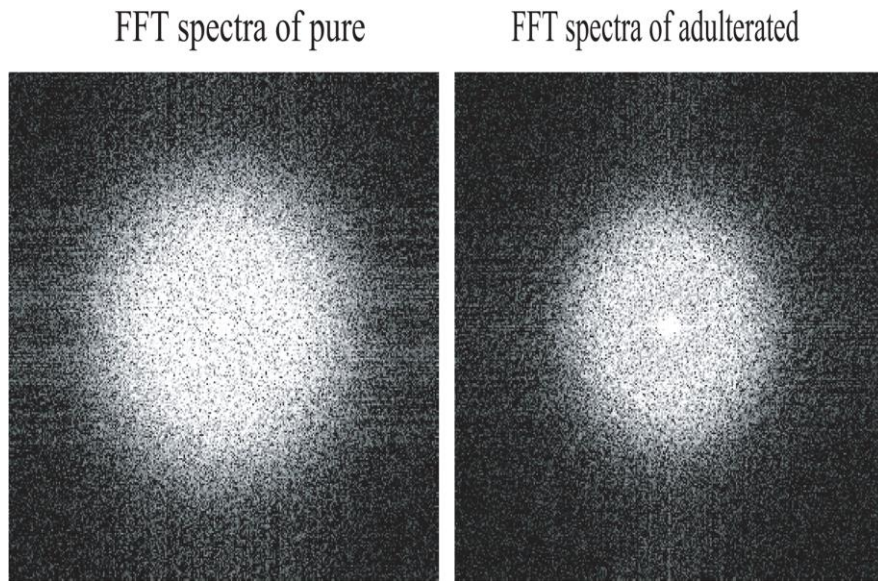
**Fig. 3.3:** Image analysis using histogram of HSV channels.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

After color space conversion the annular-mean feature (AMF) frequency domain feature extraction technique was used on the V channel. There are several reasons why frequency domain transforms are common in image processing applications; in this instance, the simpler interpretation of the difference is one of them. Figure 3.4 shows the fast Fourier transform (FFT) spectrum of images for pure and adulterated turmeric which demonstrates the variations in frequency distributions. The FFT spectrum of *pure* shows more mid-frequency components than the adulterated FFT spectrum. The possible reason is, that adding yellow *MET* will introduce flatness to the image since *MET* is quite smooth in texture and does not result in considerable fluctuations in concentration when uniformly mixed with turmeric. That causes an overall texture reduction in the adulterated sample which in turn causes the FFT spectra to concentrate more towards the low-frequency regions, that is, the center of the spectra. AMF can be a simple measure to quantify this observation. AMF due to its concentric annular partitioning of the entire spectra can easily assess the amount of frequency concentration among different frequency bands. For instance, the pure samples will have higher values in the annular rings covering higher frequency bands since the FFT spectra of adulterated samples do not show the presence of considerable frequency components there.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder



**Fig. 3.4:** FFT spectra of pure and adulterated turmeric images.

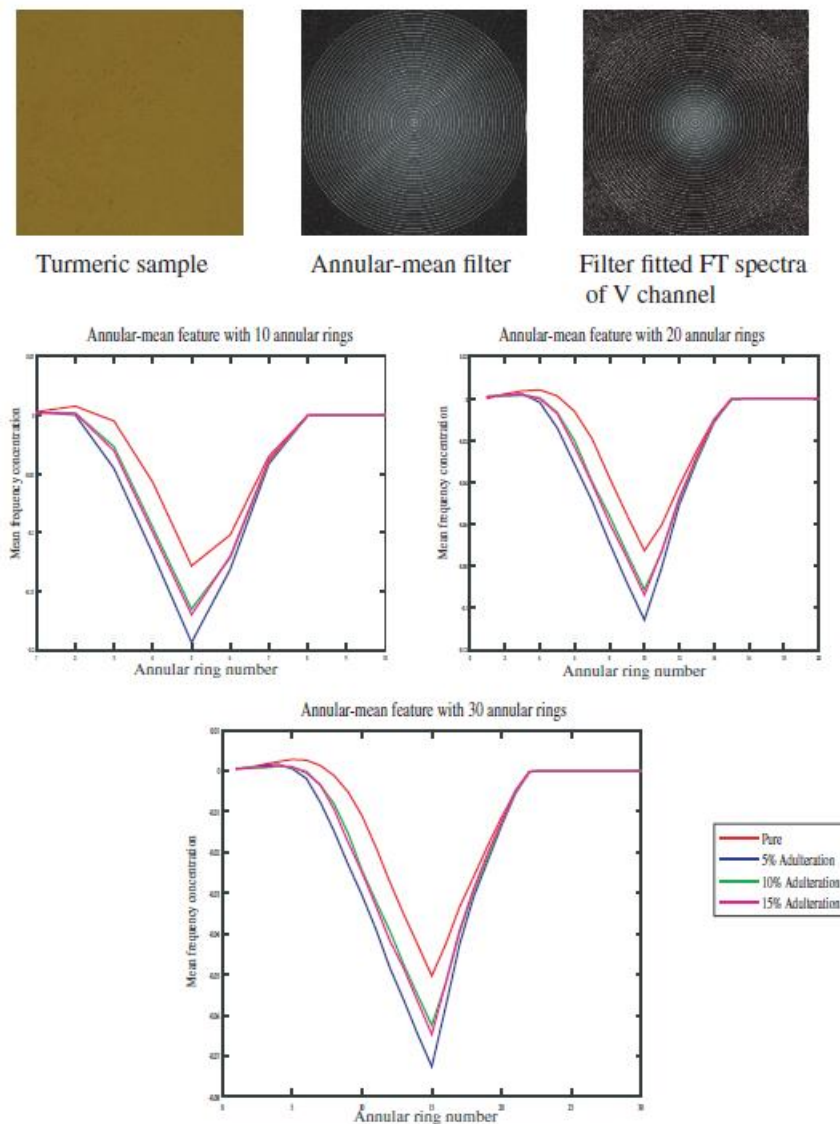
The log-transformed FT spectra were subjected to an annular mean frequency (AMF) feature extraction. AMF initially divides the FFT spectra into concentric annular rings as shown in Fig. 3.5. Then the mean value of frequency under each ring was calculated and plotted as shown in Fig. 3.5. The step size of the AMF could be user-specific where the numbers of annular rings decide the length of the feature set. It can be noted that Fig.3 .5 shows an example of a captured image from the database and AMF-masked FT spectra of V-channel along with the plots of features extracted with 10, 20, and 30 annular rings. The plots of Fig.3.5 also show how the AMF can visually distinguish between the *pure* and *adulterated* turmeric samples. These AMFs were scaled to bring all the features in the same interpretability by mean subtraction normalization as expressed in Eq. (3.1) [3.17].

$$\hat{x} = (x-u)/s \quad (3.1)$$

where,  $\hat{x}$  is the scaled value of  $x$ ,  $u$ , and  $s$  represent the mean and standard deviation of the dataset.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

Apart from distinguishability, the plots have also revealed that there were no substantial frequency concentrations at higher frequency ranges. Therefore, the higher frequency concentrations which were resulting '0' were truncated. For example, with 30 annular rings, 23<sup>rd</sup> ring onwards the values were becoming 0, hence the final featureset was prepared taking to 23<sup>rd</sup> value of the annular-mean filter output. The dimension of the final feature set for classification and prediction was 6000×23.



**Fig. 3.5:** Example of annular rings on FFT spectra and plots of extracted features with different numbers of annular rings.



## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

### 3.2.3. Data analysis models

A deep neural network (DNN) is used here for analysis. Though DNN is suitable for large data sets nowadays it is found suitable for small data sets also [3.18]. The remarkable performance with small datasets [3.19] is the reason for using DNN. The feature set as discussed in the previous section was subjected to a two-folded goal i.e. classification between *pure* and *adulterated* turmeric powder and prediction of the probable degree of adulteration present in the sample. Incorporating treatments such as data augmentation, pre-trained models [3.20], and employing the cosine loss function [3.21] are commonly employed when utilizing Deep Neural Network (DNN) with small datasets to mitigate overfitting issues. This work highlights the utilization of two robust classification and regression models: Random Forest (RF) [3.22] and Deep Neural Network (DNN) [3.23]. These models have demonstrated superior performance compared to other classifiers [3.24] such as logistic regression (LR), support vector machine (SVM), k-nearest neighbor (KNN), and decision tree (DT) [3.25] classifiers across various engineering applications.. A brief introduction to these two models has been included here.

#### 3.2.3.1. Random forest (RF) classifier and regressor

RF is one of the robust ensemble learning algorithms that has shown better results than some of the other popular algorithms in both classification and regression tasks. The base learner of RF is decision tree (DT) but it overcomes the over-fitting problem of decision tree using an ensemble of output from numbers of DTs. The over-fitting problem causes a lack of generalization and poor performance on unknown data while excellent performance with the known data. RF is widely used by different machine learning applications since it is also advantageous in terms of smaller number of tunable parameters. In particular two major tunable parameters are the number of trees ( $n_{trees}$ ) and the number of variables ( $m$ ) where  $p$  is the total number of variables i.e. 23 in our case) attached to individual trees. During

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

classification of observations, each DT in RF optimizes towards maximization of the information gain ( $IG$ ) calculated between the parent and child nodes resulted during splitting using entropy as expressed as Eq. (3.2). The basic steps of the algorithm are shown in Table 3.1.

$$IG = \text{entropy (parent)} - [\text{Weight average entropy (child)}] \quad (3.2)$$

Where, entropy for the observation is calculated using conventional formula considering as the probability of occurrence of a particular instance in other word the probability of belongingness to a particular class in case of classification.

**Table 3.1:** RF algorithm (for both classification and prediction)

---

Step1: Initiate the algorithm with $n_{trees}$
Step 2: Associate each tree with a <i>bootstrap</i> from the original dataset ( <i>bootstrap</i> refers to the process of selecting a random subset of samples as well as features from the entire dataset with replacement)
Step 3: Unpruned regression or classification trees ( $T_{bs}$ ) are grown to the <i>bootstrap</i> samples repeating following rules at each terminal node:  Randomly $m_{try}$ variables from the total $p$ variables in the model are chosen  The variable that results in best binary split of the input space is calculated.  The variable that results in the best binary split of the input space is calculated.  The input space is split with the chosen best-split variable.
Step 4: The final classification and prediction decision is made based on the output of all trees created by an algorithm. For prediction, the averages of all the predictions from individual trees are calculated as follows and are

---



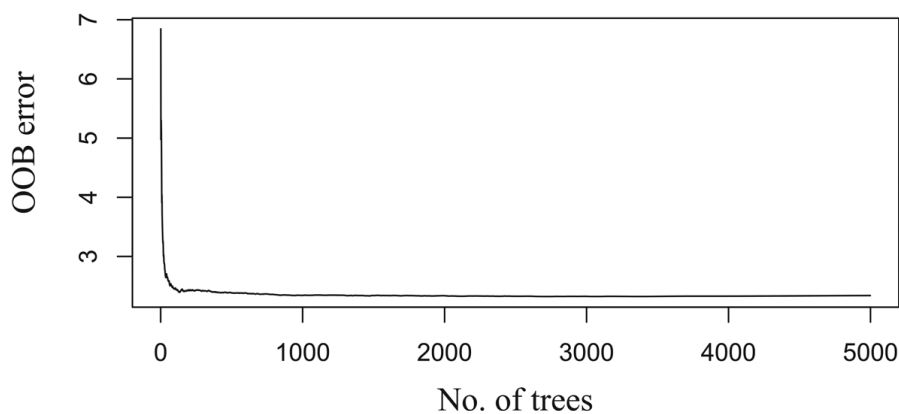
## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

declared as the final prediction.

$$\hat{Y}(x_i) = \frac{1}{B} \sum_{b=1}^B T_b(x_i) \text{ for prediction on } i^{th} \text{ sample of on feature set } x$$

For classification applications, the *majority voting* strategy is used where the class to which the majority of the trees vote is declared as the final output.

The RF model is commonly evaluated against the *out-of-bag* (OOB) error metric. In general two-thirds of the entire dataset is used for training and the remaining one-third is used for OOB error evaluation. This ensures that the trees are subjected to predict the samples with which they have not been trained. The performance of the RF in this work with the different numbers of trees and the OOB errors with different *trees* are shown in Fig.3.6. It can be seen that the OOB error does not improve after engaging 2000 trees. Hence  $n_{trees}$  parameter for our model is fixed as 2000. The trained RF model has been used for both prediction and classification purposes in this work.



**Fig. 3.6:** Oob errors plotting of RF model with different number of trees.

### 3.2.3.2. Deep neural network (DNN) classifier

DNN is one of the emerging machine learning algorithms that has shown remarkable performance across diverse fields. It is particularly important for cases where the data spread

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

is complex, and non-linear and its parent proposition artificial neural network (ANN) performs poorly to establish a satisfying correlation between input and output data pattern. DNN consists of three types of layers, input, hidden (maybe one or multiple), and output layer. Each layer consists of a given number of operational nodes. The nodes are designed with an activation function that governs the non-linearity and output of the nodes. In DNN all the nodes between different layers are connected in a weighted fashion. Output nodes provide the probabilities of belongingness to the different classes based on the activation function of the output layer nodes. The DNN reaches the training goal by minimizing the *loss* function following the specified minimization algorithm.

In the presented architecture the activation function of the hidden layer nodes is set as a rectified linear unit (ReLU). The transformation of input data with ReLU follows Eq.(3.3) [3.26] which shows that the output of the ReLU function is the same as the input  $z$  value in case  $z$  is positive else it is 0.

$$R(z) = \max(0, z) \quad (3.3)$$

The activation function of the output node was set to *softmax* which follows the principle shown in Eq. (3.4) [3.27] where  $p$  is the probability of belongingness to the  $j^{th}$  class among total  $k$  classes.

$$p(y = j | \Theta^i) = \frac{e^{\Theta_j^i}}{\sum_{k=0}^K e^{\Theta_k^i}} \quad (3.4)$$

The weight adjustment of the network is achieved by a back-propagation mechanism where the *loss* at the output layer is calculated by a *categorical-cross-entropy* function that generates a probability distribution among a defined number of classes while the

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

sum of all the probabilities becomes 1. The expression for *categorical- cross-entropy* is shown in Eq. (3.5) [3.28] considering  $y$  and the actual and predicted result on  $M$  observations and  $N$  classes.

$$L(y, \hat{y}) = - \sum_{j=0}^M \sum_{i=0}^N (y_{ij} \log (\hat{y}_{ij})) \quad (3.5)$$

The weight adjustment is achieved using *adam* [3.29] dynamics which make a balance between stochastic gradient descent (SGD) and RMSprop. The learning rate is scaled using a squared gradient and the momentum attribute of SGD. The main advantage of *adam* is its automatic adjustment of the learning rate in an adaptive manner. The weight update in *adam* is expressed as Eq.(3.6) where  $w_t$  is the weight at  $t^{th}$  iteration,  $\eta$  is the step size which depends on the number of iterations,  $\hat{m}_t$  and  $\hat{v}_t$  are the bias-corrected estimators for the first and second moments and very small constant value can be set as  $10^{-8}$ .

$$w_t = w_{t-1} \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \quad (3.6)$$

One advantage of DNN is its potential to avoid over-fitting by the use of a batch normalization mechanism [3.30]. This function re-centers and rescales the output of a hidden layer before passing it to the next hidden/output layer toward reducing the internal covariance shift of the nodes in that hidden layer. It is implemented through mini-batch statistics where a mini-batch in and the output is expressed as Eq. (3.7). The DNN model parameters as implemented for this work are listed in Table 3.2 and the general architecture is shown in Fig.3.7.

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (3.7)$$

Where,  $\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}$  ;  $\varepsilon$  a small constant for numerical stability,  $\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2$  ;

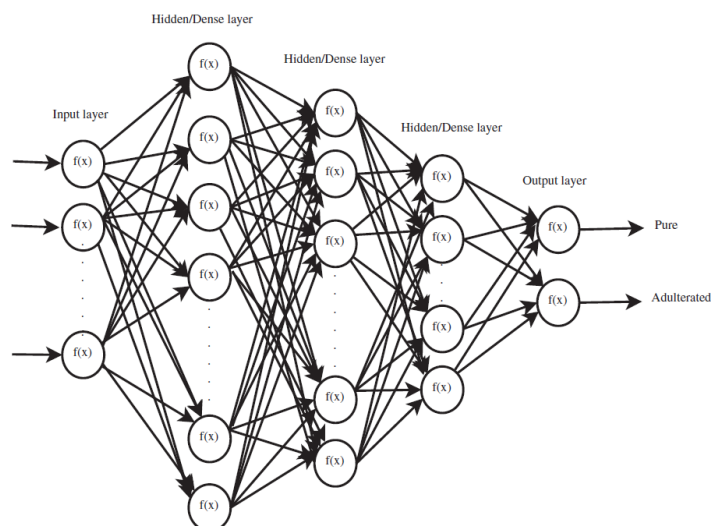
### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

the variance and  $\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$  : mean of the mini-batch  $B$  comprising  $m$  data points.

