

# **An Integrated Computer Vision and Machine Learning System for Emulsion Processing**

by

**Saritha Unnikrishnan**

Supervisors: Dr. David Tormey and Dr. John Donovan

A thesis submitted for the award of Doctor of Philosophy





Submitted to the Institute of Technology, Sligo, November 2019

## Declaration

I, Saritha Unnikrishnan, declare that this thesis, titled 'An Integrated Computer Vision and Machine Learning System for Emulsion Manufacturing' is entirely my own work and has not been submitted for any other academic award, or thereof, at this or any other education establishment. I confirm that this work was carried out by me for the award of Doctor of Philosophy under the supervision of Dr. David Tormey and Dr. John Donovan.

Signed:  \_\_\_\_\_ (Candidate)

Signed:  \_\_\_\_\_ (Supervisor)

  
Signed: \_\_\_\_\_ (Supervisor)

Date: 01/07/2020

## Publications

1. **Unnikrishnan, S.**, Donovan, J., Macpherson, R., & Tormey, D. (2020). An Integrated Histogram-Based Vision and Machine-Learning Classification Model for Industrial Emulsion Processing. *IEEE Transactions on Industrial Informatics*, 16, 5948-5955. (Impact Factor: 7.377)
2. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey, D. (2019) 'Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions', *Journal of Pharmaceutical Innovation*, pp. 1-12 (Impact Factor: 1.452).
3. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey (2018), D. 'Machine vision for the quality assessment of emulsions in pharmaceutical processing'. *4th International Conference on Universal Village (UV)*. Boston, MA, USA, IEEE, pp. 1-6.
4. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey, D. (2018) 'Automated Analysis of Micrographs for Emulsion Quality Evaluation'. *Microscopy Society of Ireland Annual Symposium*, Sligo, Ireland (**Received the best oral presentation award**).
5. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey, D. (2018). 'A Comparative Statistical Study of Micrograph Classification using Repeatability & Reproducibility Analysis'. *Conference on Applied Statistics*, Galway, Ireland (Poster presentation).
6. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey, D. (2017) 'Multi-response Optimisation of Image Processing Parameters using CCD and Desirability Function ', *International Manufacturing Conference (IMC)*. Sligo, Ireland.
7. **Unnikrishnan, S.**, Donovan, J., Macpherson, R. and Tormey, D. (2017) 'Emulsion quality evaluation using automated image analysis'. *International Conference of Innovative Design & Manufacturing (ICIDM)*, Milan, Italy.

## Patent Application

**Title:** System for Controlling an Emulsification Process.

**Inventors:** Saritha Unnikrishnan, John Donovan and David Tormey. International Patent Application number: WO 2020/079283 A1 (published). Filed on 21<sup>st</sup> October 2019.

## **Acknowledgements**

I would like to express my sincere gratitude to my research supervisors, Dr. David Tormey and Dr. John Donovan, for their valuable advice and constant encouragement for completing this thesis.

The financial support provided by the President's bursary award is acknowledged.

I wish to express my sincere thanks to Dr. Russell Macpherson and Mr. Eugene McHenry, GlaxoSmithKline (Ireland) Ltd for providing the industrial infrastructure for my research.

I would also like to thank Dr. John Bartlett, Ms. Veronica Cawley and Ms. Ursula Cox for their research support.

Many thanks to Dr. Leo Creedon, Dr. Darren Whitaker, Dr. Konrad Mulrennan and Mr. Sean Mullery for sharing their state-of-the-art knowledge and experience.

I am grateful to Ms. Una Parsons, Head of Faculty of Engineering and Design and Ms. Diane O'Brien, Head of department of Computing and Electronics for their constant support.

My sincere thanks go to all my friends and colleagues at IT Sligo.

I would also like to thank my parents (Unnikrishnan Nair and Vijayalekshmi) and my sister, Sangeetha Nair for providing constant encouragement.

Heart-felt thanks to my husband, Suresh, who has been a constant source of love and support.

## **Dedication**

To my parents, my husband, Suresh and our boys, Sreehari and Sreesekhar.

# TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>VIII</b>
<b>LIST OF TABLES .....</b>	<b>XI</b>
<b>ABSTRACT .....</b>	<b>XIII</b>
<b>ABBREVIATIONS .....</b>	<b>XIV</b>
<b>CHAPTER 1 .....</b>	<b>1</b>
1.1 SCOPE OF THIS RESEARCH .....	2
1.2 AIMS AND OBJECTIVES .....	3
1.2.1 Objectives .....	3
1.2.2 Potential Industrial Impact.....	4
1.3 OVERVIEW OF EXISTING TECHNIQUES FOR EMULSION QUALITY EVALUATION .....	5
1.3.1 Manual Evaluation .....	5
1.3.2 Other Conventional Methods .....	6
1.3.3 Offline Image Processing techniques for Droplet Detection .....	8
1.3.4 Image Segmentation .....	14
1.4 INLINE DROPLET SIZE MONITORING .....	15
1.5 CURRENT CHALLENGES IN DROPLET DETECTION FROM EMULSION MICROGRAPHS .....	18
1.6 THESIS OUTLINE .....	19
1.7 CONCLUSIONS .....	23
<b>CHAPTER 2 .....</b>	<b>24</b>
2.1 INTRODUCTION .....	24
2.2 MACHINE LEARNING .....	24
2.2.1 Principal Component Analysis .....	25
2.2.2 Linear Discriminant Analysis .....	27
2.2.3 Logistic Regression .....	28
2.2.4 Random Forest .....	30
2.2.5 Neural Networks.....	32
2.2.6 Convolutional Neural Networks .....	35
2.3 INDUSTRIAL APPLICATIONS .....	39
2.4 CONCLUSIONS .....	48
<b>CHAPTER 3 .....</b>	<b>49</b>
3.1 INTRODUCTION .....	49
3.2 OVERVIEW OF THE APPROACH.....	49
3.3 MICROGRAPH ACQUISITION.....	52
3.3.1 Process-1 .....	52
3.3.2 Process-2.....	52
3.4 IMAGE PROCESSING AND SEGMENTATION .....	54
3.4.1 Fiji, an Extended Version of ImageJ .....	54
3.4.2 Edge and Symmetry Technique (EST) .....	55
3.4.3 Histogram-Based Technique (HBT) .....	60
3.5 STATISTICAL ANALYSIS OF DROPLET CHARACTERISTICS .....	64
3.5.1 R Statistical Software .....	65
3.6 MICROGRAPH CLASSIFICATION .....	65
3.6.1 Principal Component based Linear Discriminant Analysis (PC-LDA).....	66
3.6.2 Cross-validation .....	68
3.6.3 Multinomial Logistic Regression (MLR) .....	68
3.6.4 Random Forest (RF).....	69
3.6.5 Neural Network (NN) .....	70
3.6.6 Convolutional Neural Network (CNN) .....	70
3.7 REMAINING PROCESSING TIME PREDICTION .....	71
3.8 CONCLUSIONS .....	72
<b>CHAPTER 4 .....</b>	<b>74</b>

4.1	INTRODUCTION .....	74
4.2	RESULTS AND DISCUSSION .....	74
4.2.1	<i>Process-1: Droplet Detection and Analysis</i> .....	74
4.2.2	<i>Process-2: Droplet Detection and Analysis</i> .....	90
4.3	CONCLUSIONS .....	98
<b>CHAPTER 5</b>	<b>.....</b>	<b>99</b>
5.1	INTRODUCTION .....	99
5.2	CLASSIFICATION MODELLING: RESULTS AND DISCUSSION.....	100
5.2.1	<i>PC-LDA Model</i> .....	100
5.2.2	<i>MLR Model</i> .....	109
5.2.3	<i>RF Model</i> .....	115
5.2.4	<i>VNN Model</i> .....	120
5.2.5	<i>CNN Model</i> .....	125
5.2.6	<i>Comparison between the Classification Models</i> .....	131
5.3	CONCLUSIONS .....	134
<b>CHAPTER 6</b>	<b>.....</b>	<b>135</b>
6.1	INTRODUCTION .....	135
6.2	CLASSIFICATION STUDY AND ATTRIBUTE AGREEMENT ANALYSIS (AAA) .....	137
6.3	RESULTS AND DISCUSSION .....	138
6.4	DETAILED ANALYSIS OF THE STUDY.....	148
6.5	CONCLUSIONS .....	149
<b>CHAPTER 7</b>	<b>.....</b>	<b>151</b>
7.1	INTRODUCTION .....	151
7.2	REGRESSION MODELLING: RESULTS AND DISCUSSION .....	152
7.2.1	<i>Linear Regression Model</i> .....	152
7.2.2	<i>Random Forest Regression Model</i> .....	162
7.2.3	<i>Comparison between the LR and RF models</i> .....	169
7.3	CONCLUSIONS .....	170
<b>CHAPTER 8</b>	<b>.....</b>	<b>172</b>
8.1	CONCLUSIONS .....	172
8.1.1	<i>A novel Image segmentation for automated droplet characterisation</i> .....	172
8.1.2	<i>Machine Learning Classification</i> .....	173
8.1.3	<i>Processing Time Prediction</i> .....	174
8.1.4	<i>Knowledge Contribution and Impact</i> .....	175
8.2	FUTURE WORK .....	177
8.2.1	<i>Generative Models</i> .....	178
8.2.2	<i>Ready-to-use object detection networks</i> .....	179
8.2.3	<i>Real-Time Integration</i> .....	179
8.2.4	<i>Extension of the Manual Versus Machine Learning Classification</i> .....	180
<b>BIBLIOGRAPHY</b>	<b>.....</b>	<b>181</b>
<b>APPENDIX A:</b>	<b>  FORMULAE FOR THE DROPLET SIZE CHARACTERISTICS USED IN THIS STUDY .....</b>	<b>193</b>
<b>APPENDIX B:</b>	<b>  MULTI-RESPONSE OPTIMISATION OF IMAGE PROCESSING PARAMETERS USING CCD AND DESIRABILITY FUNCTION .....</b>	<b>194</b>
<b>APPENDIX C:</b>	<b>  PLOTS OF ADDITIONAL DROPLET CHARACTERISTICS .....</b>	<b>196</b>
<b>APPENDIX D:</b>	<b>  TAMU MICROGRAPHS .....</b>	<b>201</b>

## LIST OF FIGURES

Figure 1.1. Schematic representation of the manual quality evaluation of emulsions. ....	5
Figure 1.2. Focused beam reflectance method (FBRM) for droplet size distribution analysis: (left) 2D-ORM sensor, (centre) FBR sensor and (right) working principle of the FBRM probe (Emmerich <i>et al.</i> , 2018). ....	7
Figure 1.3. Emulsion sample micrograph a) after 8 days b) after 28 days (Scherze <i>et al.</i> , 2005). ....	10
Figure 1.4. Changes in the $d_{32}$ index for the most stable and unstable O/W emulsion samples after 48 hours storage (Hosseini <i>et al.</i> 2015). ....	12
Figure 1.5 - Experimental set-up for <i>in-situ</i> droplet size monitoring (Khalil <i>et al.</i> , 2010). ....	16
Figure 1.6. Droplet detection using Hough circle transform: a) raw micrograph b) processed image with detected droplets (Khalil <i>et al.</i> , 2010). ....	17
Figure 1.7. Schematic representation of the thesis structure. ....	20
Figure 2.1. Projection of maximum variance of the three-dimensional variable space to the first principal component. The first principal component, $P_1$ is the line of best fit across the data points (Wold <i>et al.</i> , 1987) ....	26
Figure 2.2. The linear decision boundary $H$ , where $g(x) = w^T x + w_0 = 0$ , separates the feature space into two half-spaces $R_1$ (where $g(x) > 0$ ) and $R_2$ (where $g(x) < 0$ ) (Duda <i>et al.</i> , 2012). ....	28
Figure 2.3. A decision tree model using Iris data in R. ....	31
Figure 2.4. Schematic of a single hidden layer, feed-forward neural network (Friedman <i>et al.</i> , 2001). ....	34
Figure 2.5. Convolutional Layer. a) Input image of size 7 x 7 and b) Convolved output image of size 5x5. ....	36
Figure 2.6. Max pooling of a 4 x 4 data matrix down to 2 x 2. ....	38
Figure 2.7. Micrographs obtained from barley milk samples at a) 15 seconds b) 30 seconds c) 45 seconds and d) 60 seconds of blending (Kljusuric <i>et al.</i> , 2015). ....	41
Figure 2.8. Confusion matrices from 10-fold cross-validation of the PC-LDA model. Each matrix is the sum of 10 matrices from 10 test sets (Bertani <i>et al.</i> , 2017). ....	44
Figure 3.1. Schematic representation of the Research Methodology. ....	51
Figure 3.2. Detection of a circular object using edge and symmetry-based image segmentation. ....	57
Figure 3.3. Detection of oil droplets and characteristics using the Edge and Symmetry segmentation technique. ....	58
Figure 3.4. Detection of a circular object using histogram-based image segmentation. ....	60
Figure 3.5. Schematic showing the detection of oil droplets and their characteristics from an emulsion micrograph using the Histogram-Based image segmentation (HBT) macro in Fiji. ....	62
Figure 3.6. Schematic of the methodology applied for micrograph classification. ....	66
Figure 3.7. Methodology applied for optimum processing time prediction. ....	72
Figure 4.1. Micrographs acquired at 5, 10, 15, 20, 25 and 30 minutes of the 30 minutes duration emulsification process. ....	75
Figure 4.2. Droplet Detection using EST. a) A five-minute processed emulsion sample micrograph. b) Droplets detected. ....	76
Figure 4.3. Box Plots of mean droplet size characteristics and count from EST. a) area in $\mu\text{m}^2$ , b) perimeter in $\mu\text{m}$ , c) Sauter mean diameter d) maximum Feret diameter and e) minimum Feret diameter. Each box plot represents 10 micrographs. ....	78
Figure 4.4. Bar plots of the equivalent circle diameter of droplets obtained from EST. The colours represent the diameter ranges of the droplets as given in the legend. ....	80
Figure 4.5. Frequency plots of droplet size characteristics obtained from EST: a) area b) perimeter c) Feret and d) minFeret. The colours signify the process interval as shown in the legend. ....	82
Figure 4.6. HBT droplet detection. (a) A sample micrograph taken after five minutes, (b) Output image from the Histogram-based technique. ....	83
Figure 4.7. Process-1. Box plots of the mean droplet size characteristics and count from HBT: a) area, b) perimeter, c) Feret, d) minFeret, e) Sauter mean diameter and f) Droplet count. Each box plot represents 10 micrographs. ....	84
Figure 4.8. Bar plots of the equivalent circle diameter of droplets from HBT. The colours represent the different diameter ranges of the droplets as given in the legend. ....	86
Figure 4.9. Frequency plots of the droplet size characteristics obtained from HBT: a) area b) perimeter c) Feret and d) minFeret. The colours signify the process interval as given in the legend. ....	87
Figure 4.10. Box plots showing the % of oil concentration from 5 to 30 minutes of emulsification. Each box plot represents 10 micrographs. (a) Edge & Symmetry Technique, (b) Histogram-Based Technique. ....	88



Figure 4.11. Micrographs acquired from Process-2. A sample micrograph obtained at every five-minute interval is presented. The last sample micrograph taken at 90 minutes represents the discharge sample obtained at 32°C. .... 91

Figure 4.12. Process-2. Box plots of the mean droplet size characteristics and count from HBT: a) area, b) perimeter, c) Feret, d) minFeret, e) Sauter mean diameter and f) Droplet count. Each box plot represents 10 micrographs. The last box plot represents the discharge sample obtained after 90 minutes. .... 92

Figure 4.13. Box plots showing the % of oil concentration during the 90 minutes emulsification process (Process-2). Each box plot represents 10 micrographs..... 95

Figure 4.14. Micrographs obtained from process-2 categorised into TAMU based on the evolution of mean droplet area ( $\mu\text{m}^2$ ). .... 96

Figure 5.1. Scree plot of the five principal components (PCs)..... 102

Figure 5.2. The Principal Component Analysis score plots using the first two PCs. ‘U’ (stands for Unacceptable) is represented by purple circles, ‘M’ (Marginal) by green, ‘A’ (Acceptable) by red and ‘T’ (Target) by blue respectively. .... 103

Figure 5.3. The Principal Component Analysis score plots using the first three PCs. ‘U’ (stands for Unacceptable) is represented by purple circles, ‘M’ (Marginal) by green, ‘A’ (Acceptable) by red and ‘T’ (Target) by blue respectively. .... 103

Figure 5.4. Linear Discriminant Analysis classification presented by the histograms of the three discriminant functions. a) LD1, b) LD2 and c) LD3. .... 105

Figure 5.5. LD2 vs LD1 plot showing the separation between the TAMU micrographs. .... 105

Figure 5.6. Confusion matrix obtained from the stratified 5-fold cross-validation. It represents the sum of 5 confusion matrices from the 5 models. Each model classified a test set of 12 micrographs (3 from each TAMU category)..... 107

Figure 5.7. Confusion matrix showing the classification of the 150 micrographs, into TAMU, obtained from an independent batch of the emulsion. .... 108

Figure 5.8. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using Multinomial Logistic Regression. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 45 micrographs..... 112

Figure 5.9. Confusion matrix showing the TAMU classification of the 150 unseen micrographs obtained from the independent batch. .... 113

Figure 5.10. Variation in the OOB error rate of the RF model as the number of trees grow from 0 to 500. The coloured lines represent the individual categories (TAMU). The black line represents the overall OOB error estimate of the model. .... 116

Figure 5.11. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using Random Forest. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 45 micrographs. .... 118

Figure 5.12. Confusion matrix showing the TAMU classification of the 150 unseen micrographs using the Random Forest model. .... 119

Figure 5.13. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using the Vanilla Neural Network. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 30 micrographs. .... 123

Figure 5.14. Confusion matrix showing the TAMU classification of the 150 unseen micrographs using the VNN model. .... 124

Figure 5.15. Micrograph of a five-minute processed emulsion sample split into four quarters in Fiji. .... 126

Figure 5.16. Keras sequential CNN model. The loss of the model during training is presented in the top plot. The accuracy of the model, which is inversely proportional to the loss, is presented in the bottom plot. This plot was obtained for  $n = 500$  and 4 units in the convolutional layer. .... 129

Figure 5.17. Each box plot represents the average pixel intensity distribution of all the micrographs from the corresponding category. .... 130

Figure 6.1. The comparative study between manual and automated micrograph classification using Attribute Agreement Analysis (AAA)..... 139

Figure 6.2. Attribute agreement analysis (AAA) results from Minitab. .... 142

Figure 7.1. Predicted RPT of the test data (from the 10 LR models) plotted against the Specified RPT. The additional x-axis shows the current process time. The diagonal line represents the ideal case with zero error. Each box plot represents the test micrographs selected from that process interval. .... 155

Figure 7.2. Evolution of mean droplet characteristics a) area, b) perimeter, c) Feret, d) minFeret and e) droplet count from the emulsification process of the independent batch. The last box plot represents the 10 micrographs obtained from the discharge sample. .... 159

Figure 7.3. Box plots showing the Predicted RPT (minutes) of the micrographs plotted against their current process time. Each box represents 10 micrographs. .... 160

Figure 7.4. OOB error of the RF regression model plotted against the number of trees (*ntree*) ..... 163

Figure 7.5. Predicted RPT of the test data (from the 10 RF models) on the y-axis and Specified RPT on the x-axis. The additional x-axis shows the corresponding process time. The diagonal line represents the ideal case with zero error. Each box plot represents the test micrographs selected from that process time. .... 165

Figure 7.6. Predicted RPT plotted against current process time for the independent batch of 140 micrographs. Each box plot represents 10 micrographs. .... 167

Figure C.1. Box plots of average pixel intensity characteristics of droplets. a) integrated density b) raw integrated density. Each box plot represents the average of 10 micrographs. .... 196

## LIST OF TABLES

Table 3-1 Sampling protocol from Process-2.....	53
Table 3-2. Comparative Analysis of the EST and HBT Techniques.....	63
Table 4-1. Droplet characteristics obtained from the EST.....	76
Table 4-2. Median of droplet count obtained from 10 micrographs at each interval from the start of the process until 90 minutes. Orange cells represent the period of emulsification which presented an inverse variation in the count. Grey cells denote the stable droplet count.....	94
Table 4-3. Number of micrographs in each of the TAMU categories.....	97
Table 5-1. Correlation matrix of droplet size characteristics.....	101
Table 5-2. Eigenvectors obtained from the PCA for the five-variable space.....	101
Table 5-3. Proportion of total variance explained by each Principal Component (PC). The grey cells represent the significant PCs.....	102
Table 5-4. Processing times from which the TAMU micrographs for the 5-fold cross-validation were selected.....	106
Table 5-5. TAMU micrographs obtained from the independent batch of the emulsion.....	108
Table 5-6. Droplet characteristics selected as the predictor variables of the MLR model.....	109
Table 5-7. Number of TAMU micrographs from Process-2 assigned to training (70%) and test (30%) samples.....	109
Table 5-8. Results from the two tailed Z test of the MLR model trained using 70% of Process-2 micrographs.....	110
Table 5-9. Results from the 10-fold cross-validation of the MLR models.....	112
Table 5-10. Summary of the MLR model statistics obtained from the classification of the independent set of micrographs.....	114
Table 5-11. Random Forest model output showing OOB error and confusion matrix for the training micrographs.....	115
Table 5-12. Tuning results of the 10 folds for the <i>mtry</i> values, 1 to 5.....	117
Table 5-13. 10-fold cross-validation results of the RF models for <i>ntree</i> =400 and <i>mtry</i> =3.....	117
Table 5-14. Summary of the RF model statistics obtained from the classification of the independent set of micrographs.....	119
Table 5-15. The number of TAMU micrographs from Process-2 assigned to training (80%) and test (20%) samples.....	121
Table 5-16. VNN model tuning with the 80% training data.....	121
Table 5-17. The accuracy of the VNN models obtained from the 10-fold cross-validation.....	122
Table 5-18. Summary of the VNN model statistics obtained from the classification of the independent set of micrographs.....	124
Table 5-19. The number of training micrographs in the TAMU categories after splitting.....	126
Table 5-20. The rows represent the number of micrographs from each category in the data array. The response variable was encoded as either 0 or 1 based on which category the micrograph belonged to.....	127
Table 5-21. Classification results of the industrial emulsification process obtained from the models....	131
Table 6-1. Micrographs from Process-1 labelled as TAMU.....	137
Table 6-2. Classification results of the Analysts (A1 and A2) and the PC-LDA model. The orange cells represent the misclassifications.....	140
Table 6-3. Classification results of the Senior Analysts (SA1 and SA2) and the PC-LDA model. The orange cells represent the misclassifications.....	141
Table 6-4. Repeatability agreement within appraisers.....	143
Table 6-5. Accuracy Agreement of each Appraiser vs Standard.....	144
Table 6-6. Reproducibility agreement between Appraisers.....	146
Table 7-1. Predictor variables of the LR model.....	152
Table 7-2. LR Model Summary for the 70% training data.....	153
Table 7-3. ANOVA results from the LR model.....	154
Table 7-4. Summary of the median Predicted RPT compared with the Specified RPT values of the test predictions. The green cells represent reasonably good predictions and the white cells represent poor predictions.....	157
Table 7-5. Comparison of the Predicted RPT values of the independent set of micrographs with the Specified RPT values. The reasonably good predictions are highlighted in green. The white cells represent poor predictions.....	161

Table 7-6. Random Forest Regression Model with default *n*tree and *m*try values..... 163

Table 7-7. Random Forest Regression model summary with tuned *n*tree and *m*try values ..... 164

Table 7-8. Summary statistics of the Predicted RPT compared with the Specified RPT of the test data from the 10 RF models. The green cells represent reasonably good predictions and the white cells represent poor predictions. The values in bold shows the best predictions. .... 166

Table 7-9. Comparison of the Predicted RPT values of the independent set of micrographs with their Specified RPT values. The reasonably good predictions are highlighted in green..... 168

Table 7-10. RMSE and  $R^2$  values obtained for the LR and RF models for the test data predictions. .... 169

---

## Abstract

The existing quality evaluation of emulsions is typically based on subjective examination of samples under the microscope by trained analysts. The major drawbacks of such manual assessment include inter-observer variability, intra-observer variability, lack of speed, poor accuracy and it is also prone to over-processing. Other conventional droplet analysis techniques such as laser diffraction and spectroscopy, which require time-consuming sample preparation, have been verified as unreliable and introduce an additional complexity to industrial processes. In order to overcome these challenges, a novel automated approach based on image segmentation and machine learning is investigated in this research for the quality evaluation and optimisation of industrial emulsion processing.

Bright field micrographs were obtained during an industrial emulsification process. Two image segmentation techniques, Edge & Symmetry (EST) and Histogram-Based (HBT), were applied to detect the oil droplets from the micrographs. These techniques were also used to extract various morphological characteristics of the droplets. The most significant predictors were selected from these droplet characteristics for developing machine learning models. The most efficient image segmentation technique was also identified. The micrographs were grouped into four quality-based categories identified as TAMU (Target, Acceptable, Marginal and Unacceptable).

Supervised machine learning and deep learning models were developed for the TAMU classification of unknown emulsion micrographs. A comparative study was performed between manual and machine learning classification using Attribute Agreement Analysis. Regression models were developed to predict the RPT (Remaining Processing Time) required, at all stages of emulsification, to achieve the target characteristics. These prediction models were intended to avoid over-processing in emulsion manufacturing.

HBT exhibited excellent potential in droplet detection and characterisation compared to the EST approach. HBT was successful in detecting droplets with diameter as low as *ca.* 1  $\mu\text{m}$  from emulsion samples having dispersed phase fraction  $\approx 50\%$ . The machine learning classification models presented high accuracies ranging from 92% to 100%. The deep learning models demonstrated lower accuracies from 44% to 89%. The results of the comparative analysis showed that the machine learning classification is superior to manual classification with respect to speed (180 times faster), greater accuracy (10% to 40%) and repeatability. The prediction models presented an adjusted  $R^2 \approx 92\%$ .

The entire automated approach based on image segmentation and machine learning was implemented as a soft sensor. The soft sensor supports the real-time deployment of the technique into an industrial environment. The proposed approach has the potential to predict instantaneous product quality as well as the process time required to achieve the desirable droplet characteristics. This will avoid over-processing and wastage of resources leading to more efficient and sustainable emulsion manufacturing.

---

## Abbreviations

EST	Edge & Symmetry Technique
HBT	Histogram-Based Technique
TAMU	Target, Acceptable, Marginal and Unacceptable
Feret	Maximum Feret diameter
minFeret	Minimum Feret diameter
rID	Raw integrated density
ID	Integrated density
PC-LDA	Principal Component based Linear Discriminant Analysis
MLR	Multinomial Logistic Regression
RF	Random Forest
VNN	Vanilla Neural Network
CNN	Convolutional Neural Network
R&R	Repeatability and Reproducibility
AAA	Attribute Agreement Analysis
MSA	Measurement Systems Analysis
LR	Linear Regression
RPT	Remaining Processing Time
AI	Artificial Intelligence

---

## Chapter 1

### Introduction

Emulsification is the process of breaking large oil globules into a homogenous distribution of microscopic droplets. Emulsions can be either two-phase (single emulsions) or multiphase (double emulsions) dispersions. A two-phase oil in water (o/w) emulsion is formed when oil becomes homogeneously dispersed as tiny droplets in water. This can be achieved by high shear homogenisation, where oil forms the dispersed phase and water constitutes the continuous phase of the emulsion (Tontul & Topuz, 2015). Shear stress is defined as the component of stress that acts parallel to a surface or a material cross section. Emulsions include a wide range of food, pharmaceutical and cosmetic products such as skin care creams, lotions, hair products, vitamin syrups, medicines *etc.* (Bakry et al., 2016; Dao et al., 2018; Rodriguez et al., 2016). They are applied as delivery vehicles for drugs, pesticides and are also extensively used in emulsion based paints (Cho et al., 2015; Dokania & Joshi, 2015; George et al., 2018; A. K. Singh et al., 2009; Tasker et al., 2018; Vasconcelos et al., 2018).

Quality evaluation, in emulsion manufacturing, has been identified as a challenging task by food, pharmaceutical and chemical industries. This can be attributed to the increased time consumption, subjectivity and inconsistency of existing evaluation techniques. The quality of the final emulsion product is highly dependent on its droplet size distribution, which is in turn influenced by the operating conditions and process parameters such as time, temperature *etc.* (Amokrane et al., 2016; Qu et al., 2018; Zeaiter et al., 2006). Manual evaluation of emulsion samples, by microscope analysts, through physical sampling and observing the samples under the microscope is one of the techniques currently practiced in industries. The product can be over-processed, in most cases, due to the lack of an objective quality evaluation and decision-making procedure.

---

Over-processing is one of the potential challenges in high energy emulsification as it is wasteful of energy and can cause an increase in droplet size (Tontul and Topuz, 2015). This is due to several reasons including slow adsorption rate of the surface-active agent used for binding the droplets, residence time of emulsification, high rate of droplet coalescence and a high energy density (Jafari *et al.* 2007).

Other conventional techniques employed in the quality evaluation of emulsions include laser-based scattering and diffraction methods, which are found to produce unreliable droplet size measurements and require time consuming sample preparation (Abidin *et al.* 2013, Maaß *et al.* 2012, Honkanen *et al.* 2010, Greaves *et al.* 2008). There is a wide range of literature available stating the challenges and drawbacks of the current techniques in emulsion quality evaluation, which are discussed later in this chapter. The existing offline techniques used for emulsion characterisation and quality evaluation are detailed in Section 1.3. These include manual evaluation, spectroscopic methods and computer-aided image analysis techniques. Section 1.4 describes the existing inline droplet size monitoring techniques in various multiphase systems. The image processing software, which have been used previously for both offline and inline droplet detection, is also detailed in these sections.

## **1.1 Scope of this research**

This research addresses the current issue of subjective, time consuming and unreliable quality evaluation practiced in emulsion manufacturing industries. Soft sensors, for automated quality evaluation and process control, are one way of advancing the current system (Sun *et al.*, 2014). Soft sensors are inferential models developed to predict process variables (both categorical and continuous) using measured variables. These soft sensors can be integrated with emulsion processing, both offline and inline, to develop enhanced



---

intelligent systems. The scope of this research is focused on the image processing of optical micrographs (image taken from a microscope) as a technique, for emulsion characterisation, which is followed by automated emulsion classification and the prediction of remaining process time. In-process samples of a topical cream emulsion product, acquired from a pharmaceutical manufacturing industry, has been used in this study. An offline implementation of the automated techniques, as a soft sensor, has been validated. A future extension of this work is planned for the inline quality evaluation and optimisation of industrial emulsification processes.

## **1.2 Aims and Objectives**

The aim of this research is to develop a novel automated technique, for droplet characterisation and quality evaluation, to overcome the current challenges faced by emulsion manufacturing industries. The new technique is targeted on characterising the oil droplets efficiently during emulsification, classifying in-process emulsion samples into quality categories and identifying the optimum processing time. Finally, inferential models will be employed to predict the anticipated extra processing time at all stages of the emulsification process. The objectives of this research are focused on achieving this aim.

### **1.2.1 Objectives**

The following are the objectives of the project:

- Conduct a literature review of:
  - The existing droplet detection, characterisation and quality evaluation techniques in emulsion manufacturing to identify the current challenges.
  - The state-of-the-art machine learning techniques applied for vision integrated automated quality evaluation in manufacturing industries including food, pharmaceutical and chemical.

- 
- Develop an efficient image segmentation approach suitable for the automated characterisation of droplets from in-process emulsion micrographs.
  - Identify the desirable droplet quality characteristics through statistical analysis and establish the completion time of the emulsification process, *i.e.* when the droplets attain their target characteristics.
  - Categorise and label the emulsion samples, at specific processing stages, based on their droplet quality characteristics and industrial expert advice.
  - Develop supervised machine learning models to classify unknown emulsion samples/micrographs into the labelled categories.
  - Evaluate manual versus machine learning classification approaches.
  - Develop prediction models to predict the processing time required by the emulsion, at all stages of emulsification, to achieve the target droplet characteristics.
  - Validate the developed techniques using industrial data.

### 1.2.2 Potential Industrial Impact

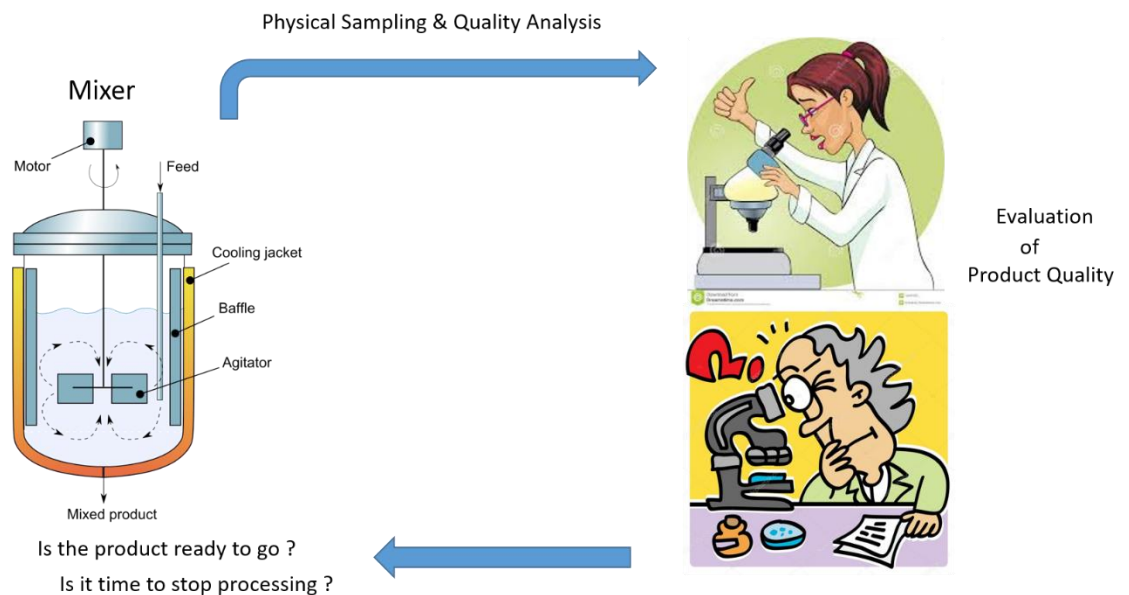
The proposed approach has significant potential impact in a wide range of industries including food, pharmaceutical, biomedical and chemical industries. The current requirements demanded by Industry 4.0, such as digitising traditional industries for improving process and product evaluation techniques, could be addressed by the new automated approach. Minimum human intervention, with increased machine to machine communication is the potential target of the proposed approach. Moreover, the current techniques used in this area demand significant use of personnel resources, time and energy, resulting in increased cost of production. The proposed automated approach, in this research, will have the potential to overcome these challenging demands leading to more efficient and sustainable emulsion manufacturing.

---

### 1.3 Overview of existing techniques for emulsion quality evaluation

#### 1.3.1 Manual Evaluation

Manual evaluation of in-process samples is one of the techniques currently employed in the industrial quality evaluation of emulsions (Figure 1.1). During the emulsification process, samples are taken from the stirring vessel (mixer). Analysts observe these samples under the microscope to evaluate the quality of the product and confirm if it is fully processed and meets the desirable characteristics. This is done in different ways, including manual counting of the oil droplets and by subjectively judging the distribution of the droplets, based on previous knowledge.