**Table 3.2:** DNN layer parameters

Layer	Output shape	No. of parameters
Dense 1	(None, 500)	3500
Batch_normalization 1	(Batch(None, 500))	2000
Dense 2	(None, 250)	125250
Batch_normalization 2	(Batch(None, 520))	1000
Dense 3	(None, 125)	31375
Batch_normalization 3	(Batch(None, 125))	500
Dense 4	(None, 50)	6300
Batch_normalization 4	(Batch(None, 50))	200
Dense 5	(None, 2)	102
Total parameters: 170,227		
Trainable parameters: 168,377		
Non-trainable parameters: 1,850		

## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

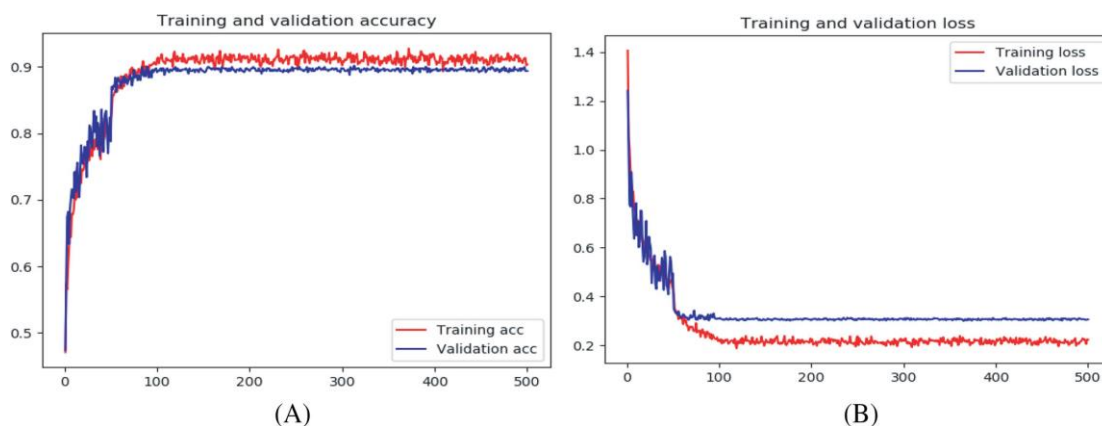


**Fig. 3.7:** General architecture of DNN.

### 3.3. Results and discussions

The entire dataset containing 6000 observations was partitioned into training, validation, and testing set in 60:20:20 ratios. The DNN model stability was verified using training-validation statistics as shown in Fig. 3.8 where the DNN model was trained for 500 epochs. The parameter values of DNN and RF as used for this work have been consolidated in Table 3.3. Figure 3.8 shows that both the accuracy and loss plots for the validation set closely match the trend of training plots with a small gap between them. Matlab was used for image acquisition and pre-processing, and Python was used for classification and prediction using Anaconda Spyder [3.31], Tensor Flow 2.0 backend [3.32], Scikit-learn [3.32] Python libraries, and keras[3.30]. Such similar trends confirm that the model does not result in over- or under-fitting and potential of generalization towards the correct identification of unknown observations. The accuracy of the training and validation set was achieved to the tune of 98% and 96%, respectively, which conveys the promising generalization potential of the model as well.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder



**Fig. 3.8:** Model validation plots of training and validation set using (A) accuracy and (B) loss.

Another popular method for system consistency evaluation is 10-fold cross-validation. The advantage of this process is that each of the observation is included as training and testing set at least once. The repeated runs of 10-fold also convey stability and consistency in terms of the standard deviation in accuracy values between the different runs. The results of 10-fold cross-validations with the RF and DNN are shown in Table 3.4. For comparison purposes 10-fold cross-validation results with four more popular classification models, namely, LR, KNN, DT, and convolutional neural network (CNN) have been included in Table 4. For LR, KNN, and DT the same feature set has been used while for CNN the images are used as it is since it does not require hand-engineered featureset. The images have been divided into training: validation: and testing partitions in 60:20:20 ratios using *keras* libraries.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

**Table 3.3:** Parameter settings of DNN and RF

DNN parameters	
Kernel initializer	Random normal
Activation function	
dense layers	ReLU
output layer	Softmax
Optimizer	Adam
Loss function	Categorical cross-entropy
Batch size	64
Number of epochs	500
RF parameters	
$n_{trees}$	2000
max_tree_depth	8
min_samples_leaf	2
Optimization metric	Mse

It can be noted that the inclusion of CNN [3.31] is to consider its emerging importance in computer vision tasks. The CNN model used here is a basic model having 1 conv\_2d layer as input layer comprising 32 kernels, 1 2d max\_pooling layer of filter dimension, 1 flatten layer followed by a dense layer with ReLU activation having output dimension of 128, and a final output layer of dimension 1 having *sigmoid* activation function. The presented scores are a mean of 10 runs of 10-fold cross-validation where each fold has a random arrangement of observations. Table 3.4 shows that RF, CNN, and DNN have performed better than the two

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

other classifiers under consideration. The mean scores for RF, CNN, and DNN are 92.83%, 95.32%, and 97.89%, respectively. The DNN performs best among the classifiers not only in terms of classification accuracy value but also has shown a much higher degree of consistency by the lowest standard deviation value among the classifiers under consideration.

**Table 3.4:** 10-fold cross-validation results (mean of 10 runs)

Classifiers						
Run	LT	KNN	DT	RF	CNN	DNN
1	0.7681	0.7527	0.8131	0.7527	0.9576	0.9795
2	0.7760	0.8461	0.8736	0.9560	0.9705	0.9949
3	0.8756	0.7472	0.9175	0.9285	0.9571	0.9745
4	0.8493	0.8397	0.8397	0.9005	0.8853	0.9566
5	0.9080	0.8618	0.7845	0.9171	0.9802	1
6	0.9004	0.9889	0.9392	0.9889	0.9418	0.9782
7	0.9134	0.9005	0.9226	0.9558	0.9217	0.9789
8	0.9065	0.9171	0.9723	0.9392	0.9657	0.9545
9	0.8197	0.9116	0.8729	0.9447	0.9922	0.976
10	0.9505	0.9558	0.9502	1	0.9594	0.9965
Mean classification accuracy	0.8668	0.8721	0.8886	0.9283	0.9532	0.9789
Highest classification accuracy	0.9505	0.9889	0.9723	1	0.9922	1
Lowest classification accuracy	0.7681	0.7472	0.7845	0.7527	0.8853	0.9545
Standard deviation	±0.0616	±0.0753	±0.0589	±0.0686	±0.0347	±0.0153

Confusion matrix is another important metric for classification performance evaluation. This matrix is constructed with 4 attributes; true positive (TP), true negative (TN), false positive



### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

(FP), and false negative (FN). From the matrix some important evaluation parameters, namely, precision, recall, F1 score, and overall accuracy are derived following Eqs. (3.8) – (3.10). Precision assesses of out of predicted positive classes (*pure* in our case) by the model how many are positive. Recall indicates out of all positive classes how many have been correctly classified. A combination of higher precision, recall, and F-measure values reflects a better model. As the name says overall accuracy indicates the overall performance of the classification model. The confusion matrix on 1200 unknown (not-included in training and validation stage) test observations with different classifiers are shown in Fig. 3.9.

$$Precision = \frac{TP}{TP + FP} \quad (3.8)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.9)$$

$$F - measure = \frac{2 \times recall \times precision}{(recall + precision)} \quad (3.10)$$

	P	A		P	A		P	A		
P	88.5	11.5		P	94.7	5.3		P	96.7	3.3
A	14.6	85.6		A	12.2	87.8		A	1.2	89.8
LR			KNN			DT				
	P	A		P	A		P	A		
P	97.3	2.7		P	96.9	3.1		P	98.1	1.9
A	2.4	97.6		A	3.7	96.3		A	1.5	98.5
RF			CNN			Proposed model				

**Fig. 3.9:** Confusion matrix of different classifiers on unknown testing set.

The metric values derived from confusion matrix have been shown in Table 3.5. It shows that DNN performs comparatively better than other classifiers under consideration with an overall

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

accuracy of about 98.5% over the unknown test dataset (1200 observations) and 99.01% over the entire dataset (6000 observations). RF and CNN also perform well in terms of overall accuracy over unknown test datasets and entire datasets with accuracy in the tune of more than 96% and 97%, respectively. The F-measure values also show that DNN can outperform other classifiers under consideration with values of more than 98% on both the unknown test dataset and the entire dataset. The higher F-measure confirms an improved balance between *precision* and *recall* with DNN. CNN also shows the tendency of overfitting with the current architecture. However, this can be improved with more optimized CNN models. It can also be noted that LR and KNN show 87% and 90% accuracy, respectively, on unknown test datasets which is less than the performance of DNN but confirms the potential of the AMF featureset proposed in this work that is small in length (only 23 features) and thus results in faster decision making by the model. The result of DT reflects the over-fitting problem of the DT algorithm which results in very good scores on the entire dataset since the training set is very well classified but when it is subjected to unknown dataset it shows lack of generalization and results poor accuracy as well as other metrics. The over-fitting problem is visible with LR results also to some extent.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

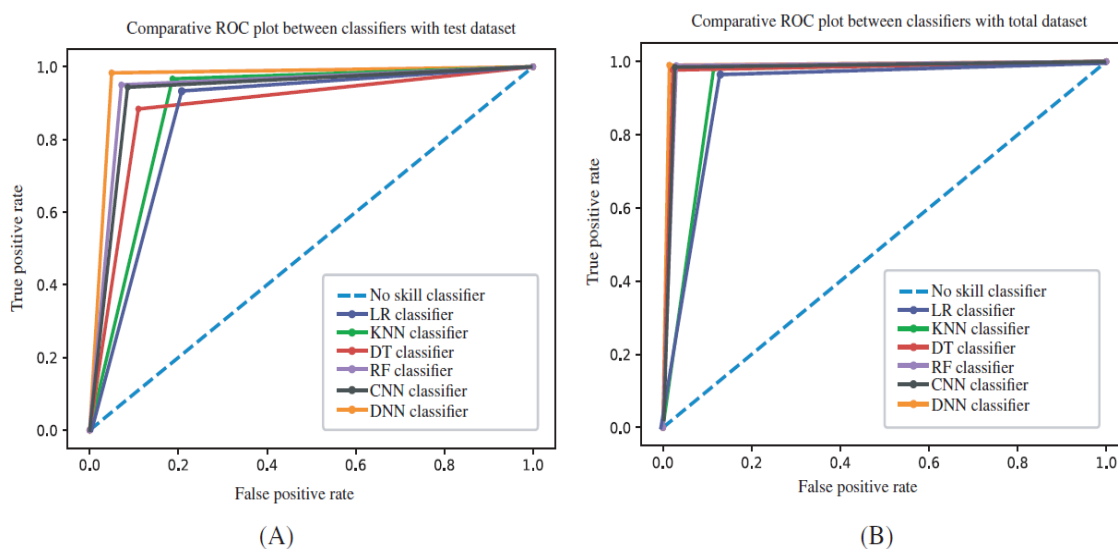
**Table 3.5:** Performance evaluation of different classifiers

Classifiers	Recall (%)		Precision (%)		F-Measure (%)		Accuracy (%)	
	Test set	Total set	Test set	Total set	Test set	Total set	Test set	Total set
LR	82.52	88.29	88.58	89.41	85.54	88.84	87.00	89.73
KNN	85.51	89.40	94.77	95.40	89.00	92.30	90.83	93.05
DT	87.71	97.63	96.71	96.97	91.98	97.29	92.75	97.33
RF	96.91	97.61	97.29	97.96	97.09	97.78	97.50	97.85
CNN	96.91	98.07	95.24	98.58	96.07	98.32	96.58	98.20
DNN	98.07	98.97	98.07	98.86	98.07	98.91	98.33	99.01

Another very popular and efficient metric in classifier assessment is area under curve of receiver operating characteristics (AUC-ROC). These metric plots the true positive rate vs. false positive rate at different threshold values and calculate the area under the curve. Conventionally it is plotted against the naïve classifier that has 0.5 AUC-ROC which indicates model has no class separation ability. Similarly, the higher AUC-ROC values reflect higher capability towards separating the classes. The ROC plots for different classifiers under consideration are shown in Fig. 3.10 for both entire dataset and unknown test dataset. The AUC-ROC values are listed in Table 3.6. The plot shows that DNN results best classification capability with about 98% classification potential while RF and DT also results better than LR and KNN with AUC-ROC values close to 98% classification potential. The same interpretation holds true for ROC plots as well. Plots for both test and entire dataset show that RF and CNN are very close to DNN in terms of performance but CNN performs better in case of entire dataset while in case of test set RF performs better than CNN. This again confirms a lack of generalization with the present implementation of CNN Nevertheless, RF,

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

CNN and DNN provides better potential to handle imbalance dataset i.e. the samples/observations under different classes are not equal. DT performs well when entire dataset is considered but not that well when unknown dataset is subjected due to tendency of over-fitting.



**Fig. 3.10:** AUC-ROC plots of different classifiers.

In Table 3.6 another important metric called *log-loss* values for each classifier have been included. These values indicate the degree of deviation between predicted and actual labels. This metric is also used for minimization of prediction errors. Values closer to 0 indicates lower difference between predicted and actual labels and thus better classifiers. The log-loss values also give consistent result as given by other metrics. DNN has provided least values of 0.6858 for the entire dataset and 1.1513 for unknown test dataset. DT, RF and CNN show almost equivalent performance in total dataset but DT performs considerably poor in case of test data. The lack of generalization in case of CNN is reflected with higher log-loss value for the test set than that of the total dataset. The high log-loss values of LR and KNN also confirm the potential of RF and DNN over those two classifiers since they have limited capability to handle non-linearity present in the dataset.

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

**Table 3.6:** AUC-ROC and log-loss values of different classifiers

Classifiers	AUC-ROC		Log-loss	
	Total dataset	Test dataset	Total dataset	Test dataset
LR	0.9081	0.8946	3.1490	4.1642
KNN	0.9329	0.9160	2.3051	3.1661
DT	0.9771	0.8671	0.7810	4.6052
RF	0.9795	0.9562	0.7048	1.8229
CNN	0.9801	0.9488	0.7136	2.0341
DNN	0.9901	0.9868	0.6858	1.1513

The regression analysis was done with the same train, validation and test partitioning of entire dataset. To evaluate the prediction capability of the regression model four popular metrics have been used, namely, mean absolute error (MAE), root mean squared error (RMSE) and R-squared ( $R^2$ ) score. MAE is calculated as the average of absolute difference between actual and predicted values represented as and across all the observations under consideration as shown in Eq. (3.11). RMSE is calculated as Eq. (3.12) and it is more appropriate compared to MAE since the squaring of error gives higher weight to large differences. Thus smaller values of MAE and RMSE are desired for better regression model.  $R^2$  is of particular interest as it depicts the ability of the predictor where the correlation coefficient  $R$  value between an independent variable  $x$  and a dependent variable  $y$  for  $n$  number of data points is calculated as Eq. (3.13).

$$MAE(y_j, \hat{y}_j) = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (3.11)$$

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

$$RMSE(y_j, \hat{y}_j) = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (3.12)$$

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3.13)$$

The  $R^2$  values closer to 1 indicate better prediction capability. The metric values for three regression models against unknown test data and the entire dataset are shown in Table 3.7. It shows that the RF regression model gives the best among the three regression models under consideration that can provide more than 94% correct prediction capability on unknown test data while on total dataset including the training data shows almost 99% prediction capability in terms of  $R^2$  values. The DT is showing the limitation of over-fitting in the case of the regression model as well. The support vector regression (SVR) is another popular regression model but in our case, it has not shown as good results as the other two regression models under consideration.

**Table 3.7:** MAE, RMSEP and  $R^2$  value comparison between different regression models

Evaluation metrics	Regression models					
	DT		RF		SVR	
	Test set	Total set	Test set	Total set	Test set	Total set
MAE	0.5694	0.1138	0.5770	0.2941	1.2815	1.2974
RMSEP	2.15	0.9643	1.4798	0.8485	2.5357	2.5337
$R^2$	0.8552	0.9715	0.9418	0.9876	0.7998	0.8035

RMSEP - root mean square error of prediction

Table 3.8 represents a comparative illustration of prediction performance in terms of  $R^2$  and RMSEP with so near-infrared sorted spectral techniques involving infrared, Raman

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

spectroscopy and near infrared spectroscopy. It shows that presented method can perform competitively. In terms of  $R^2$  the performance is poorer than PLSR method reported in reference 3.7 but in terms of RMSEP it works equivalently. The method works equivalently to the FT-Raman and FT-IR based techniques reported in reference 3.6 in terms of  $R^2$  but could not be compared in terms of RMSEP since that evaluation has not been reported in reference 3.6. The method reported in reference outperforms all other techniques under consideration in terms of both  $R^2$  and RMSEP. Nevertheless, considering the simplicity, non-invasive nature, low-cost and possibility of portable/mobile handheld device development in future like NIR handheld devices, the presented method can be considered as a promising initial step.