**Figure 1.1. Schematic representation of the manual quality evaluation of emulsions.**

Figure 1.1 shows the manual microscopic evaluation of in-process emulsion samples, performed by microscope analysts, to determine the completion point of the emulsification process. Such evaluation techniques have been found highly subjective, erroneous and time consuming in previous studies (Boxall et al., 2010; Gwyn et al., 1965; Maaß et al., 2012). Boxall *et al.* (Boxall et al., 2010) have reported an average difference of 5.1% in the mean droplet size measurement between two analysts. In a similar study,

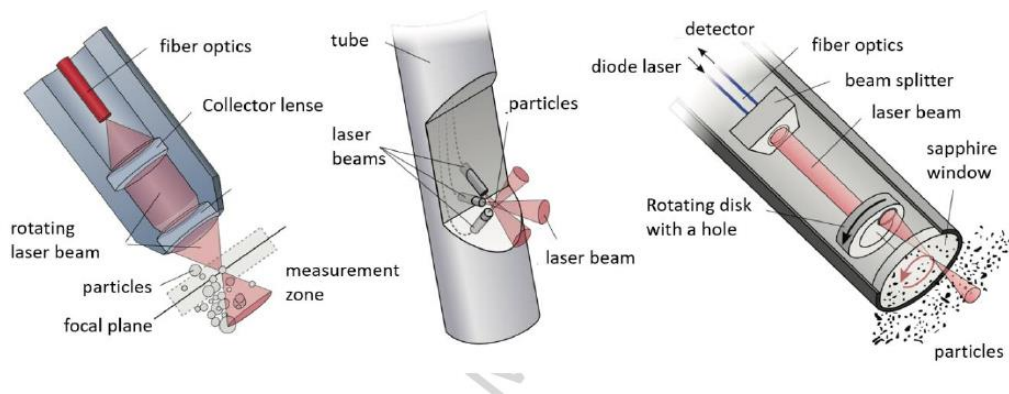
---

(Maaß et al., 2012) have found significant inter-analyst and intra-analyst errors in the droplet size and count measurement. Their study reported a  $\pm 5\%$  and  $\pm 15\%$  deviation between four analysts in the measurement of average droplet size and count respectively. In repeating the analysis with the same analysts, the deviation in the measurements almost doubled. To resolve this problem of subjectivity in manual microscopic evaluation, the previous studies have suggested the use of automated image processing techniques for droplet size monitoring.

### 1.3.2 Other Conventional Methods

Vibrational spectroscopic techniques such as mid-infrared (MIR), near infrared (NIR) and Raman spectroscopy using laser diffraction have been extensively applied in the quality evaluation of food emulsions (Berrueta et al., 2007; Kljusuric et al., 2015; P. Wang et al., 2016). These techniques have been employed mainly for the analysis of emulsion composition and to investigate areas such as adulteration in olive oil, process optimisation in barley milk production *etc.* Laser diffraction techniques have demonstrated proficiency in the development of chemometric analysis for emulsion quality evaluation (Bosque-Sendra et al., 2012; Geladi, 2003). One of the major challenges associated with such spectroscopic techniques is the requirement to dilute samples to achieve a good spectra (P. Wang et al., 2016). This introduces additional complexity in industrial processes and ensures such techniques are difficult to automate. Laser scattering had been the most widely used technique for droplet size analysis, apart from chemometric analysis, in emulsions from the 1990s (Heffels et al., 1998; Sachweh et al., 1998). In laser diffraction, the particle/droplet size is measured in terms of the angular variation in the intensity of light scattered when a laser beam passes through a dispersion sample (P. Wang et al., 2016; Y. Wang et al., 2004).

FBRM (Focused Beam Reflectance Measurement) is one of the most popular laser back scattering techniques applied for both particle and droplet size measurements (Emmerich et al., 2018). The working principle of droplet/particle size measurement using FBRM is shown in Figure 1.2. The two-dimensional optical reflectance measurement (2D-ORM) sensor (Figure 1.2, left) detects the droplets near the FBRM probe window through the reflection of an intense laser beam. The Forward-Backward Ratio (FBR) sensor (Figure 1.2, centre) measures the spatial pattern of the laser beam scattered in both directions for smaller particles/droplets with radius less than one tenth of the beam wavelength. The average diameter of a group of particles/droplets is measured from the measured light intensity ratio. The laser beam from the cylindrical probe rotates at high speed, focusing on the particles/droplets close to the sapphire window (Figure 1.2, right), and the light is scattered back by the particles/droplets. The particle/droplet sizes (chord length) are calculated as the product of the rotational speed and the sampling time.



**Figure 1.2. Focused beam reflectance method (FBRM) for droplet size distribution analysis: (left) 2D-ORM sensor, (centre) FBR sensor and (right) working principle of the FBRM probe (Emmerich et al., 2018).**

These laser based techniques have many advantages including the possibility of analysing a wide range of particles from nanometres to micrometres, rapid measurement potential and high sample throughput. However, there are certain limitations associated with such techniques, which are discussed by (Vankeirsbilck et al., 2002). These are time-

---

consuming sample preparation, sample dilution and the influence of the particle surface on the quality of the scattered light. Laser techniques were found incapable of delivering reliable droplet size measurements in the past (Abidin et al., 2013). Many authors have obtained unsatisfactory results from analysing spherical droplets in dispersion systems using laser scattering (Greaves et al., 2008; Honkanen et al., 2010; Maaß et al., 2012). Regarding smaller droplets of diameter  $< 1 \mu\text{m}$ , scattered light creates additional noise which limits the spectral resolution (Schuster *et al.* 2012). Recent studies in the literature, regarding NIR and laser diffraction techniques, have found that image processing and analysis of emulsion micrographs can give better insight into droplet size measurements compared to these techniques (Kljusuric *et al.* 2015, Abidin et al., 2013).

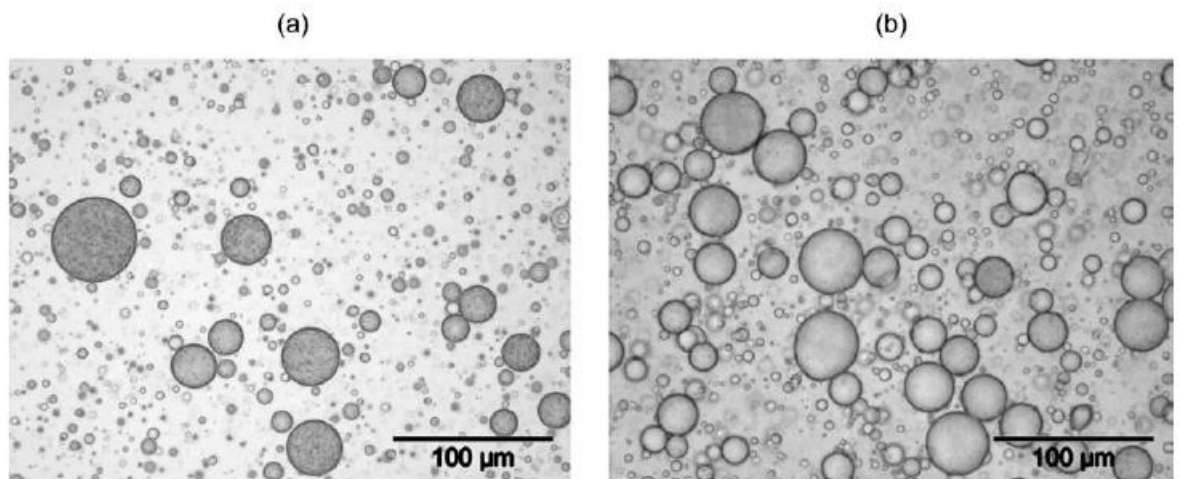
### 1.3.3 Offline Image Processing techniques for Droplet Detection

Emulsion stability studies have been performed using light microscopy in conjunction with image processing and statistical analysis from the early 21<sup>st</sup> century (Freire et al., 2005; Hosseini et al., 2015; Silva et al., 2010). These studies have identified image processing of optical micrographs as an efficient and cost effective method for the analysis of various aspects of emulsion characterisation. Freire *et al.* (2005) have analysed the aging mechanism of perfluorocarbon emulsions using image analysis over a 42 day storage period and identified temperature as a key factor influencing the evolution of the mean droplet size.

Similar studies of droplet size distributions have also been performed for the characterisation of double emulsions (Scherze et al., 2005; Schuster et al., 2012a). (Scherze et al., 2005) developed an automated image analysis methodology for optical micrographs, in the industrial preparation of double emulsions, to control the yield of the inner continuous phase. The emulsion micrographs were obtained using extended focal

---

imaging, a special module contained in the analysis software (Soft Imaging System) of the optical microscope they used (Olympus BX61). Extended focal imaging has the ability to capture images at diverse focus settings. The droplets of the external and internal phases of the emulsion were detected from the micrographs with the help of a macro, programmed using Optimas (Media Cybernetics), and the droplet characteristics were obtained (Scherze et al., 2005). The oil droplet size distribution and the average pixel intensity of each droplet were statistically analysed. Due to the larger number and smaller sizes  $< 1.5 \mu\text{m}$ , it was difficult to segment the droplets in the internal phase, as they appeared as texture. Therefore, the droplets in the internal phase were analysed based on their roughness and brightness in terms of pixel intensity. The relative brightness of the droplets was calculated as the difference between average pixel intensity of the droplets, weighted by area, and the average pixel intensity of the background. The decrease in the relative brightness of the droplets was studied over a storage period of 28 days. This is shown in Figure 1.3. The statistical evaluation of the relative brightness was used to determine the required level of filling of oil droplets in the inner aqueous phase. The results of the study showed strong positive correlation between relative brightness and filling level of the droplets. A controlled filling of droplets was achieved to obtain optimum yield of the inner phase of the emulsion (Scherze et al., 2005). Their study also found that the textural quality of the microscopic images can be improved by controlling the process parameters and the formulation of the emulsion.



**Figure 1.3. Emulsion sample micrograph a) after 8 days b) after 28 days (Scherze et al., 2005).**

The potential of optical imaging techniques such as Confocal Laser Scanning Microscopy (CLSM) combined with statistical analysis of droplet size data was investigated in both single and double emulsions by Schuster *et al.* (2012). The CLSM imaging technique had previously proven effective in the analysis of single emulsion systems (Blonk & Van Aalst, 1993; Van Dalen, 2002). The study performed by Schuster *et al.* (2012) addressed the question of how accurately the droplet size distribution (DSD) of food emulsions can be obtained from CLSM images with minimum error in the processing algorithms and the application of error correction methods. The images obtained from CLSM were processed using ImageJ software and the distribution of the volume weighted droplet diameter ( $d_{43}$ ) was analysed ("<https://imagej.net/Welcome>"). Their image processing methodology involved the following steps.

1. The images were thresholded through the red channel.
2. Converted into black and white binary images.
3. Noise removal was applied using the ‘despeckle’ function.
4. ‘Watershed’ segmentation was applied to recreate the border between inner and outer phases and also to separate overlapping droplets.



---

Finally, the droplets with circularity ranging from 0.7 to 1.00 were identified. The circularity of a droplet is  $4\pi$  multiplied by the ratio of the droplet area to the square of its perimeter. A droplet with a circularity of 1.00 is a perfect circle. Schuster *et al.* (2012) suggested analysing the maximum Feret diameter of the droplets, in addition to droplet area, in order to minimise the optical frame error. This minimises the error created by omitting larger droplets, with maximum Feret diameter less than or equal to the optical frame length, appearing at the edges of the optical frame. The maximum Feret diameter is the longest distance between any two points along the selected droplet boundary and is applicable to non-circular shapes. Their study obtained improved DSD results using image processing and raised the future possibility of integrating automated image processing techniques with statistical analysis for the investigation of emulsion agglomerates (Schuster *et al.* , 2012).

In a similar study conducted by Hosseini *et al.* (2015), droplet diameter measurements were obtained by processing emulsion micrographs, using ImageJ software, which was followed by statistical DSD analysis. The image processing and statistical analysis were performed in three major steps:

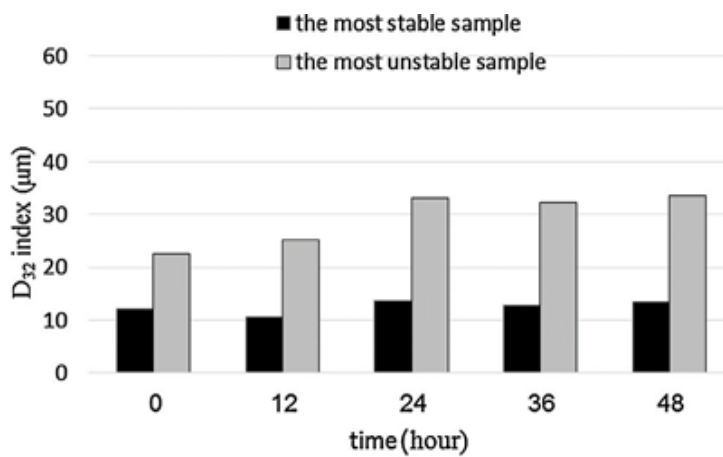
1. The image was converted into binary (black and white) and noise was reduced using the ‘despeckle’ function.
2. The identified droplets were analysed, in the image, to obtain their characteristics.
3. The droplet characteristics such as average droplet size and standard deviation were statistically analysed.

The average droplet size was measured in terms of the surface area weighted mean diameter ( $d_{32}$ ), often referred to as Sauter mean diameter as given in Equation 1.1. The Sauter mean diameter is a different way of examining the average droplet size, which

indicates the centre point of the frequency distribution in terms of the surface area and not the number of the droplets. Their study of the evolution of average droplet size over time for the most stable and the most unstable emulsion samples is shown in Figure 1.4.

$$d_{32} = \frac{\sum_{i=1}^N n_i d_i^3}{\sum_{i=1}^N n_i d_i^2} \quad \text{Equation 1-1}$$

$N$  is the total droplet count in a micrograph and  $n_i$  is the number of droplets with diameter,  $d_i$ .



**Figure 1.4.** Changes in the  $d_{32}$  index for the most stable and unstable O/W emulsion samples after 48 hours storage (Hosseini *et al.* 2015).

The emulsion samples were prepared with different emulsifier concentrations and were stored at different temperatures. The most unstable emulsion had an increasing average droplet size over time, while the most stable emulsion showed an initial decrease in the average droplet size and then remained steady. The emulsion stability study by (Hosseini *et al.*, 2015) demonstrated that the analysis of the  $d_{32}$  index is effective in finding the optimum parameters of emulsion preparation such as homogenisation time and storage temperature.

---

Freire *et al.* (2005) and Silva *et al.* (2010) applied a set of image processing techniques, in ImageJ, to obtain oil droplet characteristics from the optical micrographs of O/W emulsions as described below.

1. The original RGB image was subtracted from its background and converted into 8 bit greyscale.
2. The image was then thresholded and converted to binary.
3. Noise correction was performed using median filtering.
4. Erosion was applied to suppress the structures connected to the image border.
5. Morphological analysis was done by creating a mask over the identified droplets.

Droplet size parameters such as diameter, area, volume and standard deviation were determined and statistical analysis of the parameters was conducted. The droplet size distribution was analysed over a period of 42 days. The results showed an increase in the average droplet size of the emulsions due to coalescence over time, with an increase in the storage temperature (Freire *et al.*, 2005).

The image processing techniques, using Imagej, reported in the literature for emulsion droplet analysis have commonly applied a general processing sequence. This is based on noise filtering, thresholding, converting to binary and separating the overlapped droplets using watershed segmentation or erosion. This is followed by analysing the droplet characteristics such as size and shape (Freire *et al.*, 2005; Hosseini *et al.*, 2015; Silva *et al.*, 2010). Other image processing techniques applied for droplet detection from emulsion micrographs include Hough circle transform and edge detection algorithms based on Matlab (Maaß *et al.*, 2012; Panckow *et al.*, 2017). Maaß *et al.* (2012) used a circular edge detection algorithm for droplet size and count characterisation. Their

---

technique comprised of three steps for pattern identification. These include the correlation of pre-filtered images with search patterns, the pre-selection of possible circle coordinates in the images followed by the classification of the droplets in each image by an edge detection algorithm. They recommended a future extension of their method for detecting non-circular droplets. In summary, the robustness of the image processing technique plays a major role in the accuracy of droplet detection in dispersion systems. From the extensive literature review conducted as part of this research, there have been no studies reported to date, which have applied an intensity-based droplet detection approach for in-process emulsion micrographs. The literature strongly suggests that border/edge based techniques are more typically used for droplet detection.

#### 1.3.4 Image Segmentation

Image segmentation is the primary step involved in any image analysis procedure. The goal of image segmentation is to partition the Region Of Interest (ROI) in an image into meaningful objects/segments (Aly et al., 2011). The major applications of image segmentation include medical imaging, object detection and recognition and image classification. There are two basic types of image segmentation such as global segmentation and local segmentation (Anjna & Er, 2017). Global segmentation is useful if the ROI is spread across the whole image and the latter one is applied if the ROI is a specific part/region of the image.

Image segmentation approaches are generally classified into two major categories (Anjna & Er, 2017). These are the discontinuity detection-based approach and similarity detection-based approach. The discontinuity approach partitions the ROI of an image based on pixel intensity gradient, which detects the separation between the edges of the ROI from the background. Edge detection is the most widely used technique for the discontinuity approach (Kumar et al., 2012). Similarity detection is aimed at identifying

---

similar regions of an image using Thresholding or Histogram-based techniques, in which identical pixel values are grouped together (Haralick & Shapiro, 1985). This is also known as a region-based approach. The selection of the most appropriate image segmentation approach depends on the properties of the image and the ROI (Bora & Gupta, 2014; P. Singh & Chadha, 2013). Thresholding techniques are the most suitable for image segmentation if the ROI has indistinct edges, while edge detection techniques work well with distinguishable edges.

#### **1.4 Inline Droplet Size Monitoring**

A precise understanding of the droplet size distribution of emulsions has been identified as the key factor to control and optimise industrial processes in various fields of applications (Panckow et al., 2017). The potential of inline droplet size monitoring, using automated image processing followed by statistical analysis, has been investigated in various multiphase systems (Boxall et al., 2010; Khalil et al., 2010; Maaß et al., 2012; Pacek, Moore, et al., 1994; Pacek, Nienow, et al., 1994). These inline techniques have been employed to identify the process parameters that provide the optimal target characteristics (Crawley & Malcolmson, 2004; Khalil et al., 2010; Maaß et al., 2012).

The aim of (Khalil et al., 2010) was to investigate the efficiency of coupling imaging software with inline hardware devices to evaluate the evolution of droplet size in an emulsification process. Their experimental fixture for *in-situ* process monitoring involved a video microscope probe submerged in a laboratory reactor as shown in Figure 1.5 (Khalil et al., 2010). A pulse generator was triggered by the video camera, which in turn sets the back lighting supplied by Light Emitting Diode (LED). The back lighting resulted in capturing droplet shadows as black structures, appearing on a white background. The recorded video was retrieved by a video grabber and sent to a computer.

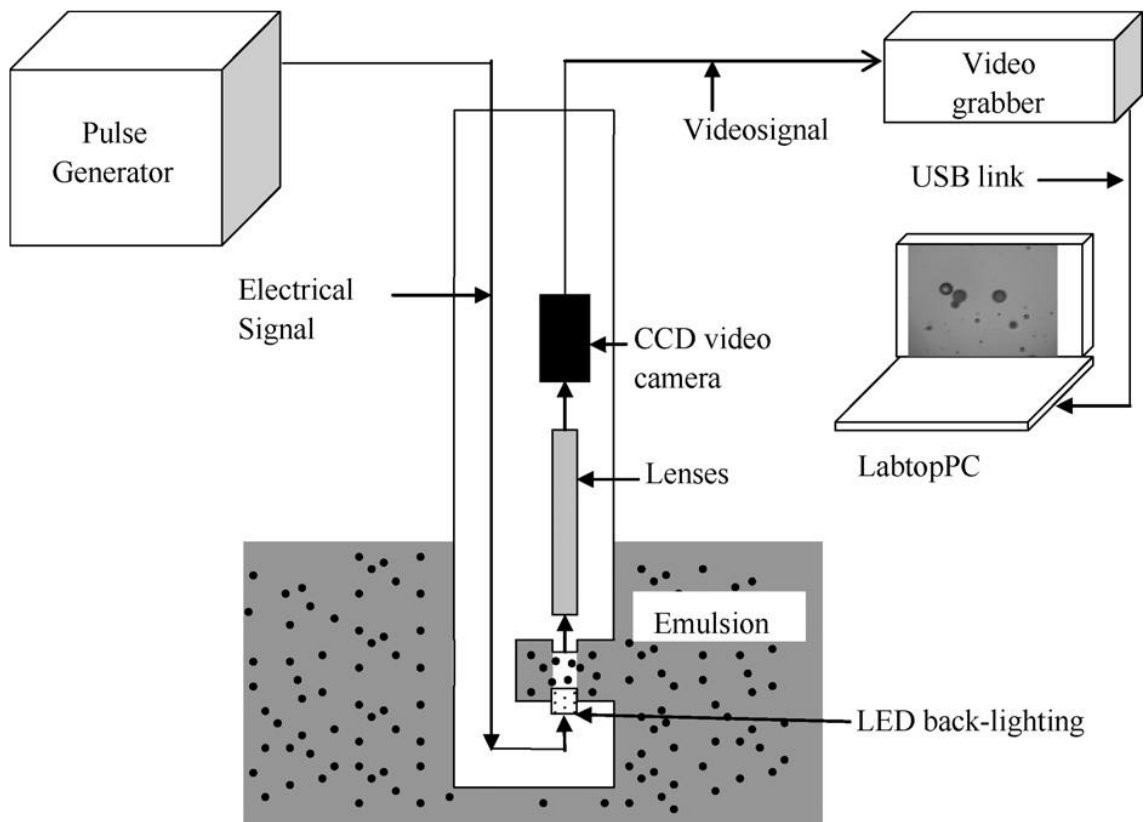
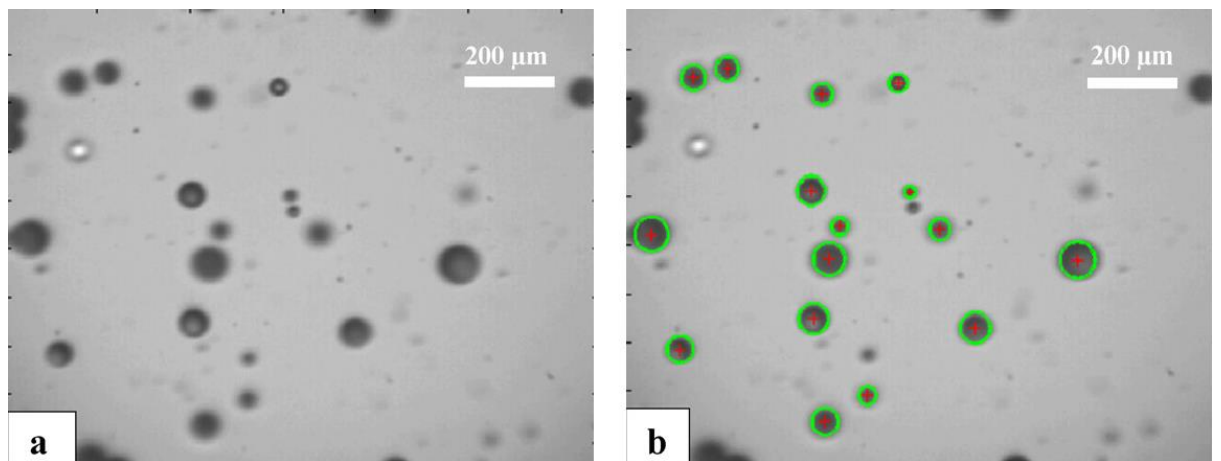


Figure 1.5 - Experimental set-up for *in-situ* droplet size monitoring (Khalil et al., 2010).

The image acquisition was performed in various steps. A 30 second video sequence was recorded periodically, and 300 frames were selected from each recording for automatic image processing. In order to eliminate duplicate detection of droplets, one out of five frames were selected for analysis and a circular Hough transformation method was applied to detect the circular droplets in the images (Figure 1.6). The droplet size range was limited to 10-90  $\mu\text{m}$ , since the accuracy of detection was weak below 10  $\mu\text{m}$ , due to a resolution of 2  $\mu\text{m}$  per pixel and was found inaccurate above 90  $\mu\text{m}$ , due to the accumulated detection of several droplets at the same time. A set of image processing parameters were required for detection including the maximum and minimum radii of the droplets in a frame. The parameters were tuned by comparing images acquired from three

---

different runs of video sequences and three sets of parameters were obtained. The mean parameter values were chosen as the inputs for the automated image analysis. The study by (Khalil et al., 2010) demonstrated that it is possible to optimise an emulsification process in a laboratory reactor using an *in-situ* experimental set-up with software and hardware devices coupled with automated image analysis software. However, their study was limited to the detection of circular droplets within a diameter range of 10  $\mu\text{m}$  - 90  $\mu\text{m}$ .



**Figure 1.6. Droplet detection using Hough circle transform: a) raw micrograph b) processed image with detected droplets (Khalil et al., 2010).**

A similar study was performed by (Maaß et al., 2012) to overcome the errors caused by manual droplet counting, physical sampling and the usage of inefficient image processing techniques. It has been found that even minute changes in the sampling time can result in substantial droplet size measurement errors. Maaß *et al.* (2012) implemented an automated image processing and droplet detection technique based on edge detection in MATLAB®. This algorithm worked in two stages, the first stage detected the edges of the droplets and created an output image, which was used in the second stage to measure the circular droplets. The proposed algorithm by (Maaß et al., 2012) was found to be faster and more efficient in droplet detection and analysis compared to human analysis.

---

Their study recommended the optimisation of the image processing parameters used for the droplet detection in order to improve the speed and accuracy of the computing software. The parameter optimisation was attained by performing several runs of the image processing software to extract the characteristics of a few manually marked droplets in a number of images. Their study also found that automated micrograph analysis of droplet counting, and characterisation is approximately fifty times faster and less erroneous compared to manual assessment.

*In-situ* droplet size measurements, in biofuel production systems, using photo optical techniques have been performed using similar circular edge detection techniques (Panckow et al., 2017). Their study detected circular droplets from O/W emulsions with a dispersed phase fraction of 1% to 5% and  $d_{32}$  ranging from 70 to 110  $\mu\text{m}$ . The detection of droplets in production systems with an expected mean droplet size  $< 10 \mu\text{m}$  is still found challenging in the existing studies. In addition, droplet detection in extremely concentrated emulsions, with a dispersed phase fraction greater than 10% to 15%, using existing edge detection based techniques, was also identified as a challenge in the literature (Brás et al., 2009; Maaß et al., 2012; Panckow et al., 2017).

### **1.5 Current Challenges in droplet detection from emulsion micrographs**

The existing image processing studies in droplet characterisation and emulsion quality evaluation have identified the following areas as current challenges.

- Detection of droplets from highly concentrated emulsions with phase fraction  $>15\%$  (Brás *et al.* 2009, Maaß *et al.* 2012).
- Detection of smaller droplets  $<10 \mu\text{m}$  from production systems (Panckow *et al.* 2017).



- 
- Detection of non-circular droplets from emulsion micrographs (Maaß *et al.* 2012).
  - Implementation of automated droplet detection techniques with minimum user input parameters (Khalil *et al.* 2010, Maaß *et al.* 2012).
    - To eliminate the parameter optimisation step.
    - To reduce the error induced by the variability in the parameter values.
    - To enable automation of the system.

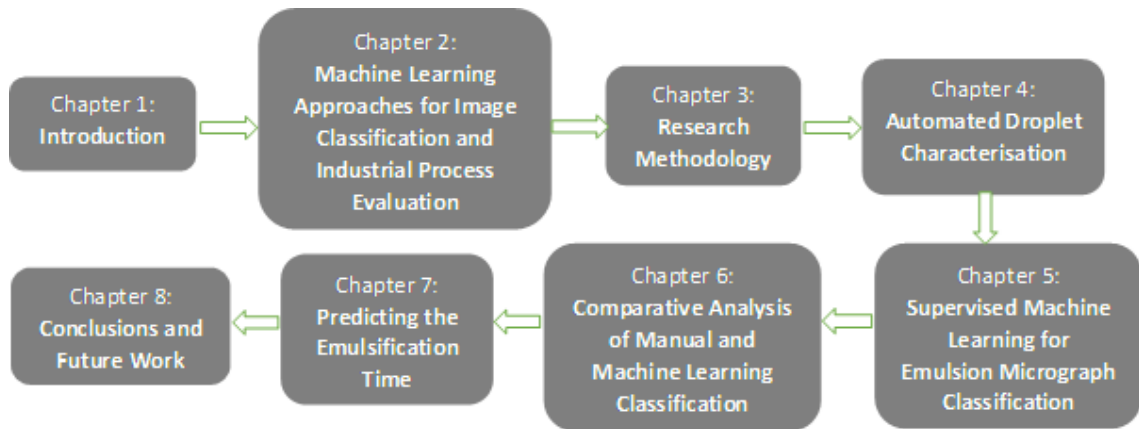
With respect to the current challenges, a novel automated approach is necessary to evaluate the quality of an emulsion product effectively during the emulsification process. This could potentially avoid over-processing, leading to efficient utilisation of resources such as time, energy and raw materials. Such an approach could also minimise manual intervention in the quality assessment procedure in manufacturing industries. However, such techniques have not been developed to overcome the present challenges within this domain.

## **1.6 Thesis Outline**

The development of a soft sensor, for the quality evaluation and optimisation of industrial emulsion processing, is identified as the major objective of this research. Inferential models developed from droplet characteristics of in-process micrographs are one way of addressing this objective. This thesis outlines the approach adopted to develop the soft sensor for improved emulsion processing. The development of an automated image segmentation technique capable of addressing the existing difficulties in emulsion droplet detection is identified as the initial focus of this research. The new image segmentation technique is aimed at the automated detection of droplets from emulsion micrographs during its industrial processing. This technique is also expected to extract various size, shape, centroid and orientation characteristics of each individual droplet in a micrograph. The droplet characteristics form the input features of the inferential models. The

---

inferential models are developed using machine learning algorithms for two main purposes, such as supervised classification of in-process emulsion samples and prediction of the processing time required for the completion of emulsification. Structure of this research thesis is presented in Figure 1.7.



**Figure 1.7. Schematic representation of the thesis structure.**

## **Chapter 1**

Chapter 1 provides a brief introduction to emulsions and the emulsification process. It reviews the techniques, which are currently employed, for the evaluation of emulsion quality and discusses their challenges. This chapter also describes the scope, the potential industrial impact and the aims and objectives of this research.

## **Chapter 2**

An extensive literature review is conducted on the state-of-the-art machine learning techniques applied for image classification and industrial process evaluation. It discusses the various unsupervised and supervised learning techniques. This chapter also identifies the machine learning models suitable for a proposed novel soft sensor approach.

---

### **Chapter 3**

This chapter provides an overview of the approach undertaken to develop a novel technique for automated droplet characterisation, from micrographs, followed by emulsion process evaluation. It illustrates the methodologies developed for the following:

- In-process micrograph acquisition.
- Image processing and droplet detection.
- Statistical analysis of droplet characteristics.
- Supervised machine learning classification of micrographs.
- Remaining processing time prediction.

### **Chapter 4**

This chapter discusses the results obtained from the image processing and droplet characterisation techniques. Two different droplet detection techniques, an edge-based and an intensity-based method, are developed and the detected droplet characteristics are compared using statistical analysis. The best image segmentation technique is identified for developing the inferential models. The statistical analysis of the droplet characteristics is also completed for selecting the most significant model predictors.

### **Chapter 5**

This chapter presents the machine learning classification models and their results. It initially focuses on unsupervised machine learning for dimension reduction and pattern identification followed by supervised classification. The supervised classification study is aimed at discriminating the droplet characteristics from ‘Unacceptable’ to ‘Target’ categories of emulsion processing. Once the ‘Target’ characteristics are achieved, the

---

product is expected to be stable and therefore, any further processing is identified as over-processing. The classification models are evaluated using cross-validation approaches and also using micrographs from independent emulsion batches.

## **Chapter 6**

A comparative analysis between manual and automated classification of micrographs is conducted to evaluate the accuracy and precision of both techniques. Micrograph analysts from the industrial partner in this research performed the manual classification. A machine learning model is used to perform the automated classification. The results of the comparison are discussed.

## **Chapter 7**

This chapter presents the regression models developed to predict the processing time required, during emulsification, to achieve the target characteristics. These prediction models are intended to avoid over-processing in emulsion manufacturing. The results obtained from the prediction models are compared and discussed.

## **Chapter 8**

This chapter presents a discussion of the conclusions derived from Chapters 4 to 7. The conclusions derived from each chapter are related back to the initial objectives of this research. It also discusses the future possibility of extending this work into other domains and reflects ideas to overcome the limitations encountered in this study. The potential for real-time integration is then finally discussed.

---

## 1.7 Conclusions

The aim of this research is to develop a novel automated technique that can potentially replace the existing traditional and automated techniques in emulsion quality evaluation. This chapter introduced the process of industrial emulsification. It also reviewed the manual and other conventional quality evaluation techniques currently deployed in industrial emulsification processes and their limitations. Potential over-processing of emulsions due to subjectivity and inconsistency of the existing traditional techniques are identified as the major challenges. The existing offline and inline image processing techniques and the associated challenges are also discussed in this chapter. The development of a novel soft sensor approach for improved droplet detection, characterisation and process evaluation in emulsion production systems is identified as a major requirement to overcome the challenges.

The objectives of this research are defined to meet Industry 4.0 requirements to develop a fast, accurate, objective and sustainable evaluation technique with minimal manual intervention. This chapter also outlined the potential impact of the proposed approach in industrial emulsion manufacturing. Finally, an outline of the thesis structure is provided with a brief introduction to the individual chapters. The following chapter details an extensive literature review of the machine learning techniques, which could be applied for the automated approach proposed in this research.

---

## Chapter 2

### Machine Learning Approaches for Image Classification and Industrial Process Evaluation

#### 2.1 Introduction

This chapter presents a detailed literature review of the various supervised and unsupervised machine learning techniques used for image classification and regression applications. A review of the existing studies, which have applied similar techniques integrated with image processing for quality evaluation in process industries, is also presented in this chapter (section 2.3). This is followed by a review summary and overview of the methodologies applied to meet the objectives of this research (section 2.4).

#### 2.2 Machine Learning

Machine learning is an application of Artificial Intelligence (AI) that builds algorithms with the ability to learn from data. These algorithms are non-linear statistical models, which can detect complex relationships between the dependent and independent variables in the data (Friedman *et al.* 2001). The dependent variable is called the response variable and the independent variables are referred to as features or predictors. Machine learning techniques are classified into unsupervised and supervised techniques based on how they learn. Unsupervised methods, which do not presume any previous knowledge of the data, provide an unbiased interpretation of the data and therefore, are most commonly used for exploratory data analysis and pattern recognition (Boutros & Okey, 2005). Supervised techniques, on the other hand, use trained data models, which predict the membership of an unknown sample based on *a priori* knowledge and offer powerful classification models (Berrueta *et al.*, 2007). Supervised techniques have a risk of overfitting, i.e., fitting the

---

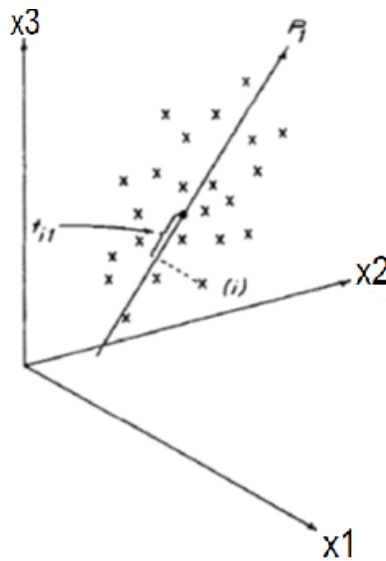
data perfectly even if there is no significant relationship in the data. It is essential to validate the models using unseen data, i.e., data independent of the training set.

The following subsections (2.2.1 to 2.2.6) present a detailed review of the unsupervised and supervised machine learning/deep learning techniques suitable for the classification and regression problems encountered in this research.

### 2.2.1 Principal Component Analysis

Principal Component Analysis (PCA), an unsupervised pattern recognition and dimension reduction technique, has been extensively applied in the field of computer vision as an efficient method for image data representation and classification from the early 2000s (Grané & Jach, 2014; Günther et al., 2018; Yang & Wu, 2006). PCA is a multivariate data projection method developed by (Pearson, 1901). It helps to reduce the dimensionality and the correlation of a multivariate feature space. This is achieved by projecting the variation in the original multivariate data set across an equal number of uncorrelated components onto an orthogonal subspace (Wold et al., 1987). For example, if PCA is applied on a multivariate feature space with  $p$  correlated variables, the resultant vector will be a set of  $p$  principal components  $(z_1, z_2, \dots, z_p)$  with zero correlation. The Principal Components (PCs) are successively calculated by computing the eigenvalues and eigenvectors based on the variance-covariance matrix of the original data (Jackson, 1980, 2003). The PCs are ordered hierarchically on the orthogonal subspace, based on their proportion of explained variance. The first principal component projects the maximum variance in the original dataset and each successive component explains the next highest percentage of the remaining variance. The projection of maximum variance

from an original set of three variables ( $X_1$ ,  $X_2$  and  $X_3$ ) on to the first principal component,  $P_1$  is shown in Figure 2.1 (Wold et al., 1987).