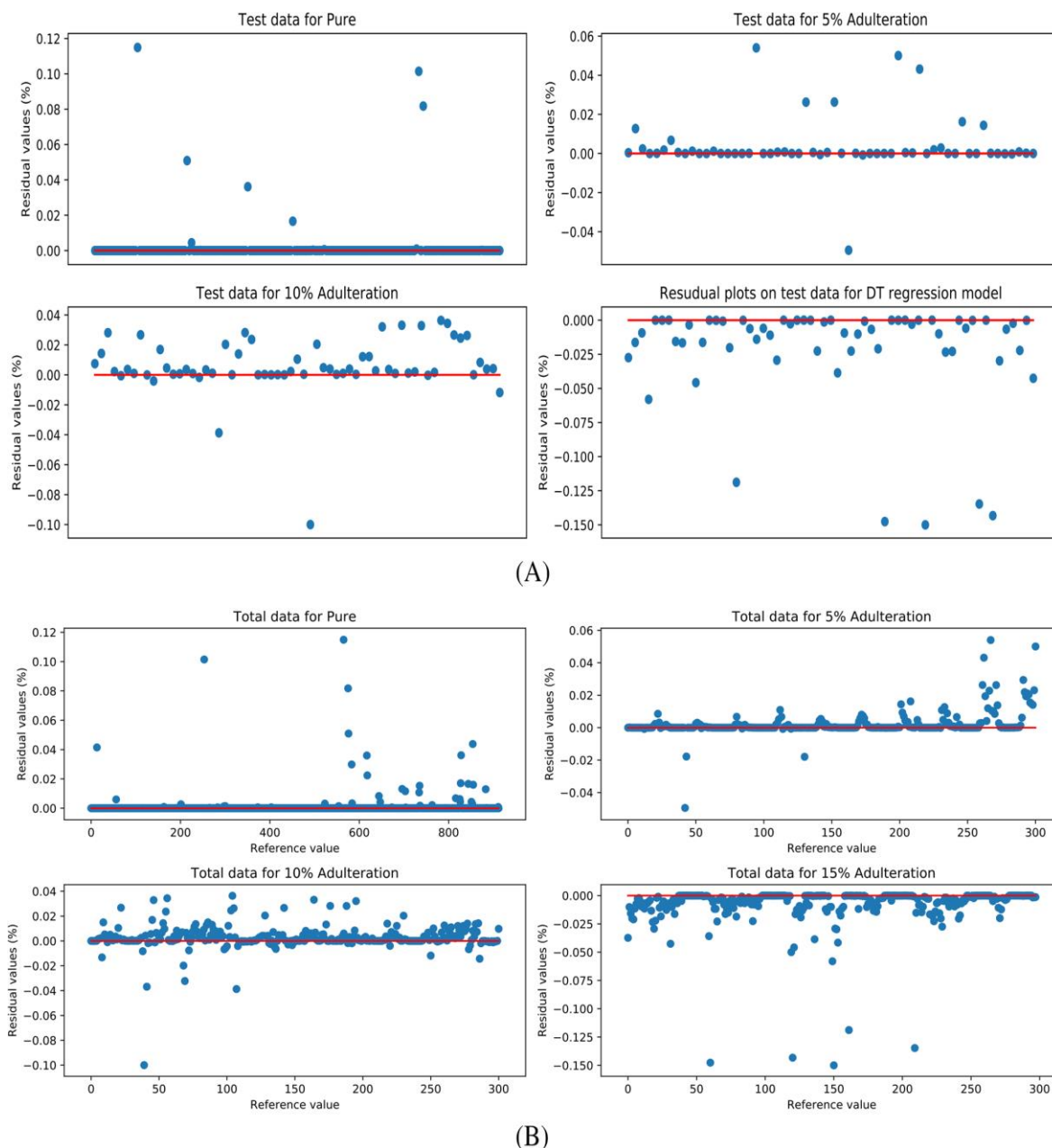
**Table 3.8:** Performance comparison against  $R^2$  and RMSEP

Evaluation parameter	FT-Raman	FT-IR	PCR	PLSR	PLSR	Presented method
$R^2$	0.930	0.950	0.965	0.974	0.980	0.942
RMSEP	-	-	0.890	0.725	1.53	1.47

Finally the assessment of the regression models was performed using residual plot. In this plot, the signed difference between predicted and actual values are plotted. The difference is plotted (generally in percentage scale) against the reference line plotted at 0. This plot can be a good interpreter of performance using adherence between predicted and actual values. It also shows the outliers' resulting from the model. The residual plots for the RF regressor with both the test and the entire dataset are shown previously in Fig. 3.9. The residual plots of Fig. 3.11 show that the RF regression model can make closer predictions to the actual values of

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

adulterations for both the test and total dataset. However, RF results in outliers but the percentage of outliers is considerably less while considering the total sample points under observation. Except for the outliers, the other points are within a very small range from the reference line at 0.



**Fig. 3.11:** Residual plots for different regression model on (a) test data and total dataset.



## Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

### 3.4. Conclusion

The chapter presented a computer vision approach to detect and estimate the quantity of MET adulteration in turmeric powder. Using an indigenous imaging chamber, photographs of pure and adulterated turmeric powder containing 5%, 10%, and 15% w/w contaminated turmeric powder were taken. From the sub-images of the acquired photos, AMF was recovered. DNN and RF models, respectively, have been used to realize the classification and prediction tasks. The findings demonstrate that the suggested approach can classify pure and adulterated turmeric powder with greater than 98% accuracy, and it can predict the degree of MET adulteration with greater than 94%.

### References

- 3.1. Nair, K. P. (2013). The agronomy and economy of turmeric and ginger: the invaluable medicinal spice crops. Newnes.
- 3.2. Osorio-Tobón, J. F., Carvalho, P. I., Barbero, G. F., Nogueira, G. C., Rostagno, M. A., & de Almeida Meireles, M. A. (2016). Fast analysis of curcuminoids from turmeric (*Curcuma longa* L.) by high-performance liquid chromatography using a fused-core column. *Food Chemistry*, 200, 167-174.
- 3.3. Sahu, P. K., Panda, J., Jogendra Kumar, Y. V. V., & Ranjitha, S. K. (2020). A robust RP-HPLC method for determination of turmeric adulteration. *Journal of liquid chromatography & related technologies*, 43(7-8), 247-254.
- 3.4. Dhakal, S., Chao, K., Schmidt, W., Qin, J., Kim, M., & Chan, D. (2016). Evaluation of turmeric powder adulterated with metanil yellow using FT-Raman and FT-IR spectroscopy. *Foods*, 5(2), 36.

### **Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder**

---

- 3.5. Kar, S., Tudu, B., Bag, A. K., & Bandyopadhyay, R. (2018). Application of near-infrared spectroscopy for the detection of metanil yellow in turmeric powder. *Food analytical methods*, 11, 1291-1302.
- 3.6. Wulandari, R., Martono, S., & Rohman, A. (2018). Liquid chromatography and fourier transform infrared spectroscopy for quantitative analysis of individual and total curcuminoid in *Curcuma longa* extract. *Journal of Applied Pharmaceutical Science*, 8(9), 107-113.
- 3.7. Shervington, L., Ingham, O., & Shervington, A. (2016). Purity determination of three curcuminoids found in ten commercially available turmeric dietary supplements using a reverse phase HPLC method. *Nat. Prod. Chem. Res*, 4, 244-249.
- 3.8. Poudel, A., Pandey, J., & Lee, H. K. (2019). Geographical discrimination in curcuminoids content of turmeric assessed by rapid UPLC-DAD validated analytical method. *Molecules*, 24(9), 1805.
- 3.9. Dixit, S., Purshottam, S. K., Khanna, S. K., & Das, M. (2009). Surveillance of the quality of turmeric powders from city markets of India on the basis of curcumin content and the presence of extraneous colours. *Food Additives and Contaminants*, 26(9), 1227-1231.
- 3.10. Rukundo, I. R., & Danao, M. G. C. (2020). Identifying turmeric powder by source and metanil yellow adulteration levels using near-infrared spectra and PCA-SIMCA modeling. *Journal of food protection*, 83(6), 968-974.
- 3.11. Rukundo, I. R., Danao, M. G. C., Weller, C. L., Wehling, R. L., & Eskridge, K. M. (2020). Use of a handheld near infrared spectrometer and partial least squares

### Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder

---

- regression to quantify metanil yellow adulteration in turmeric powder. *Journal of near infrared spectroscopy*, 28(2), 81-92.
- 3.12. Parvathy VA, et al. Detection of plant-based adulterants in turmeric powder using DNA barcoding. *Pharmaceutical Biology*. 2015; 53(12):1774-1779.
- 3.13. Sun, D. W. (Ed.). (2016). Computer vision technology for food quality evaluation. Academic Press.
- 3.14. Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, 34(4), 625-636.
- 3.15. Chen, C. H. (2013). Computer vision in medical imaging (Vol. 2). World scientific.
- 3.16. Jones, G. A., Paragios, N., & Regazzoni, C. S. (Eds.). (2012). Video-based surveillance systems: computer vision and distributed processing. Springer Science & Business Media.
- 3.17. Karakaplan, M., & Avcu, F. M. (2021). Classification of some chemical drugs by genetic algorithm and deep neural network hybrid method. *Concurrency and Computation: Practice and Experience*, 33(13), e6242.
- 3.18. Olson, M., Wyner, A., & Berk, R. (2018). Modern neural networks generalize on small data sets. *Advances in neural information processing systems*, 31.
- 3.19. Feng, S., Zhou, H., & Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Materials & Design*, 162, 300-310.

### **Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder**

---

- 3.20. Barz, B., & Denzler, J. (2020). Deep learning on small datasets without pre-training using cosine loss. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1371-1380).
- 3.21. Elith, J. (2019). 15-Machine learning, random forests, and boosted regression trees. Quantitative analyses in wildlife science, 281.
- 3.22. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
- 3.23. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
- 3.24. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.
- 3.25. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- 3.26. Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.
- 3.27. Nudel, J., Bishara, A. M., de Geus, S. W., Patil, P., Srinivasan, J., Hess, D. T., & Woodson, J. (2021). Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. Surgical endoscopy, 35, 182-191.

### **Chapter 3: Detection of Metanil Yellow Adulteration in turmeric powder**

---

- 3.28. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
- 3.29. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., ... & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238.
- 3.30. Chollet, F. K. (2015). Available online: <https://keras.io> (accessed on 14 August 2019).© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).
- 3.31. O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

# CHAPTER 4

## Detection of Sudan dye-I adulteration in turmeric powder

### 4.1. Introduction

This chapter presents computer vision-based detection of adulteration using Sudan dye-I, another commonly used adulterant [4.1]. Sudan dye-I ( $C_{16}H_{12}N_2O$ ) belongs to the azo dye family, which also comprises several coloring agents like Sudan black, Sudan red, Sudan orange, Sudan toluidine red, and more. These dyes are primarily used in different industrial and household products but are not permitted as food additives [4.2] Sudan dyes can cause toxic effects on human organs as they can produce aniline and its derivative compounds that can even cause cancer threats. Hence they have been classified as category 3 carcinogens by the International Agency for Research on Cancer (IARC). The use of Sudan dye in food and food ingredients is thus prohibited by the food safety regulatory boards of many countries.

The works presented in this chapter has been partially published in author's following publication Mandal, D., Chatterjee, A., & Tudu, B. (2022). A color channel based on multiple Random Forest coupled with a computer vision technique for the detection and prediction of Sudan dye-I adulteration in turmeric powder. *Color Research & Application*, 47(2), 388-400.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

Different analytical methods have been used for the detection of Sudan dye adulteration. Some major works have been briefed here. High-performance liquid chromatography (HPLC), the ODS III column system, and the in-line micro-matrix solid-phase dispersion (MMSPD) method [4.3] have been presented for the detection of variations of Sudan dye adulteration in various spices. Under ideal conditions, the method has demonstrated approximately 99% recovery of analytes. Regarding recovery values, a technique that combines ion mobility spectrometry and headspace gas chromatography has been studied in [4.4]. In this regard, reports of Raman spectroscopy use have also been made, whereby significant performance has been shown by applying linear discriminant analysis (LDA) [4.5] and partial least squares regression (PLSR) techniques for classification and prediction, respectively. With surface-enhanced Raman spectroscopy (SERS) [4.6], Sudan dye I have been detected using principal component analysis (PCA) and wavelet transform [4.7]. Additionally, describes the use of PLSR and Fourier transform infrared (FT-IR) spectroscopy for Sudan Red G adulteration, with noteworthy prediction performance demonstrated by  $R^2$  values greater than 0.97. Application of Raman spectroscopy has also been reported in [4.8] and has shown high accuracy in prediction. 2-directional high-performance thin-layer chromatography (HPTLC) has also been employed for the detection of metanil yellow and Sudan dye adulteration in turmeric [4.9]. Different techniques have also been reported in this connection focusing on the chemical composition of adulterants. *In-situ* FTIR spectroscopy-based study that reveals the site-specific extension of dynamic chemisorption complexity of CO [4.10]. In [4.11] a simple and convenient method for constructing substituted cycloheptenones from 1-bromoocta-1, 7-diene-3-ols has been presented. Synthesis of some hydroxycyclohexene fused carbocycles derivatives has been reported in [4.12]. Ray et al. [4.13] have presented a new Pd-catalyzed intramolecular Heck cyclization crystallization on halogenated diene scaffolds undergoing various modes of cyclization and associated

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

dependency on temperature. Studies on the adsorption of Nitrogen to a polymer of intrinsic microporosity using in situ vibrational spectroscopy have been presented in [4.14]. The investigations and results in [4.15] reveal the potential of N<sub>2</sub> vibrational spectroscopy to identify the presence of porosity in case the sample quality is very small.

This chapter presents a database preparation with pure and adulterated turmeric powders, implementing the classification and regression models for prediction and results along with pertaining discussions on a computer vision model developed for Sudan dye-I detection in turmeric powder. Computer vision has promising solutions in many engineering applications [4.16]. Computer Vision has a remarkable application in the assessment of food quality [4.17]. Among various machine learning algorithms, Random Forest (RF) [4.18] has been used here for its inherent advantages of better potential to avoid over or under fitting problems.

### 4.2. Materials and methods

#### 4.2.1. *Experimental in-house database preparation*

Keeping synergy with the work reported in the previous chapter in this case also 5 variants of *pure* turmeric powders were taken for experimentation. Among those 5 variants, 4 were branded organic turmeric powders that claim 100% purity and were procured from the local supermarket and 1 pure variant was prepared in-house following the process involving boiling the raw turmeric roots followed by oven drying and grinding as described in [4.19]. The Sudan dye of the Sigma Aldrich brand was procured from a lab chemical supplier. Each of the 5 variants of *pure* turmeric was divided into 4 equal portions that resulted total of 20 portions. Among the 20 portions, 1 portion of each variant was kept unadulterated while the rest 3 portions were adulterated with 3 different percentages of concentration (w/w) that are

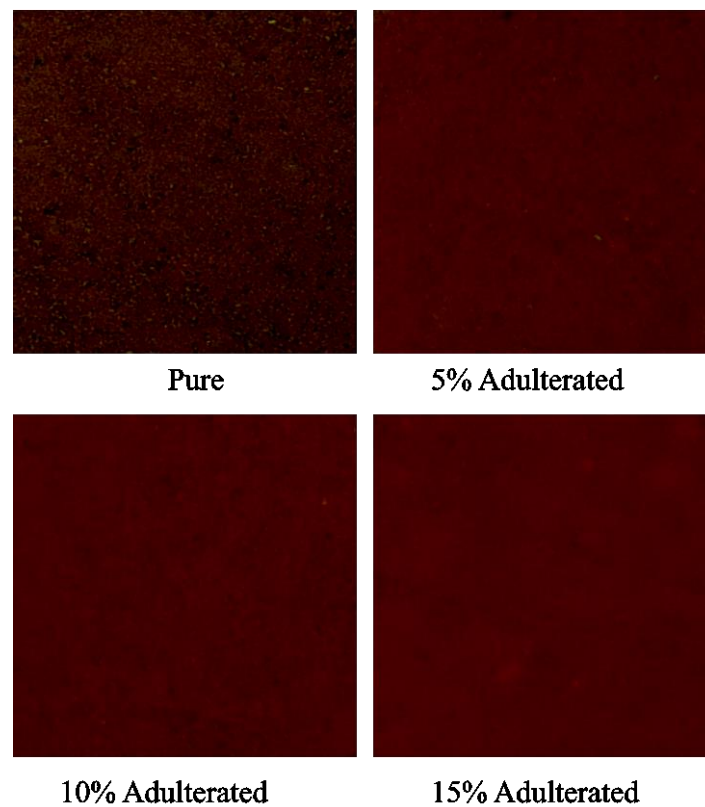


## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

5%, 10%, and 15%. Such mixing resulted in 5 *pure* and 15 *adulterated* samples comprising of 5 adulterated samples for each concentration level. All the samples were preserved in air-tight containers to avoid humidification until imaging.

Each of the prepared samples was divided into 5 equal portions and subjected to imaging. The individually captured images were further cropped into 10 sub-images of size  $300 \times 300$  pixels. The centers of these sub-images were randomly selected within the image boundary. That resulted in total  $20 \times 5 \times 10 = 1000$  (no. of samples  $\times$  no. of portions in each sample for imaging  $\times$  no. of sub-images) images. The features were extracted from these 1000 images. Some of the sample images are shown in Fig. 4.1.



**Fig. 4.1:** Samples of images from the image dataset

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

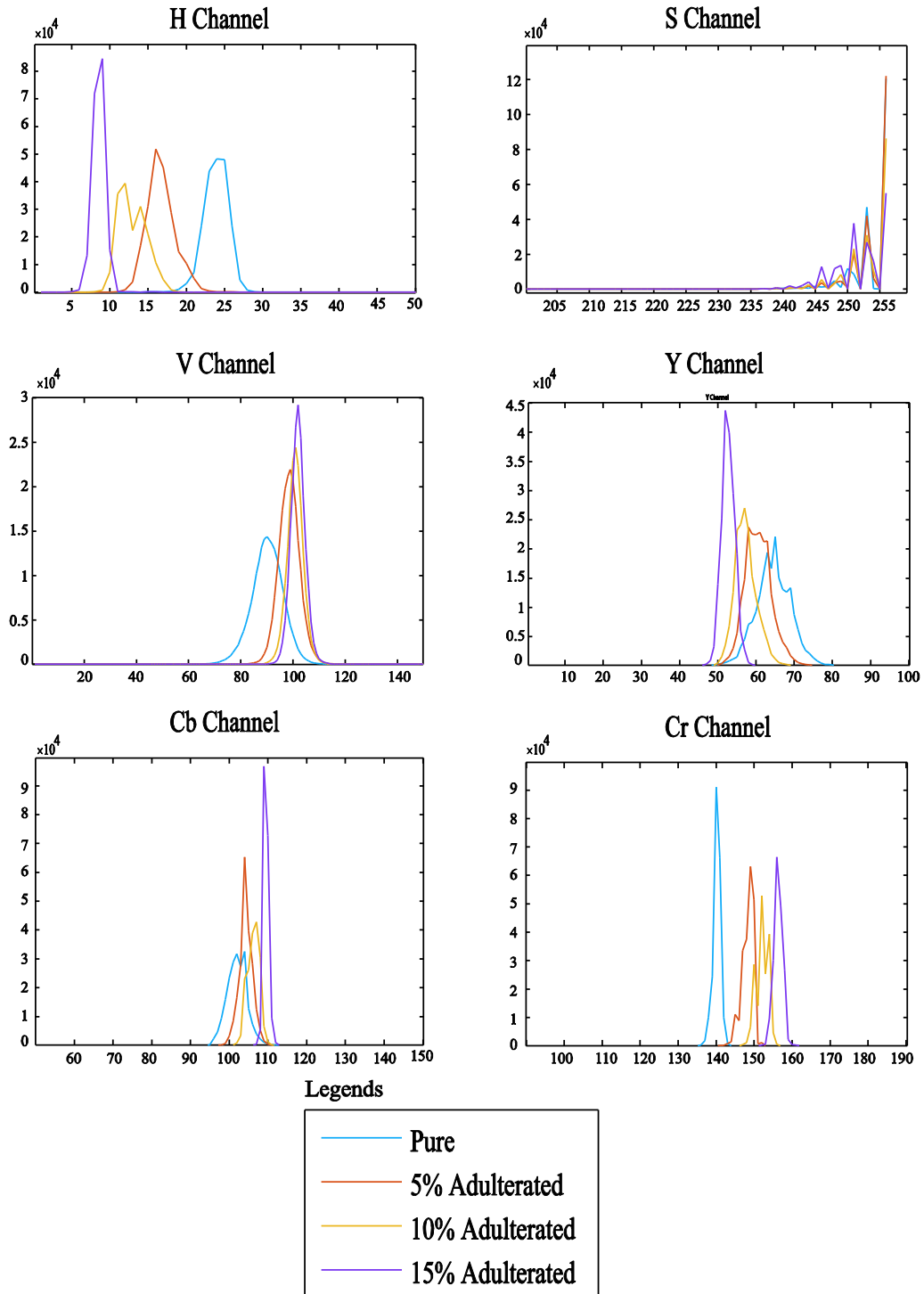
---

### 4.2.2. Color feature extraction

One of the most crucial features for adulterant detection can be color features. However, the native RGB color space cannot convey much about different aspects of color. Hence, it is a common practice in computer vision research to convert the captured images into device-independent color spaces. Hue, saturation, and value (HSV) and YCbCr are two popular color spaces [4.20] in this context that provide deeper insight to color interpretations.

The captured images were converted to HSV and Y<sub>Cb</sub>Y<sub>Cr</sub> color spaces from the RGB color space using standard conversion equations [4.21]. Although there are three different color channels in each of the stated color spaces all color channels may not convey noticeable differences. This motivated a channel-wise distinguishability analysis. To examine the separability present between *adulterated* and *pure* samples the histograms of the captured images were taken as features. Histogram has advantages like rotation invariance, simplicity, and fixed length (256 in the case of 8-bit images). The histograms for different channels are shown in Fig. 4.2.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder



**Fig. 4.2:** Color channel based distinguishability analysis for pure and adulterated samples

Figure 4.2 clearly shows that all channels do not show distinguishability between *pure* and *adulterated* samples equivalently. H channel shows the difference very clearly as the yellowness of the sample images changes which directly causes the change of value distribution in the H channel. Similarly, Cr channel also shows a very clear difference since

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

this channel represents the presence of a red component. The color has a yellow appearance and the mixing of Sudan dye contributes towards higher redness which eventually results shifting of the histogram towards the higher side of the x-axis. It can be also seen that the shift is linear as well i.e. both in Cr and H channel histogram *pure*, 5%, 10%, and 15% *adulterated* samples' peaks shift linearly towards the higher value of the x-axis. This phenomenon reduces the complexity of the feature as well.

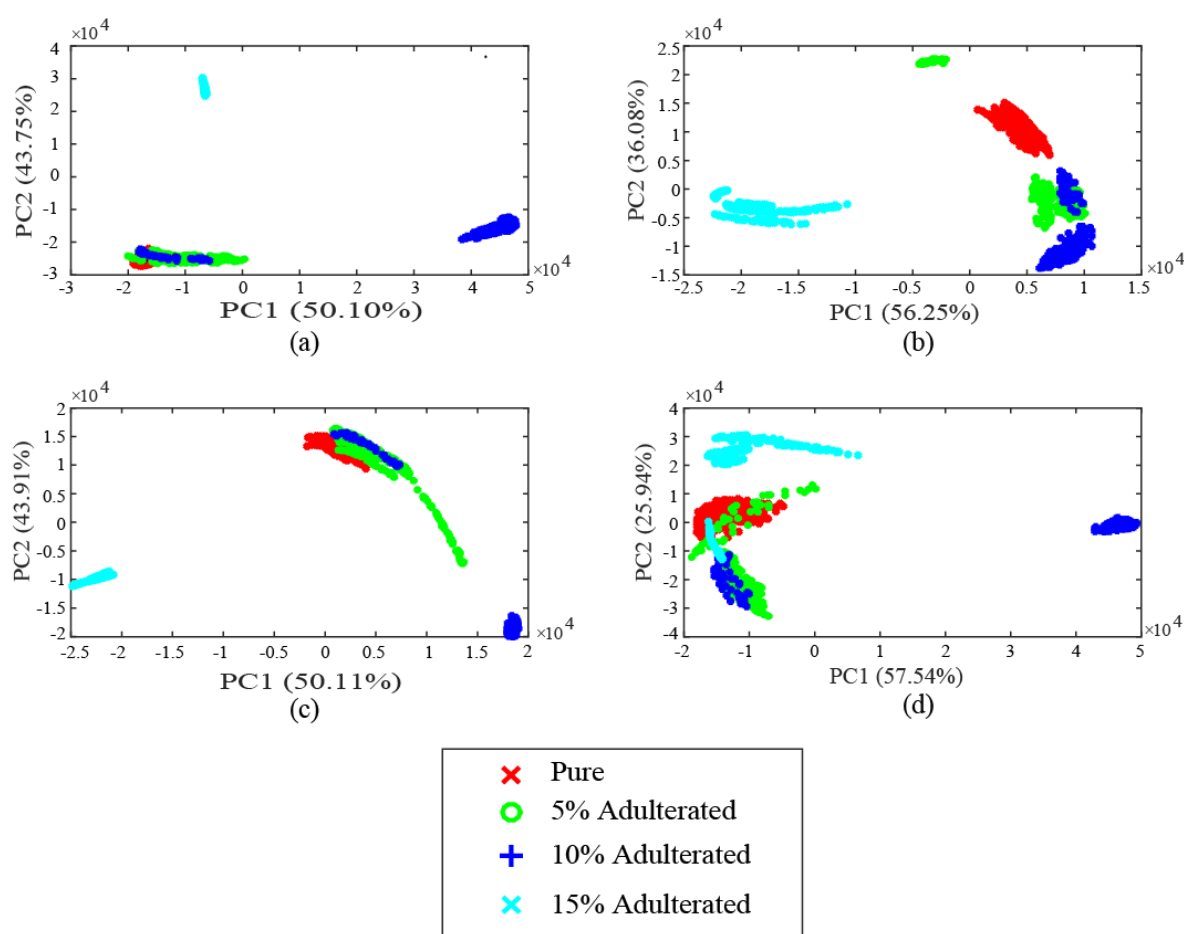
V and Y channels also show distinguishability but not to the extent of H and Cr channels. The reason for this is Sudan dye I is not a bright material, hence the luminance of the images does not change significantly. The S channel does not show much difference since all the individual hues or shades have a similar degree of saturation. Similarly, the Cb channel represents the blueness of the color which is also not a very differentiating factor as the yellow color is a combination of different degrees of red and green.

It can also be observed that significant information is not present across the entire stretch of the histogram rather it is concentrated in a particular region. For instance, the H channel information is between 0 – 50 since the colors of the turmeric samples are inclined toward yellow-red regions rather than the higher side representing the blue hues. Similarly, for Cr, the information is concentrated in the range of 140 – 180. This way in the final featureset the fragmented histogram of H, Cb, V, and Y channels was included to reduce the dimension of the featureset. The dimension of the final featureset was  $1000 \times 190$  (no. of observation  $\times$  no. of features).

The feature suitability was examined using principal component analysis (PCA) [4.22]. It can be noted that traditionally PCA is used for dimension reduction, but in this case, it has been used only to visualize the distinguishability that can be achieved using the presented feature. In this section first analysis of feature suitability i.e. whether features are

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

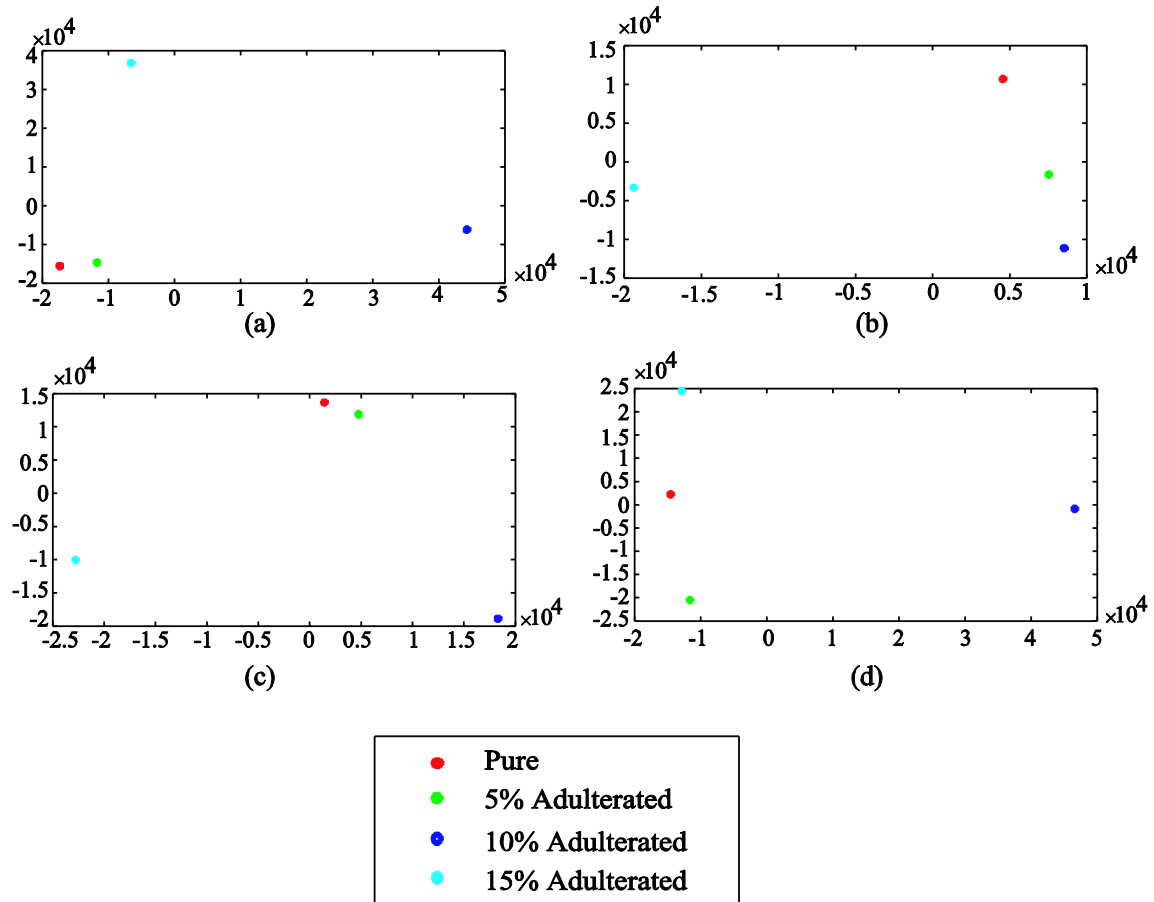
distinguishable or not is presented. The centroids of the clusters formed by PCA were obtained using the K-means clustering technique [4.23]. Generally more the centers are away from each other better the possibility of classification with the features. This is true for PCA as well. A featureset that provide higher separability can show better distinguishability in terms of clustering of points under different groups. The clusters of different classes and corresponding centroids have been shown in Fig. 4.3 and Fig. 4.4, respectively.



**Fig. 4.3:** PCA clustering of different classes for different color channels a) H, b) V, c) Y and d) Cr.

The legends are shown in the bottom.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder



**Fig. 4.4:** The centroids of different classes using K-means clustering for color channels a) H, b) V, c) Y, and d) Cr. The legends are shown at the bottom.

Figures 4.4 and 4.5 show that in terms of separability between clusters and the centroids of various classes, the V and Cr color channels are most noticeable. It is also aligned with the observations from the histograms shown in Fig.4.2. The H and Y channels exhibit commendable separability about centroid locations; however, the degree of overlap between the Pure and 5% adulteration clusters is greater than that of the V and Cr channels. Overall, the figures demonstrate that the centroids migrate away as the amount of adulteration rises. The centroids of the pure and 5% adulteration observations, for example, are closer together than the ones between the pure and the 10% and 15% adulterations. This distinguishability attests to the color characteristics' potential for classification and prediction tasks.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

### 4.2.3. Data analysis model

In the previous chapter, the Random Forest (RF) model was introduced and found to be competitive with DNN with the advantage of a much lesser number of tunable parameters. This motivated to use RF alone as a classifier as well as a regressor for this work. However, in this work, the ensemble characteristic of RF was extended to another level.

The RF algorithm has been used to classify and predict each of the H, V, Y, and Cr channels. Then the results of each RF were ensemble using a majority voting strategy for the final output of the model. In cases where both classes have got equal votes (that is both *pure* and *adulterated* have got 2 votes) the decision of the Cr channel has been used for the final decision. For prediction also majority voting strategy has been used i.e. highest number of votes for one of the 0%, 5%, 10% and 15% values has been declared as final prediction. This ensemble strategy can also help to handle outliers resulted in individual RF predictions as shown later in the result section. The entire method presented in above section has been shown in Fig. 4.5.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

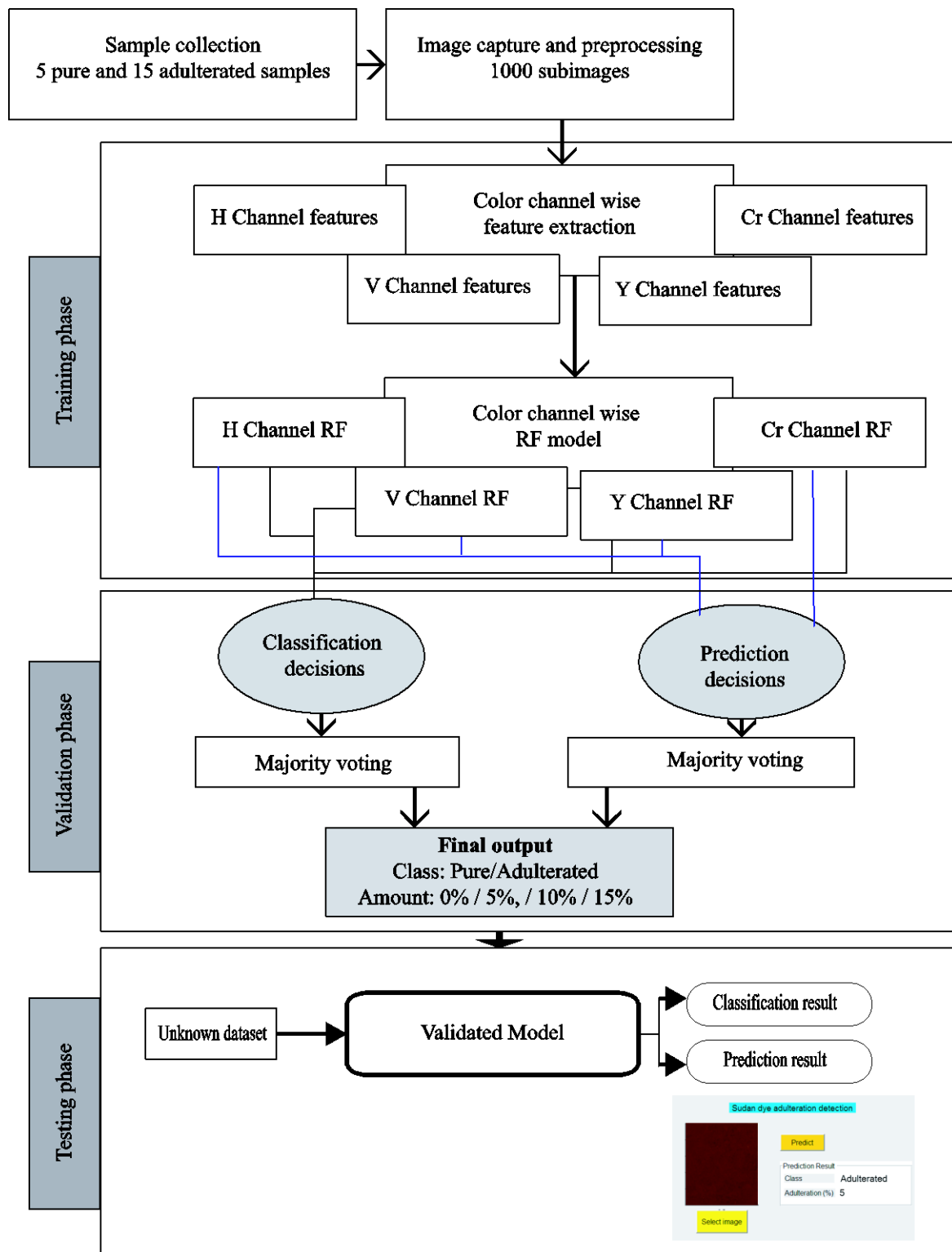


Fig. 4.5: Ensemble mechanism with multiple RF



## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

RF is a supervised machine learning algorithm that is popular for its simplicity due to the lesser number of parameters to be tuned. The conceptual forest is made up of several decision trees (DTs) where each DT performs classification probabilistically. It uses a bagging/bootstrapping concept [4.24] where each tree is subjected to different bootstrapping samples in training. Before finalizing the model the trees are cross-validated with different bootstrapping to which the DT was not subjected during training. The optimality of the model can be assessed by out-of-bag (OOB) error [4.25]. Lower OOB error reflects the minimum number of trees required for optimal model performance. It is also to be noted that RF is an ensemble learning algorithm so each DT in the forest contributes towards the final decision. The model decision is made based on the maximum voting concept for classification and probability averaging for prediction tasks. The RF algorithm [4.26] pseudocode can be provided in Table 4.1.