**Figure 2.1. Projection of maximum variance of the three-dimensional variable space to the first principal component. The first principal component,  $P_1$  is the line of best fit across the data points (Wold et al., 1987)**

A Scree plot represents a Pareto chart of the percentage of variance explained by each PC. The significant principal components are selected based on the proportion of variance explained. This reduces the dimensionality and correlation of the original variable space. Equation 2.1 shows how to apply PCA to a scaled set of original multivariate data matrix and obtain the  $z$  scores. Identifying the systematic subspace and residual subspace of the PCA latent space is represented by Equation 2.2 (Jackson 1980).

$$X = \bar{X} + Uz \quad \text{in matrix terms of } z \text{ scores} \quad \text{Equation 2-1}$$

$$X = \sum_{i=1}^a U_i z_i + \sum_{i=a+1}^n U_i z_i = \hat{X} + E \quad \text{Equation 2-2}$$

Where  $X$  represents the original data matrix,  $\bar{X}$  is the mean of  $X$ ,  $U$  stands for the eigenvector,  $z$  stands for the principal components,  $a$  is the number of principal components selected, while  $\hat{X}$  represents the estimate of  $X$  and matrix  $E$  represents the product sum of all the remaining principal components in the residual subspace.



---

### 2.2.2 Linear Discriminant Analysis

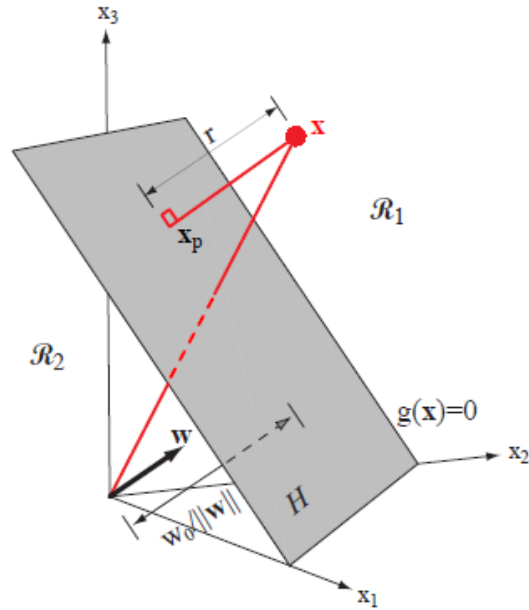
Linear Discriminant Analysis (LDA) is a supervised machine learning method commonly used for multi-level classification of a categorical response (Xanthopoulos et al., 2013). An LDA model is trained using a set of observations, of predictors and responses, taken from a multivariate data set to build a classification model. The trained model is then used to predict the category of unknown (new) data based on *a priori* knowledge (P. Wang et al., 2016). This is a simple and very powerful method of multilevel classification (James et al., 2013).

The standard implementation of the model assumes a Gaussian distribution of its independent variables and a new observation is classified to the category with the closest centroid (Friedman *et al.* 2001). LDA assumes each independent variable has equal variance and it is essential to standardise the input data for these models. LDA presents low-dimensional clusters of the response data very well and in most cases, presents the best classification results due to its simplicity and low variance (Friedman *et al.* 2001). LDA estimates the centroid of the data points, for each class, along with a common covariance matrix, which gives the spread of the data. It creates discriminant functions as hyperplanes, which forms the decision boundaries separating the response classes. A discriminant function is a linear combination of the components of the predictor variables,  $X$ . This is represented by Equation 2.3 (Duda et al., 2012) .

$$\mathbf{g}(X) = \mathbf{W}^T X + w_0 \quad \text{Equation 2-3}$$

Where  $W$  is the weight vector and  $w_0$  is the bias term.

Figure 2.2 presents the geometric representation of a linear discriminant hyperplane for a two-category case.



**Figure 2.2.** The linear decision boundary  $H$ , where  $g(x) = w^T x + w_0 = 0$ , separates the feature space into two half-spaces  $\mathcal{R}_1$  (where  $g(x) > 0$ ) and  $\mathcal{R}_2$  (where  $g(x) < 0$ ) (Duda et al., 2012).

In Figure 2.2, the orientation of the hyperplane is determined by the vector  $W$  and the location of the surface is given by the bias  $w_0$ . The discriminant function,  $g(X)$ , is proportional to the signed distance from the feature vector,  $X$  to the hyperplane. The distance from the origin to the hyperplane is given by  $w_0 / \|w\|$ . If  $w_0 > 0$ , the origin is on the positive side of the hyperplane and if  $w_0 < 0$ , it is on the negative side. The hyperplane passes through the origin if  $w_0 = 0$  (Duda et al., 2012).

There are situations in which LDA models can give poor performance. These include complex non-linear classification problems, where linear decision boundaries are insufficient to separate the response classes and also in situations where there are too many correlated predictor variables (Friedman *et al.* 2001).

### 2.2.3 Logistic Regression

Logistic regression models have categorical responses and are either binomial (two levels of response variable such as 1 and 0) or multinomial (more than two levels of the response variable). These models are primarily designed for binomial classification problems. A

---

binomial logistic regression model predicts the probability of the class membership of the response variable for one of the two response categories depending on the predictor variables given in Equations 2.4 and 2.5 respectively (Dreiseitl & Ohno-Machado, 2002).

$$P(\mathbf{y} = \mathbf{1}) = \frac{1}{1+e^{-(\alpha \cdot \mathbf{x})}} \quad \text{Equation 2-4}$$

$$P(\mathbf{y} = \mathbf{0}) = \mathbf{1} - P(\mathbf{y} = \mathbf{1}) \quad \text{Equation 2-5}$$

Where  $y$  stands for the response category,  $x$  stands for the predictor variable set,  $\alpha$  stands for the model parameter vector and  $\alpha \cdot x$  is the linear combination of the predictor variable set. Binomial logistic regression models can be extended to multinomial models when there are more than two responses. In logistic regression, one of the response categories is nominated as a reference category and the membership probability of the other categories is calculated by comparing them to that of the reference category. Suppose there are  $k$  categories in the classification problem. If one category is selected as the reference, then there are  $k-1$  equations created to predict the membership probability of each observation relative to the reference category. The probability of an observation falling in the  $j^{\text{th}}$  category is represented by Equation 2.6 (Puntanen, 2013).

$$P_j(\mathbf{x}_i) = \frac{\exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \dots + \beta_{pj}x_{pi})}{1 + \sum_{j=1}^{k-1} \exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \dots + \beta_{pj}x_{pi})} \quad \text{Equation 2-6}$$

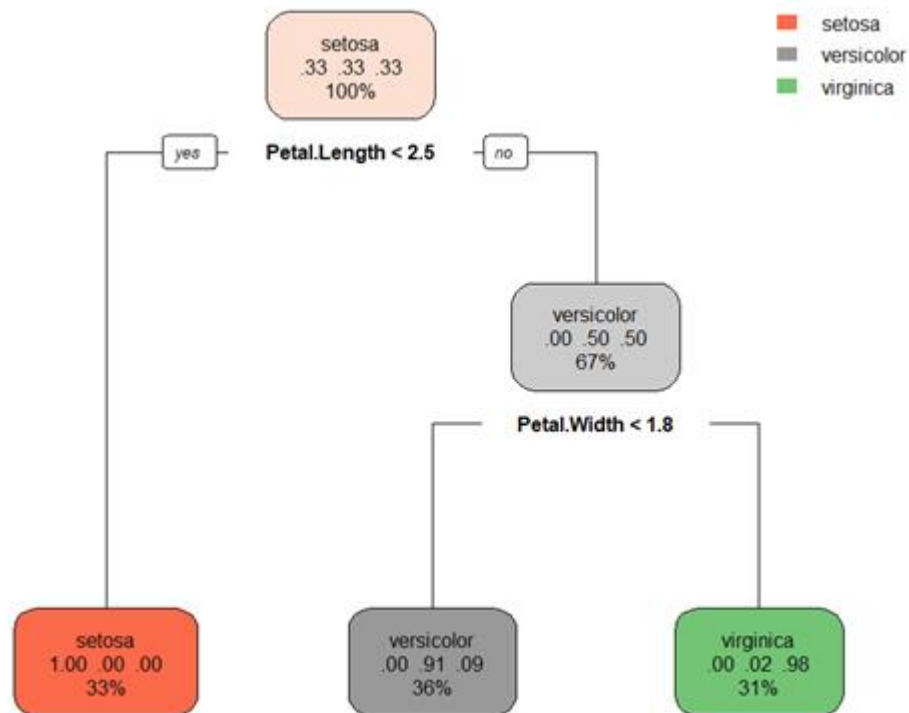
For  $j = 1, 2, \dots, (k-1)$  and  $i = 1, 2, \dots, n$ , where  $n$  is the number of independent observations with  $p$  predictor variables. Multinomial Logistic Regression (MLR) models use maximum likelihood estimation to estimate the model parameters, through a set number of iterations, until the model converges to a minimum final value.

---

Logistic regression has a number of limitations compared to linear discriminant analysis (Friedman *et al.* 2001). Logistic regression models can be unstable under certain circumstances e.g. when there are more than two response categories, if the categories are well separated and also in cases with a limited number of observations (James *et al.*, 2013).

#### 2.2.4 Random Forest

Random Forest (RF) is a decision tree based machine learning model, widely used for classification and regression. Decision trees are created from bootstrap samples selected from the model training data. Each decision tree is made of internal nodes (nodes) and terminal nodes called leaves. Nodes represent a test or condition on one or more independent variables, which further branches into leaves based on the test outcome (yes or no). The leaves of each node carry the final classification probability of the response categories. The logic of a decision tree is explained using the well-known Fisher's iris dataset. This dataset consists of three species of flowers which are setosa, versicolor and virginica and four independent variables such as Sepal.Length, Sepal.Width, Petal.Length and Petal.Width. A decision tree is formed from the whole iris dataset as shown in Figure 2.3.



**Figure 2.3. A decision tree model using Iris data in R.**

The first box represents the true percentage of each species (0.33 each) in the dataset. The first node condition uses the independent variable, Petal.Length. If the condition is true, the tree branches to the terminal node/leaf shown in orange, which has a probability of 1.00 for setosa and 0.0 for the other two species. If the condition is false, the tree branches to the light grey coloured leaf, which has 0.5 probability for both versicolor and virginica. The second node condition is formed using the variable, Petal.Width. If the condition is true, i.e., if  $\text{Petal.Length} \geq 2.5$  and  $\text{Petal.Width} < 1.8$ , the tree branches to the dark grey leaf, which has 0.91 probability for versicolor and 0.09 for virginica. If false, i.e., if  $\text{Petal.Length} \geq 2.5$  and  $\text{Petal.Width} \geq 1.8$ , the tree branches to the green node, which has a probability of .98 for virginica and 0.02 for versicolor.

RF generates a number of randomised decision trees, each of which makes predictions based on the randomised conditions at each node, and finally aggregates the predictions

---

(votes). The RF algorithm requires two main tuning parameters. These are the number of decision trees (*n<sub>tree</sub>*) and the number of variables split (*m<sub>try</sub>*) at each node of the trees. RF creates *n<sub>tree</sub>* bootstrap samples from the training data set (Leo Breiman, 1996a). Bootstrapping is a technique used to generate random samples with replacement (Efron and Tibshirani 1994). A decision tree is grown from each bootstrap sample choosing the best *m<sub>try</sub>* on a random set of predictors at each node. The model accuracy is estimated from the Out Of Bag (OOB) error, obtained through a technique called bagging (Leo Breiman, 1996b). When bagging, the RF model retains approximately one third of the training samples out of the bag for validation and the remaining two third of the samples are used for constructing the trees. Bagging is repeated on a random basis and the final OOB estimate is calculated by aggregating all the individual decision tree predictions. The model predicts the outcome based on the majority votes from the *n<sub>trees</sub>* for classification, while the model outcome is calculated as the average of *n<sub>tree</sub>* votes for regression.

### 2.2.5 Neural Networks

Neural Networks (NNs) are non-linear statistical models, which can be explained as a two-stage classification or regression technique (Friedman et al., 2001). A basic neural network model, such as a Vanilla Neural Network (VNN), consists of an input layer, a hidden layer and an output/response layer. Equation 2.7 represents the different layers of a VNN. Derived features/derivatives are created from linear combinations of the input variables,  $X$ , as represented by  $Z_m$  in Equation 2.7. The activation function commonly used to create the derivatives is a sigmoid function (Friedman et al., 2001), which compresses the output value between 0 and 1. The hidden layer is then formed as a linear combination of these derivatives ( $T_k$  in Equation 2.7). The response,  $Y$ , is modelled as a

---

non-linear function of the linear combinations of  $Z_m$ . The number of units in each hidden layer and the weights used to train the derivatives of the hidden layer can all be tuned according to the complexity of the data and the problem at hand (Friedman et al. 2001). The basic principle behind the tuning of the weights of different layers of VNNs is based on a feed-forward and back-propagation technique, which improves through a series of iterations. The network diagram representing a VNN is presented in Figure 2.4. For a  $K$ -class classification problem, there are  $K$  response units ( $Y_k, k= 1, \dots, K$ ) at the top layer of the model with the  $k$ th unit ( $Y_k$ ) modelling the probability of class  $k$ .

$$z_m = \sigma(\alpha_{0m} + \alpha_m^T X), m = 1, \dots, M$$

$$T_k = \beta_{0k} + \beta_k^T Z, k = 1, \dots, K$$

$$Y = f_k(X) = g_k(T), k = 1, \dots, K$$

**Equation 2-7 (Friedman et al., 2001)**

Where  $X = (X_1, X_2, \dots, X_p)$  for  $p$  input variables,  $Z = (Z_1, Z_2, \dots, Z_M)$  for  $M$  units in the hidden layer and  $Y = (Y_1, Y_2, \dots, Y_k)$  for  $k$  responses. The unknown parameters,  $(\alpha_{0m}, \alpha_m, \dots)$  and  $(\beta_{0k}, \beta_k, \dots)$  of a NN model are called weights. The output activation function,  $g_k(T)$  of a  $K$ -class classification problem often employs a Softmax function, which estimates the probability of the  $K$ th class (Equation 2-8). This is the function which maps the weighted derivatives of the input variables with the output layer and predicts the probability of an input image falling in each of the response categories.

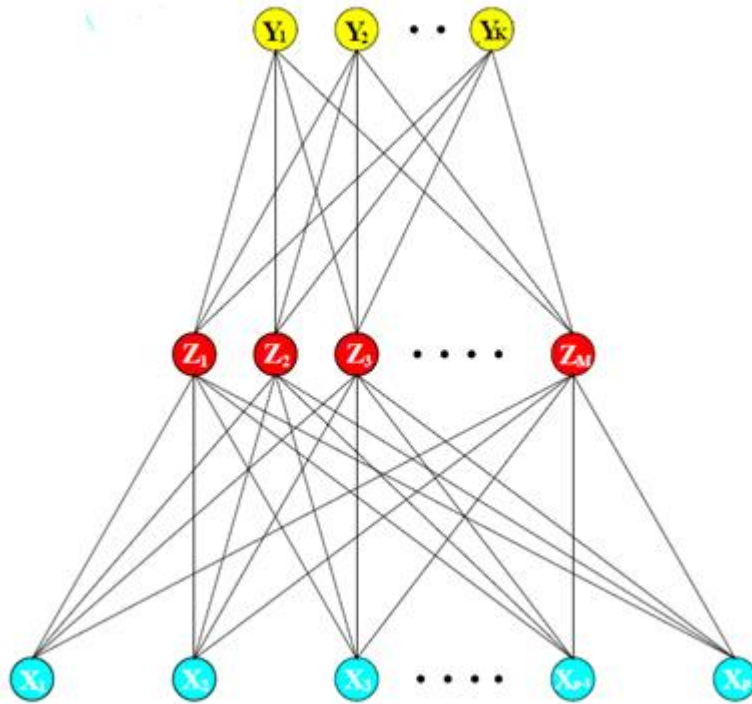


Figure 2.4. Schematic of a single hidden layer, feed-forward neural network (Friedman et al., 2001).

$$g_k(T) = \frac{e^{T_k}}{\sum_{l=1}^K e^{T_l}} \quad \text{Equation 2-8 (Friedman et al., 2001)}$$

The output obtained from the Softmax function is further analysed using a loss function to evaluate the goodness of the classifier. The error function, used to calculate the loss/deviance of a NN model, is called the cross-entropy function, which is commonly used in multinomial logistic regression models. The weight values assigned to the input variables in a NN model are chosen internally in such a way as to reduce the model error and to improve the performance of the model with the training data. The weights feed forward through the hidden layer to the output layer to predict the response classes and based on the deviance calculated by the cross-entropy function, the model propagates the information back to adjust the weights in the hidden layer. This happens repeatedly until the model error is minimised after a set level of iterations.



---

### 2.2.6 Convolutional Neural Networks

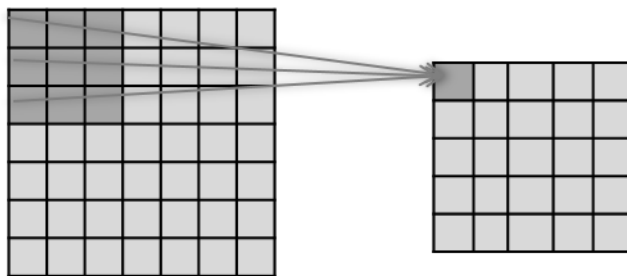
Convolutional Neural Networks (CNNs) are deep learning NNs which are widely used to solve complex image classification problems. CNNs can perform image processing and feature extraction followed by multi-layer NN based classification or prediction. Some of the most popular areas where CNNs are applied include face recognition, biometric identification, vehicle detection, signature verification etc. (Das et al., 2019; Jain et al., 2019; C. Wang & Xi). Computers process and store images as arrays of pixel values. Colour images are stored as three dimensional (3D) arrays and greyscale images are stored as two-dimensional (2D) arrays or matrices. A colour image of resolution 50 x 50 pixels is stored as a 3D array of size 50 x 50 x 3 pixels, where 3 represents the RGB channels. A greyscale image of the same resolution is stored as a 2D array of size 50 x 50 x 1 pixels. The value of each pixel, in the array, ranges from 0 to 255, which represents the intensity of that pixel. CNNs are trained with the image pixel data given in a 2D (for greyscale images) or 3D (for colour images) array format called tensors and it extracts both low level and high level features of the objects in the image through a series of hidden layers to predict the category of the image (Nielsen, 2015). The layers commonly used in a CNN model are:

1. Convolutional
2. Non-linear
3. Pooling
4. Dropout
5. Flatten
6. Fully Connected

---

### 1. Convolutional Layer

The Convolutional layer forms the first layer of a CNN model. This layer identifies the key features of the input images and reduces the input feature space to the key pixel values (weighted pixels of the identified features). The input image must be of square resolution to perform convolution operation. The image is scanned using filters, also known as kernels. A filter is an array of weights having the same depth (1 for greyscale image and 3 for colour image) as the input image but with a smaller width and height. For example, for a greyscale image of size  $50 \times 50 \times 1$ , the size of the filter used for convolution can be  $3 \times 3 \times 1$ . Figure 2.5 represents the convolution operation of an input image of size,  $7 \times 7$  using a filter of size,  $3 \times 3$  resulting in an output image of size  $5 \times 5$ .



**Figure 2.5. Convolutional Layer. a) Input image of size  $7 \times 7$  and b) Convoluted output image of size  $5 \times 5$ .**

Scanning starts from the top left corner of the input image and slides 'n' strides to the right every time until the whole image is processed. The stride represents the number of pixels a filter moves towards the right in each step of convolution. When the stride is 1, the filter moves one pixel at a time. At each move, the filter scans a 2D area of the input image, which is of the same size as the filter. A matrix dot product of the input array values (a square matrix) and the filter values is resulted from each scan. A summation of the values obtained at each scan corresponds to each weighted pixel in the output image.

---

The size of the output image after convolution is calculated using the formula given in Equation 2.9.

$$O = \left(N - \frac{F}{S}\right) + 1 \quad \text{Equation 2-9}$$

$O$  represents the size of the output image,  $N$  represents the input image size (e.g. 7 x 7),  $F$  represents the filter size (e.g. 3 x 3) and  $S$  represents the stride (in this case,  $S = 1$ ). In this example, the output image size, after convolution, will be 5 x 5. A technique called padding can be used in cases where the dimension of the output matrix should be maintained the same as the dimension of the input matrix. Padding is a method of adding zeroes (as pixels of filler) to the edges of the input matrix in a symmetrical manner. The value of padding depends on the size of the convolution filter as given in Equation 2-10.

$$P = \frac{F-1}{2} \quad \text{Equation 2-10}$$

$P$  represents padding and  $F$  denotes the filter size. The default value for padding in a convolutional layer is zero, which means no fillers are added to the input image. Convolutional layers produce feature maps which represent the key features detected from the image. A CNN model uses more than one convolutional layer and the feature maps from the previous layer forms the input of the following layer. The filters used in the initial convolution layer of a CNN model are designed to detect low level features of images, such as edges and curves, while later filters are used to detect high level features such as specific objects.

## 2. *Non-Linear/Dense Layer*

This layer takes in the summed inputs obtained from the feature maps of the convolutional layer connected to it. Activation functions are used in the dense layer units of CNN to understand the complex non-linearity in the input and to provide an optimised output

---

through backpropagation of errors. Rectified Linear Unit (ReLU) is the most commonly used activation function in the dense layer of a CNN (Nair & Hinton, 2010). The mathematical definition of ReLU is shown in Equation 2-11.

$$y = \max(0, x) \quad \text{Equation 2-11}$$

$x$  stands for the input feature map of the dense layer that is passed through the ReLU function and  $y$  represents the output feature map obtained from the dense layer after activation using ReLU. ReLU is linear for all positive input values and zero for all negative values.

### 3. Pooling Layer

This layer is a down sampling layer of CNN. It reduces the input size of the feature maps filtered through the convolutional layers. Downsampling reduces computation cost and controls overfitting. This is achieved through pooling. Max pooling is the most popular technique used in CNN. The input feature maps are filtered using filters of equal size and stride (usually 2) and outputs the maximum value in each filtered area of the image as shown in Figure 2.6.

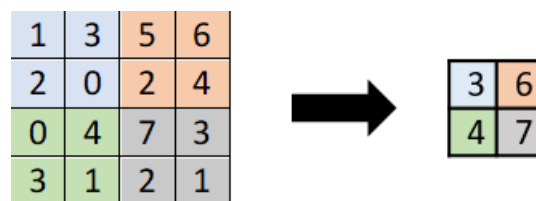


Figure 2.6. Max pooling of a 4 x 4 data matrix down to 2 x 2.

### 4. Dropout Layer (DL)

A random set of feature maps are dropped out in this layer. The DL layer is used during the training stage of a CNN to reduce overfitting. This layer ensures that the model is not tuned too precisely to the training data and performs better even after dropping out some of the feature maps.

---

### 5. *Flatten Layer*

This layer flattens the feature map, obtained from the previous layer, into a long vector. The Flatten Layer is implemented just before the final layer, which is the Fully Connected layer.

### 6. *Fully Connected Layer*

This is the final layer of a CNN model, which is connected to all the units of the previous layer. The Fully Connected (FC) layer processes the output of the previous layer, which contains the key high-level features extracted from the input image, through the various filtering layers. It then predicts the final response using activation functions which are relevant to the defined problem. Softmax is the most commonly used activation function in the FC layer for image classification problems with multiple response categories. This function calculates the correlation of the feature maps with the response categories and classifies them based on the highest correlation. The function assigns suitable weights to the feature maps, so that the probability of predicting the correct category increases.

## **2.3 Industrial Applications**

Image processing techniques integrated with machine learning classifiers have a long history of success in product quality evaluation in a wide range of industries such as automotive, electronics, medical, biopharmaceutical, food *etc.* (Bertani et al., 2017; Du & Sun, 2006; Gosselin et al., 2016; Jousse, 2008; Manak et al., 2018; Pfeil et al., 2018; Venora et al., 2009; Zhang et al., 2018). However, such techniques are currently under investigated in the area of industrial emulsification. The literature has identified that development of fundamental predictive models using particulate/droplet data can benefit the optimisation of emulsification processes (Doyle III et al., 2003; Jousse, 2008; Shi et

---

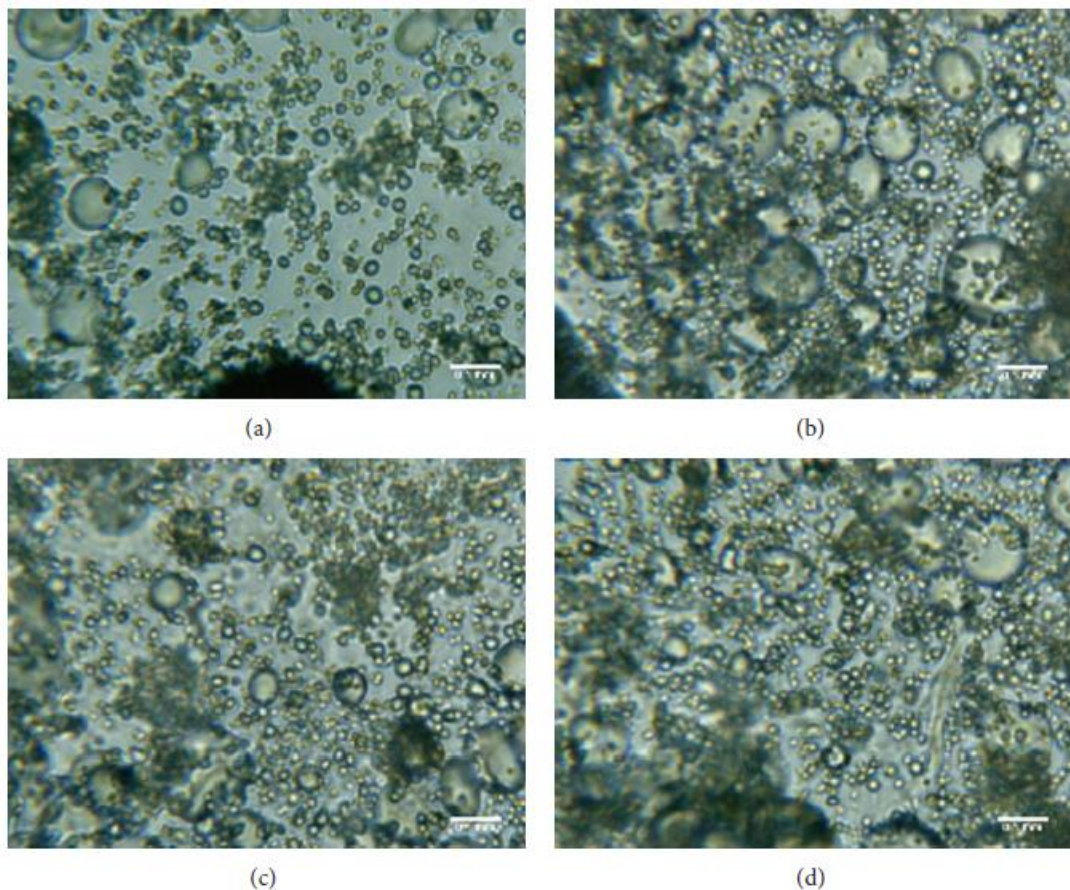
al., 2006). Only a limited number of studies have reported the application of classification or prediction modelling, using droplet data, for the quality evaluation of dispersion systems (Kljusuric et al., 2015; P. Wang et al., 2016).

Wang *et al.* (2016) have provided a detailed review of various supervised and unsupervised machine learning models applied for the quality evaluation of olive oil using chemical data obtained through spectroscopical techniques. Their study investigated the authentication of olive oil by extracting a data matrix of its chemical constituents using techniques such as mid-infrared (MIR), near-infrared (NIR) and Raman spectroscopy. They applied chemometric analysis, *i.e.* the statistical study of chemical composition to identify quality parameters, to differentiate olive oil samples from other edible oils using supervised and unsupervised classification techniques. PCA and LDA were described as the two most effective unsupervised and supervised methods used for pattern recognition and classification respectively. PCA, Hierarchical Cluster Analysis (HCA) and K-means were identified as the most commonly used unsupervised methods for identifying patterns in a data set. PCA was described as the most effective of these three techniques (Wang et al., 2016).

A similar classification study of barley milk samples, obtained by blending barley grain, was conducted by Kljusuric *et al.* (2015) to find the optimum processing time in barley milk production. The barley grain was blended for a total of 60 seconds to produce barley milk, which is an emulsion. Samples were obtained at 15 seconds intervals. The samples, at each stage of blending, were filtered to separate the grain from the milk and were analysed using NIR spectroscopy. PCA was applied to the NIR absorbance spectral data and the score plots were used to classify both the grain and the milk samples acquired at 15, 30, 45 and 60 seconds of blending (Grané & Jach, 2014). The PCA score plots

---

followed the same trend of the absorbance spectra presenting a good separation between the samples. Particle/droplet size measurements, such as median diameter, Sauter mean diameter and specific surface area, were obtained using laser diffraction. Micrographs of the milk samples were also acquired to get a better insight into the droplet size at each stage of blending and to confirm the laser diffraction measurements. The droplets in the dispersed phase of the barley milk samples, obtained at 15, 30, 45 and 60 seconds of blending, are shown in the micrographs of Figure 2.7. Their study identified 45 seconds as the optimal blending time, as the droplets appeared to form aggregates after that time according to the NIR, PCA, laser diffraction and micrograph analyses.



**Figure 2.7. Micrographs obtained from barley milk samples at a) 15 seconds b) 30 seconds c) 45 seconds and d) 60 seconds of blending (Kljusuric *et al.*, 2015).**

However, it is noted that their study used a limited set of barley milk samples and did not use any independent samples to validate their findings. In addition, their PCA results were

---

not extended to any supervised learning techniques to develop a predictive classification model and to optimise the process on an industrial scale.

PCA is extensively applied for identifying patterns in data in many areas like biostatistics and chemometrics (Garcia & Filzmoser, 2015). It is highly useful in chemometrics, where numerous variables, of very high collinearity, are analysed. At the same time, PCA is also found to be very effective in any manufacturing process industry, where a large number of correlated variables are analysed for multivariate process control, to reduce the dimensionality and correlation in the original feature space (Kourti et al., 1996; Kourti & MacGregor, 1995, 1996). (Kirby & Sirovich, 1990) first proved that human faces can be effectively represented and reconstructed by PCA models. (Turk & Pentland, 1991) presented the well-established Eigenfaces method for facial recognition. Since then, PCA has been widely accepted as one of the most effective methods for facial recognition (Dagher & Nachar, 2006; Vaswani & Chellappa, 2006; Xudong & Kin-Man, 2006).

The PCA technique aids in reducing the dimensionality of the image data and at the same time retains the maximum variability of the image features in a set of eigenvectors. In the early years, conventional PCA based pattern recognition methods used to store the two dimensional image data matrices into one-dimensional vectors and this made the multidimensional vector space too large to calculate the covariance matrix when there is a lack of training samples. In order to improve the speed and accuracy of calculating the covariance matrix and the eigenvectors, a two dimensional PCA (2DPCA) was proposed (Jian et al., 2004). In 2DPCA, the original image data matrix does not need to be transformed into one dimensional vectors and the covariance matrix and the corresponding eigenvectors are calculated directly from the image data matrix.



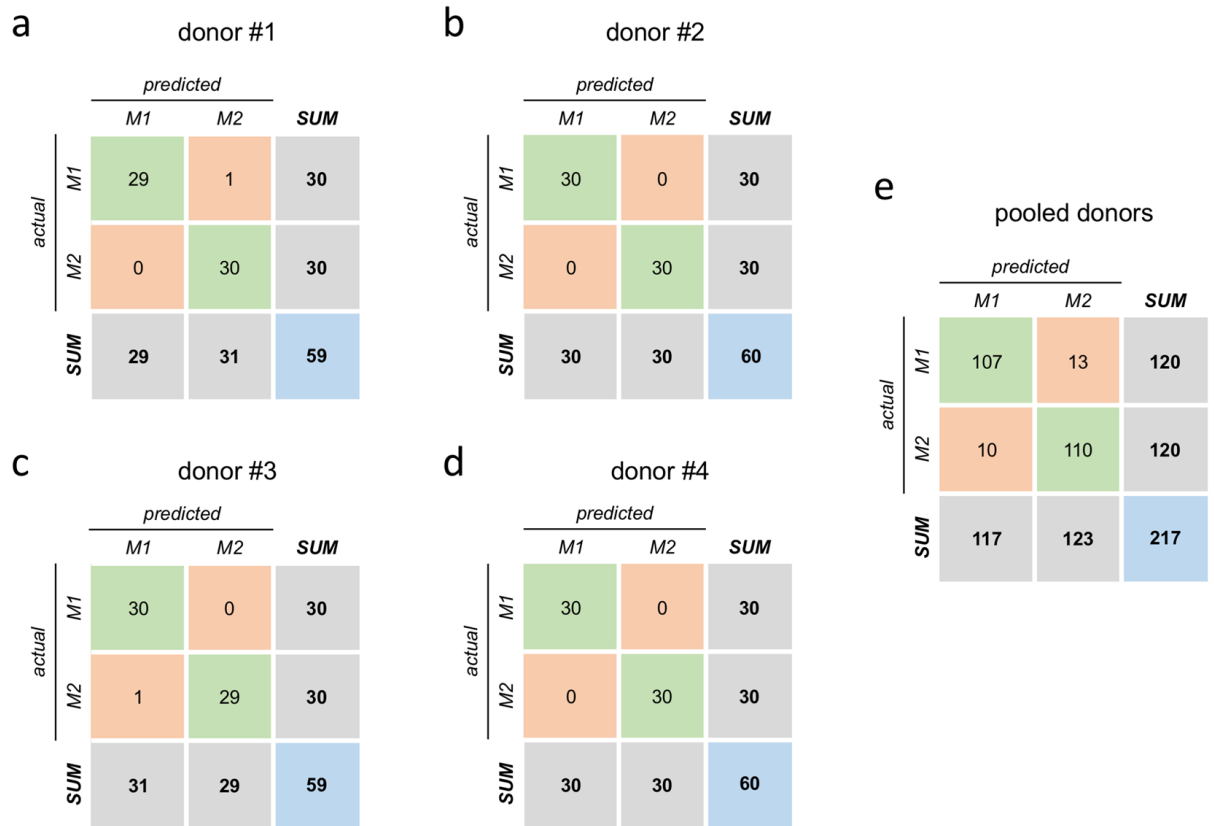
---

Applications of PCA in food science and technology are also widely discussed in the literature (Grané & Jach, 2014). Robust classification models have been developed, in the past, by performing supervised learning techniques such as Linear Discriminant Analysis on a PCA transformed space (Bertani et al., 2017; Venora et al., 2009).

A machine vision integrated LDA-Bayesian classification model was investigated for the automated quality evaluation of durum wheat (Venora et al., 2009). The model achieved good accuracy in the classification of wheat kernels into starchy and shrunken categories. Size, shape and intensity distribution features obtained from images of wheat kernels were the predictors selected for developing the classification model. Image analysis macros were developed for extracting the features of each kernel category, which were automatically integrated with the classification model. A training set of 100 grain samples were classified manually by inspectors and LDA into four categories (starchy, shrunken, vitreous and piebald). The results were used to train an online Bayesian classifier. The Bayesian classifier, trained using the LDA results, classified a test set of 30 grain samples that achieved 96.03% to 99.58% accuracy for the shrunken and starchy categories.

Bertani *et al.* (Bertani et al., 2017) achieved a partly-automated multivariate classification of hyperspectral micrographs of living cells using PCA followed by LDA. Cells from two different polarizations, M1 and M2, were selected as the classification categories. A multi-dimensional data set of 26 different spectral wavelengths was used to obtain the cell micrographs. Using PCA, the high dimensional dataset was reduced to a set of 14 variables, which were selected as their LDA model predictors. Samples were obtained from four different donors. A total of 60 samples (30 from each category) from each cell donor were then used for cross validating their PC-LDA model. Figure 2.8a to 2.8d shows the confusion matrix obtained from the 10-fold cross-validation conducted for the

individual donors. The confusion matrix obtained by cross-validating the model using the 240 pooled samples from the four donors is presented in Figure 2.8e.