**Table 4.1:** Pseudocode of RF

---

<i>Initialization</i>
Number of DT ( $N_{trees}$ ), optimization function, maximum number of features per tree ( $m_{try}$ )
<i>Training of individual tree</i>
Bagging the samples ( $P$ ) for each tree by random selection of $M$ observations and $N$ features from the entire training dataset
$P \in X^{M \times N}$ where $X$ is the entire dataset
Binary splitting the observations subjected to the DT using the cutoff point [24] derived using the optimization function.
Higher values than cutoff point to the right node/leaf and lower values to the left node/leaf.

---

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

Continuing till the terminal node is reached.

*Classification and prediction*

Classification using mode value:  $Y_{classification} = \text{mod}(y_1, y_2, \dots, y_{N_{trees}}) Y_t$

where  $Y_t$  is the terminal node of DT

Prediction using averaging:  $Y_{prediction} = \frac{1}{N_{trees}} \sum_{i=1}^{N_{trees}} Y_i$  where  $Y$  is the prediction

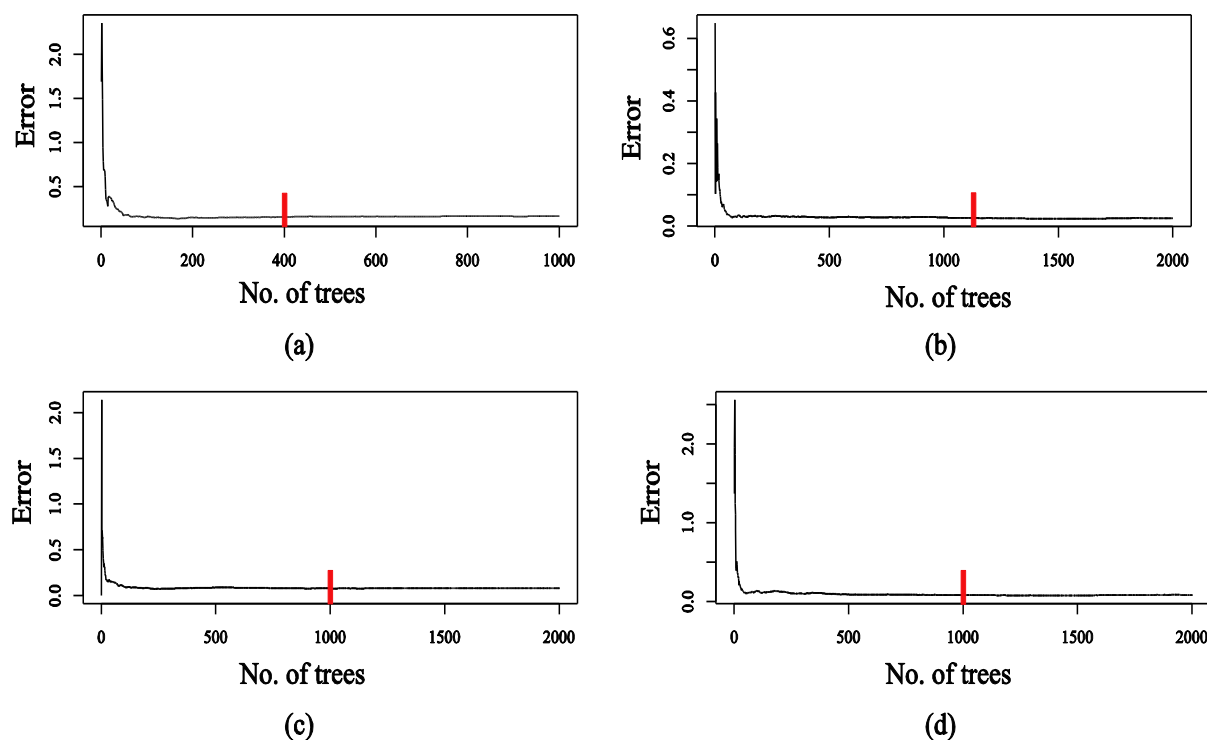
of individual DT.

As different RF models have been used for different color channels the parameters of individual models have been tuned separately. Table 4.2 shows the parameter setting for the color channels of different models. These settings were arrived at by model evaluation using the mean squared error of OOB ( $MSE_{OOB}$ ) as shown in Fig. 4.6.

**Table 4.2:** Parameter settings for different RF models

Parameter	RF_H	RF_V	RF_Y	RF_Cr
$N_{trees}$	400	1200	1000	1000
Optimization criterion	Entropy	Entropy	Entropy	Entropy
Total feature-length	19	49	46	12
Mtry	4	4	8	12
Maximum leaf node	2	2	2	4

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder



**Fig. 4.6:** The OOB error plots for the RF model of color channels a) H, b) V, c) Y, and d) Cr. The red line shows the chosen number of trees chosen for the model.

### 4.3. Results and discussions

A dataset of 1000 observations for each color channel, comprising 250 observations for each of the pure and three classes of adulteration, was used for the experiments. Without any overlap, the dataset was split into train, validation, and test sets in a 60:20:20 ratio. Such non-overlapping sets can result in the testing of the model with unknown samples that were not part of the training set. This in turn justifies the assessment of the generalization potential of the model.

The classification models were first validated using a 10-fold cross-validation method. This method is important as it divides the entire test and validation dataset into 10 folds and subjected the folds to the model in an iterative manner. This ensures each data is subjected to

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

training and testing for at least once. The result of 10-fold cross-validation shows the accuracy attained for different folds. It also conveys the consistency of the model. Table 4.3 shows the 10-fold cross-validation results for all 4 channel-wise RF models. The Table also includes the mean and standard deviations for each model. The standard deviation conveys consistency and a lower standard deviation conveys higher consistency. As can be seen, all the RF models can provide more than 90% accuracy. The standard deviations are also considerably less across all the channels. Among the 4 RFs the RF\_V and RF\_Cr provide higher average accuracy and lower standard deviations. Also for these two RF, the accuracy reached 100% for some folds.

**Table 4.3:** Results of 10-fold cross-validation for RF models

Folds	Accuracy (%)			
	RF_H	RF_V	RF_Y	RF_Cr
1	95.57	95.76	81.47	98.60
2	90.35	97.05	90.57	93.18
3	94.91	95.71	82.69	99.69
4	93.39	100	91.33	100
5	97.87	99.02	83.23	100
6	95.77	94.18	89.75	92.34
7	74.31	100	97.84	99.48
8	89.22	98.57	94.68	94.70
9	85.54	99.22	95.75	95.02
10	97.11	95.94	96.48	100
Average	91.40	97.55	90.38	97.30
Standard Deviation	± 7.15	± 2.08	± 6.07	± 3.12
Highest value	97.87	100	97.84	100
Minimum value	74.31	94.18	81.47	92.34

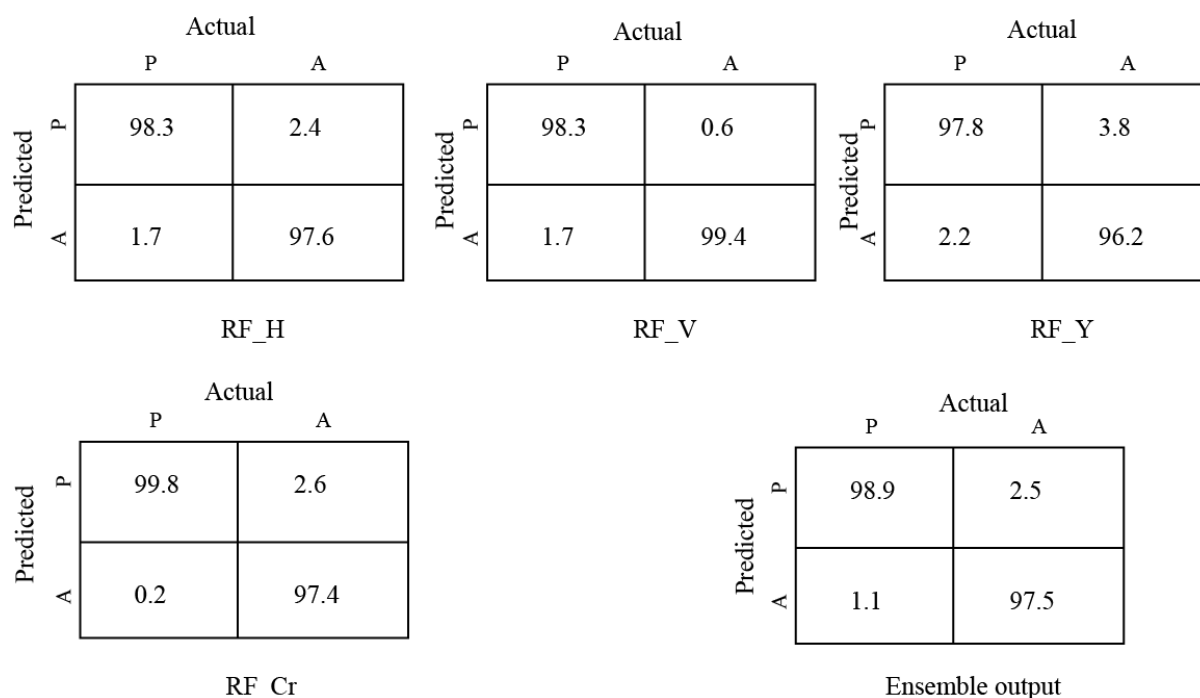
## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

The classification results were further analyzed using different assessment metrics obtained from the confusion matrix as shown in Table 4.4. It clearly shows that all the channels have the potential to classify the *pure* and *adulterated* turmeric samples with more than 96% accuracy. However, V and Cr channel classifiers have improved potential. The recall, precision, and F1 scores also show the consistency and generalization potential of the presented model. The recall rate of RF\_Cr is highest but precision is not as good as that of RF\_V, the reason is Cr channel can classify the adulterated samples well but results in the misclassification of *pure* samples which causes higher false positive cases. Python was used to implement the algorithms using Scikit-learn [4.27] and Keras [4.28] libraries. The ROC\_AUC value shows that the classifiers can provide maximum accuracy that is more than 98% for RF\_V and a minimum of 95% for RF\_Y. The Jaccard score/similarity measure [4.29] also confirms high i.e. more than 97% similarity between the predicted and actual classes. The parameters calculated on the final output of the model with the test set are shown in Table 4. It conveys the advantages that the overall results address the shortcomings of individual channel RF models. For instance, RF\_V has good overall performance but a lower recall rate while RF\_Cr has a high recall value but lower precision. The ensemble of different color channel RFs provides a balance between these variations as can be seen from the last row of the table. It conveys better generalization and a better balance between correct true positive and true negative predictions as well.

**Table 4.4:** Classification results using metrics of confusion matrix

<b>Classification model</b>	<b>Accuracy (%)</b>	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1 score</b>	<b>ROC_AUC (%)</b>	<b>Jaccard score</b>
RF_H	97.91	98.83	98.26	98.55	97.64	97.14
RF_V	98.33	99.41	98.26	98.83	98.38	97.70
RF_Y	96.25	97.67	97.10	97.40	95.56	94.91
RF_Cr	98.33	97.74	1.00	98.85	97.01	97.74
Final output	98.75	97.05	98.50	97.77	98.67	97.71

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder



**Fig. 4.7:** Confusion matrix for different RF models on the test set

A vital metric for evaluating classification performance is the confusion matrix, which comprises four attributes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The confusion matrix for test observations utilizing various Random Forest (RF) models is depicted in Figure 4.7.

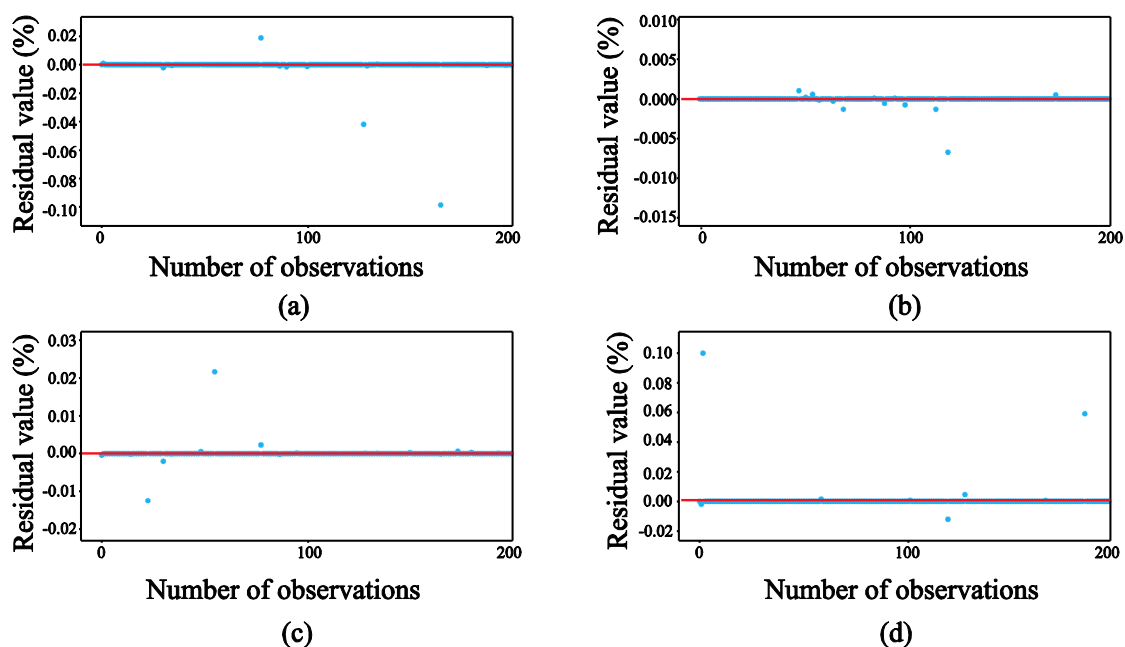
The prediction potentials of the RF models were evaluated using three important metrics, namely, root mean squared error (RMSE), coefficient of determination ( $R^2$ ), and mean absolute error (MAE) [4.30]. The standard expressions of these parameters are included in the Table 4.5. It shows that RF\_Cr provides the lowest RMSE and MAE while the highest  $R^2$  values, RF\_V perform equivalently and the rest of the color channel RFs also provide good performance. The higher  $R^2$  value of the final output-based assessment confirms that the model can explain more than 99% of the variance present in the dataset which in turn provides higher accuracy in prediction. The prediction accuracies are well observed in the residual plots shown in Fig. 4.8 as well. The residual plots show how the predicted values are

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

away from the actual values. A red line has been drawn at zero which is also called the reference line. As it can be seen most of the residual values between the actual and predicted amount of adulteration fall on this line. The outliers are also not high in numbers, for example in RF\_Cr prediction only 2 out of 200 observations (1%) have caused outliers. Therefore the residual plots also confirm the potential of the presented model for the prediction task.