**Figure 2.8. Confusion matrices from 10-fold cross-validation of the PC-LDA model. Each matrix is the sum of 10 matrices from 10 test sets (Bertani et al., 2017).**

A classification accuracy of 98-100% was achieved for the individual donors (Bertani *et al.* (Bertani et al., 2017). The classification accuracy was reduced to 90% for the general prediction model using a pool of different donors.

Logistic regression models have also been widely used for data classification and prediction purposes in the medical field for diagnostic and prognostic tasks (Dreiseitl & Ohno-Machado, 2002). Logistic regression was the most popular classification model with 28,500 publications in the early 21<sup>st</sup> century when compared to other prediction models such as artificial neural networks, k-nearest neighbours, decision trees and

---

support vector machines, as indexed in MEDLINE (Dreiseitl & Ohno-Machado, 2002). Recent studies in predictive modelling in areas such as physiology and geology have proved that logistic regression models are successful in data classification and prediction (W. Chen et al., 2017; Heung et al., 2017; Slanzi et al., 2017). High performance classification models have been developed recently, for medical imaging tasks, using Multinomial Logistic Regression (MLR) and feature ranking-based random forest algorithms (Yu et al., 2018). The study by (Yu et al., 2018) has demonstrated that these conventional classification models were able to present the same level of accuracy comparable to pretrained Convolutional Neural Network (CNN) models in the classification of liver fibrosis stages. A similar image analysis integrated classification study was conducted recently in the industrial optimisation of the castor seed oil extraction process (Isaza et al., 2018). Classification algorithms such as MLR, RF, naive-bayes and multilayer perceptron (feed forward neural network with multiple layers) were investigated to classify castor seed samples using morphological and colour features. Their study concluded with the highest accuracy of 100% presented by the RF classification model.

The Random Forest (RF) algorithm was first proposed by (Leo Breiman, 2001). Bagging, variable importance measures and resampling capability are some of the advantages of the RF algorithm that makes it stand out from other machine learning models (Biau & Scornet, 2016; Leo Breiman, 1996a, 1996b; Genuer et al., 2010). RF classification and regression approaches have been applied in a wide range of areas such as food industry, medical imaging, chemoinformatics, bioinformatics and others. The approach is noted for its success with small sample sizes and high dimensional feature spaces (Díaz-Uriarte & De Andres, 2006; Joly et al., 2014; Svetnik et al., 2003).

---

A robust rain drop classification model was developed using RF classification based on droplet features such as shape, colour and texture (Webster & Breckon, 2015). Random Forest is capable of handling unequal sample sizes of the response classes and is also proficient in handling missing data by calculating the proximity matrix from pairs of observations (L Breiman & Cutler, 2003; C. Chen et al., 2004; Kuhn & Johnson, 2013). Recent studies in medical image classification and in the food industry have found RF classification models as equally good and in some cases, as even better classifiers compared to CNN and other similar multilayer perceptrons (Isaza et al., 2018; Yu et al., 2018).

In recent years, neural network models have become the most prevalent approach for classification problems in research, which involve massive amount of training data that require significant computing capability. This can be justified by the high accuracy results demonstrated by these models in a diverse range of applications including image processing and classification, face recognition, speech recognition and natural language processing (*Cetron et al., 2019; Geuther et al., 2019; Oviedo et al., 2019*). The performance of neural network models depend solely on the quality and quantity of the training data on a big data scale (Dehghani & Kamps, 2018).

Convolutional neural networks (CNN), which fall under the neural network family, are multilayer perceptrons with convolutional layers and are commonly used for solving complex image classification problems (Heo et al., 2017; Jain et al., 2019; Krizhevsky et al., 2012; Ounkomol et al., 2018; Spier et al., 2019; Zhao et al., 2019). The CNN models, which have reported good accuracy in the existing literature, have mostly utilised significantly large datasets from publicly available databases like ImageNet, CIFAR-10

---

and others (Krizhevsky et al., 2012; Soldati et al., 2018). A range of pretrained ready-to-use CNN models, tuned for various image classification problems, such as GoogleLeNet Inception v3, ConvNet, AlexNet, ResNetXt and other custom CNNs are also publicly available (Krizhevsky & Hinton, 2009; Krizhevsky et al., 2012; Soldati et al., 2018; Szegedy et al., 2017; Szegedy et al., 2016).

A recent study conducted by Soldati *et al.* (2018) in the classification of microfluidic droplet content for the detection of tumor cells has demonstrated good accuracy using ready-to-use object detection CNN networks. Their study attempted the classification of tumor cell images, encapsulated in microfluidic water-in-oil droplets, from the images of empty/debris-containing droplets. The major objective of their study was to develop an automated classification technique to eliminate the limitations, such as subjectivity, erroneous nature and increased time consumption, of the current manual classification of such microfluidic droplets. Different approaches have been applied, by Soldati *et al.* (2018), for their classification study. These include pretrained CNN models, other custom CNNs with different layer options, which were trained using images from public databases (MDA-MB-231 cancer cell-line), and two object detection CNNs (MobileNet and Inception-v2). The object detection networks, combined together, presented the highest accuracy of 96% compared to the other trained networks, which provided an accuracy of 84% to 90%. The training set used for the various models consisted of 500 (manually labelled images) to 10,698 images (from public databases). Additionally, as a proof of concept, the NIH ImageJ was used to perform image segmentation, object detection and cell characterisation (Soldati *et al.*, 2018). The size, shape and texture characteristics of the tumor cells in the training set of images were extracted by ImageJ. Cluster analysis was performed in Python using PCA and three classification models were

---

developed such as logistic regression, Naïve Bayes and a multilayer NN. These classification models were tested using independent datasets (micrographs of cancer cells) acquired from a pool of cancer patients and achieved an accuracy of 77% to 90%.

## **2.4 Conclusions**

Recent advances in industrial automation have imposed significant competitive pressure on pharmaceutical industries to rapidly improve the techniques applied for product analysis and process optimisation (Isaksson et al., 2018; Islam et al., 2012; Panckow et al., 2017). The literature review conducted in the area of machine learning/deep learning integrated with automated imaging techniques has demonstrated a very promising application of such techniques for product quality evaluation in a range of industries (Du & Sun, 2006; Patel et al., 2012; Pfeil et al., 2018; Schulze et al., 2013; Venora et al., 2009). However, there have been no systematic studies reported previously in the use of such techniques for monitoring a complete industrial emulsion processing. This can be attributed to the current challenges of droplet detection from emulsion micrographs discussed in Chapter 1. This research proposes to address these challenges through the development of a novel approach of automated image segmentation integrated with machine learning for more enhanced and intelligent emulsion manufacturing. The next chapter discusses the overall methodology and the associated techniques developed in this research.

---

## Chapter 3

### Research Methodology

#### 3.1 Introduction

This chapter presents the various techniques that were developed and applied, in this research, to overcome the existing challenges identified in emulsion quality evaluation. This study involved the industrial production of a topical cream emulsion. Two different emulsification processes, Process-1 and Process-2, were evaluated. Process-1 was a 30 minutes long trial process used as a proof of concept to evaluate the techniques. Process-2 was a 90 minutes long industrial process of the same emulsion. The steps involved in the approach are as follows:

- Acquisition of in-process emulsion micrographs
- Image processing and segmentation for droplet detection
- Statistical analysis of droplet characteristics
- Supervised classification of micrographs
- Prediction of remaining processing time (time required to attain the target characteristics) of the emulsion.

The micrographs acquired from Process-1 were initially evaluated in order to identify the best image processing and segmentation technique. The selected technique was applied to evaluate Process-2.

#### 3.2 Overview of the Approach

A schematic overview of the approach undertaken is presented in Figure 3.1. Micrographs were acquired from emulsion samples obtained at five-minute intervals during the emulsification process. Automated image processing and segmentation was performed to

---

detect the oil droplets and their corresponding characteristics from these micrographs. Two image segmentation techniques were developed and evaluated for detecting the droplets and their characteristics; an edge detection-based technique and an intensity histogram-based technique. Statistical analysis was performed to determine the evolution of droplet characteristics throughout the emulsification process and their impact on emulsion quality. The best technique was selected for further evaluation of the industrial process. The selected droplet characteristics, which presented progressive variation throughout the emulsification process, were modelled using unsupervised and supervised machine learning algorithms. The models were applied to classify the micrographs at each interval into a predefined set of categories. Regression models were developed using the selected droplet characteristics as predictors. The objective of these regression models were to predict the completion time required by the emulsion, during emulsification, to attain the desirable/target droplet characteristics.



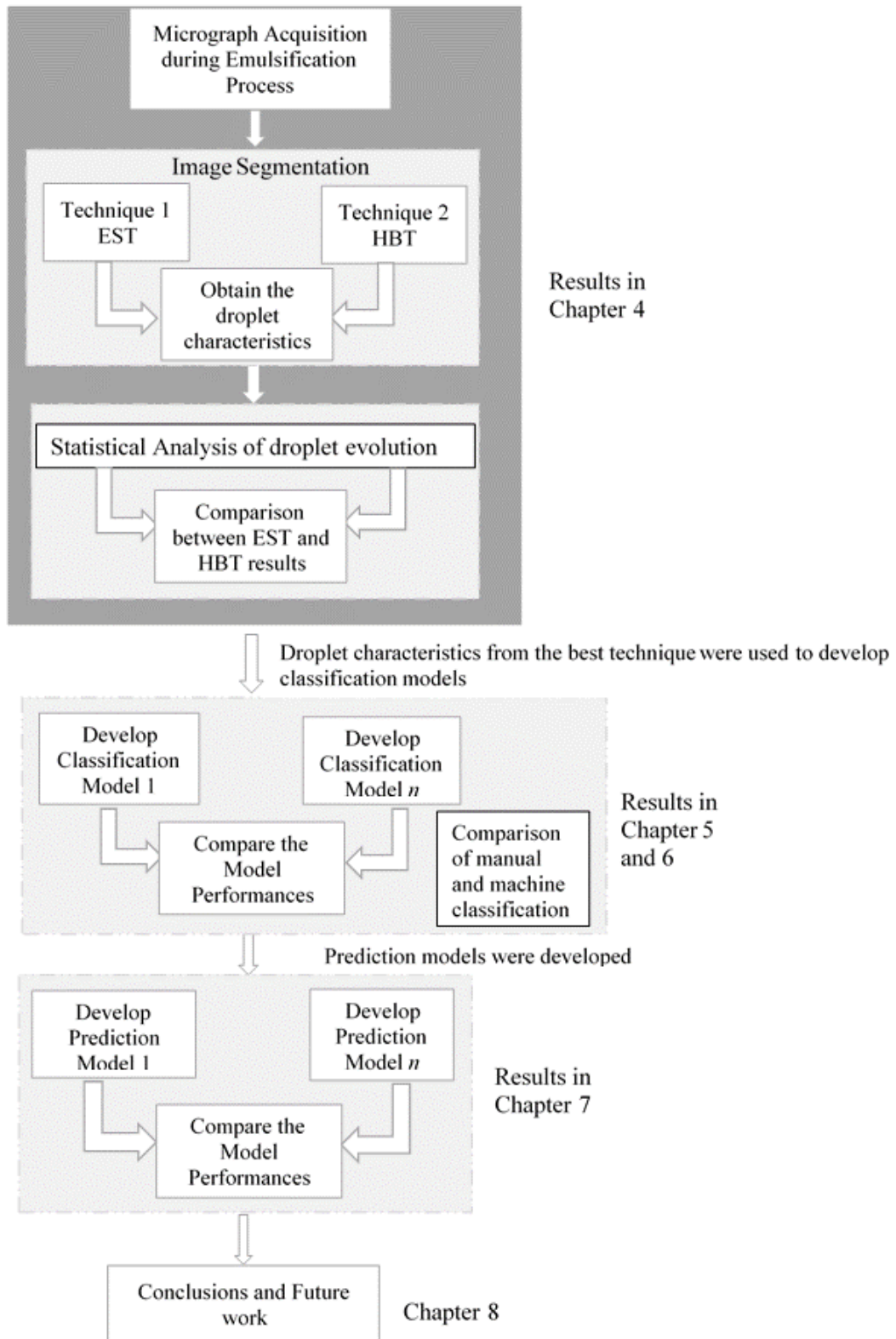


Figure 3.1. Schematic representation of the Research Methodology.

---

### 3.3 Micrograph Acquisition

During the manufacturing process of a skincare emulsion, the homogeniser (mixer) was stopped to take samples at every five-minute interval. A standardised method of sample slide preparation was followed for consistency. The sample material was placed on the microscope slide and a cover slip was placed on top without any pressure. Another microscope slide was placed on top of the cover slip and pressed down evenly at the four corners. A Zeiss Microscope Axio imager A2m was used for microscopy and the software version used for micrograph viewing was Axio Vision SE64. Bright Field (BF) micrographs of 40x magnification were acquired from each sample at a standard illumination of 7V. The micrographs were saved as TIFF (Tagged File Formats) files. The Microscopy society of America recommends that scientific images should be saved as TIFF files to avoid the loss of raw data through compression (Rossner & O'Donnell, 2004). The micrograph acquisition from Process-1 and Process-2 is explained in the following sections, 3.3.1 and 3.3.2.

#### 3.3.1 Process-1

Process-1 represents a 30 minutes emulsification process of a cream emulsion. The emulsion was continuously mixed for 30 minutes using a homogeniser at a tip speed of 25 m/s for the first 15 minutes and at a tip speed of 15 m/s for the last 15 minutes. Ten BF 40x micrographs were acquired at five-minute intervals of the emulsification process. The process was stopped after 30 minutes. Any further processing was considered as over-processing. This represented a total of 60 micrographs.

#### 3.3.2 Process-2

Process-2 represents a 90-minute industrial emulsification process of the same emulsion. Sampling was started at 55°C before emulsification commenced. Samples were taken at

five minutes intervals. Ten BF 40x micrographs were obtained from each sample. A total of 160 micrographs were obtained at 16 different times and temperatures, as shown in the sampling protocol (Table 3.1). The homogeniser was stopped after 70 minutes. The final bulk discharge was obtained after 20 minutes when the product was cooled down to 32°C. The sampling time and temperature was recorded every five minutes. The temperature is not measured in the typical emulsion manufacturing process. As a result, it was agreed not to include temperature in the current modelling approach. The processing time of the emulsion was the parameter of interest for the industry.

**Table 3-1 Sampling protocol from Process-2.**

<b>Batch 1: Cream Emulsion</b>	
<b>Sampling Time starting at 55°C</b>	<b>Sampling Temperature</b>
<b>0 Minutes</b>	55°C
<b>5 Minutes</b>	52.6°C
<b>10 Minutes</b>	50.5°C
<b>15 Minutes</b>	48.5°C
<b>20 Minutes</b>	47.2°C
<b>25 Minutes</b>	45.4°C
<b>30 Minutes</b>	44.2°C
<b>35 Minutes</b>	42.7°C
<b>40 Minutes</b>	41.5°C
<b>45 Minutes</b>	40.2°C
<b>50 Minutes</b>	39.2°C
<b>55 Minutes</b>	38.1°C
<b>60 Minutes</b>	37°C
<b>65 Minutes</b>	36.1°C
<b>70 Minutes</b>	35.2°C
<b>Final bulk discharge @ 90 Minutes</b>	32°C

An independent set of 150 micrographs were also obtained from a different batch during the industrial processing of the same product. This batch was run on a different day and

---

took 65 minutes to process. The micrographs from this batch were also obtained at five-minute intervals from the start until processing stopped after 65 minutes and the final discharge sample was acquired when the product was cooled down to 32°C. This took an additional 35 minutes. 10 BF 40x micrographs were obtained at each interval from 15 different times and temperatures. This resulted in a total of 150 micrographs.

### **3.4 Image Processing and Segmentation**

The micrographs were processed using a scientific image processing software called Fiji (version 1.51h), which is an extended version of ImageJ (J Schindelin, 2008; Johannes Schindelin et al., 2012). Two different automated image segmentation techniques were developed to detect the ROI (Region Of Interest), which are the oil droplets, in the micrographs. The techniques were edge and symmetry-based detection and histogram-based detection. The droplet characteristics were also extracted and saved to a file. The major objective of image processing and segmentation was to evaluate the technique that best detects the oil droplets and their characteristics, presenting a progressive droplet evolution during emulsification. The ultimate goal was to develop a suitable machine vision technique capable of classifying the micrographs and determining the industrial process equilibrium (a point at which the emulsification process becomes stable). The two image segmentation techniques are explained in sections 3.4.2 and 3.4.3.

#### **3.4.1 Fiji, an Extended Version of ImageJ**

ImageJ, formerly known as NIH (National Institutes of Health) Image, is a scientific open-source image processing software package, which has a proven track record of image analysis in areas such as light microscopy, scanning electron microscopy, crystallography, cell biology and medical imaging (Schneider et al., 2012; Sendra et al., 2015). The utilities of ImageJ have been extended to several variants including Fiji, a

---

“batteries-included” distribution of ImageJ. Fiji is a prominent example of how ImageJ has been accepted by the research community (J Schindelin, 2008; Johannes Schindelin et al., 2012; Schneider et al., 2012). It provides an extensive range of plugins for image processing and analysis. User-defined programs called macros can be written in various scripting languages to perform specific tasks in Fiji. The macros can be programmed to call built-in Fiji plugins. In addition to this, Fiji also provides versatility to the user in integrating new user-defined plugins, developed in other programming languages such as Java, to its plugins package.

### 3.4.2 Edge and Symmetry Technique (EST)

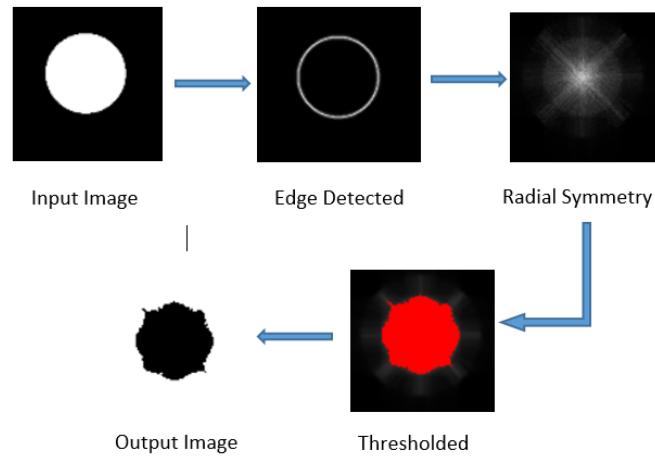
This technique is based on the Edge and Symmetry Filter (ESF) in Fiji (Boudier; Ollion et al., 2013). ESF is a multistage algorithm, which operates in two major steps such as edge detection and radial symmetry detection. The first step is edge detection, which is based on the Canny-Deriche edge detection algorithm (Canny, 1987). Edge detection includes the identification of the edges of the Region of Interest (ROI) in the image, centralising the detected edge points by preserving the local maxima and finally, minimising the false edge (noise) detection. The edge detection phase works in five different stages which can be summarised as the following:

1. Smooth the original image by applying a Gaussian filter kernel, which is convolved with the image. This step removes noise in the image.
2. Find the pixel intensity gradients of the image using the Sobel operator. Sobel is an edge detection operator which calculates the direction of the edges in an image and returns a value for its first derivative in the horizontal and vertical directions.

- 
3. Apply non-maximum suppression to sharpen the broad edges. This is an edge narrowing technique by suppressing all the gradient values except the local maxima.
  4. Apply double thresholding by setting minimum and maximum threshold limits to obtain a more accurate detection of real edges. The edge pixel gradients which are higher than the maximum threshold limit, are marked as strong and those which are smaller than the higher limit and larger than the lower limit will be marked as weak. The edge pixel values smaller than the lower limit will be suppressed.
  5. Track the edges using 8-connected neighbourhood Binary Large Object (BLOB) analysis. This technique tracks the weak edge pixels which are connected to strong edge pixels by looking at its 8-connected neighbourhood pixels. If a weak edge pixel is found connected to at least one strong edge pixel in BLOB, it is retained, otherwise, it is discarded.

Edge detection is followed by radial symmetry detection, which applies a radius parameter to identify the symmetry of the objects to be detected. In summary, the ESF algorithm uses four input parameters, the alpha Canny (Gaussian smoothing parameter), the radius of the object being detected and two internal parameters such as normalisation ( $n$ ) and scaling ( $s$ ).

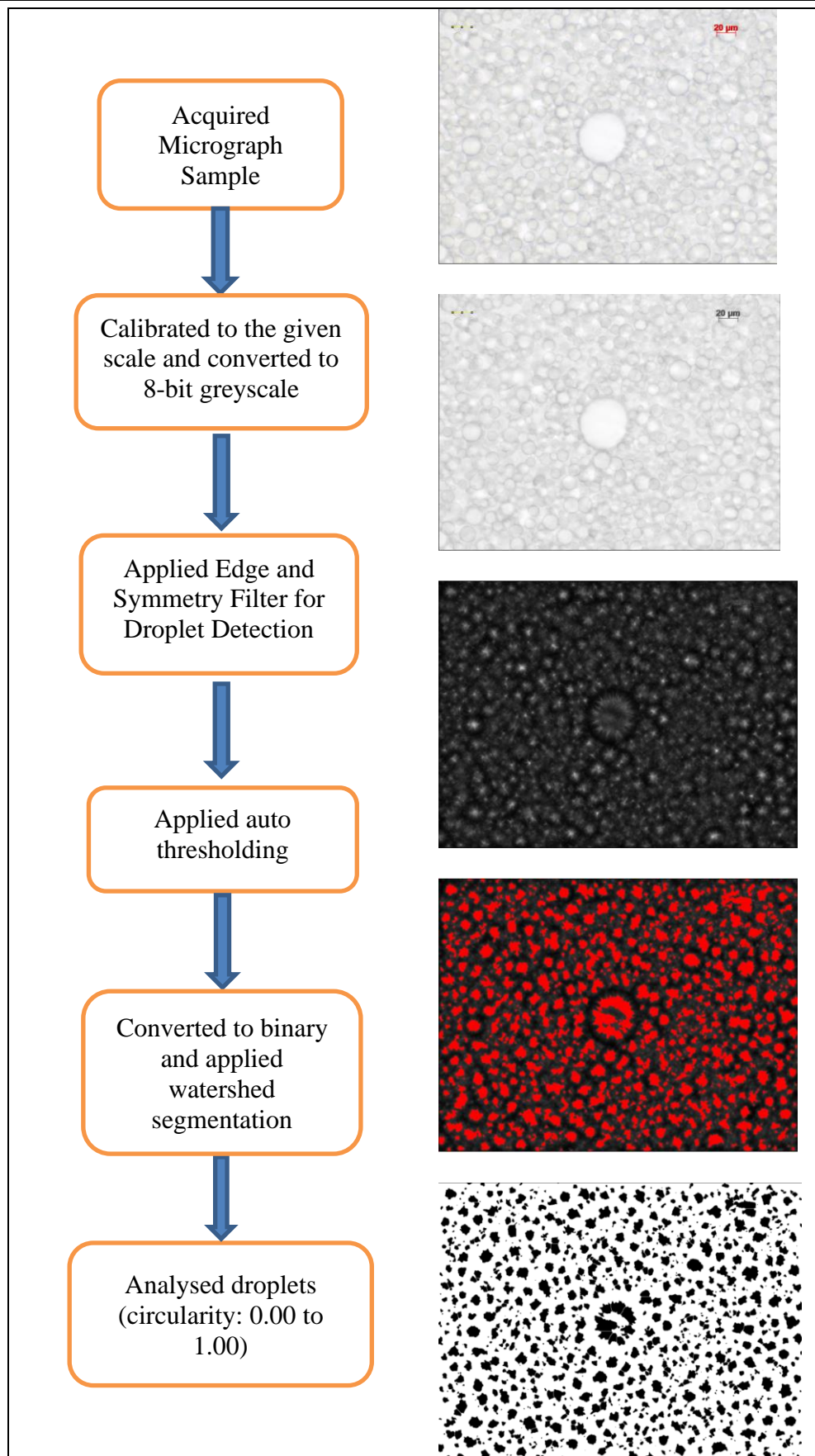
A sample detection of a circular object using the ESF algorithm is shown in Figure 3.2. The detected object is auto-thresholded (i.e. the default threshold value is set by Fiji) to obtain the output image.



**Figure 3.2. Detection of a circular object using edge and symmetry-based image segmentation.**

The optimum set of parameter values for the ESF algorithm, to detect the oil droplets in the micrographs, were obtained through design of experiments/DoE (Unnikrishnan et al., 2017). A central composite design was used and a desirability function approach was applied to detect the maximum droplets with minimum error (Lenth, 2009). The optimum input parameter values obtained were  $\alpha = 0.8$ ,  $n = 9.55$  and  $s = 2$  and these were held constant for the detection of droplets in all the 60 micrographs. The radii ( $r$ ) range was obtained as 4 to 40 pixels, through manual calibration, using trial and error.

A macro was programmed, in Fiji, to dynamically execute a series of image processing and segmentation steps to detect the droplets from the micrographs. The steps executed by the macro are schematically represented in Figure 3.3. Multiple runs of the macro, with different radii in the range of 4 to 40 pixels, were required to detect the various droplet sizes in the micrographs.



**Figure 3.3. Detection of oil droplets and characteristics using the Edge and Symmetry segmentation technique.**



---

When the macro is executed, a dialog box appears for the user to specify the input and the output directories.

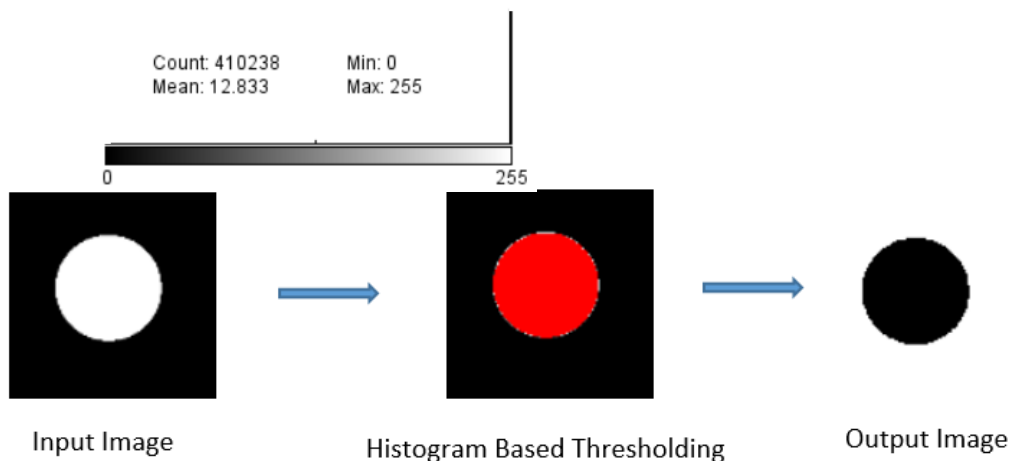
1. The micrographs were opened in a sequence from the user-specified directory.
2. Each micrograph was calibrated to a scale of 4 pixels/ $\mu\text{m}$ .
3. The micrographs were converted to 8-bit greyscale to enable further processing.
4. Edge and Symmetry filter (ESF) was applied using the optimum parameters to detect the droplets in the images.
5. The filtered images were auto-thresholded through the red channel to enhance the detected droplets.
6. The thresholded images were converted into binary and applied 'watershed segmentation' to separate the droplets that touch/overlap each other.
7. The detected droplets were analysed for a size range  $\geq 1 \mu\text{m}^2$  and a circularity range of 0.00 to 1.00, where 1.00 represents a perfect circle. Fifteen characteristics were extracted for each droplet:
  - Size features: area ( $\mu\text{m}^2$ ), perimeter ( $\mu\text{m}$ ), maximum Feret diameter (Feret in  $\mu\text{m}$ ), minimum Feret diameter (minFeret in  $\mu\text{m}$ )
  - Pixel Intensity measures: integrated density (ID) and raw Integrated density (rID)
  - Centriod coordinates: X and Y
  - Starting coordinates of the Feret (FeretX and FeretY)
  - Orientation feature: Feret angle
  - Shape features: circularity, roundness, aspect ratio and solidity.
8. The droplet characteristics were automatically exported to a CSV file in the user-specified output directory.

---

### 3.4.3 Histogram-Based Technique (HBT)

HBT is an alternative image segmentation technique developed in Fiji to detect the droplets in the micrographs. The technique is based on computing the histogram of the grey values (pixel intensity distribution) in the image and thresholding the image using the mean grey value. Histogram based methods have always shown potential in image segmentation compared to other segmentation techniques (Shapiro, 2001). In most cases, the technique requires only one pass through the pixels, which reduces the computing cost and time as well as making it viable to automate. Pixel intensity measures of droplets in micrographs have previously shown promising potential in detecting the filling level of the dispersed phase in emulsions (Schuster et al., 2012b).

A sample detection of a circular object using HBT is shown in Figure 3.4.



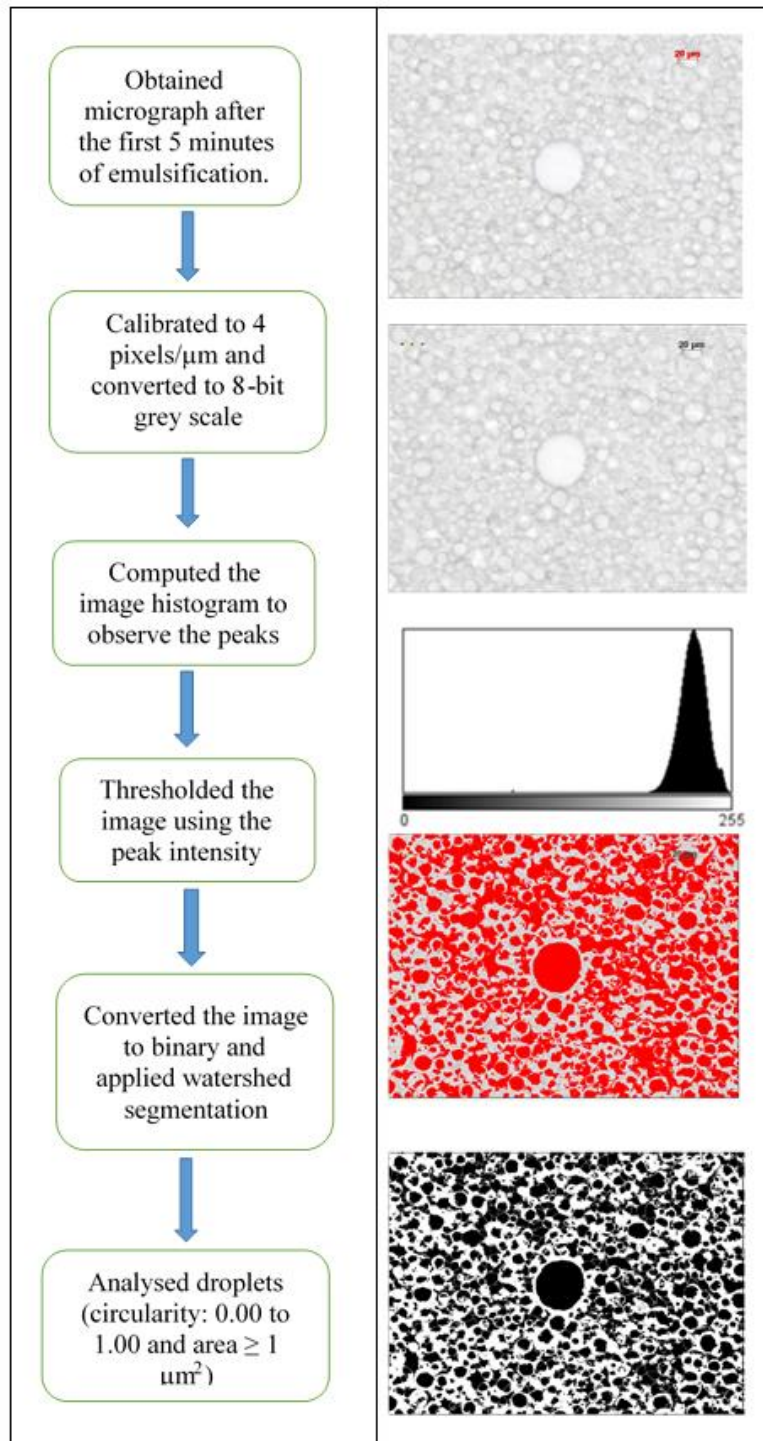
**Figure 3.4** Detection of a circular object using histogram-based image segmentation.

A macro was programmed in Fiji to execute the HBT segmentation. When the macro is executed, the following steps are performed:

1. A dialog box appears for the user to specify the input and the output directories.
2. The micrographs were opened from the user-specified directory in a sequence.
3. Each micrograph was calibrated to a micrometer scale of 4 pixels/ $\mu\text{m}$ .
4. The images were then converted to 8-bit greyscale.

- 
5. A histogram was computed from the pixel intensity values of each image.
  6. Each image was thresholded using the peak intensity value.
  7. The thresholded images were converted to binary and applied watershed segmentation to separate overlapping droplets and obtain the output image.
  8. The detected droplets are analysed for a size range  $\geq 1 \mu\text{m}^2$  and a circularity range from 0.00 to 1.00. Thirteen characteristics were extracted for each droplet similar to the EST and saved to a CSV file.

A diagrammatic representation of the macro execution is provided in Figure 3.5 and the formulae for the droplet characteristics are given in the Appendix A.



**Figure 3.5. Schematic showing the detection of oil droplets and their characteristics from an emulsion micrograph using the Histogram-Based image segmentation (HBT) macro in Fiji.**

A comparative analysis of the steps involved in the EST and HBT techniques, their advantages and limitations are presented on Table 3.2.

**Table 3-2. Comparative Analysis of the EST and HBT Techniques.**

<b>Image Segmentation Techniques</b>	<b>Steps Involved in the execution of EST and HBT</b>	<b>Advantages</b>	<b>Limitations</b>
EST	Micrographs were calibrated to 4 pixels/ $\mu\text{m}$ . These were converted to 8-bit greyscale. Edge and Symmetry filter (ESF) was applied for a set of optimum parameters. These were the estimated radii of the droplets to be detected, the Gaussian smoothing parameter and two internal parameters such as scaling and normalisation, which were kept constant. The filtered images were auto-thresholded through the red channel to enhance the detected droplets. The thresholded images were then converted into binary (black/white). Watershed segmentation was applied to separate the droplets that touch/overlap each other. The characteristics of the detected droplets were obtained for a specified size range $\geq 1 \mu\text{m}^2$ and a circularity range of 0.00 to 1.00.	This approach is effective in cases where the ROI has distinguishable edges with a strong intensity gradient.	The EST algorithm demands four input parameters such as radius of the ROI, a Gaussian smoothing parameter ( $\alpha$ ), two internal parameters such as normalisation ( $n$ ) and scaling ( $s$ ). The optimum values calculated for the three inputs except radius were $\alpha = 0.8$ , $n = 9.55$ and $s = 2$ . The radius value considerably varied for different droplet sizes throughout the emulsification process. The manual estimation of radii for smaller droplets was difficult.
HBT	Micrographs were calibrated to 4 pixels/ $\mu\text{m}$ and converted to 8-bit greyscale. For each image, a histogram of the pixel intensity values was computed, and the peak intensity value was calculated automatically in the code. The images were then thresholded using the calculated peak intensity value. The thresholded images were converted to binary and applied watershed segmentation to separate overlapping droplets and obtain the output image. The characteristics of the droplets were obtained for a given size range $\geq 1 \mu\text{m}^2$ and a circularity range from 0.00 to 1.00.	This is a similarity detection approach, which is quite useful if the ROI is a group of similar objects. In this technique, identical pixel values are grouped together.	The HBT algorithm does not demand any user input variables. The intensity calculation and thresholding are completely automated. Therefore, there are no limitations identified for this technique in relation to the droplet detection and characterisation.

Each technique has its own advantages. However, the limitations presented in Table 3.2 have identified HBT as a promising approach for a consistent detection of droplets through the emulsification process. The droplet characteristics obtained from both the techniques were also investigated using statistical plots to help identify the best technique.