**Table 4.5:** Prediction performance by different RF models

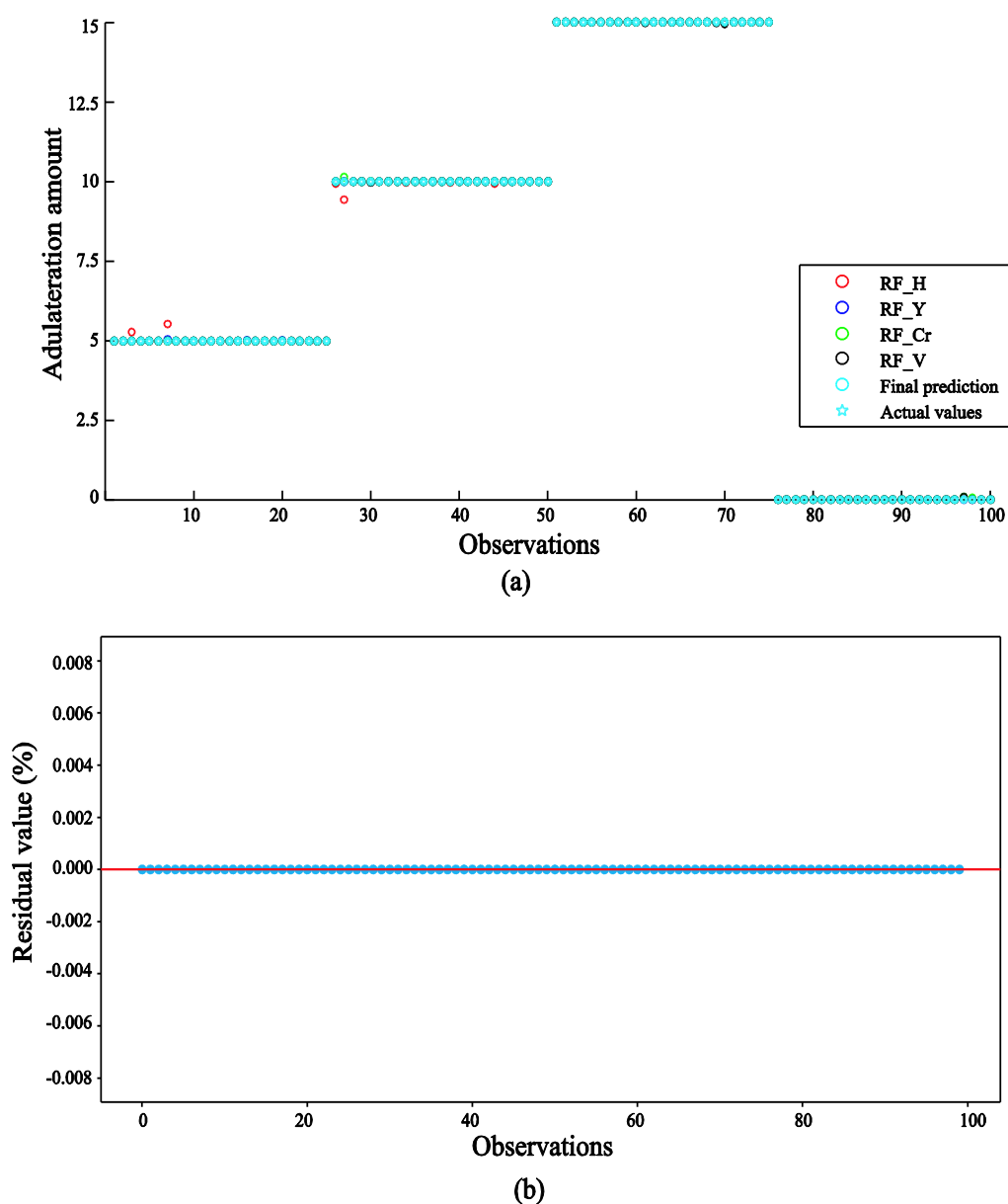
Prediction model	RMSE	$R^2$	MAE
RF_H	0.4932	0.9960	0.0705
RF_V	0.0270	0.9997	0.0170
RF_Y	0.5683	0.9954	0.0751
RF_Cr	0.0021	0.9998	0.0056



**Fig. 4.8:** Residual plots for RF prediction models for a) H, b) V, c) Y and d) Cr color channels. The red line at 0 is for reference.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

Figure 4.9 shows the advantage of ensemble multiple RFs. For this 100 random samples were chosen and subjected to different RFs and final predictions were made using the majority voting strategy. As can be seen in Fig. 4.9(a) some of the samples are not accurately predicted for 5% and 10% adulterations (the red points not lying on the blue line). But in Fig. 4.9(b) the final predictions show no such deviations from the reference line (shown in red color). This observation conveys the improved potential of ensemble RF for avoiding outliers and higher prediction accuracy.



**Fig. 4.9:** Residual plot for a) different RF models and b) final prediction output.



## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

Table 4.6 provides a comparison between different classifiers and predictors performance with the featureset. In this presentation, 3 popular classifiers, namely logistic regression (LogR), support vector machine (SVM), and a 3-layer deep neural network (DNN) model implemented using standard Python libraries (sequential model of *keras* library) have been included. Similarly, for prediction linear regressor (LinR) and support vector regressor (SVR) models have been used for comparison. It can be noted that in this table the final performance of the model has been included instead of channel wise performance.

**Table 4.6:** Performance comparison between different classifiers and prediction models

Metrics	Classification				Prediction		
	LogR	SVM	DNN	Presented model	LinR	SVR	Presented model
Overall accuracy	95.83	96.25	97.50	98.75	-	-	-
Precision	89.04	90.27	94.11	97.05	-	-	-
Recall	97.01	97.01	95.52	98.50	-	-	-
RMSE	-	-	-	-	1.8180	1.5759	0.0053
R <sup>2</sup>	-	-	-	-	0.9833	0.9857	0.9998
MAE	-	-	-	-	1.0114	0.4966	0.0032

Table 4.6 shows that for both classification and prediction, the presented method provides better performance than the other standard techniques under consideration. It can also be noted that the presented model has a balance between precision and recall values like DNN. It conveys the potential of ensemble strategy towards handling unbalanced datasets like ours which comprised 250 pure and 750 adulterated observations. At the same time, the classification accuracy to the tune of 95% by the LogR confirms the potential of the feature in

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

terms of distinguishability between *pure* and *adulterated* samples. The same interpretation holds for prediction as well. It can be seen that the presented method can provide high accuracy in terms of  $R^2$  value while lower MAE and RMSE confirm the consistency and repeatability of the presented model.

Finally, a performance comparison in terms of  $R^2$  between different reported techniques and the proposed method has been presented in Table 4.7. In this case, also the potential of the presented method is prominent. Despite simplicity and inexpensiveness, the presented method is competitive with the other techniques under consideration.

**Table 4.7:** Prediction performance comparison in terms of coefficient of determination ( $R^2$ )

	<b>MMPSD</b>	<b>Gas</b>	<b>FT Raman</b>	<b>FT_IR</b>	<b>Raman</b>	<b>Presented</b>
		<b>Chromatography</b>			<b>and IR</b>	<b>method</b>
<b><math>R^2</math></b>	0.9370	0.9900	0.9130– 0.9900	0.9700	0.9700	0.9990

### 4.4. Conclusion

This chapter has presented a computer vision framework powered by random forest predictor and regressor for detecting Sudan dye-I adulteration in powdered turmeric. Histogram-derived color channel-based features have been applied to prediction and classification. Tested against an internal experimental database, the results have demonstrated a promising potential of nearly 99% accuracy. Therefore, the method's simplicity, non-invasiveness, and low cost of operation make it a potential addition to the current analytical and instrumental screening approaches.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

### References

- 4.1. Nisa, A., Zahra, N., & Butt, Y. N. (2016). Sudan dyes and their potential health effects. *Pak. J. Biochem. Mol. Biol*, 49(1), 29-35.
- 4.2. Tripathi, M., Khanna, S. K., & Das, M. (2007). Surveillance on use of synthetic colours in eatables vis a vis Prevention of Food Adulteration Act of India. *Food Control*, 18(3), 211-219.
- 4.3. Rajabi, M., Sabzalian, S., Barfi, B., Arghavani-Beydokhti, S., & Asghari, A. (2015). In-line micro-matrix solid-phase dispersion extraction for simultaneous separation and extraction of Sudan dyes in different spices. *Journal of Chromatography A*, 1425, 42-50.
- 4.4. Maestroni, B., Besil, N., Rezende, S., Liang, Y., Gerez, N., Karunarathna, N., ... & Cesio, M. (2021). Method optimization and validation for multi-class residue analysis in turmeric. *Food control*, 121, 107579.
- 4.5. Erasmus, S. W., van Hasselt, L., Ebbinge, L. M., & van Ruth, S. M. (2021). Real or fake yellow in the vibrant colour craze: Rapid detection of lead chromate in turmeric. *Food Control*, 121, 107714.
- 4.6. Di Anibal, C. V., Marsal, L. F., Callao, M. P., & Ruisánchez, I. (2012). Surface Enhanced Raman Spectroscopy (SERS) and multivariate analysis as a screening tool for detecting Sudan I dye in culinary spices. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 87, 135-141.
- 4.7. Dhakal, S., Schmidt, W. F., Kim, M., Tang, X., Peng, Y., & Chao, K. (2019). Detection of additives and chemical contaminants in turmeric powder using FT-IR spectroscopy. *Foods*, 8(5), 143.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

- 4.8. Chao, K., Dhakal, S., Schmidt, W. F., Qin, J., Kim, M., Peng, Y., & Huang, Q. (2020). Raman and IR spectroscopic modality for authentication of turmeric powder. *Food chemistry*, 320, 126567.
- 4.9. Dixit, S., Khanna, S. K., & Das, M. (2008). A Simple 2-Directional High-Performance Thin-Layer Chromatographic Method for the Simultaneous Determination of Curcumin, Metanil Yellow, and Sudan Dyes in Turmeric, Chili, and Curry Powders. *Journal of AOAC International*, 91(6), 1387-1396.
- 4.10. Wang, C. Y., Ray, P., Gong, Q., Zhao, Y., Li, J., & Lueking, A. D. (2015). Influence of gas packing and orientation on FTIR activity for CO chemisorption to the Cu paddlewheel. *Physical Chemistry Chemical Physics*, 17(40), 26766-26776.
- 4.11. Ray, J. K., Paul, S., Ray, P., Singha, R., Rao, D. Y., Nandi, S., & Anoop, A. (2017). Pd-catalyzed intramolecular sequential Heck cyclization and oxidation reactions: a facile pathway for the synthesis of substituted cycloheptenone evaluated using computational studies. *New Journal of Chemistry*, 41(1), 278-284.
- 4.12. Chaudhuri, S., Maity, S., Roy, M., Ray, P., & Ray, J. K. (2016). A vinyl radical cyclization route to hydroxycyclohexene fused carbocycles. *Asian Journal of Chemistry*, 28(1), 233-234.
- 4.13. Ray, J. K., Singha, R., Ray, D., Ray, P., Rao, D. Y., & Anoop, A. (2019). Palladium-catalyzed expedient Heck annulations in 1-bromo-1, 5-dien-3-ols: Exceptional formation of fused bicycles. *Tetrahedron Letters*, 60(13), 931-935.
- 4.14. Ray, P., Gidley, D., Badding, J. V., & Lueking, A. D. (2019). UV and chemical modifications of polymer of Intrinsic Microporosity 1 to develop vibrational spectroscopic probes of surface chemistry and porosity. *Microporous and Mesoporous Materials*, 277, 29-35.

## Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder

---

- 4.15. Ray, P., Xu, E., Crespi, V. H., Badding, J. V., & Lueking, A. D. (2018). In situ vibrational spectroscopy of adsorbed nitrogen in porous carbon materials. *Physical Chemistry Chemical Physics*, 20(22), 15411-15418.
- 4.16. Ray, P., Xu, E., Crespi, V. H., Badding, J. V., & Lueking, A. D. (2018). In situ vibrational spectroscopy of adsorbed nitrogen in porous carbon materials. *Physical Chemistry Chemical Physics*, 20(22), 15411-15418.
- 4.17. Sun, D. W. (Ed.). (2016). Computer vision technology for food quality evaluation. Academic Press.
- 4.18. Breiman, L. (2001). Random forests *Mach Learn* 45 (1): 5–32.
- 4.19. Kar, S., Tudu, B., Jana, A., & Bandyopadhyay, R. (2019). FT-NIR spectroscopy coupled with multivariate analysis for detection of starch adulteration in turmeric powder. *Food Additives & Contaminants: Part A*, 36(6), 863-875.
- 4.20. Gonzalez, R. C. (2009). Digital image processing. Pearson education india.
- 4.21. Thanki, R. M., & Kothari, A. M. (2019). Digital image processing using SCILAB. Springer International Publishing.
- 4.22. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202
- 4.23. Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90-95
- 4.24. Breiman, L. (1996). Bagging predictors. *Machine learning*, 24, 123-140....Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020).

## **Chapter 4: Detection of Sudan dye-I adulteration in turmeric powder**

---

- 4.25. Payen, F. T., Sykes, A., Aitkenhead, M., Alexander, P., Moran, D., & MacLeod, M. (2021). Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression. *Cleaner Environmental Systems*, 2, 100024
- 4.26. Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102051.
- 4.27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- 4.28. Bisong, E., & Bisong, E. (2019). Tensorflow 2.0 and keras. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 347-399.
- 4.29. Hancock, J. M., & Bishop, M. J. (2004). HMMer. *Dictionary of bioinformatics and computational biology*.
- 4.30. Xie, X., Wu, T., Zhu, M., Jiang, G., Xu, Y., Wang, X., & Pu, L. (2021). Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecological Indicators*, 120, 106925.

# CHAPTER 5

## Detection of starch adulteration in turmeric powder

### 5.1. Introduction

This chapter explores the scope of applying computer vision for starch adulteration detection in turmeric powder. Unlike previously explained MET and Sudan Dye adulteration [5.1] starch is used for volume-wise adulteration. Apart from starch, flour, wheat, and corn are also used as volumetric adulterants [5.2]. The addition of such adulterants is although less harmful as these do not include hazardous colors but they cause substantial loss in food and medicinal values of turmeric [5.3].

This chapter presents the application of CNN for turmeric adulteration detection. The use of CNN in flour adulteration detection has been reported by (Lanjewar et al., 2022) where about 95% classification accuracy has been reported with four levels of adulteration, and the performance of the reported CNN was found to be better than the performance of some of the

**The works presented in this chapter has been partially published in author's following publication Mandal, D., Chatterjee, A., and Tudu, B., Identifying the degree of cornstarch adulteration in turmeric powder using optimized convolutional neural network, Intelligent Decision Technologies, Article in Press, DOI 10.3233/IDT-240656**

## Chapter 5: Detection of starch adulteration in turmeric powder

---

pre-trained models like VGG16, DenseNet and MobileNetV2. In this work, tartrazine-colored rice was used as an adulterant which causes color adulteration. On the other hand, our work was focused on volumetric adulteration using white wheat flour which is mixed in much higher quantity.

Various analytical techniques have been employed to identify starch adulteration. Below are summarized some significant methodologies utilized for this purpose. Corn starch was mixed with turmeric in varying proportions and analyzed using FTIR [5.4] spectroscopy. The overall efficiency of the analysis was approximately 94%. NIR spectroscopy was employed to detect starch adulteration. Reflectance spectra were acquired within the 900 – 1700 nm range, and subsequent signal processing involved filtering with a Savitzky–Golay (SG) filter [5.5] The potential of IR spectroscopy in addressing this issue has been explored, with data analysis conducted using PLS-DA and SIMCA techniques [5.6].IR spectroscopy has also been applied in this context, wherein a genetic algorithm-based variable selection mechanism was utilized for the PLS algorithm. This approach demonstrated an accuracy of approximately 99% [5.7]. There has been the application of FT\_NIR spectroscopy, and the results show that when PLSR was employed as the regression model, the  $R^2$  value was between 0.91 and 0.99[5.8]. The loop-mediated isothermal amplification method [5.9] is used for adulteration detection. A near-infrared (NIR) spectroscopy for detecting Sudan Red and Metanil yellow adulterations achieved over 90% accuracy using Soft independent modeling of class analogy (SIMCA) and partial least squares (PLS) techniques[5.10]. The detection of Sudan Red and yellow in turmeric powder using Raman imaging and Fourier transform infrared radiation (FT-IR) spectroscopy using partial least square regression (PLSR) method results in  $R^2$  values of 0.97 and 0.95 respectively[5.11]. FTIR and micro FTIR imaging with orthogonal partial least squares-discriminant analysis (OPLS-DA) [5.12] yield more than 98% accuracy. Laser-induced breakdown spectroscopy (LIBS) has also been informed [5.13]. Sensor-based detection has shown notable potential in terms of recovery rate and relative



## Chapter 5: Detection of starch adulteration in turmeric powder

---

standard deviation [5.14]. For the authentication of *C. longa*, a real-time PCR assay has shown significant potential with an  $R^2$  value of 0.99 [5.15].

Among the different advantages of CNN, the major one is the elimination of feature extraction steps that were explained in previous chapters. With the help of several filters loaded in CNN layers, it can extract abstract features from the images by itself. Such features make the model sensitive to tiny changes and that in turn provides its power to classify between different classes that are difficult to detect by human eyes and some of the conventional classifiers too. Often CNN is attributed with the disadvantage of a huge training dataset, but recent research has shown that a small strategically developed dataset along with an optimized model can even give significant results. The chapter also includes a binary genetic algorithm (BGA) [5.16] hyperparameter optimization scheme for improved performance and consistency of the CNN model.

### 5.2. Materials and methods

#### 5.2.1. *Experimental in-house database preparation*

5 variants of *pure* turmeric powders were taken for experimentation. Among those 5 variants, 4 were branded organic turmeric powders that claim 100% purity and were procured from the local supermarket and 1 pure variant was prepared in-house following the process involving boiling the raw turmeric roots followed by oven drying and grinding. Starch manufactured by Sigma Altrich was procured from lab supplier. Mixed samples were stored in air-tight containers in dry condition where room temperature and humidity were varying within 20 - 30°C and 75 – 80%, respectively.