---

### 3.5 Statistical Analysis of Droplet Characteristics

This section briefly describes the statistical approaches and the tools used for the analysis of the droplet characteristics extracted from the in-process micrographs. The statistical analysis was performed in RStudio version 1.1.383, which is an integrated development environment for R language (R Core Team, 2019). R version 3.4 was used in this study. The utilities of the R software are explained in subsection 3.5.1.

The evolution of each droplet characteristic throughout the emulsification process was investigated to identify the most suitable image segmentation technique, which gives a progressive evolution of the droplet formation. This was performed by plotting box plots of the mean droplet characteristics obtained using the EST and HBT techniques in order to observe the evolution. The droplet characteristics which varied significantly throughout the emulsification process were identified.

This was followed by a probability distribution analysis of the significant characteristics to find out if the droplets attain desirable characteristics towards the end of emulsification. Bar charts of droplet size characteristics were also analysed over the processing time to investigate the proportion of variation in the number of droplets from various size range. These plots are presented and analysed in detail in Chapter 4. The statistical variation of each droplet characteristic and the correlation between the characteristics were also investigated. The ggplot2 data visualisation package was used to create all the plots (Wickham, 2016).

The droplet characteristics, which presented significant variation throughout the emulsification process, were considered as the important features for classifying the

---

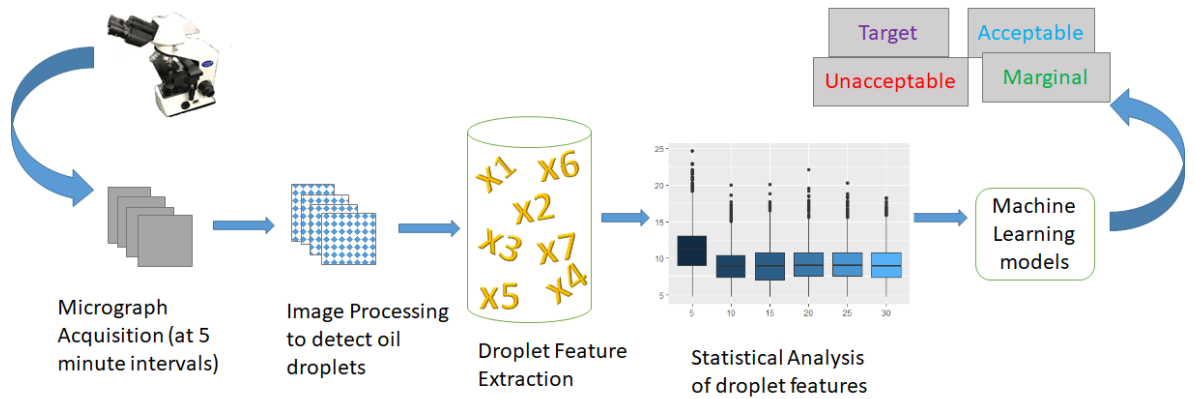
micrographs obtained at various intervals. The guidance obtained from the industrial collaborator was applied to categorise and label the micrographs into quality categories. The results of the statistical analyses and the criteria for micrograph categorisation are discussed in detail in Chapter 4.

### 3.5.1 R Statistical Software

R is a language and environment for statistical computing and graphics (R Core Team, 2019). It provides a wide range of packages for statistics, graphics, exploratory data analysis, linear and non-linear modelling, machine learning, deep learning etc. It also provides efficient data handling and storage facilities. RStudio is an integrated development environment built in order to make R programming more user friendly. User-defined R functions can be programmed and integrated with built-in R packages in RStudio. It provides programming, debugging and visualisation tools (<https://www.rstudio.com/>).

## 3.6 Micrograph Classification

Supervised and unsupervised machine learning models as well as deep learning models were developed to classify the in-process micrographs obtained from Process-2. The significant droplet characteristics, obtained from the statistical analysis, were considered as the original input feature space of the models. The nomenclature used for the response categories of Process-2 was TAMU (Target, Acceptable, Marginal and Unacceptable). The naming convention of the micrograph categories was recommended by the industrial collaborator based on the manual assessment of the emulsion quality and droplet characteristics. A schematic representation of the classification methodology is given in Figure 3.6.



**Figure 3.6. Schematic of the methodology applied for micrograph classification.**

The classification models developed in this study are:

- Principal Component based Linear Discriminant Analysis
- Multinomial Logistic Regression
- Random Forest
- Feed Forward Neural Network
- Convolutional Neural Network

Each model was evaluated using the micrographs obtained from Process-2. Additionally, the 150 micrographs obtained from the independent batch of Process-2 were also used to evaluate the accuracy of the models in classifying unseen data.

A detailed review of each of these machine learning techniques and their recent applications in industry have been discussed in Chapter 2 (section 2.2). The methodology of developing these models in this research is described in the following subsections, while the results of each model are explained in Chapter 5.

### 3.6.1 Principal Component based Linear Discriminant Analysis (PC-LDA)

A PC-LDA model was developed to classify the micrographs, obtained at various stages of emulsification, into the TAMU categories. The `prcomp` routine in the R base package



---

and the lda routine from the Mass package were used to develop the model (R Core Team, 2019; Ripley, 2002).

The droplet characteristics identified from the statistical analyses were considered as the input features of the model. An unsupervised grouping of the micrographs from the four categories was performed, using PCA, to observe the pattern and the separation between the categories. As the selected droplet characteristics were correlated, PCA was applied to achieve a reduced set of uncorrelated components. The selected characteristics of each individual droplet, obtained from the TAMU micrographs, formed the PCA observations. The resultant set of principal components (PCs), which explained a significant percentage of variance in the original features were selected to perform the LDA. This resulted in a set of discriminant functions, as the canonical variables, that classified each droplet into the TAMU categories. The final classification of each micrograph was based on the highest percentage of droplets classified from that micrograph. The PC-LDA model accuracy was evaluated using a stratified 5-fold cross-validation. Cross-validation is explained in section 3.6.2. Stratified cross-validation was applied because of the approach used in this model. The observations of the PC-LDA model are individual droplets rather than micrographs. This resulted in a large difference in the number of observations obtained in each TAMU category, as the droplet count increased significantly from ‘U’ to ‘T’ during the emulsification process. Stratified cross-validation was used to achieve a balanced set of observations across the TAMU categories. The model was also evaluated using the independent set of 150 micrographs. The results obtained from the PC-LDA approach are presented in Chapter 5.

A manual versus machine learning classification study was performed to evaluate the PC-LDA model developed using the EST segmentation approach (Figure 3.1). The

---

methodology and the results associated with this comparison study are presented in Chapter 6.

### 3.6.2 Cross-validation

Cross-validation (CV) is a repeated sampling technique used to assess the performance of a machine learning model on a limited data sample (Refaeilzadeh et al., 2009; Stone, 1974). The most commonly used technique is  $k$ -fold cross-validation, where  $k$  stands for the number of groups/folds that the original data set is split into. Each fold is further split into a test and a training data set. The percentage splits of the training and the test sets are defined according to the sample size. A model is created for each fold, trained using the training set and then validated using the test set. The model is discarded after recording the classification score. Finally, the overall classification score is calculated as the summation of the individual model scores, which is represented by a confusion matrix (Bertani et al., 2017). Cross-validation can be performed using either stratified or random sampling. Stratified sampling is applied in cases where the response categories have quite an imbalanced frequency (Bertani et al., 2017). In stratified sampling, an equal number of data sets should be selected from all the response categories for each fold created for the CV. The PC-LDA model, in this study, used stratified sampling because the individual observations of the model were the PCs corresponding to the selected characteristics of each droplet. All the other classification models, in this research, used random sampling, which adopts a random percentage split for each fold as appropriate for the size of the dataset.

### 3.6.3 Multinomial Logistic Regression (MLR)

The multinom routine from the R caret package was used to develop the MLR model (Max, 2020). The mean of the selected droplet size characteristics and the droplet count,

---

of each micrograph, were used as the input feature space of this model. This was different from the approach taken with the PC-LDA model. A single micrograph was represented by an individual observation in the MLR model. The four categories of micrographs formed the response classes. The Target category was selected as the reference. The probability of each micrograph being in one of the three remaining categories was calculated, relative to the reference category. A 10-fold cross-validation was applied to evaluate the classification accuracy of the model. A 70/30 split of the training and the test data sets was used in each fold. 10 models were created from the training sets of the 10 folds. Each model was then evaluated using the corresponding test data. The classification accuracies and the confusion matrices of the 10 models were recorded. A summation of the 10 confusion matrices was obtained to evaluate the cross-validation results. Finally, the model was trained using the complete set of micrographs and evaluated using the independent set of micrographs. The MLR model results are discussed in Chapter 5.

#### 3.6.4 Random Forest (RF)

The RF classification model was developed using the same five predictor variables as that of the MLR model. A single micrograph was represented by an individual observation in the RF model. The `rf` routine from the R `caret` package was used to build this model (Max, 2020). The model parameters, *n*tree and *m*try, were tuned in a step by step process. An initial model was trained using 70% of the 150 micrographs using the default parameter values. The OOB error estimate of the model was used as the metric to determine the optimum *n*tree value. The optimum *m*try value was selected through cross-validation. A 10-fold cross-validation was performed by creating 10 random 70/30 folds. For each fold, *m*try values from 1 to 5 were evaluated to select the best *m*try that gives the minimum OOB error. This was repeated for all the 10 folds to find the optimum *m*try. A model was developed from each fold using the optimum *n*tree and *m*try values. Each model was

---

evaluated using its test set. The classification accuracy and the confusion matrix were recorded for the 10 models. Finally, the model was trained using the complete set of micrographs and evaluated using the independent set.

### 3.6.5 Neural Network (NN)

A Vanilla Neural Network (VNN) classification model was developed as a deep learning approach to classify the TAMU micrographs. This is a single layer backpropagation and feed-forward neural network model. The `nnet` routine from the R `caret` package was used to build the model (Max, 2020). The VNN model used the same set of five predictor variables that were used in the MLR and the RF models. The number of units in the hidden layer and the weight decay parameter are the two hyper-parameters that require tuning in the VNN. A 10-fold cross-validation was applied for tuning these hyper-parameters. An 80/20 split was applied for each fold. Each model was evaluated using its test set. The classification scores were recorded and the sum of the 10 confusion matrices was obtained to evaluate the overall accuracy of cross-validation. The model was trained on the complete data set and evaluated on the independent batch. The classification results of the models are explained in detail in chapter 5.

### 3.6.6 Convolutional Neural Network (CNN)

CNNs are deep learning networks, which are widely applied to solve complex image classification problems. CNNs perform image feature extraction followed by multi-layer neural network-based classification or prediction. The `keras-2.2.4` and `tensorflow-hub-0.3.0` packages were used to train the CNN model (JJ & François, 2019; JJ & Yuan, 2019). Keras is a high-level neural network API (Application Programming Interface) used for developing and training deep learning models. The Keras framework comprises three major steps for model training:

- 
- Configuration of the model architecture.
  - Compilation of the model.
  - Fitting the model.

Keras runs the framework code on a CPU or GPU on top of multiple back-ends such as TensorFlow, Theano and others. The CNN model, employed in the current study, ran on a GPU using TensorFlow. Each of the 150 micrographs were split into four to obtain a larger dataset. These were used for developing the model. The split micrographs had a resolution of 694 x 520 pixels. An R function was written to perform the following steps:

- Resize the split micrographs to a square resolution ( $n \times n$ ) to perform convolution.
- Flatten the micrographs and extract the pixel intensity values.
- Store the pixel values in an input matrix.
- Obtain the micrograph labels, encode as binary and store in a response matrix.

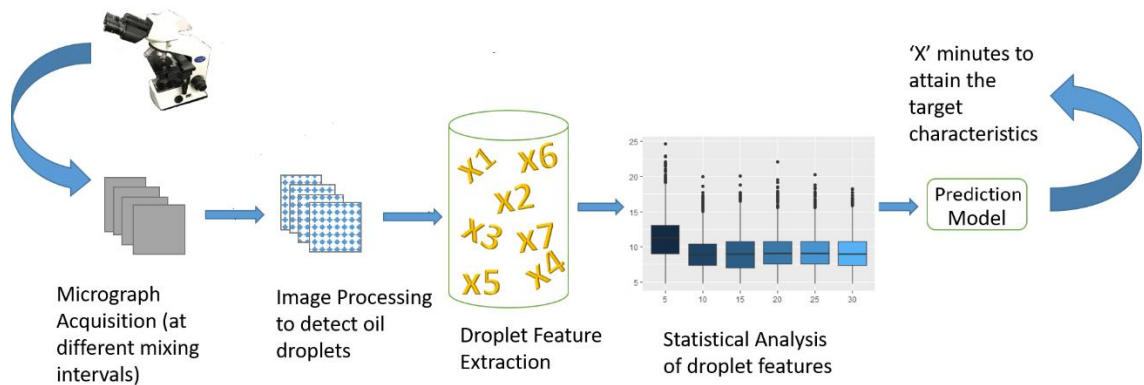
A Keras sequential CNN model was configured using a set of the specified layers. 80% of the whole dataset was used to train the model. The model results are discussed in Chapter 5.

The results obtained from the classification models were compared and the approaches that presented the best TAMU classification were considered to develop regression models. The purpose of the regression model is to predict the remaining processing time, at any stage of the industrial processing, to complete the emulsification process. The following section describes the regression modelling approaches used in this research.

### **3.7 Remaining Processing Time Prediction**

The objective of the prediction model is to overcome the existing challenges in industry, including over-processing, excessive energy utilisation, wastage of resources and ultimately, the cost of production. This can be achieved by predicting ‘how much more

time is required by the emulsion to achieve the target droplet characteristics’ during an emulsification process. A schematic representation of the prediction methodology is presented in Figure 3.7.



**Figure 3.7. Methodology applied for optimum processing time prediction.**

The processing time required by the emulsion to attain the target droplet characteristics was identified as the optimum processing time of Process-2. For each micrograph, the difference between its “current process time” and the “optimum processing time identified” was calculated. These values were used as the response data to train the prediction models. The predictor variables selected for these models were the same five variables, which were found significant during classification. The models were trained with 70% of the micrographs obtained from Process-2. The remaining 30% was retained as test data for validation. The potential of the models to predict the processing time required by a new (unseen) set of emulsion samples to achieve the Target was also validated. The prediction model results are discussed in Chapter 7.

### 3.8 Conclusions

This chapter presents the overall approach undertaken in this research to develop a novel automated method for evaluating in-process emulsion quality as well as identifying the optimum processing time. Emulsion samples were obtained, at five-minute intervals,

---

from two different emulsification processes, a trial process called Process-1 and an industrial process called Process-2. Process-1 was initially investigated as a proof of concept, to evaluate the image segmentation techniques.

Bright Field micrographs of 40x magnification were acquired from the samples. An edge and symmetry-based technique and a histogram-based technique were evaluated for image segmentation and subsequent droplet detection from the micrographs. A tabular analysis of the advantages and the limitations of the techniques was performed. The emulsion droplet characteristics determined by both techniques were statistically analysed. The results were compared to find the technique that best characterised the droplet evolution throughout the emulsification process. Those droplet characteristics which showed progressive variation over the emulsification process were identified and selected as the feature space for developing the classification models. The micrographs obtained from Process-2 were labelled as TAMU under the guidance of industrial experts who collaborated on this research work. Supervised machine learning models were developed and trained using a certain percentage of these micrographs, which were partitioned as the training set and the remaining percentage was used to validate the classification accuracy of the models. Cross-validation was used to evaluate the models. The optimum emulsification time of the industrial process was identified. Prediction models were developed and trained to predict the processing time required to reach the optimum completion point from any stage of the emulsification process. The results of the applied methodologies are discussed in the following chapters.

---

## **Chapter 4**

### **Automated Droplet Characterisation**

#### **4.1 Introduction**

This chapter discusses the results obtained from image processing of micrographs of a topical cream emulsion. This is followed by a discussion of droplet characterisation and their statistical analysis. The Edge & Symmetry (EST) and the Histogram-Based techniques (HBT) were the two image segmentation procedures applied for droplet detection. As a proof of concept, the micrographs obtained at five-minute intervals from Process-1 were processed using the two image segmentation macros, EST and HBT, to detect droplets and their characteristics. The droplet characteristics were then analysed to investigate their evolution throughout the emulsification process. Previous studies identified droplet size as a key factor in the control and optimisation of emulsification processes (Panckow et al., 2017). The results from the analysis, obtained from both segmentation techniques, were compared to identify the most effective technique for detecting droplets, of various sizes, from a range of micrographs. The technique identified from this proof of concept phase was applied for the investigation of Process-2.

#### **4.2 Results and Discussion**

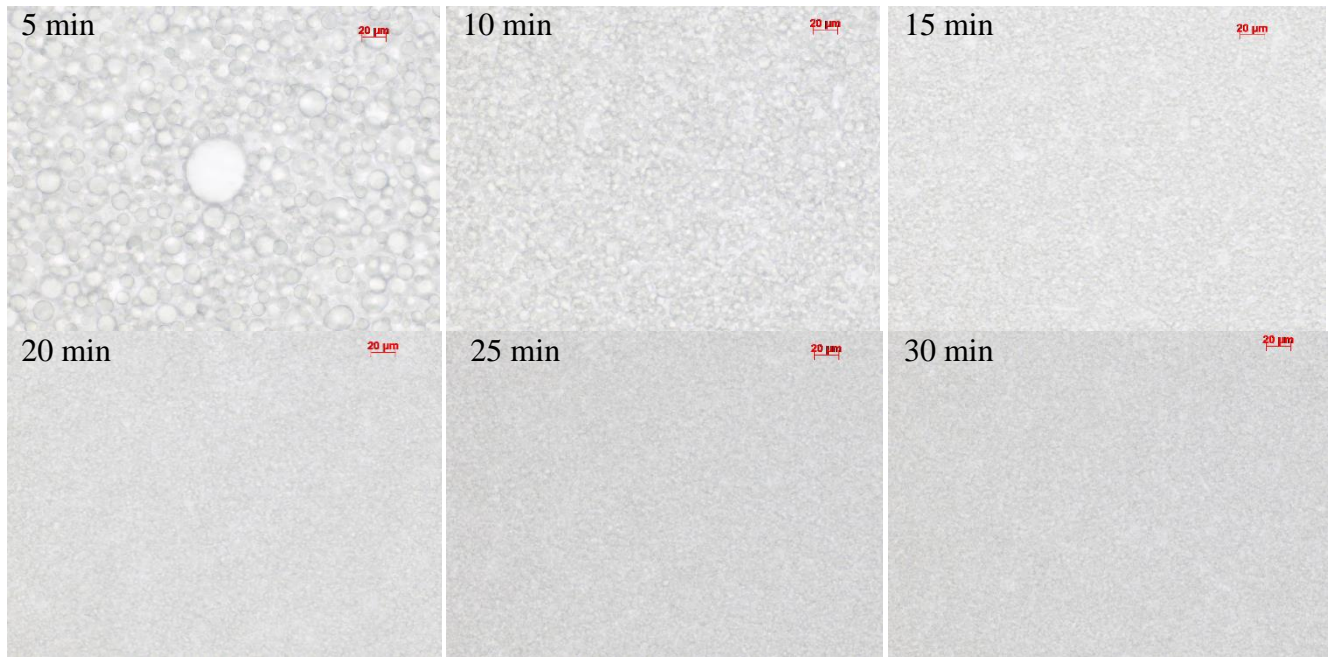
##### **4.2.1 Process-1: Droplet Detection and Analysis**

A total of 60 Bright Field micrographs of 40x magnification, 10 every five minutes, were acquired during the emulsification process. One sample micrograph obtained at each interval, from 5 to 30 minutes, is shown in Figure 4.1. The 60 micrographs were processed in Fiji using the EST and HBT macros. The droplet characteristics were automatically exported by the macros to a CSV file created in the user-specified directory. In addition to



---

this, a summary CSV file was also saved, which contained a summary of the droplet characteristics of each micrograph. The summary consists of droplet count, average droplet area, % concentration of droplets and the average of perimeter, circularity, minimum and maximum Feret diameters respectively.



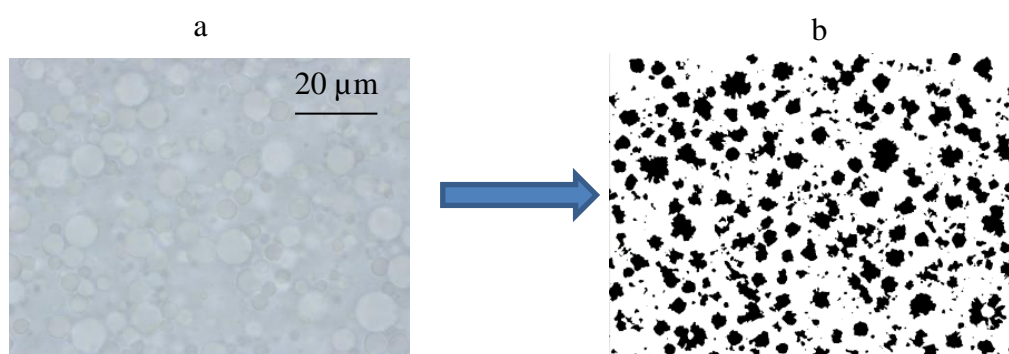
**Figure 4.1. Micrographs acquired at 5, 10, 15, 20, 25 and 30 minutes of the 30 minutes duration emulsification process.**

#### **4.2.1.1 Analysis of EST Results**

The Edge and Symmetry Technique (EST) and its input parameters are explained in section 3.4.2. The four input parameters required by this algorithm are the alpha Canny ( $\alpha$ ), radius ( $r$ ), normalisation ( $n$ ) and scaling ( $s$ ). These parameter values were selected through multi-response optimisation using a central composite design and desirability function (Unnikrishnan et al., 2017). Figure 4.2 shows the droplets detected from a sample micrograph, using the EST method, after the initial five minutes of emulsification. Fifteen characteristics were extracted for each droplet as shown in Table 4.1.

**Table 4-1. Droplet characteristics obtained from the EST.**

Size	Area ( $\mu\text{m}^2$ ), perimeter ( $\mu\text{m}$ ), Feret ( $\mu\text{m}$ ) and minFeret ( $\mu\text{m}$ )
Pixel Intensity	Integrated density (ID) and raw Integrated density (rID)
Centroid Coordinates	X and Y
Coordinates of Feret	FeretX and FeretY
Orientation	Feret angle
Shape	circularity, roundness, aspect ratio and solidity



**Figure 4.2. Droplet Detection using EST. a) A five-minute processed emulsion sample micrograph. b) Droplets detected.**

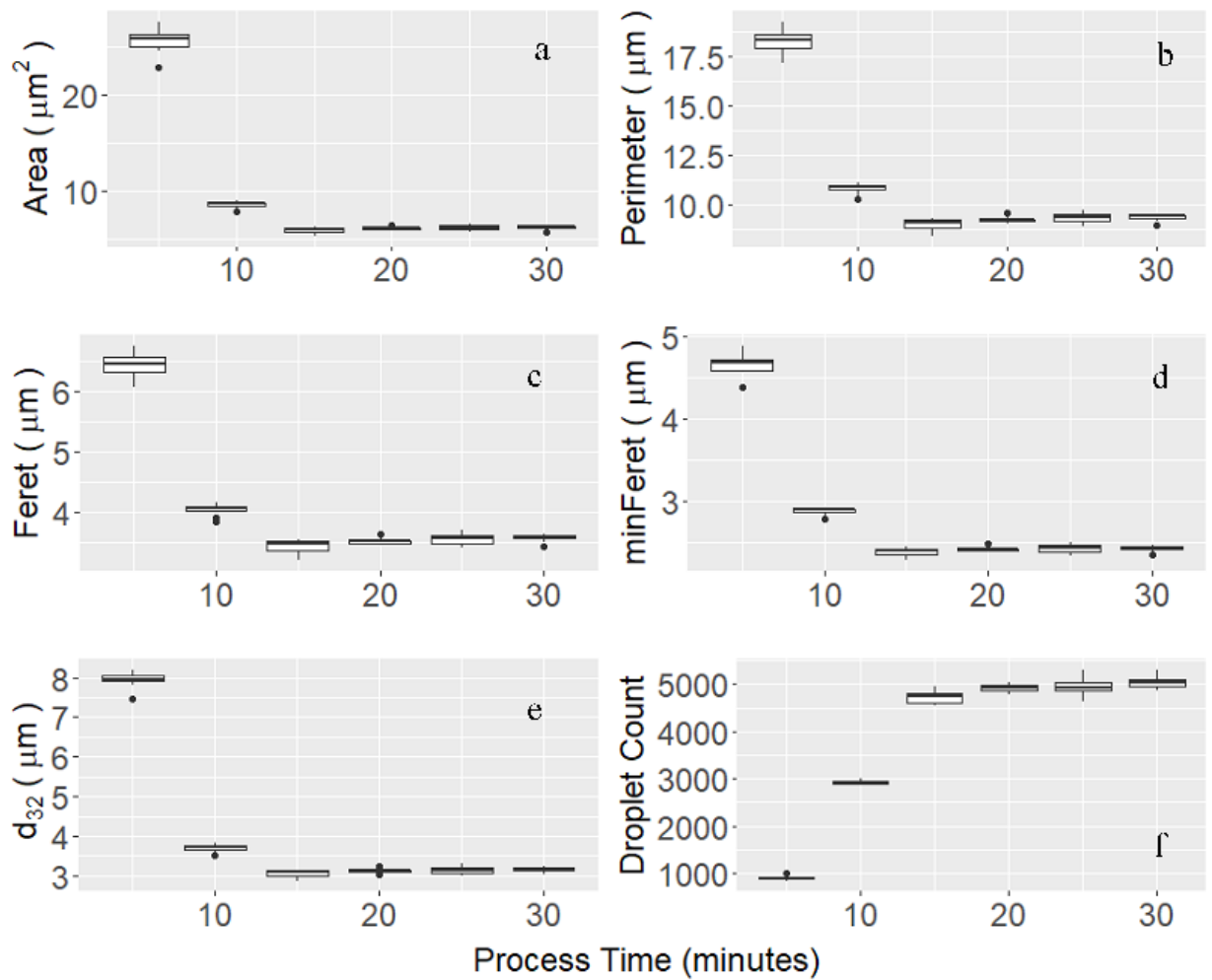
The droplets identified by the EST method appear to have detected the core of the droplets, to a certain extent, but not their actual size (Figure 4.2). The technique was highly dependent on the radius parameter ( $r$ ) of the ROI specified by the user for the droplet detection. This required multiple iterations of the macro, with varying values of  $r$ , to analyse a single micrograph and detect all droplet sizes.

Approximately 800 to 1,000 droplets were detected from a typical micrograph acquired after the initial five minutes of emulsion processing. This increased to about 5,000 droplets in micrographs acquired after 20 minutes of processing. The individual droplet

---

characteristics were analysed using box plots. Among the 15 characteristics extracted from each droplet, the four size features including area, perimeter, Feret and minFeret along with the droplet count showed significant variation during the emulsification process. The pixel intensity characteristics, ID and rID, were found redundant as they are a function of droplet area. The remaining characteristics such as shape, centroid and orientation presented no variation throughout the manufacturing process and therefore, were not considered any further for this study. The box plots of these droplet characteristics, which were found irrelevant are presented in Appendix C.

The box plots presenting the droplet size characteristics and droplet count are shown in Figure 4.3a to 4.3f. The equivalent circle diameter of each droplet, i.e., the diameter of a circle with the same area as the droplet, was calculated as  $d$  using the formulae,  $A = \pi r^2$  and  $d = 2r$  (where,  $A$  is the area,  $\pi$  is a constant and  $r$  is the radius). This value was used to compute the Sauter mean diameter ( $d_{32}$ ), otherwise known as area weighted average diameter, with the formula given in Equation 1.1. Figure 4.3e shows the  $d_{32}$  analysis and 4.3f presents the droplet count.



**Figure 4.3. Box Plots of mean droplet size characteristics and count from EST. a) area in  $\mu\text{m}^2$ , b) perimeter in  $\mu\text{m}$ , c) Sauter mean diameter d) maximum Feret diameter and e) minimum Feret diameter. Each box plot represents 10 micrographs.**

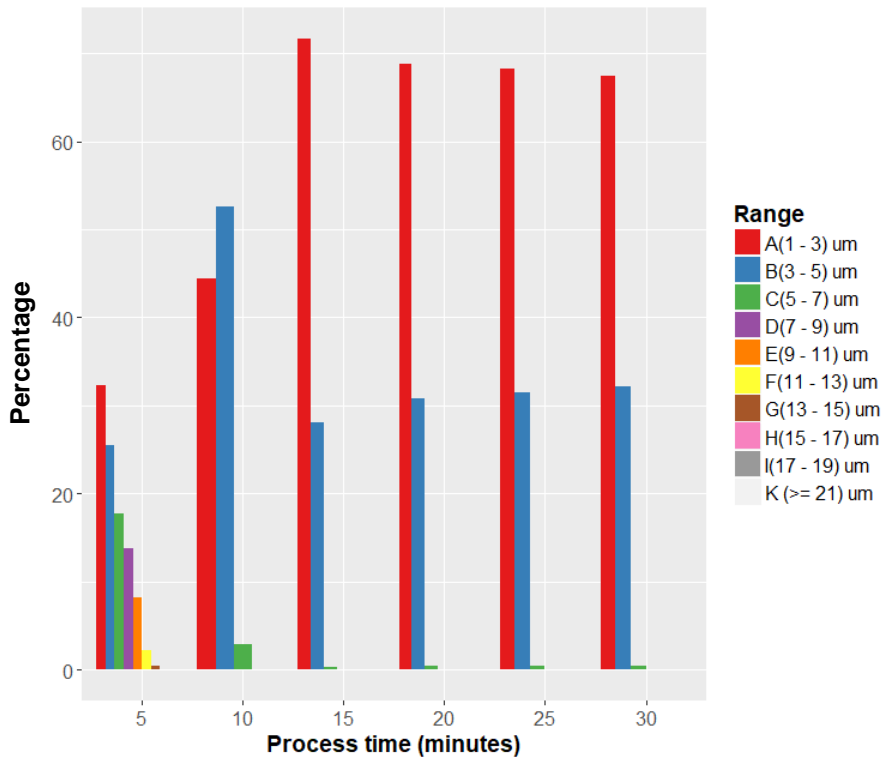
The average droplet size presented a sharp decrease (Figure 4.3a to 4.3e), as expected, in the initial 10 minutes and a smaller decrease in the following five minutes. After a further five minutes of processing, i.e. after 20 minutes in total, the average droplet size showed a slight unexpected increase. During the final 15 minutes of processing, there was minimal variation in the droplet size. The droplet count showed a progressive increase during the initial 15 minutes of emulsification and remained almost steady for the last 15 minutes (Figure 4.3f).

---

The slight increase in the average droplet size after 20 minutes followed by minimal variation afterwards could be due to either droplet coalescence (smaller droplets joining together) or the lack of detection of smaller droplets by the algorithm. However, observing the droplet count, which increased slightly after 20 minutes, eliminates the possibility of coalescence. The overall variation in the mean droplet characteristics during the 30-minute process can be summarised as follows:

- Area decreased from 25.6 to 6.2  $\mu\text{m}^2$ .
- Perimeter decreased from 18.3 to 9.4  $\mu\text{m}$ .
- Sauter mean diameter ( $d_{32}$ ) decreased from 7.95 to 3.2  $\mu\text{m}$ .
- Maximum Feret diameter (Feret) decreased from 6.4 to 3.6  $\mu\text{m}$ .
- Minimum Feret diameter (minFeret) decreased from 4.7 to 2.4  $\mu\text{m}$ .
- Droplet count increased from 840 to 5000.

The variation in the percentage of individual droplets ranging from 1  $\mu\text{m}$  to  $\geq 21 \mu\text{m}$  diameter, at five-minute intervals, was also observed. Bar plots representing this variation are presented in Figure 4.4.



**Figure 4.4. Bar plots of the equivalent circle diameter of droplets obtained from EST. The colours represent the diameter ranges of the droplets as given in the legend.**

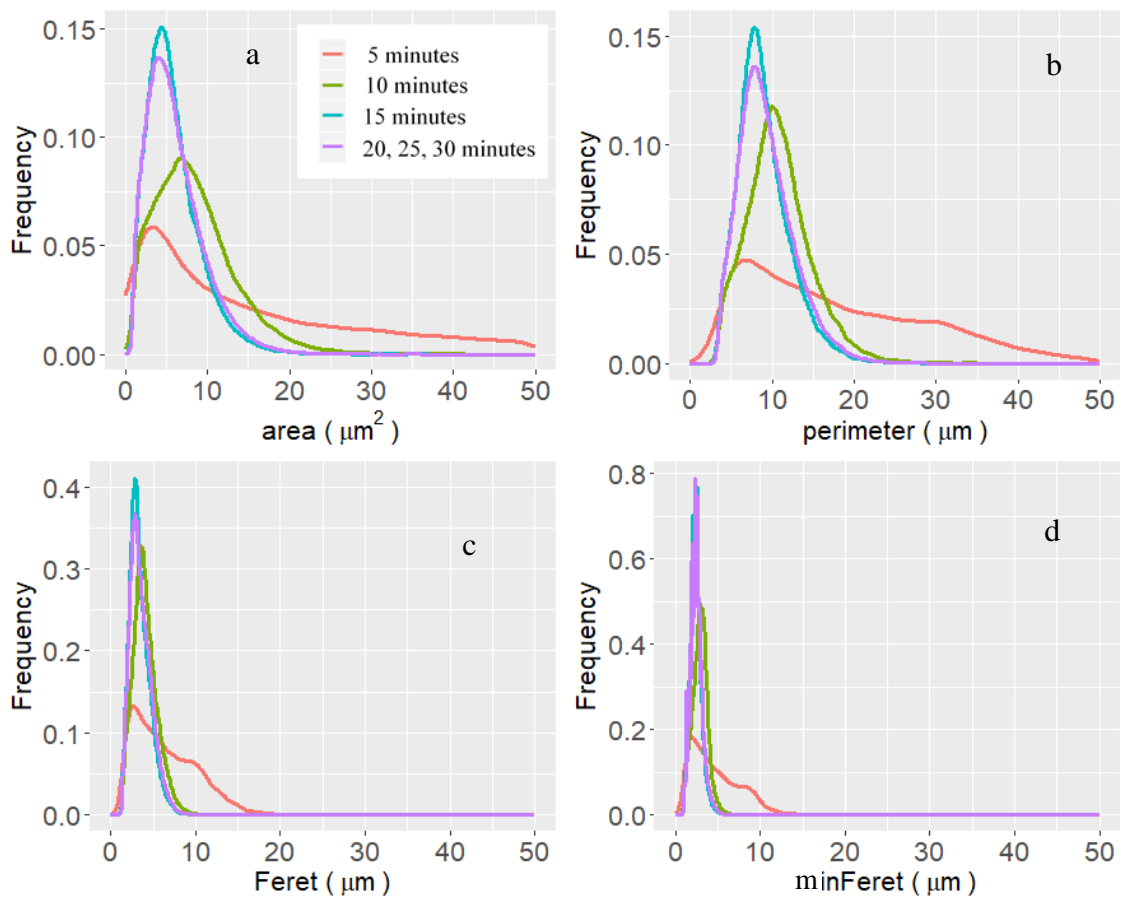
During the initial 15 minutes of processing, the bigger droplets with  $d > 7 \mu\text{m}$  disappeared and the percentage of smaller droplets (1-5  $\mu\text{m}$ ) increased to *ca.* 99%. However, the percentage of the smallest droplets, in the 1 to 3  $\mu\text{m}$  range, unexpectedly decreased after 20 minutes, while the droplets in the 3 to 5  $\mu\text{m}$  range presented a slight increase. This observation is consistent with the marginal increase previously noticed in the average droplet size after 20 minutes (see box plots from Figure 4.3a to 4.3e).

The droplets appeared as image texture rather than discrete particles in the micrographs obtained after 20 minutes of processing (Figure 4.1). This created difficulty in identifying an accurate set of values for the droplet radii ( $r$ ) towards the end of the emulsification process. This was the major limitation of the EST technique. The unexpected increase in

---

the average droplet size after the initial 20 minutes and the negligible variation observed in the average size and count in the last 10 minutes of the process could be attributed to the lack of detection of smaller droplets by the EST segmentation. In addition to this, the micrographs acquired after the initial 5 to 10 minutes of emulsification appeared to have a varying range of droplet sizes (Figure 4.4). In such cases, multiple detections with varying values of radii ( $r$ ) were required for a single micrograph. This was considered inefficient in terms of the overall droplet detection time and accuracy. In addition, the requirement for multiple detections also presented difficulty in achieving an optimum set of radii to automate the image segmentation process.

Frequency/density plots were also developed to study the variation in the overall droplet size distribution during emulsification. These are presented in Figure 4.5.



**Figure 4.5. Frequency plots of droplet size characteristics obtained from EST: a) area b) perimeter c) Feret and d) minFeret. The colours signify the process interval as shown in the legend.**

The frequency distribution of droplet size in terms of area, perimeter, Feret and minFeret are shown in Figure 4.5. It is noted that their distributions narrowed towards the end of the emulsification process. However, a slight decrease in the distribution peak was observed for the 20 to 30 minutes curve (purple colour). This indicated that the frequency/number of droplets started dropping after 20 minutes of emulsification.

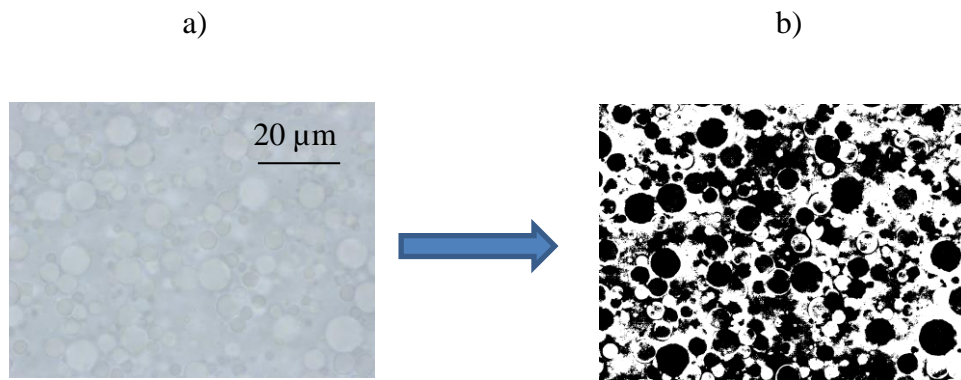
#### **4.2.1.2 Analysis of HBT Results**

The Histogram Based image segmentation technique (HBT) is explained in detail in section 3.4.3. Unlike EST, HBT is an entirely automated procedure, which is not dependent on any user input parameters other than the input and output directories. The droplets and their corresponding characteristics, detected using HBT, were analysed and



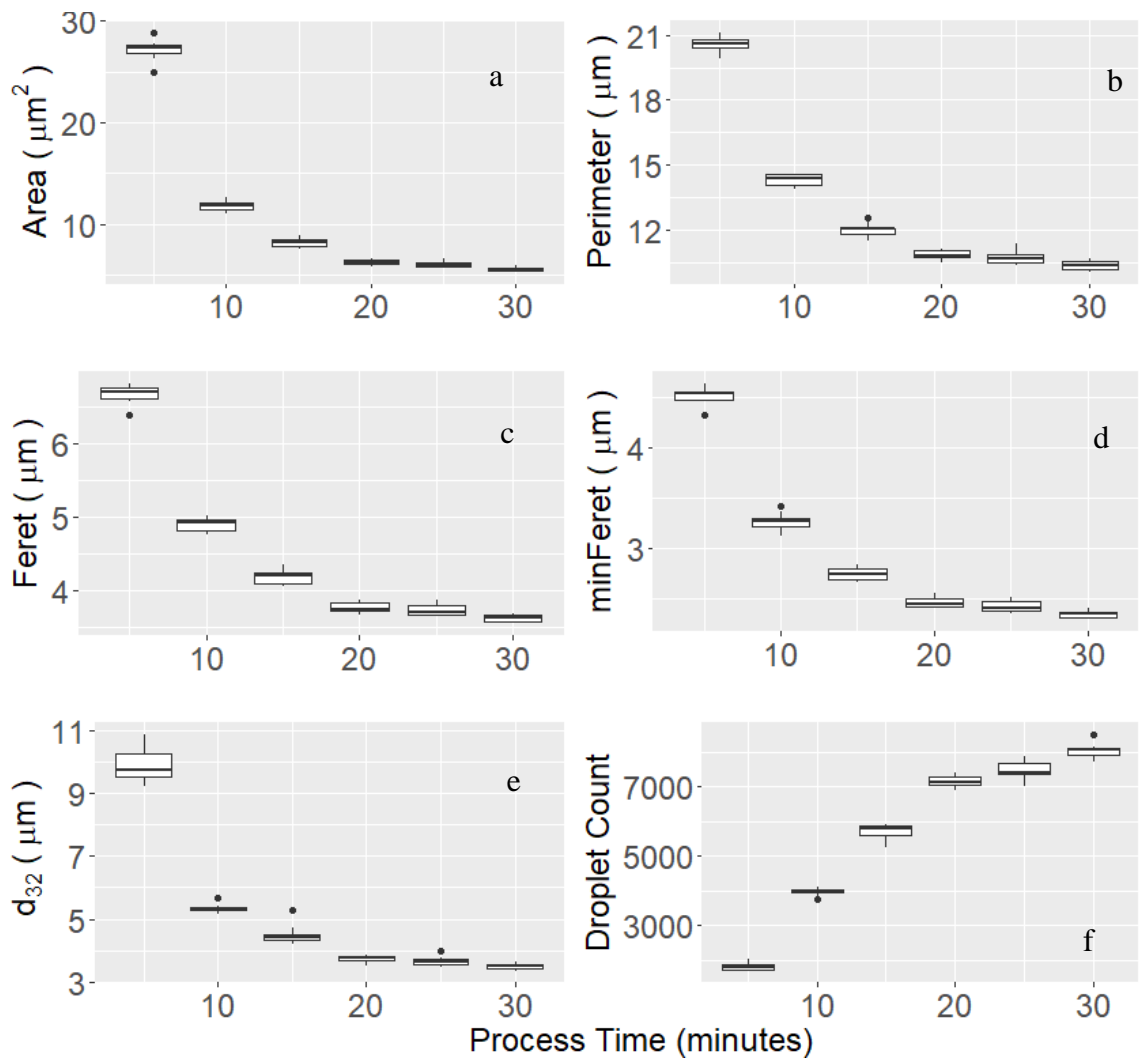
---

compared with the EST result. The droplets detected from a five-minute processed sample micrograph are shown in Figure 4.6.



**Figure 4.6. HBT droplet detection. (a) A sample micrograph taken after five minutes, (b) Output image from the Histogram-based technique.**

Thirteen characteristics were extracted from each droplet and were saved as a CSV file. These droplet characteristics were the same as those of the EST (Table 4.1). The HBT technique demonstrated better detection of varying droplet sizes in a single pass (Figure 4.6). Approximately 1,500 to 2,000 droplets were detected from a single micrograph after the initial five minutes of the emulsification process. The droplet count went up to ca. 8,000 towards the end of the process (i.e., 30 minutes). The variation in the droplet count and the average of droplet size characteristics, such as area, perimeter, Feret, minFeret, and Sauter mean diameter ( $d_{32}$ ), throughout the process were analysed using the box plots shown in Figure 4.7. The remaining characteristics such as shape, centroid and orientation presented no variation throughout the emulsification process and therefore, were not considered significant for this study. The box plots of those characteristics, that were found irrelevant for this study, are presented in Appendix C.



**Figure 4.7. Process-1. Box plots of the mean droplet size characteristics and count from HBT: a) area, b) perimeter, c) Feret, d) minFeret, e) Sauter mean diameter and f) Droplet count. Each box plot represents 10 micrographs.**

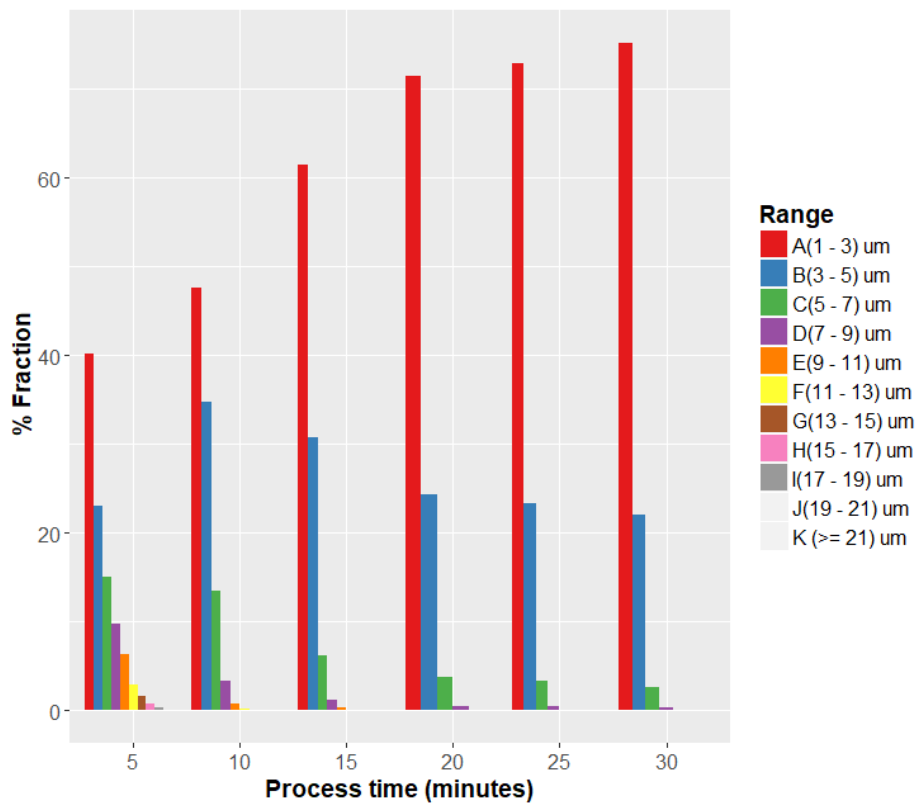
Each box plot, given in Figure 4.7a to 4.7e, represents the average droplet size in terms of area, perimeter, Feret, minFeret and  $d_{32}$  obtained from the 10 micrographs acquired at every five-minute interval of the process. The average droplet size presented a sharp decrease after the first five minutes of processing followed by a progressive decrease up to 20 minutes. Minimal variation was observed afterwards. The average droplet area

---

diminished from  $27.1 \mu\text{m}^2$  to  $5.6 \mu\text{m}^2$  during the 30-minute process, while the droplet count increased from 1,500 to 8,500. During the final 10 minutes of processing, minimal variation was observed in the average droplet area ( $6.2 \mu\text{m}^2$  to  $5.6 \mu\text{m}^2$ ) and count (7,000 to 8,000). This indicated that the process was approaching a steady equilibrium state. The overall variation in the mean droplet characteristics during the whole 30-minute process can be summarised as follows:

- Area decreased from 27.1 to  $5.6 \mu\text{m}^2$ .
- Perimeter decreased from 20.6 to  $10.3 \mu\text{m}$ .
- Sauter mean diameter ( $d_{32}$ ) decreased from 9.9 to  $3.5 \mu\text{m}$ .
- Maximum Feret diameter (Feret) decreased from 6.7 to  $3.6 \mu\text{m}$ .
- Minimum Feret diameter (minFeret) decreased from 4.5 to  $2.3 \mu\text{m}$ .
- Droplet count increased from 1,500 to 8,500.

Additional analysis was performed to investigate variation in the droplet diameter distribution during the 30-minute process. Figure 4.8 shows the distribution of the equivalent circle diameter of the droplets, over various ranges from A ( $1-3 \mu\text{m}$ ) to K ( $\geq 21 \mu\text{m}$ ).

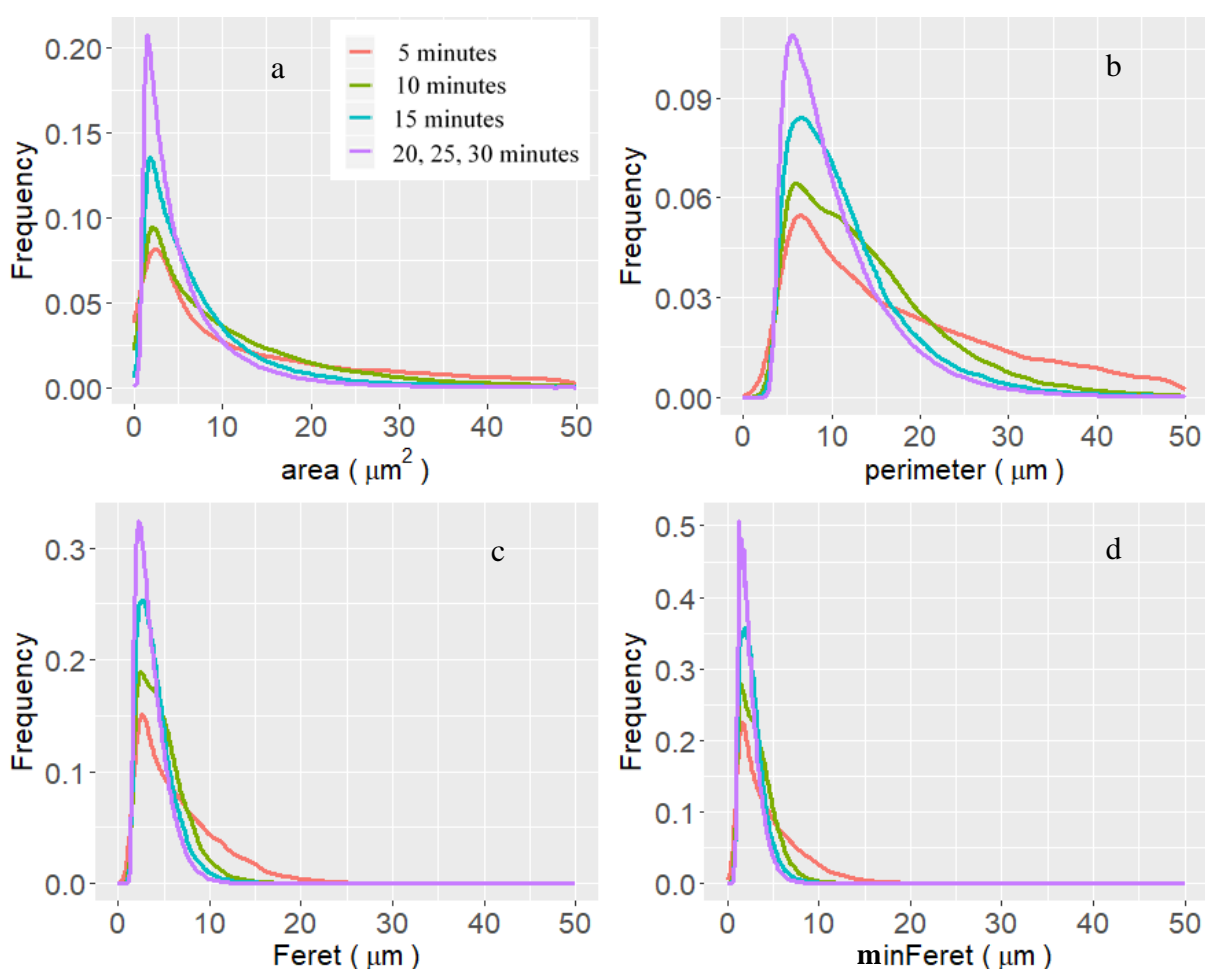


**Figure 4.8. Bar plots of the equivalent circle diameter of droplets from HBT. The colours represent the different diameter ranges of the droplets as given in the legend.**

The micrographs acquired after the initial five minutes of emulsification had droplets detected from a wide range of diameters ranging from 1  $\mu\text{m}$  to  $\geq 21 \mu\text{m}$ . 40% of these droplets were in the 1 to 3  $\mu\text{m}$  range, 23% were between 3 to 5  $\mu\text{m}$ , 25% between 5 to 9  $\mu\text{m}$ , 11% between 9 to 15  $\mu\text{m}$  and the remaining 1% was  $\geq 15 \mu\text{m}$ . As the emulsification process approached 20 minutes, the % of droplets in the 1 to 3  $\mu\text{m}$  range increased to 72% and simultaneously, the diameter ranges  $> 9 \mu\text{m}$  started diminishing. The maximum diameter range that existed after 20 minutes was 9 to 11  $\mu\text{m}$ , which contributed only 0.1% of the total. After a further 10 minutes of processing, the lowest range of droplets (1 to 3  $\mu\text{m}$ ) exhibited a steady rise along with a decrease of those in the 3 to 5  $\mu\text{m}$  range. As a result, the droplets in the 1 to 3  $\mu\text{m}$  diameter range increased to 75% of the total population after 30 minutes of processing. The percentage of droplets, with diameters  $\geq 1 \mu\text{m}$ , at the end of the emulsification process can be summarised as follows:

- 1 to 3  $\mu\text{m}$ : 75%
- 3 to 5  $\mu\text{m}$ : 22%
- 5 to 7  $\mu\text{m}$ : 2.5%
- 7 to 9  $\mu\text{m}$ : 0.25%

Frequency plots were also prepared to investigate the variation in the overall droplet size distribution during emulsification. These plots are presented in Figure 4.9. The droplet size characteristics obtained after 20 to 30 minutes of processing were grouped together, as they presented minimal variation.



**Figure 4.9.** Frequency plots of the droplet size characteristics obtained from HBT: a) area b) perimeter c) Feret and d) minFeret. The colours signify the process interval as given in the legend.

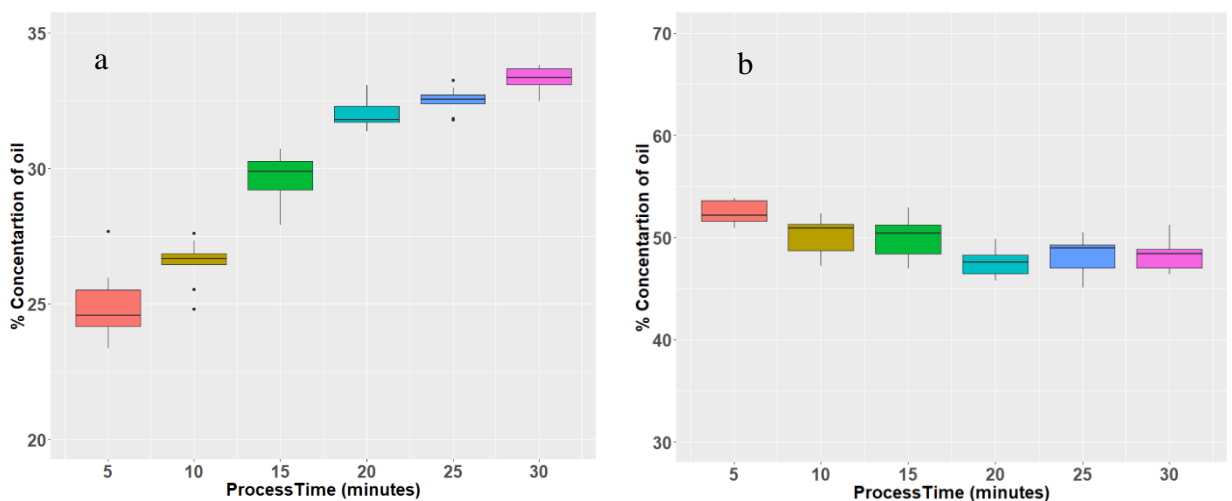
The frequency plots for area, perimeter, Feret and minFeret, narrowed to a sharp curve as the emulsification process approached 20 to 30 minutes. This was a good indication of

---

the droplet size approaching a narrow distribution as expected in an emulsification process.

#### 4.2.1.3 Comparison of EST and HBT

The HBT approach demonstrated a very promising evolution of droplet size and count throughout the process compared to EST. The HBT method was found capable of detecting both larger and smaller droplets thus providing a better representation of droplet evolution. The two image segmentation techniques, EST and HBT, were also evaluated by comparing the total oil concentration detected in the micrographs, as this was expected to be reasonably consistent throughout the emulsification process. The percentage of oil detected in the micrographs was obtained from Fiji as a summary characteristic of each micrograph. This was calculated as the ratio of the total area of droplets detected in a micrograph to the total area of the micrograph. The box plots showing the percentage of oil concentration detected from EST and HBT over the processing period are presented in Figure 4.10.



**Figure 4.10. Box plots showing the % of oil concentration from 5 to 30 minutes of emulsification. Each box plot represents 10 micrographs. (a) Edge & Symmetry Technique, (b) Histogram-Based Technique.**

---

The box plots, obtained using EST, showed notable inconsistency in the oil concentration throughout the process. The values on the y-axis, in Figure 4.10a, varied from 23% to 33%, which did not agree with the oil to water ratio of the product according to its formulation. This can be explained by the limited detection of oil droplets of varying sizes at each stage of emulsification due to the difficulty in estimating accurate radii values. In contrast, the oil concentration obtained using the HBT method presented a minimal variation from 47% to 53%, which was in close agreement with the oil content of the emulsion product formulation ( $\approx 50\%$ ). This is shown in Figure 4.10b. The slight drop in the percentage of oil, as seen in Figure 4.10b, can be explained by the amount of oil attaching to the sides of the stirring vessel during the process.

In summary, the evolution of droplet size, count and the oil concentration results demonstrated that the HBT method, in comparison to EST, was more proficient in the detection of oil droplets from the in-process micrographs. In addition to this, the HBT technique does not require any user-defined input parameters, enabling its automated implementation as a soft sensor. The EST method, in contrast, had a distinct disadvantage of recalibrating the radius parameter as droplet sizes varied. This affected the overall accuracy and speed of droplet detection and presented additional difficulty in automating the EST technique.

A precise understanding of droplet size distribution has been identified, in the literature, as the key factor to control and optimise industrial emulsification processes (Panckow et al., 2017). The existing droplet detection algorithms have focused on identifying the droplet border, using edge detection, at the phase change between oil and water (Hosseini et al., 2015). These techniques have been unsuccessful in detecting droplets from emulsions with a dispersed phase fraction greater than 10 to 15% (Khalil et al., 2010;

---

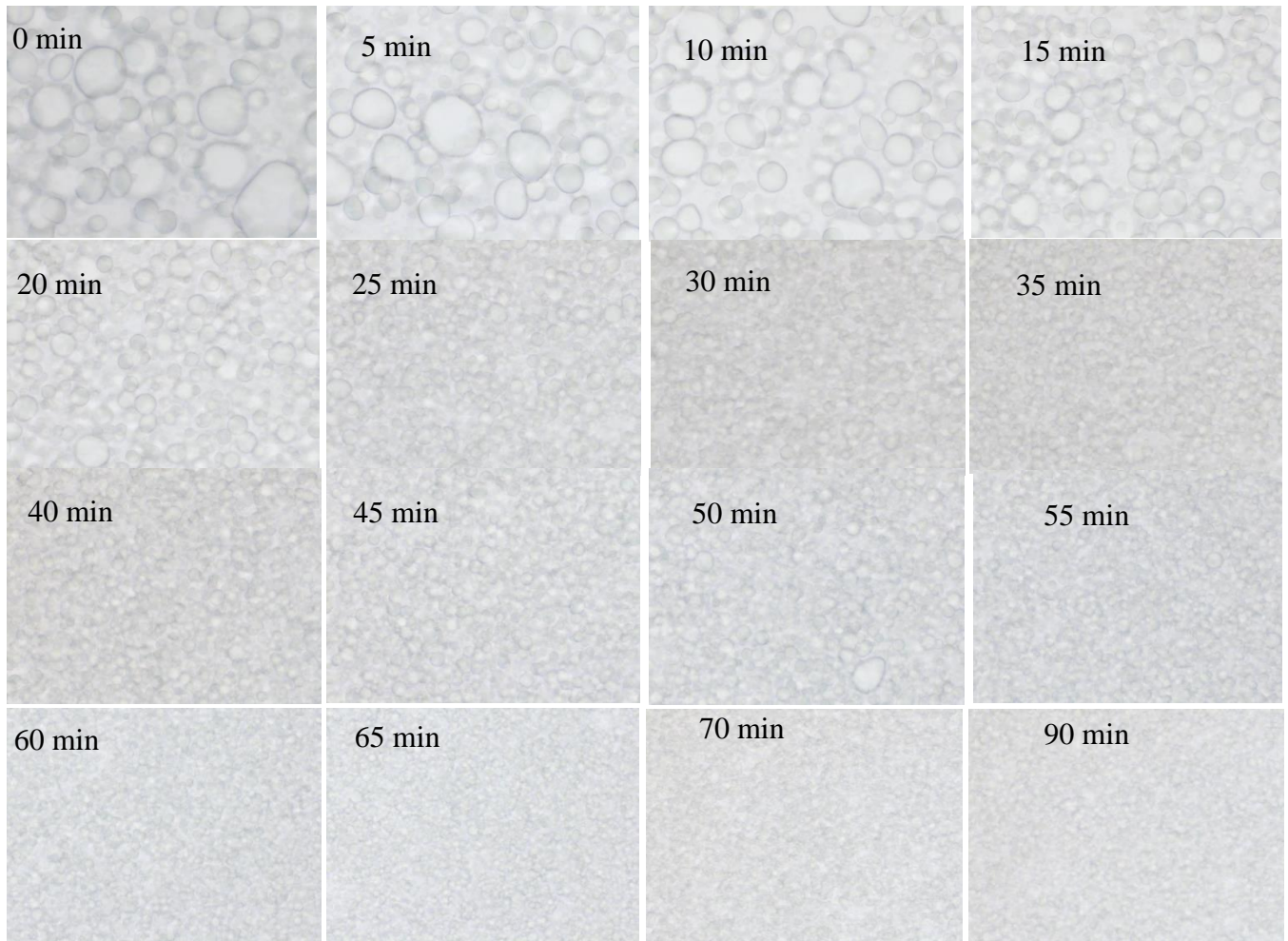
Maaß et al., 2012; Mickler et al., 2011). Panckow *et al.* (2017) were able to detect circular droplets from fermentation broths with  $d_{32}$  ranging between 70 and 110  $\mu\text{m}$ .

However, the detection of droplets in production systems of emulsions with an expected mean drop size  $< 10 \mu\text{m}$  was not feasible in the previous studies (Panckow et al., 2017). The HBT, presented in this study, was successful in detecting droplets with diameter as low as *ca.* 1  $\mu\text{m}$  from emulsion samples having dispersed phase fraction  $\approx 50\%$  (Unnikrishnan *et al.* 2018).

#### 4.2.2 Process-2: Droplet Detection and Analysis

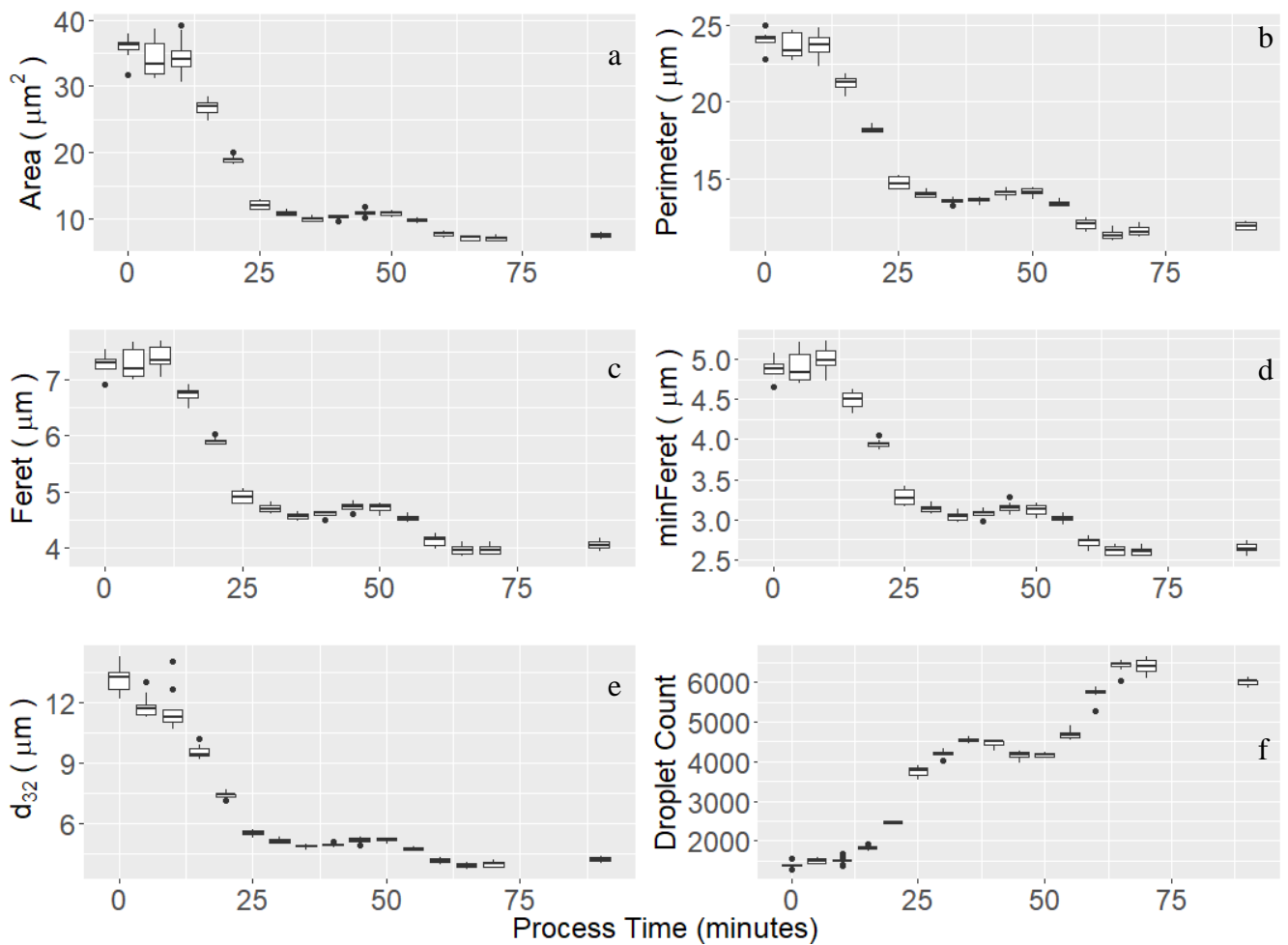
The results obtained from the statistical analysis of Process-1, discussed in section 4.2.1, demonstrated the proficiency of the HBT approach in extracting the details of droplet evolution throughout an emulsification process. The HBT approach was therefore applied for the analysis of Process-2, which is an industrial emulsification process that had a duration of 90 minutes. A total of 160 micrographs were acquired from Process-2. A sample micrograph, obtained at each interval, is shown in Figure 4.11. These micrographs were processed with the HBT macro. Approximately 1,500 droplets were detected from a micrograph sample after five minutes of processing. As the emulsification process approached 70 minutes, the number of detected droplets increased to *ca.* 6,500. The same 13 droplet characteristics were extracted as previously identified.





**Figure 4.11. Micrographs acquired from Process-2. A sample micrograph obtained at every five-minute interval is presented. The last sample micrograph taken at 90 minutes represents the discharge sample obtained at 32°C.**

The evolution of mean droplet size characteristics and droplet count throughout Process-2 were investigated using box plot analysis. The droplet characteristics extracted from the 10 micrographs at every process interval were used to create the box plots shown in Figure 4.12a to Figure 4.12f.



**Figure 4.12. Process-2. Box plots of the mean droplet size characteristics and count from HBT: a) area, b) perimeter, c) Feret, d) minFeret, e) Sauter mean diameter and f) Droplet count. Each box plot represents 10 micrographs. The last box plot represents the discharge sample obtained after 90 minutes.**

Each box plot (Figure 4.12a to 4.12e) represents the average droplet size characteristics of the 10 micrographs taken at five-minute intervals. The average droplet size, measured in terms of area, perimeter, minimum and maximum Feret diameters and  $d_{32}$ , was found to decrease gradually as emulsification progressed until the process reached 70 minutes. The decrease in the size characteristics over the 70-minute period are detailed quantitatively as follows:

- 
- Area decreased from 35.8 to 6.7  $\mu\text{m}^2$ .
  - Sauter mean diameter ( $d_{32}$ ) decreased from 13.2 to 3.8  $\mu\text{m}$ .
  - Maximum Feret diameter (Feret) reduced from 7.3 to 3.9  $\mu\text{m}$ .
  - Minimum Feret diameter (minFeret) reduced from 4.9 to 2.5  $\mu\text{m}$ .
  - Perimeter decreased from 24.0 to 11.3  $\mu\text{m}$ .

However, there was minimal variation in these droplet size characteristics after 65 minutes of processing. The discharge sample micrographs obtained after 90 minutes presented a slight increase in the average droplet size with minimal variation in most characteristics (Figure 4.12). This indicated that the emulsification process attained the target size characteristics after 65 minutes of processing. This was further reaffirmed by the industrial process experts. Any further processing was identified as over-processing.

The droplet count, shown in Figure 4.12f, presented a sharp increase from *ca.* 1,500 at the start of the process to *ca.* 6,500 after 70 minutes. Minimal variation was observed in the middle of the process between 30 and 50 minutes. This was identified as the period of the emulsification process, which required significant energy to break down the droplets further. This is reflected in the boxplots of mean droplet size and count. After 55 minutes of processing, the count started increasing marginally and remained steady until the process reached 70 minutes. The final bulk discharge samples obtained after 90 minutes (last box plot of Figure 4.12f) presented a slight reduction in the count. The variation in the median droplet count from the start of the process until 90 minutes is summarised in Table 4.2.

---

**Table 4-2. Median of droplet count obtained from 10 micrographs at each interval from the start of the process until 90 minutes. Orange cells represent the period of emulsification which presented an inverse variation in the count. Grey cells denote the stable droplet count.**

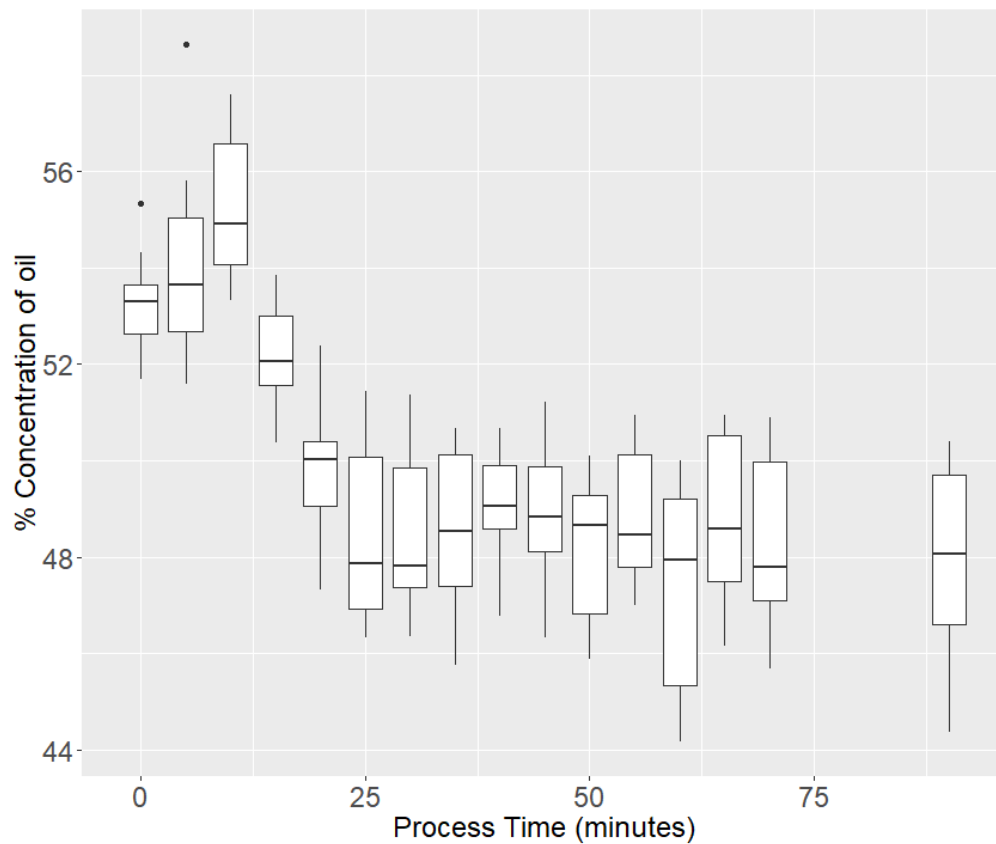
Process Time (min)	Droplet Count
0 (start)	1383
5	1498
10	1505
15	1814
20	2483
25	3769
30	4208
35	4534
40	4496
45	4182
50	4168
55	4652
60	5740
65	6456
70	6386
90	6034

The slight inverse variation in the count observed from 40 to 50 minutes (orange cells) can be attributed to the formation of agglomerates or bigger droplets through coalescence. The droplet count attained its maximum value after 65 minutes of processing. Only a slight reduction was observed after a further five minutes (grey cells). This also indicated that the process attained a stable droplet count after 65 minutes and therefore, the additional 5 minutes of processing was identified as over-processing.