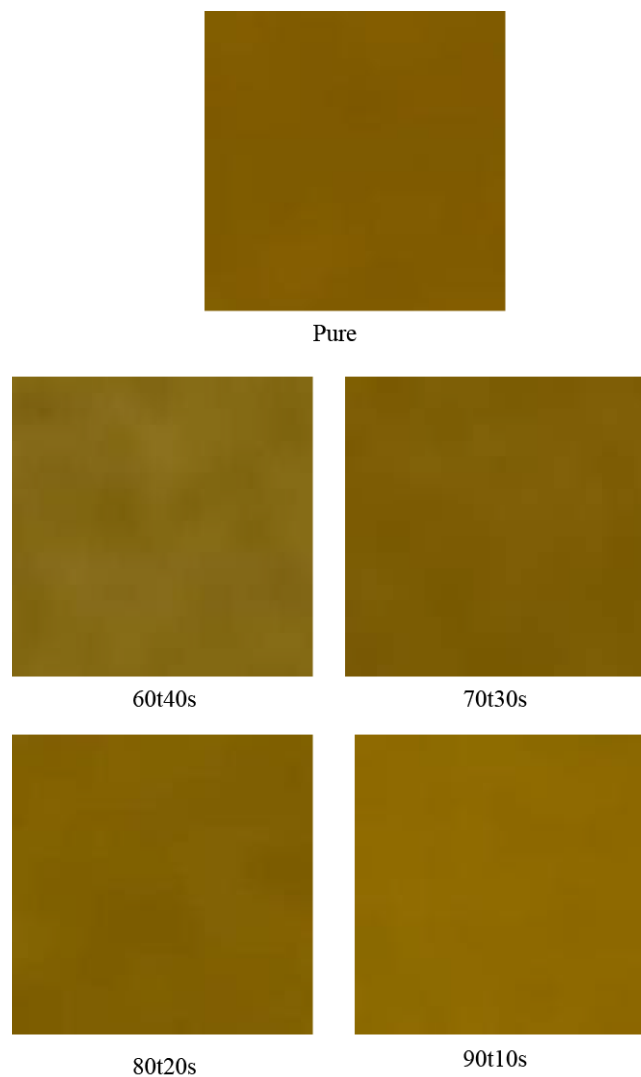
The experiment was performed by adulterating the turmeric powder samples with starch in a controlled manner. The starch was added with the turmeric powder in four different percentages by volume, i.e. 10%, 20%, 30%, and 40% and those were labeled as %t% s i.e. if

## Chapter 5: Detection of starch adulteration in turmeric powder

---

the sample is made of 80% turmeric and 20% starch it was labeled as 80t20s. For this work cameras of 5 different mobile phones having resolutions ranging from 48MPixel to 64MPixel were used. The reason was to get a closer approximation of the real-life consumer behavior. No camera filters or external light sources (like flashlight) were used during imaging.

Once captured the images were cropped into  $300 \times 300$  pixel dimension. The cropping was conducted automatically by taking random centers in the entire image. For each of the turmeric samples 1000 such images were considered. Since there were total five classes (pure and four levels of adulteration) the total images in the experimental dataset resulted into 5000. Some of the cropped image samples are shown in Fig. 5.1.



**Fig. 5.1:** Samples of images from the image dataset

## Chapter 5: Detection of starch adulteration in turmeric powder

### 5.2.2. Development of CNN model

CNN models comprise three fundamental layers; convolution (conv\_2d), pooling, and fully connected (FC) layers [5.17]. The combination of conv\_2d and pooling performs the feature extraction task from the images while the fully FC layer is similar to conventional artificial neural network (ANN) architecture made of input, hidden and output layers. The filters specified in the convolution layers transform the image as per the mathematical dynamics of the filters while the pooling layer commonly extracts either mean or max values from the filtered images. Numbers of conv\_2d and pooling layer combinations can be used in sequence depending on the nature of input images.

In this work initially 2 conv\_2d and max\_pooling layer combinations were taken. It can be noted here that a shallow architecture of CNN has been adopted where the dimension of each conv\_2d layer will be half of its previous conv\_2d layer. As the inclusion of more layers can make the model heavier so inclusion of the normalization method namely, batch-normalization were also added to observe the performance. The adam optimizer [5.18] was used as it can automatically modulate the learning rate. The activation function was ReLU for all the layers other than output, which is expressed as Eq. (5.1).

$$f(x) = \max(0, x) \quad (5.1)$$

Softmax activation function, as expressed as Eq. (2) [5.19], was used for output layer where K is the number of classes,  $\vec{z}$  is the input vector and  $\sigma$  is the softmax function.  $e^{\vec{z}_i}$  and  $e^{\vec{z}_j}$  are the standard exponential function for input and output vector, respectively.

$$\sigma(\vec{z})_i = \frac{e^{\vec{z}_i}}{\sum_{j=1}^K e^{\vec{z}_j}} \quad (5.2)$$

## Chapter 5: Detection of starch adulteration in turmeric powder

The loss function for optimizing the model was categorical cross-entropy and the initial model parameter setting is consolidated in Table 5.1.

**Table 5.1:** Initial model parameter settings

Layer type	Output dimension	Number of parameters
Conv_2d	148, 148, 32	896
Max_pooling	74, 74, 32	0
Conv_2d	72, 72, 32	9248
Max_pooling	36, 36, 32	0
Flatten	41472	0
Hidden	128	5308544
Output	5	516

### 5.2.3. Hyperparameter optimization of the CNN model

Hyperparameter settings optimization is a frequently practices step to enhance CNN performance in terms of accuracy and consistency. Different approaches for this are reported. In this work, a simple binary genetic algorithm (BGA) based approach was adopted. The choice of BGA was its simplicity, and proven efficiency and as the model architecture does not include many layers hence a simple mathematical optimization can be computationally less expensive to realize. Like other heuristics, BGA also starts with random solutions and reaches to the final optimized solutions through evaluation and generation of solutions against the fitness function. Crossover and mutation are two nature-inspired operations in BGA that results in the generation of superior offspring compared to the parents. Based on the literature reviews on GA applications the crossover probability was initially set at 0.5 and with iterations reduced to 0.1, the mutation probability was set at 0.05 and started with 20 initial randomly generated binary solutions. A single-point cross-over operation was adopted where the cross-over point was randomly generated in [1, 2,..., 9]. The fitness

## Chapter 5: Detection of starch adulteration in turmeric powder

function[5.20]used in this work is presented in Eq. (5.3) The ranges of different model parameters as mentioned in Table 5.2 were taken for initiation of the BGA algorithm. Each solution was represented as a binary string of length of 9 bit where the bits were decoded as shown in Fig. 5.2.

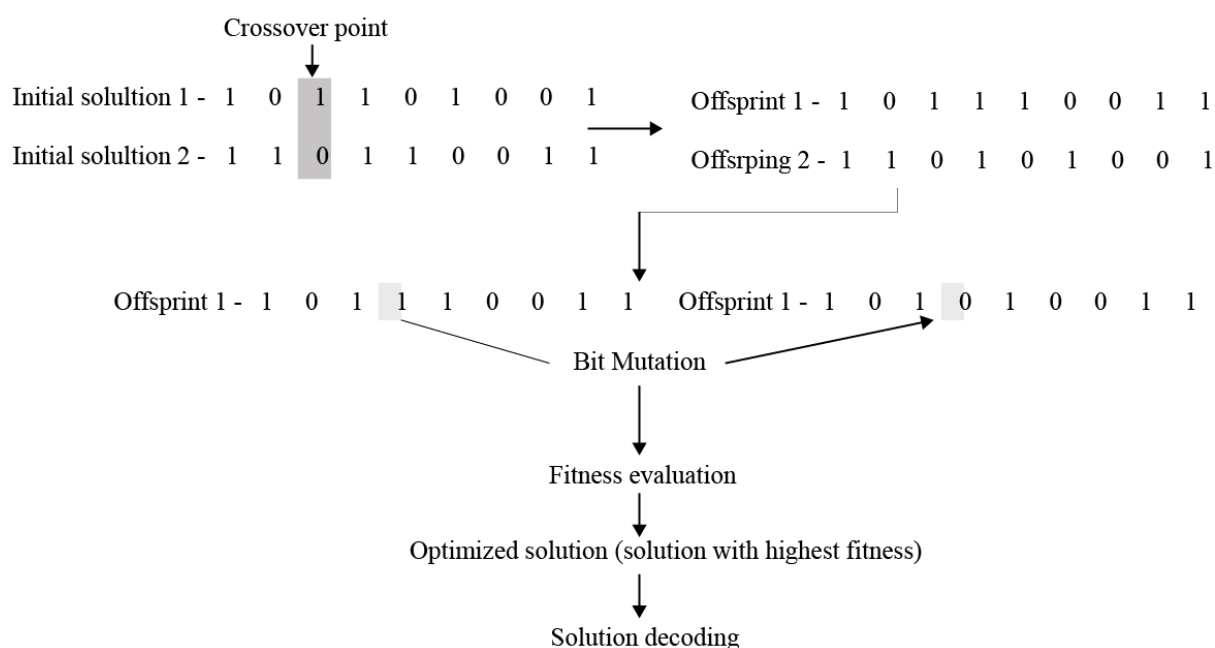
$$O = \frac{1}{2} \left( \frac{\sum_N (y - \hat{y})^2}{N} \right)^{1/2} \quad (5.3)$$

Where  $y$  and  $\hat{y}$  are the expected and actual classification accuracy of the model over  $N$  number of training samples. The algorithm was iterated till the stopping criterion was met which either was exhausting the defined number of iterations or achieving the desired value of  $O$  that is 0 in our case.

**Table 5.2:** Binary code representations of optimization parameters

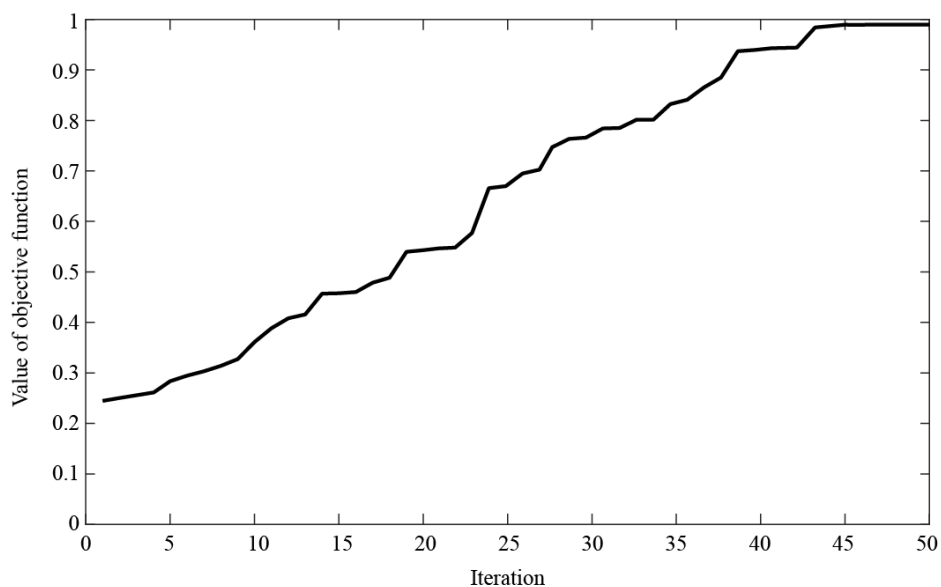
Parameter	Possible values	Binary representation	code	Code length
No. of conv_2d layers	2, 3, 4, 6	00,01, 10, 11	2	
No. of kernels in first conv_2d layer	16, 32, 64, 128	00, 01, 10, 11	2	
Filter dimension (square matrix) for conv_2d layers	3, 5, 7, 9	00,01,10, 11	2	
Stride	1, 2	0,1	1	
No. of hidden layers in FC layer	1, 2, 3, 4	00,01,10, 11	2	
Total length			9	

## Chapter 5: Detection of starch adulteration in turmeric powder



The convergence plot for the CNN parameter optimization is shown in Fig. 5.3 which shows that at the end of 50 iterations the accuracy of 99% is obtained by the model with the optimized parameter settings as provided in Table 5.3. The Convolutional Neural Network (CNN) model implementation was achieved on the Python platform within the Spyder IDE, utilizing standard libraries such as Keras [5.21] and Scikit-learn [5.22] in Windows environment.

## Chapter 5: Detection of starch adulteration in turmeric powder



**Fig. 5.3:** Convergence plot of BGA optimization

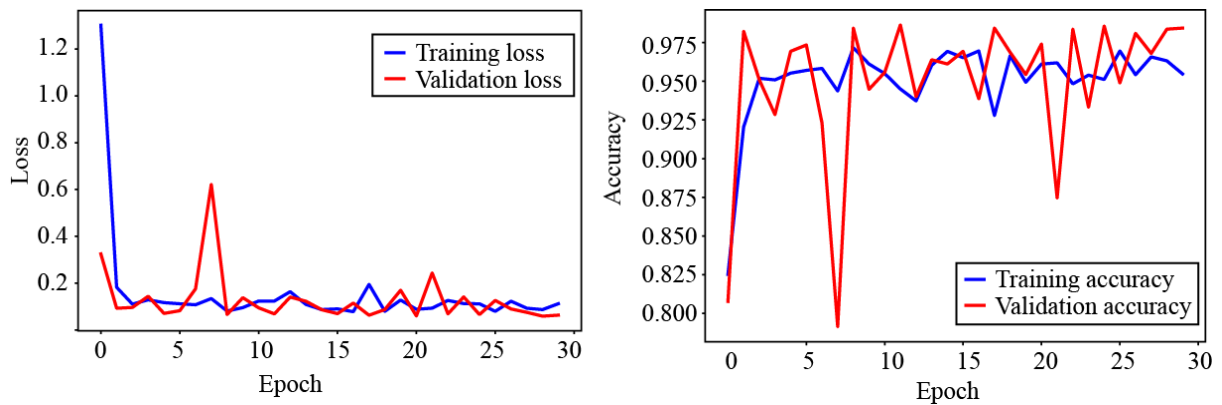
**Table 5.3:** Optimized hyperparameter settings

Layer (type)	Shape	Parameters
conv2d (Conv2D)	(300, 300, 32)	896
max_pooling2d (MaxPooling2D)	(150,150,32)	0
conv2d_1 (Conv2D)	(150,150,64)	18496
max_pooling2d_1 (MaxPooling2D)	(75,75,64)	0
conv2d_2(Conv2D)	(75,75,64)	36928
max_pooling2d_2 (MaxPooling2D)	(35,35,64)	0
conv2d_3 (Conv2D)	(33,33,128)	73856
max_pooling2d_3 (MaxPooling2D)	(16,16,128)	0
flatten (Flatten)	32768	0
dense (Dense)	256	8388864
dense_1 (Dense)	5	1285
<b>Total Parameters</b>		<b>8520325</b>
<b>Trainable Parameters</b>		<b>8520325</b>

## Chapter 5: Detection of starch adulteration in turmeric powder

### 5.3. Results and discussions

Like the previous chapters in this case also the database was divided into training, validation and testing set in 60:20:20 ratios. Before finalizing the model its generalization potential was verified against accuracy and loss values for the training and validation sets. It can be noted here that a model with poor generalization potential can cause poor performance on unknown test cases while performing brilliantly with training cases. The validation plots are shown in Fig. 5.4. It shows that the model has considerable potential since the training and validation plots are not resulting any visible difference at the ends. Plots also show that the model has balanced performance capability avoiding the chances of over- and under-fitting. It can also be notes that in both cases the loss is considerably low in the order of  $10^{-2}$  while the accuracy is in the tune of 98%. At the end of epochs the blue line of validation accuracy is almost meeting with the red line for training accuracy having small gap. This reflects that the model performance is towards better conformance to the desired low bias and high variance nature.



**Fig. 5.4:** CNN model validation plots for (a) accuracy and (b) loss between training and validation set.

The model consistency was tested against 10-fold cross validation method. The accuracy and loss values for different folds are presented in Table 5.4. It can be seen that across all the folds the accuracy is more than 90% except one. The maximum accuracy has gone to the



## Chapter 5: Detection of starch adulteration in turmeric powder

extent of more than 99% which is almost 100%. Regarding loss there are two folds where the loss is comparatively high. For other folds it is in the range of 4 – 18%.

**Table 5.4:** Result of 10-fold cross validation

Fold Number	Validation Loss (%)	Validation Accuracy (%)
1	4.00152	99.2188
2	22.8981	88.8021
3	12.5592	95.8333
4	6.51735	98.4375
5	18.3899	91.4062
6	15.161	94.0104
7	6.43669	97.1354
8	10.5456	97.9167
9	11.3576	95.8333
10	31.8529	94.5312

The confusion matrix for the model performance with testing data has been shown in Fig. 5.5 and the associated metric values have been consolidated in Table 5.5. This figure shows that for all the classes the proposed model provide 100% accuracy except for 80t20s and 90t10s, nevertheless in these case too none of the adulterated samples have been identified as pure. I these two cases also the accuracy is in the tune of 95% and on an average about 7% wrong classification have occurred.