The consistency of the oil concentration during the entire process was investigated by plotting the percentage of oil detected from the 10 micrographs at each process interval.

---

The box plots showing the oil concentration are presented in Figure 4.13. The median of the box plots is used for the comparative discussion.



**Figure 4.13. Box plots showing the % of oil concentration during the 90 minutes emulsification process (Process-2). Each box plot represents 10 micrographs.**

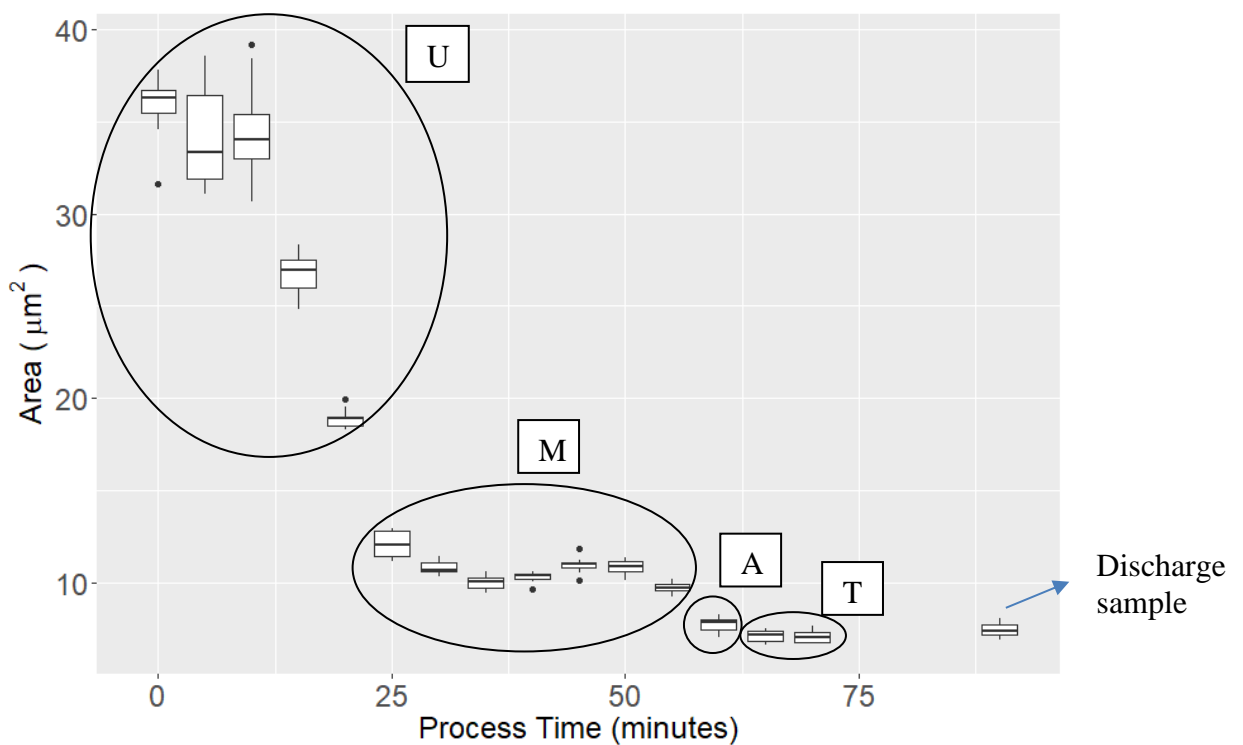
The percentage of oil concentration varied between 52% and 55% in the initial 15 minutes. This was followed by a 2% decrease in the oil content after 20 minutes while remaining stable at around 48% to 50% throughout the rest of the process. The initial variation observed in the total oil content can be attributed to the oil attaching to the sides of the vessels during mixing. Overall, the oil concentration analysis remained reasonably consistent throughout the process and, was in close agreement with the oil content of the emulsion formulation ( $\approx 50\%$ ).

The results clearly demonstrate the proficiency of the HBT approach in the detection of oil droplets and their characteristic evolution throughout the emulsification process.

---

#### 4.2.2.1 TAMU Labelling

The 150 micrographs acquired from Process-2 (excluding discharge samples) were labelled, by expert analysts from industry, into four quality-based categories named TAMU (Target, Acceptable, Marginal and Unacceptable). Laminated printouts of the micrographs were produced and were compared to the observed droplet size and count characteristics. The final categorisation was made by the experts based on the visual aspects of the micrographs as well as the observed droplet evolution. The TAMU labelling with respect to the mean droplet area is represented in Figure 4.14.



**Figure 4.14. Micrographs obtained from process-2 categorised into TAMU based on the evolution of mean droplet area ( $\mu\text{m}^2$ ).**

The following summarises the TAMU categorisation of the 150 micrographs:

- Samples obtained in the first 20 minutes were categorised as Unacceptable (U)
- Samples acquired from 25 to 55 minutes were categorised as Marginal (M)
- Samples obtained in the following 5 minutes were categorised as Acceptable (A)

- 
- Samples obtained from 65 to 70 minutes of processing were categorised as Target (T)

The number of micrographs belonging to the TAMU categories are presented in Table 4.3. A sample set of TAMU micrographs is presented in Appendix D.

**Table 4-3. Number of micrographs in each of the TAMU categories.**

Label	Number of micrographs
Target (T)	20
Acceptable (A)	10
Marginal (M)	70
Unacceptable (U)	50

These TAMU labelled micrographs were employed for developing the supervised classification models, which will be presented in Chapter 5.

---

### 4.3 Conclusions

This chapter presented the evaluation of two emulsification processes, Process-1 and Process-2. Process-1 was a trial process used as a proof of concept to evaluate the two image segmentation techniques, EST and HBT, and their competency in detecting droplets and eliciting their characteristics during emulsification. The droplet characteristics that progressively varied during emulsification were considered significant and included area, perimeter, Feret, minFeret and count. The HBT approach was identified as the most proficient technique compared to the EST. The HBT was able to detect droplets of diameter down to *ca.* 1  $\mu\text{m}$  from the emulsion product with a dispersed phase fraction  $\approx$  50%. The HBT also presented a consistent detection of oil content throughout the process.

Process-2 was evaluated using HBT. The results demonstrated the potential of the technique in detecting droplets of varying sizes from the start of the process until the target characteristics were achieved. The technique was also successful in identifying the optimum processing time of the emulsification process as 65 minutes when the mean droplet size and the count attained a stable state. The emulsion was found to be over-processed after that point. The approach indicated that the final 5 minutes of processing was unnecessary, which resulted in an increase in the droplet size and decrease in the count due to potential droplet coalescence. The micrographs obtained from Process-2 were labelled by industrial experts into four quality categories known as TAMU.

Classification models, based on machine learning algorithms, were developed as the next stage of this study and are discussed in the following chapter.



---

## Chapter 5

# Supervised Machine Learning for Emulsion Micrograph Classification

### 5.1 Introduction

This chapter discusses the machine learning models developed for the classification of micrographs acquired from Process-2 and compares their results. The evolution of droplet characteristics during emulsification was best represented by the Histogram-Based Technique (HBT). Therefore, the droplet characteristics obtained from HBT were considered for developing the classification models. Droplet area, perimeter, maximum Feret diameter, minimum Feret diameter and count were identified as the significant characteristics in the classification of the micrographs. The classification models evaluated in this study are:

- Principal Component based Linear Discriminant Analysis
- Multinomial Logistic Regression
- Random Forest
- Vanilla Neural Network (VNN)
- Convolutional Neural Network (CNN)

There were four categories of micrographs in Process-2. These categories were determined based on the expert opinion of the micrograph analysts from industry and also by assessing the variation in the droplet size characteristics and droplet count. The nomenclature used for the micrograph categories were Target, Acceptable, Marginal and

---

Unacceptable (TAMU). The number of micrographs in each category is presented in Table 4.3.

## **5.2 Classification Modelling: Results and Discussion**

The classification methodology is explained in Chapter 3, section 3.6. The results obtained from the machine learning models, developed for the TAMU classification of the Process-2 micrographs, are discussed in this chapter.

### **5.2.1 PC-LDA Model**

The approach used for developing the PC-LDA model is explained in detail in Chapter 3. The PCA approach was considered due to the high correlation observed between the individual droplet characteristics. PCA was applied to reduce the correlated droplet characteristics into a reduced set of Principal Components (PC). Each PC score corresponded to the characteristics of a single droplet. Therefore, the number of observations (PC scores) in a micrograph category was equivalent to the total droplet count of all the micrographs selected from that category.

The significant Principal Components were then selected to perform the supervised classification of micrographs using Linear Discriminant Analysis. A micrograph was classified into a TAMU category, by the model, based on the highest percentage of droplets classified from that micrograph. Stratified cross-validation was applied in order to obtain an equivalent number of observations from the four categories and thus avoid bias to any category. This was achieved by selecting an equal number of micrographs (10) from all the four response categories.

The original feature space consisted of the five selected characteristics, droplet area, perimeter, Feret, minFeret and count. The droplet size characteristics were found to be highly correlated ( $r \approx 0.84$  to  $0.98$ ). The correlation matrix is presented in Table 5.1.

**Table 5-1. Correlation matrix of droplet size characteristics.**

	<b>Area</b>	<b>Perimeter</b>	<b>Feret</b>	<b>minFeret</b>
<b>Area</b>	1.00	0.85	0.84	0.87
<b>Perimeter</b>	0.85	1.00	0.98	0.96
<b>Feret</b>	0.84	0.98	1.00	0.94
<b>minFeret</b>	0.87	0.96	0.94	1.00

PCA was performed using the droplet size and count data of the 150 micrographs and transformed the correlated feature space into an equal set of uncorrelated components. The eigenvectors obtained from the PCA for the five-variable space are presented in Table 5.2.

**Table 5-2. Eigenvectors obtained from the PCA for the five-variable space.**

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>
<b>Area</b>	-0.467	0.089	-0.875	0.087	0.023
<b>Perimeter</b>	-0.500	0.107	0.272	0.154	-0.801
<b>Feret</b>	-0.496	0.096	0.346	0.569	0.549
<b>minFeret</b>	-0.496	0.111	0.202	-0.803	0.238
<b>Count</b>	0.201	0.979	-0.007	0.010	0.004

The eigenvalues show the contribution of each variable to the five principal components. The negative eigenvalues signify that a variable has a negative correlation with a PC and the magnitude represents how significant is the contribution of a variable to a PC. PC1 has an approximately equal contribution from all the droplet size variables and are negatively correlated. PC2 has all the eigenvalues positive with the count variable showing the maximum contribution of 0.979. This means that droplet count was the most significant variable for PC2, while it presented a very little contribution of 0.2 for PC1.

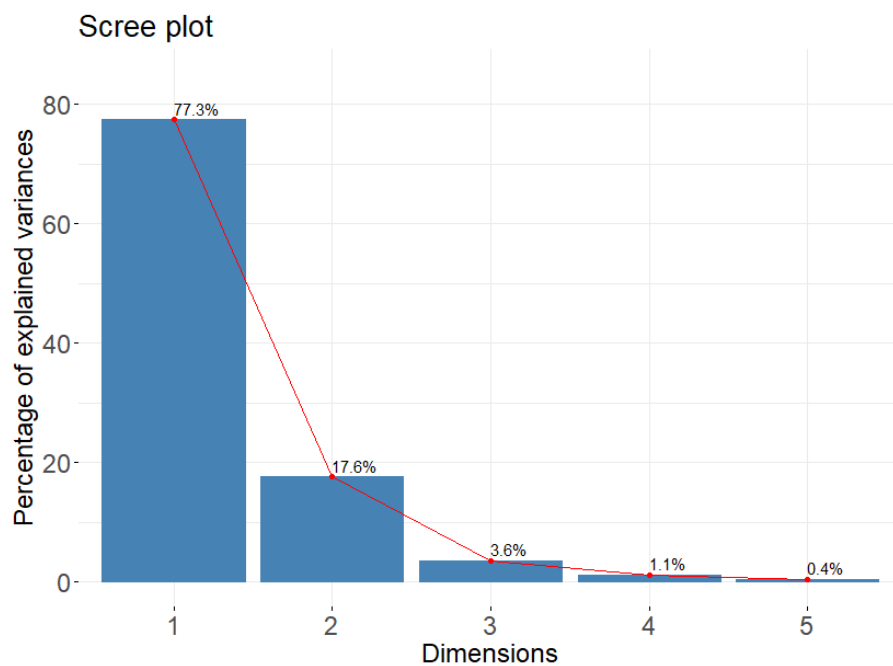
The eigenvalues of PC3, PC4 and PC5 present a varying contribution from the five variables with the droplet count showing minimal contribution.

The resultant set of five principal components (PCs) obtained from the eigenvectors are listed hierarchically, in Table 5.3, in relation to their proportion of explained variance.

The scree plot of the PCs is also shown in Figure 5.1.

**Table 5-3. Proportion of total variance explained by each Principal Component (PC). The grey cells represent the significant PCs.**

Principal Components	PC1	PC2	PC3	PC4	PC5
Proportion of Variance	0.773	0.176	0.036	0.011	0.004
Cumulative Proportion	0.773	0.949	0.985	0.996	1.000



**Figure 5.1. Scree plot of the five principal components (PCs).**

The first three PCs explained 98.5% of the total variance. PC4 and PC5 explained the remaining variance. The 2D and 3D score plots of PC1, PC2 and PC3 are presented in Figure 5.2 and Figure 5.3 respectively. These plots show the separation between the TAMU categories of the micrographs.

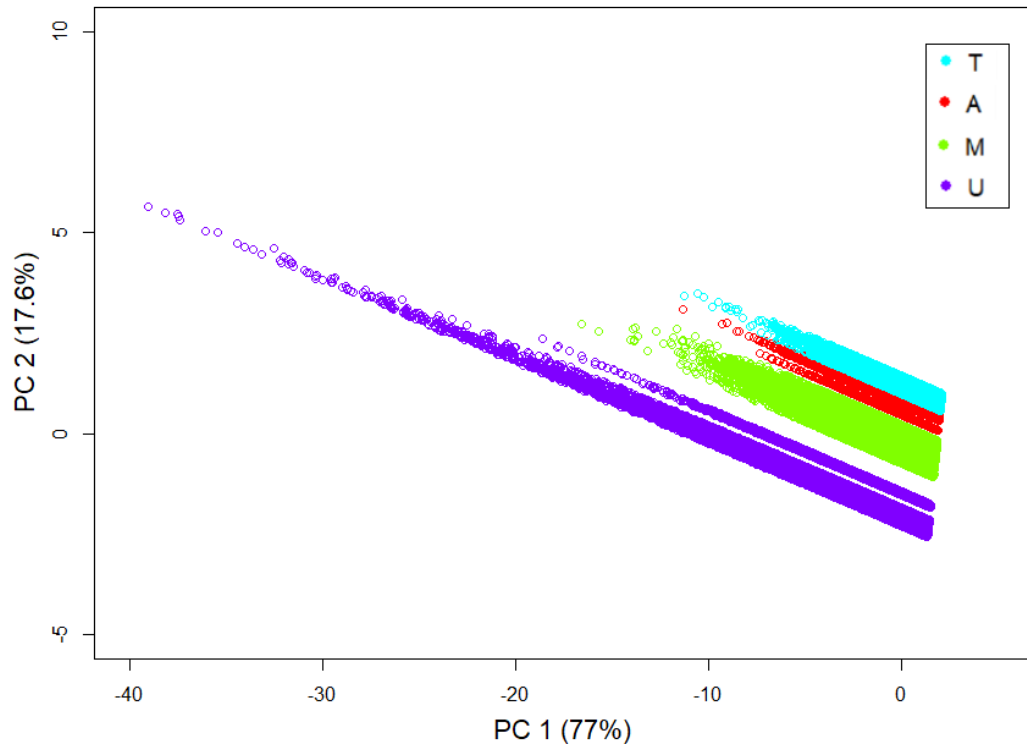


Figure 5.2. The Principal Component Analysis score plots using the first two PCs. 'U' (stands for Unacceptable) is represented by purple circles, 'M' (Marginal) by green, 'A' (Acceptable) by red and 'T' (Target) by blue respectively.

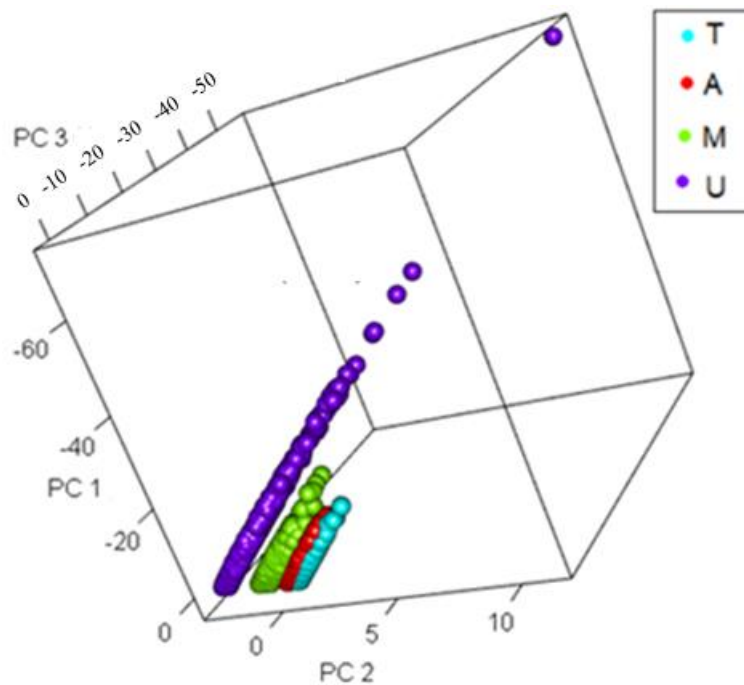


Figure 5.3. The Principal Component Analysis score plots using the first three PCs. 'U' (stands for Unacceptable) is represented by purple circles, 'M' (Marginal) by green, 'A' (Acceptable) by red and 'T' (Target) by blue respectively.

---

The category separation is very evident along the PC1-PC2 plane of Figures 5.2 and 5.3. The first three PCs were selected as the predictor variables of the PC-LDA model. The model was formed as a linear combination of these three predictors, which resulted in three discriminant functions as given in equations 5.1, 5.2 and 5.3. The number of predictive discriminant functions derived from the model is calculated as the minimum of  $G-1$  and  $p$ , where  $G$  is the number of response categories (*i.e.*, four) and  $p$  is the number of predictor variables (*i.e.*, three). In the current case, the value of  $p$  and  $G-1$  are equal and therefore, the PC-LDA model resulted in three discriminant functions, LD1, LD2 and LD3. The model formula is represented by equation 5.4.

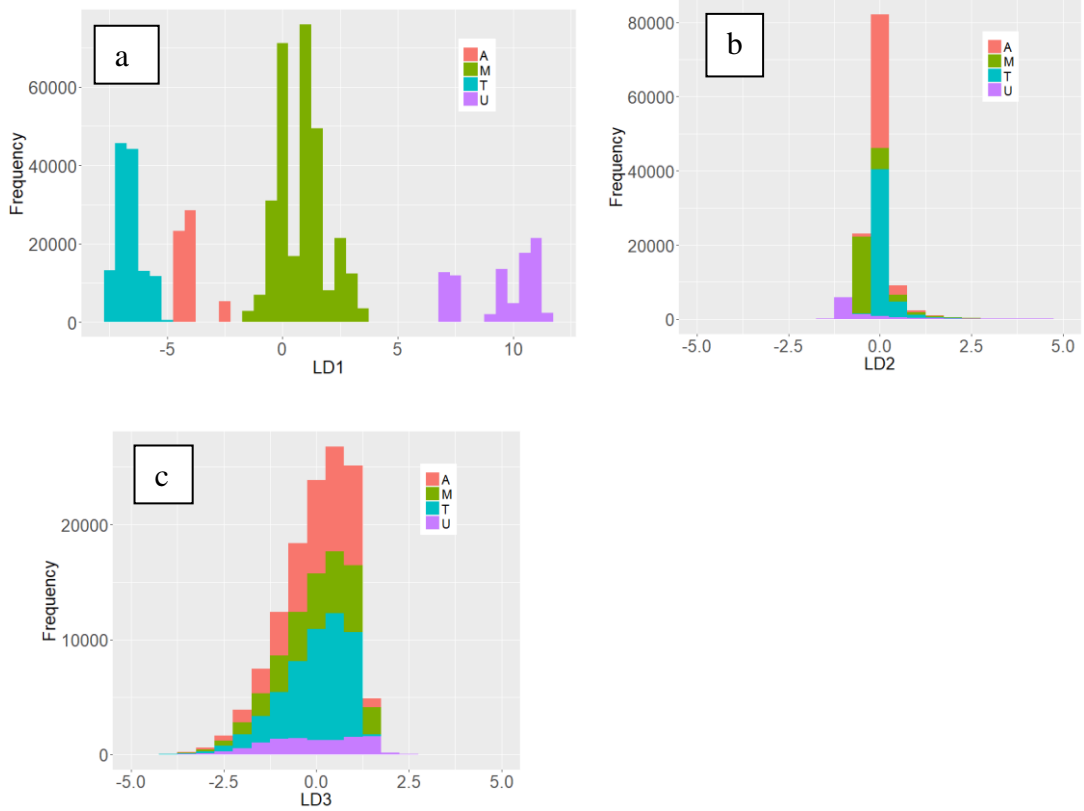
$$\mathbf{LD1 = -1.44 PC1 - 0.37 PC2 + 0.29PC3} \quad \mathbf{Equation\ 5-1}$$

$$\mathbf{LD2 = 7.06 PC1 - 0.32 PC2 + 0.28PC3} \quad \mathbf{Equation\ 5-2}$$

$$\mathbf{LD3 = -0.33 PC1 + 1.39 PC2 - 1.72PC3} \quad \mathbf{Equation\ 5-3}$$

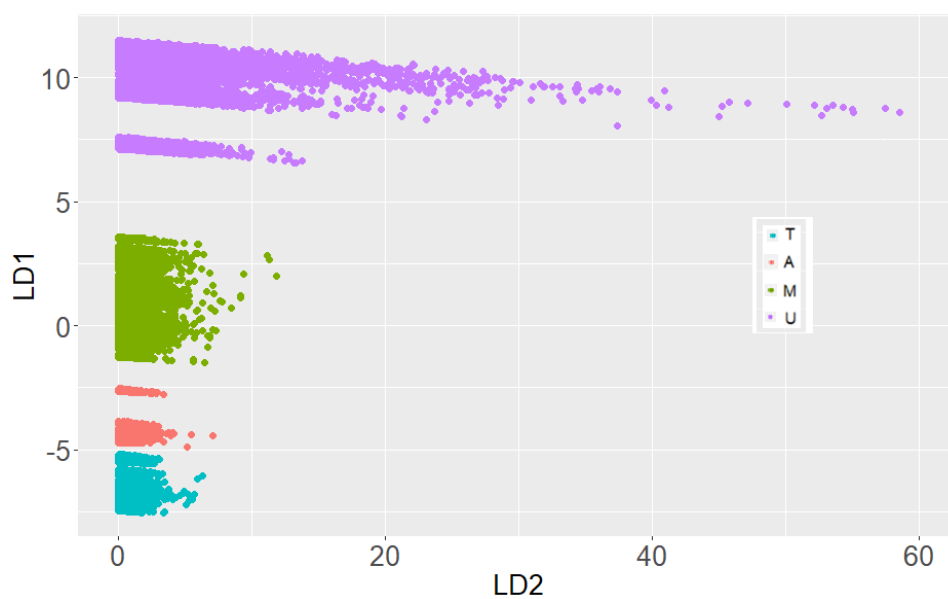
$$\mathbf{Category \sim LD1+LD2+LD3} \quad \mathbf{Equation\ 5-4}$$

The response category of the PC-LDA model was populated with the TAMU labels, for each individual droplet in a micrograph, such as ‘T’ for Target, ‘A’ for Acceptable, ‘M’ for Marginal and ‘U’ for Unacceptable respectively. The model was trained using the first three PCs obtained for the 150 micrographs. The trained model was used to predict the TAMU categories using the predict routine in R. The percentage separation between the categories explained by each discriminant function is given by its proportion of trace, which was the highest for LD1 (99.96 %). The successive discriminant functions, LD2 and LD3, together explained the remaining variance of 0.04%. Figure 5.4a to 5.4c presents the histograms of the three individual discriminant functions. A two-dimensional scatter plot of LD2 vs LD1 was also plotted to observe the classification achieved between the TAMU categories. This is presented in Figure 5.5.



**Figure 5.4. Linear Discriminant Analysis classification presented by the histograms of the three discriminant functions. a) LD1, b) LD2 and c) LD3.**

The histograms confirmed the separation between the TAMU categories achieved by LD1, LD2 and LD3.



**Figure 5.5. LD2 vs LD1 plot showing the separation between the TAMU micrographs.**

---

A well-defined separation was observed between the TAMU categories along the LD1 axis (Figure 5.5). This resulted in a reduction of the final set of canonical variables, used for the model prediction, to a single variable, LD1.

### 5.2.1.1 Cross-Validation

The model validation was performed using stratified 5-fold cross-validation. 10 micrographs from each category were selected for cross-validation. The reason for selecting 10 micrographs from each category was because of the maximum number of micrographs identified in ‘the ‘Acceptable’ category was 10 (Table 4.4). In order to maintain a balanced sample size, 10 micrographs from random processing times were chosen from the other three categories. The processing times from which the micrographs of each category were selected are presented in Table 5.4.

**Table 5-4. Processing times from which the TAMU micrographs for the 5-fold cross-validation were selected.**

Category	Processing Time (minutes)	Micrographs
U	0, 5, 10, 15, 20	10
M	25, 30, 35, 49, 45, 50, 55	10
A	60	10
T	65, 70	10

For each of the five models, 30% of the data (12 micrographs) was retained for testing and the remaining 70% (28 micrographs) was used for training. The training set corresponded to approximately 7000 to 40,000 droplets as observations from the ‘U’ to ‘T’ categories respectively. The training and the test sets of each model were different, ensuring that the training data never overlapped with the test data for a model. The



classification score was recorded, and confusion matrices were created for the five models. The matrices were summed to provide the final classification score. This is presented in Figure 5.6. The green boxes, in Figure 5.6, represent the number of correctly classified micrographs from each category, while the red boxes represent the incorrect classifications. The blue box at the bottom right corner represent the total number of correctly classified micrographs. The light blue box at the end of each row carries the sum of the cells in that row. Similarly, the light blue box at the end of each column carries the sum of the cells in that column.

		Predicted				SUM
		T	A	M	U	
Actual	T	15	0	0	0	15
	A	0	15	0	0	15
	M	0	0	15	0	15
	U	0	0	0	15	15
	SUM	15	15	15	15	60

**Figure 5.6. Confusion matrix obtained from the stratified 5-fold cross-validation. It represents the sum of 5 confusion matrices from the 5 models. Each model classified a test set of 12 micrographs (3 from each TAMU category).**

The test sets of micrographs, in all the five folds, achieved 100% correct classification of their corresponding droplets. Thus, all five models classified their test sets into the correct categories thereby demonstrating 100% accuracy. The 12 micrographs (three from each category) from the test set of each of the five models were summed to 60 micrographs in the final confusion matrix.

The model was also evaluated, for its classification efficiency on unseen data, using the independent set of 150 micrographs obtained from a different batch of the same emulsion. These micrographs were processed using the HBT macro and the droplet characteristics

were extracted. The number of micrographs and their TAMU categories for this independent batch are given in Table 5.5.

**Table 5-5. TAMU micrographs obtained from the independent batch of the emulsion.**

Category	Micrographs
T	29
A	61
M	10
U	50
Total	150

The confusion matrix showing the classification of these 150 micrographs is presented in Figure 5.7.

		Predicted				SUM
		T	A	M	U	
Actual	T	29	0	0	0	29
	A	0	61	0	0	61
	M	0	0	10	0	10
	U	0	0	0	50	50
	SUM	29	61	10	50	150

**Figure 5.7. Confusion matrix showing the classification of the 150 micrographs, into TAMU, obtained from an independent batch of the emulsion.**

The PC-LDA model classified the 150 micrographs into their correct TAMU categories. In summary, the PC-LDA model was successful in classifying the in-process micrographs, acquired from two batches of the industrial process (Process-2), with 100% accuracy.

---

### 5.2.2 MLR Model

The approach used for developing the Multinomial logistic regression model (MLR) model is explained in detail in Chapter 3. The MLR model was developed using the mean of the four droplet size characteristics and droplet count as the predictor variables. An individual observation in the MLR model corresponded to a single micrograph. The five droplet characteristics, which were used as the predictor variables in the model, are presented in Table 5.6.

**Table 5-6. Droplet characteristics selected as the predictor variables of the MLR model.**

$x_1$	Mean droplet area
$x_2$	Mean Feret
$x_3$	Mean minFeret
$x_4$	Mean perimeter
$x_5$	Droplet count

A random 70/30 training/test sample split was applied to the 150 micrographs. This resulted in a training set consisting of 105 micrographs and a test set consisting of 45 micrographs (Table 5.7). The model was trained using the 105 TAMU micrographs from the training set.

**Table 5-7. Number of TAMU micrographs from Process-2 assigned to training (70%) and test (30%) samples.**

Categories	Training	Test
T	14	6
A	7	3
M	49	21
U	35	15
Total	105	45

The ‘M’ category consisted of 46.7% of the training set, while the ‘A’ category contained only 6.7%. The ‘T’ category was set as the reference class for the model. The probabilities of the other three categories (U, A and M) were predicted, by the model, with reference to the ‘T’ category. The model coefficients are presented by Equations 5.5, 5.6 and 5.7.

$$\ln\left[\frac{p(A)}{p(T)}\right] = 0.335 + 0.171x_1 + 0.889x_2 + 1.844x_3 + 0.579x_4 - 0.005x_5 \quad \text{Equation 5-5}$$

$$\ln\left[\frac{p(M)}{p(T)}\right] = 0.142 + 1.822x_1 + 0.803x_2 + 2.278x_3 + 0.585x_4 - 0.008x_5 \quad \text{Equation 5-6}$$

$$\ln\left[\frac{p(U)}{p(T)}\right] = 0.088 + 2.763x_1 + 0.680x_2 + 2.161x_3 + 0.450x_4 - 0.012x_5 \quad \text{Equation 5-7}$$

Each of the equations predict the log odds of the probability of the class of interest with reference to the Target class. A two-tailed Z test was also performed to investigate the statistical significance of the predictor variables. The model summary and the *p*-values obtained from the Z test are presented in Table 5.8.

**Table 5-8. Results from the two tailed Z test of the MLR model trained using 70% of Process-2 micrographs.**

<b>Category</b>	<b>Intercept</b>	<b>x<sub>1</sub></b>	<b>x<sub>2</sub></b>	<b>x<sub>3</sub></b>	<b>x<sub>4</sub></b>	<b>x<sub>5</sub></b>
<b>Coefficients:</b>						
A	0.335	0.171	0.889	1.844	0.579	-0.005
M	0.142	1.822	0.803	2.278	0.585	-0.008
U	0.088	2.76	0.680	2.161	0.450	-0.012
<b>Standard Errors:</b>						
A	0.014	0.242	0.086	0.260	0.059	0.001
M	0.007	0.152	0.043	0.133	0.030	0.001
U	0.006	0.202	0.050	0.158	0.033	0.003
Residual Deviance: 13.726						
<b>p-values obtained from the two tailed Z test:</b>						
A	0	0.479	0	1.410e <sup>-12</sup>	0	9.917e <sup>-08</sup>
M	0	0.000	0	0.000	0	0.000
U	0	0.000	0	0.000	0	2.219e <sup>-05</sup>

---

The model gave a residual deviance of 13.7, which is a reasonably low value. An MLR model uses maximum likelihood estimation to minimise the model error. The model converges to a minimum negative log-likelihood value after a set of iterations. Twice the negative log-likelihood value is presented as the residual deviance of the model. Residual deviance of an MLR model, in general, shows how good is the model in predicting the probability of the correct category when the predictor variables are included. The model gets better as the residual deviance gets smaller. The model coefficients represent the relationship of each predictor variable with the response variable. These coefficients were used to form the model equations (5.5, 5.6 and 5.7) that predict the probability of the response categories. For e.g., a one unit increase in  $x_1$  increases the log odds of being in ‘A’ compared to ‘T’ by 0.335. The negative coefficients stand for a negative correlation between the predictor and the response. The standard error values are related to the model coefficients and these are used to test if the estimated values of the coefficients are significantly different from zero. The two-tailed Z-test was performed by dividing the model coefficients with their standard errors and the corresponding  $p$ -values represented the statistical significance of each coefficient for a 95% confidence interval ( $\alpha = 0.05$ ). The Z test results presented low  $p$ -values for the five the predictor variables except for  $x_1$ , when predicting ‘A’ with reference to ‘T’. This exception can be explained by the marginal difference observed in the mean droplet size ( $x_1$ ) between the ‘A’ and ‘T’ categories. The model was then validated using the 30% test data, which consisted of 45 micrographs and obtained 100% correct classification.

#### **5.2.3.1 Cross-Validation**

A 10-fold cross-validation was performed using the total 150 micrographs to evaluate potential overfitting and underfitting of the model. Each fold consisted of a random 70/30

split of training and test micrographs. 10 MLR models were developed from the 10 folds. The accuracy of each model in classifying their test data set was evaluated. The accuracy of the 10 models is presented in Table 5.9.

**Table 5-9. Results from the 10-fold cross-validation of the MLR models.**

Model	Testing Accuracy (%)
1	100
2	100
3	100
4	97.8
5	100
6	100
7	93.3
8	95.6
9	97.8
10	95.6

The average accuracy of the 10 models was calculated as 98.01% and the standard deviation was 2.44%. The classification score was recorded for each model and confusion matrices were created. The matrices were summed to provide the final score. Each model classified 45 micrographs. In total, 450 micrographs were classified by the 10 models. The summed confusion matrix is shown in Figure 5.8.

		Predicted				SUM
		T	A	M	U	
Actual	T	55	5	0	0	60
	A	2	27	1	0	30
	M	0	0	209	1	210
	U	0	0	0	150	150
	SUM	57	32	210	151	441

**Figure 5.8. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using Multinomial Logistic Regression. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 45 micrographs.**

All the Unacceptable micrographs were correctly classified by the 10 models, while one Marginal micrograph was incorrectly classified as Unacceptable. The Acceptable category had three misclassifications, one classified as Marginal and the other two as Target. The most misclassifications occurred for the Target category, which had five of the micrographs incorrectly classified as Acceptable. The 10 models correctly classified 441 micrographs out of the 450 thus giving an overall accuracy of 98%.

An additional evaluation of the MLR model on unseen data was also conducted. The model was trained with the total set of 150 micrographs obtained from the first batch of Process-2. This model was then evaluated using the independent set of 150 micrographs. The classification results obtained are shown in the confusion matrix of Figure 5.9.

		Predicted				SUM
		T	A	M	U	
Actual	T	29	0	0	0	29
	A	5	49	7	0	61
	M	0	0	10	0	10
	U	0	0	0	50	50
	SUM	34	49	17	50	138

**Figure 5.9. Confusion matrix showing the TAMU classification of the 150 unseen micrographs obtained from the independent batch.**

Out of the 150 micrographs, 138 were correctly classified by the MLR model. The Target, Marginal and Unacceptable micrographs were 100% correctly classified. The misclassifications occurred in the Acceptable category, in which five micrographs were classified as Target and seven were classified as Marginal. Overall, 92% accuracy was

achieved by the model in classifying the independent data set. The overall statistics given by the model are presented in Table 5.10.