## Chapter 5: Detection of starch adulteration in turmeric powder

	Predicted class				
	60t40s	70t30s	80t20s	90t10s	Pure
Actual class	60t40s	1	0	0	0
	70t30s	0	1	0	0
	80t20s	0	0.01	0.98	0.01
	90t10s	0	0	0.06	0.94
	Pure	0	0	0	1

**Fig. 5.5:** The confusion matrix of CNN classification

**Table 5.5:** Comparative evaluation of classification performance

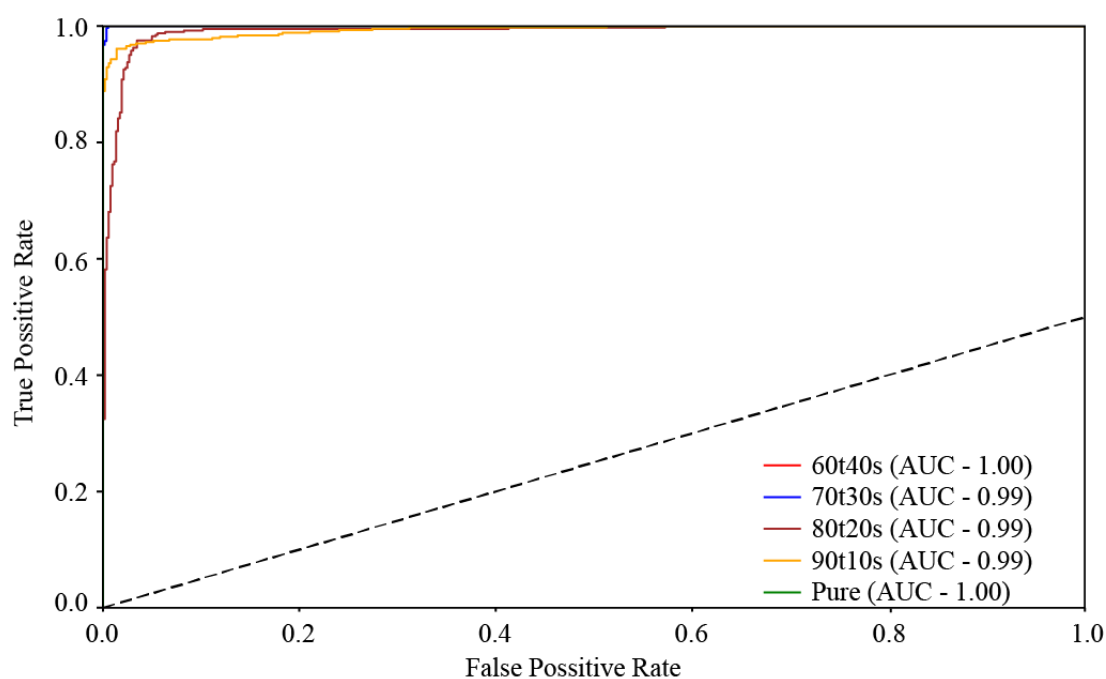
	VGG16	MobileNetV2	Resnet50	EfficientNetB0	Proposed model
<b>Precision</b>	0.8816	0.8871	1.0000	0.8816	1.0000
<b>Recall</b>	1.0000	1.0000	1.0000	1.0000	1.0000
<b>Accuracy</b>	0.2152	0.5785	0.9356	0.2152	0.9771
<b>F-Score</b>	0.9371	0.9402	1.0000	0.9371	1.0000

In Table 5.5 some other pre-trained models have been included for performance comparison purposes. The weights of these models have already been trained but some fine-tuning occurs while these are trained with application-specific datasets. The associated metric values i.e. precision, recall, accuracy, F-score [5.23], and are depicted in Table 5.6 All these models VGG16[5.25], MobileNetV2[5.26], Resnet50[5.27], EfficientNetB0[5.28] are considerably heavier in terms of the number of layers, depth, and connection weights. The objective of this comparison was to understand the competitive potential of our CNN model having a lesser number of layers and tunable parameters against the pre-trained models. Table 5.6 shows that the proposed method can achieve more than 97% accuracy for the unknown test data. The recall value of 1 reflects the potential of high accuracy in identifying the true positive

## Chapter 5: Detection of starch adulteration in turmeric powder

samples that means a pure or adulterated sample will be classified as that with high perfection.

The plots for AUC-ROC [5.29] presented in Fig. 5.6 show that the presented model is potential across all the classes of adulterations considered in this study. Although for few classes the AUC is not 1 but it is in the tune of 0.99 which indicates about 99% correct classification in those classes. It is also important to note that the model performs about 100% prediction accuracy for pure and highest degree of adulteration (60t40s) in this work which in turns confirms that the possibility of classifying unadulterated sample as adulterated and highly adulterate samples as unadulterated or pure is very small.



**Fig 5.6:** ROC-AUC plot for classification results in different classes

Finally the speed of the CNN based implementations have always been a major concern since the training and validation process is computationally expensive. The speed depends on the hardware capability as well however all the models under comparison were implemented in

## Chapter 5: Detection of starch adulteration in turmeric powder

same hardware set-up as explained in chapter 2 which has been used throughout the works presented in this thesis. Table 5.7 shows the comparison in terms of time.

**Table 5.6:** Comparison on processing time

	VGG16	MobileNetV2	Resnet50	EfficientNetB0	Proposed CNN
<b>Model</b>	5056.77	1607.12	3575.71	2351.12	1380.05
<b>training</b>					
<b>time (in</b>					
<b>seconds)</b>					

Table 5.6 shows that our model is advantageous over the other pre-trained model as it results in faster training. The main reason behind this is fewer layers and fewer tunable parameters that too been optimized using BGA.

### 5.4. Conclusion

A CNN-based implementation of adulteration detection in starch has been presented in this chapter. Volume-based adulteration has been considered in this work and starch has been opted as adulterator. The samples were prepared with 4 degrees of adulteration and one portion was left unadulterated. The results confirm more than 97% accuracy with high valued recall and precision values. A BGA optimization framework has also been presented in this chapter for the optimization of the CNN hyperparameter. The optimization resulted in a lighter, consistent and efficient model which was faster as well compared to the pre-trained CNN models. Overall the work shown promising potential of the CNN based model in detection of starch adulteration in turmeric.

## Chapter 5: Detection of starch adulteration in turmeric powder

---

### References

- 5.1. Moyer, D. C., DeVries, J. W., & Spink, J. (2017). The economics of a food fraud incident—Case studies and examples including Melamine in Wheat Gluten. *Food Control*, 71, 358-364.
- 5.2. Kaur, T., Gamez, A., Olvera, J. L., Schaefer, B. C., & Corona-Chavez, A. (2023). I-TAINTED: Identification of turmeric adulteration using the cavity perturbation technique and technology optimized machine learning. *IEEE Access*.
- 5.3. Kaur, T., Gamez, A., Olvera, J. L., Schaefer, B. C., & Corona-Chavez, A. (2023). I-TAINTED: Identification of turmeric adulteration using the cavity perturbation technique and technology optimized machine learning. *IEEE Access*.
- 5.4. Ballesteros, J. I., Lim, L. H. V., & Lamorena, R. B. (2024). The feasibility of using ATR-FTIR spectroscopy combined with one-class support vector machine in screening turmeric powders. *Vibrational Spectroscopy*, 130, 103646.
- 5.5. Lanjewar, M. G., Morajkar, P. P., & Parab, J. S. (2024). Portable system to detect starch adulteration in turmeric using NIR spectroscopy. *Food Control*, 155, 110095.
- 5.6. Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2021). An evaluation of IR spectroscopy for authentication of adulterated turmeric powder using pattern recognition. *Food Chemistry*, 364, 130406.
- 5.7. de Macêdo, I. Y. L., Machado, F. B., Ramos, G. S., do Carmo Costa, A. G., Batista, R. D., Galvao Filho, A. R., ... & de Souza Gil, E. (2021). Starch adulteration in turmeric samples through multivariate analysis with infrared spectroscopy. *Food Chemistry*, 340, 127899.

## Chapter 5: Detection of starch adulteration in turmeric powder

---

- 5.8. Kar, S., Tudu, B., Jana, A., & Bandyopadhyay, R. (2019). FT-NIR spectroscopy coupled with multivariate analysis for detection of starch adulteration in turmeric powder. *Food Additives & Contaminants: Part A*, 36(6), 863-875.
- 5.9. Sheu, S. C., Wu, Y. C., Lien, Y. Y., & Lee, M. S. (2021). Specific, sensitive and rapid *Curcuma longa* turmeric powder authentication in commercial food using loop-mediated isothermal nucleic acid amplification. *Saudi Journal of Biological Sciences*, 28(10), 5931-5936. Sheu, S. C., Wu, Y. C., Lien, Y. Y., & Lee, M. S. (2021). Specific, sensitive and rapid *Curcuma longa* turmeric powder authentication in commercial food using loop-mediated isothermal nucleic acid amplification. *Saudi Journal of Biological Sciences*, 28(10), 5931-5936.
- 5.10. Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2022). Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: The role of preprocessing methods in analysis. *Vibrational Spectroscopy*, 120, 103372.
- 5.11. Chao, K., Dhakal, S., Schmidt, W. F., Qin, J., Kim, M., Peng, Y., & Huang, Q. (2020). Raman and IR spectroscopic modality for authentication of turmeric powder. *Food chemistry*, 320, 126567.
- 5.12. Shannon, M., Lafeuille, J. L., Frégière-Salomon, A., Lefevre, S., Galvin-King, P., Haughey, S. A., ... & Elliott, C. T. (2022). The detection and determination of adulterants in turmeric using fourier-transform infrared (FTIR) spectroscopy coupled to chemometric analysis and micro-FTIR imaging. *Food Control*, 139, 109093.

## Chapter 5: Detection of starch adulteration in turmeric powder

---

- 5.13. Kumar, T., Rai, A. K., Dwivedi, A., Kumar, R., Azam, M., Singh, V., ... & Rai, A. K. (2022). Chemical characterization for the detection of impurities in tainted and natural curcuma longa from india using LIBS coupled with PCA. *Atoms*, 10(3), 91.
- 5.14. Nag, S., Das, D., Naskar, H., Tudu, B., Bandyopadhyay, R., & Roy, R. B. (2022). Detection of metanil yellow adulteration in turmeric powder using nano nickel cobalt oxide modified graphite electrode. *IEEE Sensors Journal*, 22(13), 12515-12521.
- 5.15. Oh, S. H., & Jang, C. S. (2020). Development and validation of a real-time PCR based assay to detect adulteration with corn in commercial turmeric powder products. *Foods*, 9(7), 882.
- 5.16. Xue, Y., Zhu, H., Liang, J., & Słowik, A. (2021). Adaptive crossover operator based multi-objective binary genetic algorithm for feature selection in classification. *Knowledge-Based Systems*, 227, 107218.
- 5.17. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019.
- 5.18. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- 5.19. Wang, F., Cheng, J., Liu, W., & Liu, H. (2018). Additive margin softmax for face verification. *IEEE Signal Processing Letters*, 25(7), 926-930.
- 5.20. Rere, L. M., Fanany, M. I., & Arymurthy, A. M. (2016). Metaheuristic algorithms for convolution neural network. *Computational intelligence and neuroscience*, 2016.
- 5.21. Ketkar, N., & Ketkar, N. (2017). Introduction to keras. *Deep learning with python: a hands-on introduction*, 97-111.

## Chapter 5: Detection of starch adulteration in turmeric powder

---

- 5.22. Fabian, P. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, 2825.
- 5.23. Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval* (pp. 345-359). Berlin, Heidelberg: Springer Berlin Heidelberg.
- 5.24. Vieira, S. M., Kaymak, U., & Sousa, J. M. (2010, July). Cohen's kappa coefficient as a performance measure for feature selection. In *International conference on fuzzy systems* (pp. 1-8). IEEE.
- 5.25. Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Almansour, N. A. (2021, July). Deep CNN model based on VGG16 for breast cancer classification. In *2021 International conference on information technology (ICIT)* (pp. 805-810). IEEE.
- 5.26. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
- 5.27. Theckedath, D., & Sedamkar, R. R. (2020). Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Computer Science*, 1(2), 79.
- 5.28. Pramudhita, D. A., Azzahra, F., Arfat, I. K., Magdalena, R., & Saidah, S. (2023). Strawberry Plant Diseases Classification Using CNN Based on MobileNetV3-Large and EfficientNet-B0 Architecture. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9(3), 522-534.
- 5.29. Muschelli III, J.: ROC and AUC with a binary predictor: a potentially misleading metric. *Journal of Classification*.37(3), 696-708 (2020).



# CHAPTER 6

## Concluding remarks

### 6.1. Introduction

This chapter summarizes the major findings and observations noted during this thesis work. A brief comparison between the techniques reported in this thesis has as well been presented for ready reference. The research questions which were clearly framed at the beginning of the thesis have been addressed. Finally, the chapter includes the limitations of the presented methods and thereafter the possible future research directions of the presented works.

### 6.2. Major findings

The major findings of the thesis can be mentioned as follows;

- There was no chemical reactions involved during the adulteration detection process hence the methods were non-invasive.
- No skilled man-power was required for sample preparation unlike the conventional spectroscopic methods.
- The in-house imaging systems with controlled illumination could be helpful for consistency in illumination which in-turn can result good accuracy with a comparatively smaller database.

## Concluding remarks

- All the implementations were inexpensive as those required only basic computation and imaging facilities. Even mobile phone cameras could give promising results.
- Overall accuracies across the different techniques are in the range of 97 – 99% which is motivating and competitive to the existing expensive instrumental and analytic methods.
- Although slightly reduced accuracies have been occurred in some cases but for pure and adulterated samples this has been always in the tune of 99% or more which confirms that the chances of detecting an adulterated samples as pure and vice-versa is notably less.
- As shown in Table 6.1 the performance of the CNN model is visibly better than the performance of conventional classifiers. CNN also eliminates the feature extraction step which in-turn makes the system faster.

**Table 6.1:** Performance comparison between the proposed methods in this thesis

Adulteration	Detection method	Classes	Feature	Accuracy
Metanil yellow	RF and DNN	Pure and 3 degrees of adulteration	Projection features on Fourier spectra of color channels	98%
Sudan dye I	PCA and ensemble RFs	Pure and 3 degrees of adulteration	Color channel features	98.75%
Cornstarch	BGA optimized CNN	Pure and 4 degrees of adulteration	Automated feature extraction by convolution layers of CNN	98%

- Optimization of the hyper parameters for CNN can also enhance the performance of the CNN models where a comparatively lighter CNN model could perform potentially.

## Concluding remarks

---

- The potential of Random Forest was also remarkable considering its ease of implementation and computational inexpensiveness.
- The ensemble implementation of multiple RF models can also be used for faster decision making.
- Color channel based analysis and feature extraction plays significant role for conventional classifiers.
- For all the cases the decision making was visibly fast.

### 6.3. Responses to the research questions

The observations recorded during the thesis works reveal following responses to the RQs framed in the beginning of the thesis.

- Computer vision can be a potential alternative to the existing methods of screening for turmeric powder. In all the cases the results show considerable generalization potential and accuracy in the range of 97 – 100%. It can be also helpful in faster and less expensive detection of turmeric adulteration. This addresses the RQ1.
- Deep learning-based classification with color features extracted from images has shown promising potential over conventional machine learning algorithms. However, the random forest has also performed competitively, particularly in the case of value prediction. This addresses the RQ 2.
- CNN, particularly optimized CNN has also shown remarkable performance with the advantage of automatic feature extraction. An application oriented CNN can work competitively to the pre-trained models which are heavier and computationally expensive. This addresses RQ 3.

## Concluding remarks

---

- Computer vision models due to its connectivity to internet and possibility of development in terms of apps and handheld devices opens up the huge possibility to integrate the reported techniques in IoT based implementation with much ease compared to existing instrumental methods. This can be also extended to cloud computing based implementation which further vouches the advantage of IoT integration for the proposed methods. This addresses RQ 4.

### 6.4. Future scope of work

Although the reported method shown promising potential there were some limitations as well. Based on such limitations the following avenues may be listed towards extension of the presented works.

- Considering the varieties n types of adulteration in turmeric additional types of adulterants which are commonly used to be studied. However, the presented methods can be extended to those once the image datasets for those adulterants have been prepared.
- Currently the imaging chamber is having a single illuminant type while in real-life the illumination can be of different types. Hence, more variations n illumination and well as distortions due to illumination and other imaging conditions can be covered in the dataset preparation.
- As RF has shown significant potential and simplicity this can be further explored with several extracted features and model optimization paradigms so that a faster training and validation can be achieved along with equivalent performance like DNN or CNN can be achieved. Different illumination invariant features may also be explored in this context.

## Concluding remarks

---

- Different learning mechanisms of CNN like transfer learning, reinforcement learning, etc. can be explored with the application specific CNN development to achieve more robustness in decision making.
- More varieties in terms of degree of adulteration can be explored. Currently all our proposed methods have 4-5 variants in this regard which can be extended to more.
- Hardware and handheld device based deployment remains a very promising area to work for real-life implementation of the presented methods. This can also be useful for on-spot adulteration detection in the supply chain as well as in the market by the consumers themselves.

### 6.5. Conclusion

The thesis has presented different proposals for computer vision based adulteration detection in turmeric powders. The results were promising and motivating. The RQs framed at the beginning of the thesis also have been responded logically based on the major findings. Considering the different possible future research direction to address the limitations observed during this thesis work can be considered as a fruitful initial yet significant step towards development of non-invasive, inexpensive and easily IoT deployable computer vision based adulteration detection system for powdered turmeric.

Dipankar Mandal

  
**Associate Professor**  
Department of Printing Engineering  
Jadavpur University, Salt-lake Campus  
Kolkata - 700 098