**Table 5-10. Summary of the MLR model statistics obtained from the classification of the independent set of micrographs.**

<b>Overall Statistics:</b>				
Accuracy (Acc)	0.92			
95% CI	(0.864, 0.958)			
No Information Rate (NIR)	0.333			
P-Value [Acc > NIR]	$< 2.2 e^{-16}$			
<b>Statistics by Class:</b>				
	Class: T	Class: A	Class: M	Class: U
Sensitivity	0.853	1.000	0.588	1.000
Specificity	1.000	0.881	1.000	1.000
Detection Rate	0.193	0.327	0.067	0.333
Balanced Accuracy	0.927	0.941	0.794	1.000

Acc represents the accuracy of the model and NIR denotes No Information Rate in Table 5.10. NIR is the largest proportion of the TAMU categories, which is 0.333 for the ‘U’ category with the largest number of micrographs. Accuracy greater than NIR confirms the significance of the model. The  $p$ -value less than 0.01 also demonstrated that the model results are significant. Sensitivity, also known as the true positive rate, is the proportion of true positive samples that are classified positive by the model. Specificity, also known as the true negative rate, is the proportion of true negative samples that are classified as negative. The A and U categories had sensitivity equal to 1.0, as all the positive classifications in those categories were truly positive. The specificity of the T, M and U categories were equal to 1.0, as all the negative classifications in those categories were truly negative. The detection rate signifies the percentage of micrographs in each category. The balanced accuracy column provides a measure of each individual class accuracy by taking the true positives, the false positives and the false negatives into account. This is a good measure to evaluate imbalanced data sets. In summary, the model



presented a reasonably good accuracy of 92% in the classification of the unseen micrographs.

### 5.2.3 RF Model

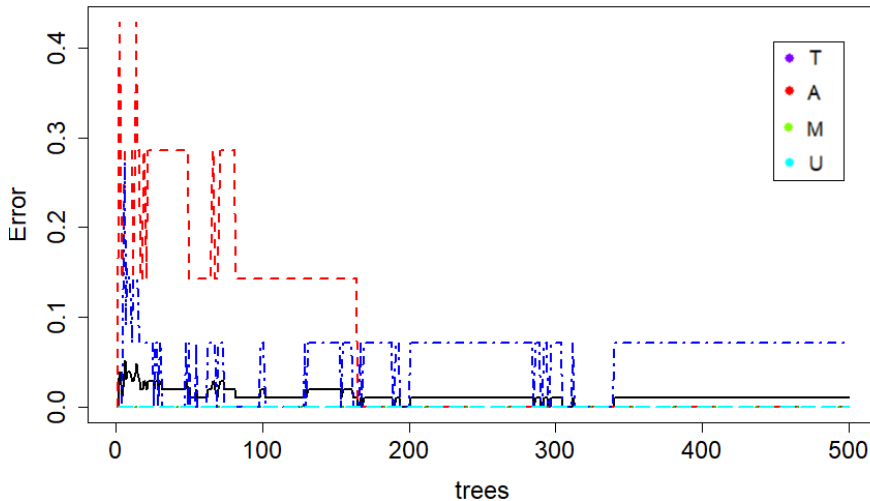
The approach used for developing the Random Forest (RF) model is explained in detail in Chapter 3. The RF model was built and tuned in a step by step process. The same set of five predictor variables ( $x_1$  to  $x_5$ ), presented in Table 5.5, were the model predictors. The model was trained for the TAMU classification. A single observation of the RF model represents an individual micrograph.

The initial model was built using the default *n*tree and *m*try values chosen by the RF classification algorithm. The default *n*tree was 500. The default *m*try, *i.e.*, the number of variables to be split at each node of the 500 trees, was calculated as the square root of the number of predictor variables, *i.e.*,  $\sqrt{5} = 2.2$ , which was rounded to 2 by the algorithm. The 150 micrographs were split into a random 70/30 for training and testing. The initial model was trained with 70% of the data. The OOB error rate and the confusion matrix obtained from training the model are given in Table 5.11.

**Table 5-11. Random Forest model output showing OOB error and confusion matrix for the training micrographs.**

OOB error rate: 0.95%					
Confusion Matrix:					
	A	M	T	U	Class.error
A	7	0	0	0	0.00
M	0	49	0	0	0.00
T	1	0	13	0	0.07
U	0	0	0	35	0.00

The model presented a low OOB error estimate of 0.95%. The classification error was 0% for the ‘U’, ‘M’ and ‘A’ categories. The ‘T’ category presented an error rate of 7% (one out of 14 micrographs was misclassified as ‘A’). The variation observed in the OOB error rate during the training of the model is presented in Figure 5.10.



**Figure 5.10.** Variation in the OOB error rate of the RF model as the number of trees grow from 0 to 500. The coloured lines represent the individual categories (TAMU). The black line represents the overall OOB error estimate of the model.

The black line in Figure 5.10 corresponds to the OOB error rate of 0.0095 (0.95%). The error rate of the model stabilised after approximately 400 trees. Therefore, the optimal *ntree* value was chosen as 400.

### 5.2.3.1 Tuning and Cross-Validation

The 150 micrographs were split into 10 folds. Each fold consisted of a random 70/30 split of training and test micrographs. For each fold, the *mtry* was tuned, from 1 to 5, to select the optimal model with the lowest OOB error. The *ntree* was kept constant at 400. The OOB error estimate of each *mtry* was recorded. The optimal *mtry* values, which gave the minimum OOB error from among the five trials, for each fold are presented in Table 5.12.

---

**Table 5-12. Tuning results of the 10 folds for the *mtry* values, 1 to 5.**

Fold	Optimal <i>mtry</i>	OOB Error Estimate (%)
1	1,2,3,4,5	0.95
2	3,4,5	0
3	3,4,5	0
4	3,4,5	0.95
5	2,3,4,5	0.95
6	3,4,5	0
7	3,4,5	0.95
8	2,3,4,5	0
9	2,3,4,5	0
10	3,4,5	0

The *mtry* values of 3, 4 and 5 provided the minimum OOB error estimate consistently in all the 10 folds. The maximum OOB error rate obtained from among the 50 trials (5 *mtry* trials for the 10 folds) was 1.9%. The average OOB error rate from all the 50 trials was 0.95%. The optimal *mtry* was chosen as 3 (closer to the default *mtry*) for the final model in each fold. A 10-fold cross-validation was performed. 10 models were developed from the 10 folds with *mtry* = 3 and *ntree* = 400. The accuracy of each model in classifying its test data set was evaluated. The testing accuracy of the 10 models is presented in Table 5.13.

**Table 5-13. 10-fold cross-validation results of the RF models for *ntree*=400 and *mtry*=3.**

Model	Accuracy (%)
1	100
2	100
3	97.8
4	100
5	100
6	100
7	100
8	100
9	100
10	100

The average accuracy of the 10 models was calculated as 99.78% and the standard deviation was 0.7%. The classification score was recorded for each model and confusion matrices were created. The matrices were summed to provide the final score. Each model classified 45 micrographs. Thus, in total, 450 micrographs were classified by the 10 models. The overall confusion matrix is shown in Figure 5.11.

		Predicted				SUM
		T	A	M	U	
Actual	T	59	1	0	0	60
	A	0	30	0	0	30
	M	0	0	210	0	210
	U	0	0	0	150	150
	SUM	59	31	210	150	449

**Figure 5.11. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using Random Forest. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 45 micrographs.**

The Unacceptable, Marginal and Acceptable micrographs were correctly classified by all the 10 models. Only one Target micrograph was misclassified as Acceptable. The RF model presented a very good overall accuracy of 99.78%.

The model's performance on the classification of unseen data was evaluated using the independent batch. The complete set of 150 micrographs from the first batch was used to train the RF model by setting  $n_{tree} = 400$  and  $m_{try} = 3$ . The model was then used to classify the 150 micrographs from the independent set. The confusion matrix is presented in Figure 5.12.

		Predicted				SUM
		T	A	M	U	
Actual	T	29	0	0	0	29
	A	4	55	2	0	61
	M	0	0	10	0	10
	U	0	0	0	50	50
	SUM	33	55	12	50	144

**Figure 5.12. Confusion matrix showing the TAMU classification of the 150 unseen micrographs using the Random Forest model.**

The model correctly classified 144 out of the 150 unseen micrographs. The Unacceptable, Marginal and Target micrographs were classified with 100% accuracy. Out of the 61 Acceptable micrographs, four were misclassified as Target and the other two as Marginal. The misclassifications, in Figure 5.11 and Figure 5.12, occurred mainly between the Acceptable and Target micrographs. This is due to the minimal variation exhibited by the droplet characteristics between these two categories (Figure 4.14). The overall statistics given by the model are presented in Table 5.14.

**Table 5-14. Summary of the RF model statistics obtained from the classification of the independent set of micrographs.**

<b>Overall Statistics:</b>				
Accuracy (Acc)	0.96			
95% CI	(0.915, 0.985)			
No Information Rate (NIR)	0.407			
P-Value [Acc > NIR]	$< 2.2 e^{-16}$			
<b>Statistics by Class:</b>				
	Class: T	Class: A	Class: M	Class: U
Sensitivity	0.967	1.000	0.986	1.000
Specificity	1.000	0.902	1.000	1.000
Detection Rate	0.193	0.367	0.067	0.333
Balanced Accuracy	0.984	0.951	0.992	1.000

---

The model presented an overall accuracy of 96% in the classification of the 150 unseen micrographs and the  $p$ -value shows that the model performance is statistically significant by giving an accuracy much higher than the NIR. The individual class statistics of the U category in the RF model was almost identical to the MLR model results. The major difference was that, the total number of misclassifications in the RF model was 50% less than that of the MLR model. The sensitivity of the M and the T categories were 10% and 40% higher respectively and the specificity of the A category was 2% higher in the RF model compared to the MLR model. All the four categories presented very high balanced accuracies due to reduced misclassifications.

#### 5.2.4 VNN Model

The approach used for developing the Vanilla Neural Network (VNN) model is explained in detail in Chapter 3. VNN is a deep learning model with a single hidden layer. The same five droplet characteristics,  $x_1$  to  $x_5$ , were used as the predictor variables (Table 5.5). An individual observation in the VNN model corresponded to a single micrograph. The VNN model has two hyper-parameters, such as size and weight decay, which should be tuned to obtain the maximum likelihood fitting of the model. The size parameter represents the number of units in the hidden layer of the network and the weight decay is the regularisation parameter, which helps reduce the overfitting of the model with the training data.

The 150 micrographs, from the first batch, were split into a random 80/20 for training and testing respectively. The training split was increased by 10%, compared to the previous models, due to the limited number of samples in the Acceptable and Target categories and the knowledge that the VNN model presented very poor accuracy in predicting the

low frequency categories. The number of samples contained in the training and test splits are shown in Table 5.14.

**Table 5-15. The number of TAMU micrographs from Process-2 assigned to training (80%) and test (20%) samples.**

Categories	Training	Testing
T	16	4
A	8	2
M	56	14
U	40	10
Total	120	30

The ‘M’ category consisted of the maximum percentage, 46.7%, of the training set, while the ‘A’ category contained the minimum, 6.7%.

#### **5.2.4.1 Tuning and Cross-Validation**

The VNN model was trained using 80% of the micrographs and was tuned for a default set of size and weight decay parameters to obtain the optimal model. Accuracy was the metric used by the model to find the optimum parameter values. The model tuning results are presented in Table 5.16.

**Table 5-16. VNN model tuning with the 80% training data.**

Size	Weight Decay	Accuracy	Kappa
1	0e+00	0.446	0.000
1	1e-04	0.446	0.000
1	1e-01	0.628	0.316
3	0e+00	0.446	0.000
3	1e-04	0.446	0.000
3	1e-01	0.890	0.814
5	0e+00	0.446	0.000
5	1e-04	0.464	0.024
5	1e-01	0.904	0.842

---

The model achieved 90% training accuracy for size = 5 and weight decay = 0.1 as seen from Table 5.16. The kappa values denote the significance of the model accuracy. The highest kappa values  $\approx 1.0$  are the most significant. The remaining 20% test data was evaluated using the optimal model and obtained 93% accuracy.

Ten random 80/20 folds of training and test data sets were created from the 150 micrographs. The size and the weight decay values were tuned for all the 10 folds and the optimum values, for all the folds, were obtained as 5 and 0.1 respectively. Ten models were created from the 10 folds with these optimum hyper-parameter values. A 10-fold cross-validation was performed and the classification accuracy of the of the test data by the 10 models were recorded. This is presented in Table 5.17.

**Table 5-17. The accuracy of the VNN models obtained from the 10-fold cross-validation.**

Model	Accuracy (%)
1	90.0
2	89.6
3	88.0
4	85.2
5	89.8
6	89.0
7	79.3
8	89.2
9	88.2
10	87.9

The 10 models presented an average accuracy of 87.62% and a standard deviation of 3.2%. Each model was evaluated using its test data set and the classification scores were recorded in a confusion matrix. These confusion matrices were combined to provide an overall classification score and is presented in Figure 5.13.



		Predicted				SUM
		T	A	M	U	
Actual	T	30	2	8	0	40
	A	8	7	5	0	20
	M	0	0	140	0	140
	U	0	0	10	90	100
SUM		38	9	163	90	267

**Figure 5.13. Confusion matrix obtained from the 10-fold cross-validation of the 150 micrographs using the Vanilla Neural Network. It represents the sum of the confusion matrices from the 10 models. Each model classified a test set of 30 micrographs.**

Each model classified 30 micrographs. Therefore, in total, 300 micrographs were classified by the 10 models. The Marginal micrographs were correctly classified by all the 10 models. Most of the misclassifications occurred with the Acceptable micrographs (13 misclassifications). These were incorrectly classified as either Target or Marginal. The Target and the Unacceptable category also had 10 misclassifications. The 10 models correctly classified 267 micrographs out of the 300 thus giving an overall accuracy of 89%.

The VNN model performance on the classification of unseen data was evaluated using the independent batch. The 150 micrographs obtained from the first batch of Process-2 was used to train the model by setting the hyper-parameters, size and weight decay, to 5 and 0.1 respectively. The model was then used to classify the 150 micrographs from the independent batch. The confusion matrix showing the classification score is presented in Figure 5.14.

		Predicted				SUM
		T	A	M	U	
Actual	T	29	0	0	0	29
	A	11	39	11	0	61
	M	0	0	10	0	10
	U	0	0	0	50	50
	SUM	40	39	21	50	128

**Figure 5.14. Confusion matrix showing the TAMU classification of the 150 unseen micrographs using the VNN model.**

The model correctly classified the Target, the Marginal and the Unacceptable micrographs. The misclassifications occurred in the Acceptable category. Of the 61 Acceptable micrographs, 22 were incorrectly classified. 11 were classified as Target and the other 11 as Marginal. Overall, 128 micrographs out of the 150 were classified correctly. This represents an accuracy of 85.3%. The very limited number of training samples is the reason for the poor classification accuracy of the VNN models, especially for the Acceptable category, which constituted only 6.7% of the total training dataset. The model accuracy can potentially be improved by obtaining an increased sample size for the low frequency categories. The overall statistics obtained from the model are presented in Table 5.18.

**Table 5-18. Summary of the VNN model statistics obtained from the classification of the independent set of micrographs.**

<b>Overall Statistics:</b>				
Accuracy (Acc)	0.853			
95% CI	(0.786, 0.906)			
No Information Rate (NIR)	0.333			
P-Value [Acc > NIR]	$< 2.2 e^{-16}$			
<b>Statistics by Class:</b>				
	Class: T	Class: A	Class: M	Class: U
Sensitivity	0.725	1.000	0.476	1.000
Specificity	1.000	0.802	1.000	1.000
Detection Rate	0.193	0.260	0.067	0.333
Balanced Accuracy	0.863	0.901	0.738	1.000

---

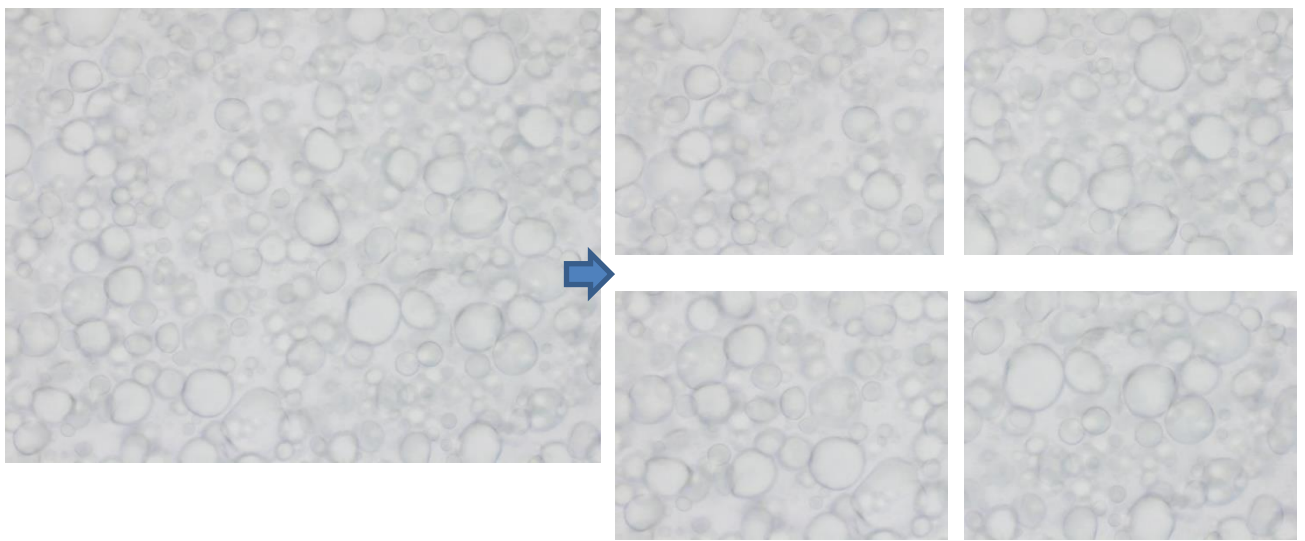
The Unacceptable, Marginal and the Target categories had 100% specificity, which was similar to the MLR and the RF model results. This is explained by the zero misclassifications from these categories. The Acceptable category had a sensitivity equal to 1.0, similar to the MLR and RF models. This explains that all the micrographs classified as Acceptable were truly acceptable. However, the specificity of the Acceptable category was reduced by 10% and 12% compared to the RF and MLR models respectively. This can be explained by the increased number of Acceptable micrographs being incorrectly classified. The balanced accuracy of the individual categories was also reduced, compared to the MLR and the RF models, except for the Unacceptable category. Overall, the VNN model presented an accuracy of 85.3%, which was very low compared to the other classification models discussed in this chapter.

### 5.2.5 CNN Model

The approach used for developing the Convolutional Neural Network (CNN) model is explained in detail in Chapter 3. The CNN model used a different methodology from the previous classification approaches discussed in this chapter. The input feature space used previously was based on the droplet characteristics acquired from the micrographs by means of Histogram-Based image segmentation. The pixel intensity values of the micrographs (0 to 255) formed the input features of the CNN model. The packages used for training and validating the CNN model are keras-2.2.4 and tensorflow-hub-0.3.0.

The 150 micrographs obtained from the first batch of Process-2 were a very limited dataset for developing a CNN model. Therefore, splitting the micrographs, into four quarters, was considered as an option to obtain an increased dataset. Each micrograph was split into four and 600 micrographs were obtained in total. A macro was written in

Fiji for splitting the micrographs. The splitting of a sample micrograph is shown in Figure 5.15. The resolution of the split micrographs was 694x520 pixels. Approximately 80% of the micrographs from each of the TAMU categories were considered as the training set and the remaining 20% was withheld for testing. The breakdown of the training set is shown in Table 5.19. The percentage of micrographs, from each category, can be summarised as ‘M’ = 44%, ‘U’ = 35%, ‘T’ = 14% and ‘A’ = 7%.



**Figure 5.15. Micrograph of a five-minute processed emulsion sample split into four quarters in Fiji.**

**Table 5-19. The number of training micrographs in the TAMU categories after splitting.**

Category	Micrographs
T	64
A	32
M	200
U	160
Total	456

---

The steps involved in the CNN approach are:

1. The micrographs were transformed to a square resolution of  $n \times n$  pixels, where  $n = 500$ .
2. These were then converted to greyscale.
3. The pixel values were extracted from each micrograph and stored in a 2D matrix of size  $n \times n$ .
4. Step 3 was repeated for the 456 micrographs and finally, a 3D data array of size  $456 \times n \times n$  was developed.
5. The response (independent variable) data was encoded in the format shown in Table 5.20.

**Table 5-20. The rows represent the number of micrographs from each category in the data array. The response variable was encoded as either 0 or 1 based on which category the micrograph belonged to.**

Rows	T	A	M	U
1 - 160	0	0	0	1
161 - 360	0	0	1	0
361 - 392	0	1	0	0
393 - 456	1	0	0	0

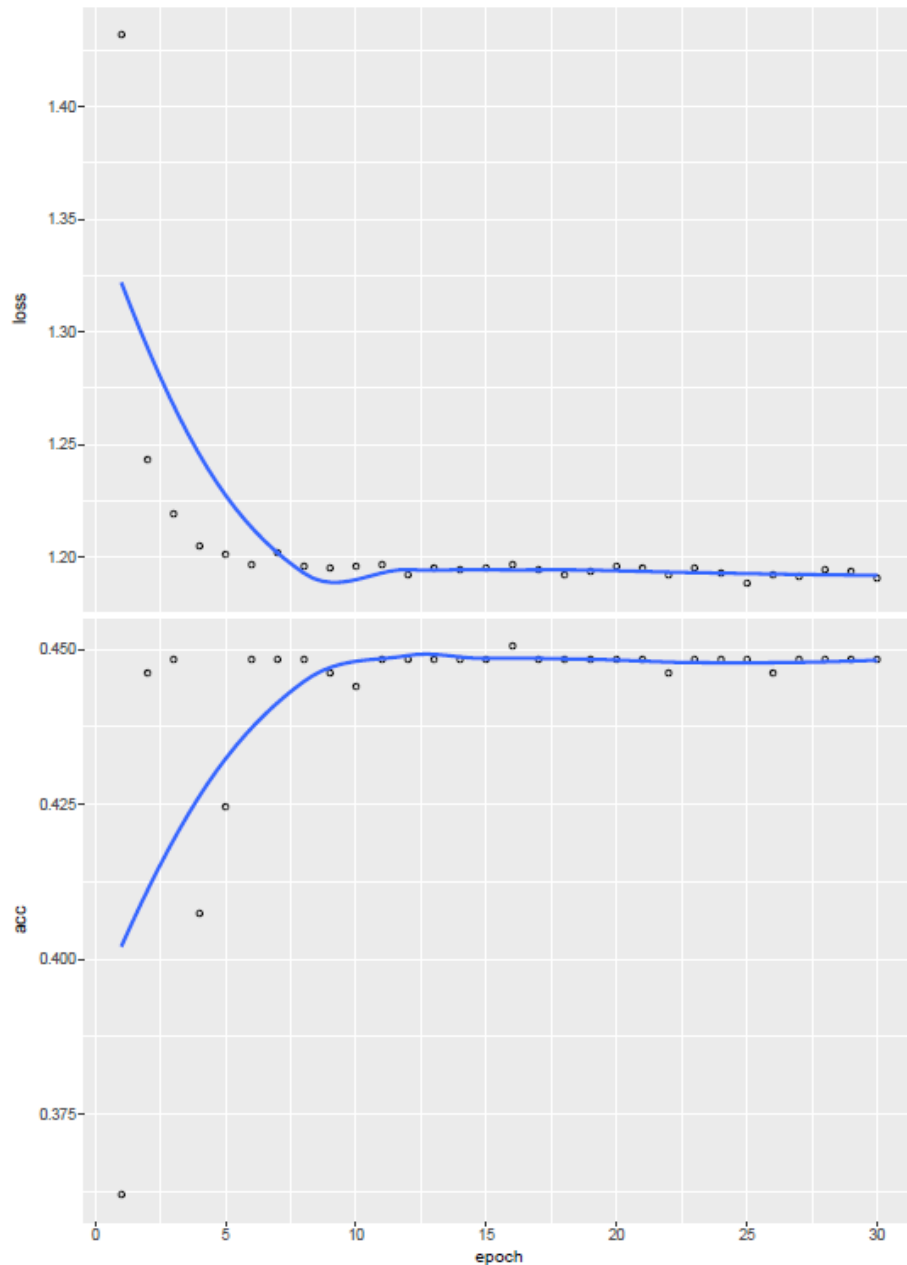
6. The model was trained using the 456 micrographs.
7. The training error and accuracy were evaluated.

The Keras sequential model was built with the following layers:

- Two convolutional layers of four units, each with 3x3 filters, one stride and zero padding.
- Two max pool layers of 4 units, each with 2x2 filters.
- Two dropout layers.

- 
- A flatten layer
  - The fully connected output dense layer with four units.

Softmax was used as the activation function and ‘categorical\_crossentropy’ was the loss function used in the output layer. Softmax is the activation function used for classification problems and categorical\_crossentropy is the loss function for multilevel (more than two response classes) classifications. The number of units in the convolutional layer varied between 1 and 5, by trial and error, to investigate any change in the loss and accuracy of the model. The model consistently presented very poor performance in all the trials. A sample plot of the loss and accuracy obtained from the model training, for  $n = 500$  and 4 units in the convolutional layer, is shown in Figure 5.16. This represents the best fit obtained from the overall model training.



**Figure 5.16. Keras sequential CNN model. The loss of the model during training is presented in the top plot. The accuracy of the model, which is inversely proportional to the loss, is presented in the bottom plot. This plot was obtained for  $n = 500$  and 4 units in the convolutional layer.**

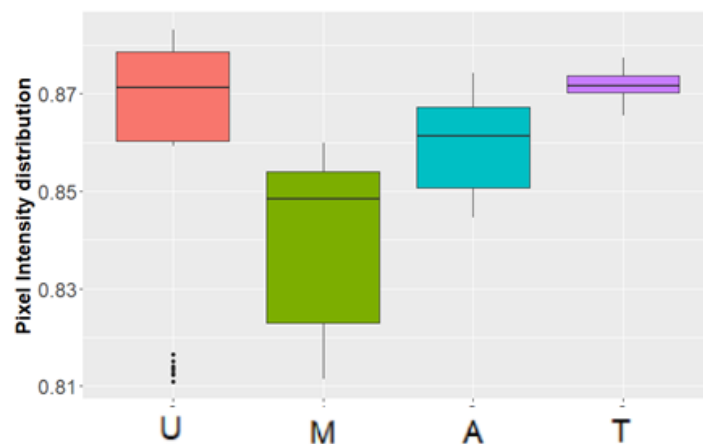
The training accuracy of the model did not show any progress after 10 epochs and remained almost steady throughout. After 30 epochs, the minimum loss (cross entropy) and the maximum accuracy of the model were evaluated as 1.18 and 0.448 (44.8%) respectively. The loss is expected to be less than 1 for a good model. An evaluation of

---

the test data was irrelevant, as the model presented a very poor learning performance and a training accuracy less than 50%.

The poor performance of the model can be attributed to the following reasons:

- Very limited data set of 456 micrographs.
- Minimal inter-class variation observed between the TAMU categories in terms of the average pixel intensity data (Figure 5.17).



**Figure 5.17.** Each box plot represents the average pixel intensity distribution of all the micrographs from the corresponding category.

The average pixel intensity values, as shown in Figure 5.17, presented only a marginal variation between the TAMU micrographs. In addition to the limited number of micrographs, the minimal variation in the pixel intensity values between the four categories is another reason for the poor accuracy of the model. The CNN model was neither cross-validated using the Process-2 micrographs nor evaluated using the independent data set due to the poor training accuracy presented by the model.

The existing image classification studies using convolutional neural networks have used large image databases on a big data scale to train their models (Das et al., 2019; Ounkomol et al., 2018). Real-time image acquisition is another practical approach to obtain a larger dataset for training such models in order to achieve better accuracy (Heo et al., 2017).



---

### 5.2.6 Comparison between the Classification Models

The accuracies obtained from the classification models, except CNN, are presented in Table 5.21.

**Table 5-21. Classification results of the industrial emulsification process obtained from the models.**

Model	Cross-Validation	Independent Batch
PC-LDA	100%	100%
MLR	98%	92%
RF	99.78%	96%
VNN	89%	85.3%

The ‘Cross-Validation’ column, in Table 5.21, represents the classification results of the data obtained from the first batch of Process-2. The ‘Independent Batch’ column represents the classification results of unseen micrographs acquired from the independent batch of the emulsion. The PC-LDA and the RF models presented the highest accuracies in both cross-validation and in classifying the unseen data. However, the PC-LDA approach presented the best results (100%) with the cross validated data as well as the unseen data set. This can be explained by the balance achieved in the bias-variance trade-off in the PC-LDA model. The entire range of droplets in a micrograph was used to train the PC-LDA model instead of the summary characteristics of each micrograph such as the mean droplet area, mean Feret, mean minFeret, mean perimeter and droplet count. This approach was successful in achieving a low bias and aided in recognising the underlying pattern of the training data. Moreover, the dimensionality reduction of the input feature space using PCA and LDA also helped to achieve low variance in the model predictions.

---

Among the deep learning models, the VNN model performed better than the CNN. This can be attributed to the different approaches used in the two models. Droplet characteristics, such as size and count, were the predictors of the VNN model, while the CNN model used the image pixel intensity values as the predictors. The droplet characteristics varied significantly between the four TAMU categories (Figure 4.12), while the pixel intensity values did not show much variation between the four categories (Figure 5.17). In addition to this, a much larger data set, should increase the classification accuracy of both the neural network approaches.

The current study has investigated the application of microscopic image analysis combined with machine learning algorithms in the classification of in-process emulsion samples. The classification is based on the variation in the droplet characteristics observed during the emulsification process. Computer vision analysis, integrated with machine learning, has gained wide attention in a range of industries to achieve fast, accurate and objective evaluation of product quality (Bertani et al., 2017; Caballero et al., 2018; Manak et al., 2018; Pfeil et al., 2018). However, there have been very few systematic studies reported so far for the quality evaluation of emulsions using machine-vision techniques (Amokrane et al., 2016; Khalil et al., 2010; Maass et al., 2012; Panckow et al., 2017). The existing literature has focused only on the inline extraction of droplet characteristics from emulsions using various image processing techniques. The extension of such techniques using machine learning algorithms to develop an automated classification tool is currently unexplored in emulsion manufacturing.

The PC-LDA and the Random Forest models, developed in this research, have achieved 96% to 100% accuracy in the TAMU classification of emulsion micrographs. Both these

---

models presented high accuracy when validating micrographs from a different production batch of the same emulsion. This demonstrated the potential of the models to classify the micrographs at various stages of emulsification and to identify the optimum point of emulsion processing, where the droplet characteristics attain the target distribution.

A similar study of droplet size characterisation was performed in barley milk samples by (Kljusuric et al., 2015) to identify the optimum processing time. However, their study has not attempted any supervised classification approaches. A previous study in machine vision developed an LDA based classification model to classify wheat kernels into four categories based on their vitreousness (Venora et al., 2009). Their model has achieved an overall classification accuracy of 96%. There are very few studies reported in the literature, which have used image processing for particle characterisation and image classification using models such as LDA and RF (Calderon et al., 2017; C. Cumbaa & Jurisica, 2005; C. A. Cumbaa & Jurisica, 2010). Those studies were conducted in the area of protein crystallization and their classification models obtained an overall classification accuracy of 80% to 95%.

The TAMU classification models, using RF and PC-LDA, developed in the current study were successful in achieving better accuracy compared to previous models. These classification models were extended to regression models to predict the remaining processing time, the time required to achieve the target characteristics, in emulsification. These prediction models are discussed in Chapter 7.

---

### 5.3 Conclusions

This chapter discusses the machine learning and the deep learning models developed for the TAMU classification of the in-process micrographs obtained from Process-2. Three machine learning models, PC-LDA, MLR and RF, and two deep learning models, VNN and CNN, were developed. The Principal Component based Linear Discriminant Analysis and the Random Forest classification approaches presented 100% and 96% accuracy respectively. The PC-LDA, which used a different approach from the other models, demonstrated 100% accuracy in the classification of the independent batch. The deep learning models presented poor performance compared to the machine learning models. The limited amount of training data as well as the imbalanced data set may be the reason for the low classification accuracy presented by the VNN model. The low frequency classes, Acceptable and Target, presented most of the misclassifications in the VNN approach. Due to the limitation in the data size, a split micrograph approach was used for the CNN model. The input features of the CNN model were the pixel intensity values of the micrographs, which was different from all the other model approaches. However, the CNN model demonstrated a very poor training accuracy of 44.8%. The availability of a much larger data set should further increase the classification accuracy of both neural network approaches.

The next chapter, Chapter 6, discusses a comparative analysis performed between manual and machine learning based micrograph classification to investigate the performance of both approaches.

---

## Chapter 6

### Comparative Analysis of Manual and Machine Learning Classification

#### 6.1 Introduction

The objective of this chapter is to compare the classification of micrographs performed manually, by trained analysts, to that of a machine learning approach. This comparison studies the accuracy, repeatability, reproducibility and the time taken to perform the analysis by both classification methodologies. This work was conducted, at an early stage in this research, before the improved HBT segmentation was investigated and before obtaining the Process-2 micrographs. Therefore, the results discussed in this chapter are completely based on Process-1 micrographs acquired with BF x10 magnification and appraised by EST into their TAMU categories.

The droplet characteristics obtained using the EST segmentation were used to develop a Principal Component based Linear Discriminant Analysis model (PC-LDA). This chapter assesses the PC-LDA model in classifying 20 random micrographs into TAMU categories and compares the results with manual classification performed by four micrograph analysts. The comparative study was performed using a technique called Attribute Agreement Analysis (AAA) with Minitab software.

Studies have previously been performed to assess and improve the capability of operators in the quality evaluation of manufacturing processes using Measurement System Analysis (MSA) (Marques et al. 2018, Al-Refaie and Bata 2010, Browne *et al.* 2010, Murphy *et al.* 2009, Vago and Kemeny 2012). MSA, being a key element in manufacturing process

---

control, has gained significant attention in industries (Lyu and Chen 2008, Chen and Lyu 2011, de Mast and van Wieringen 2010). MSA is performed by analysing the accuracy (bias) and precision of the measurement system used in industries. Accuracy is measured as the percentage of agreement of the evaluation with the standard. Precision is measured in terms of the repeatability and reproducibility (R&R) of the measurement system. MSA can be performed on both quantitative as well as attribute data. The statistical technique used to investigate the accuracy and precision of an evaluation system using attribute data is called Attribute Agreement Analysis (AAA) (Minitab, Attribute Agreement Analysis, n.d.).

AAA is mostly performed in industries to evaluate manual measurement systems that classify machine parts into defective and non-defective categories. The same parts are classified by several inspectors at different times and on each subsequent classification, their previous classification is unknown. AAA allows one to determine the measurement system accuracy by comparing the results to the known status or category. It also allows the repeatability and reproducibility to be assessed. Repeatability is a measure of the consistency in the individual assessments when repeated. Reproducibility is a measure of the agreement between the inspectors. The AAA results indicate if the measurement system requires any improvement or replacement. The evaluation of the AAA results is based on the percentage of agreement values as well as Fleiss's Kappa statistic ( $\kappa$ ), which is a measure of the performance of each inspector in the evaluation of each attribute category. It compares the amount of agreement between inspectors purely by chance. Fleiss's Kappa statistic is defined by Equation 6.1, where the denominator is the degree of agreement achievable above chance and the numerator is the degree of agreement actually achieved above that of chance (Nichols *et al.*, 2010).  $\bar{P}$  is the mean of the proportion of the evaluations, *i.e.* the mean of the  $p$  values, which represent the extent to

---

which the inspectors agree with each other on each category.  $\bar{P}_e$  is the sum of the squared values of the proportion of all evaluations to each category (Nichols *et al.*, 2010).

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad \text{Equation 6.1}$$

The larger the Kappa statistic, the better the agreement. A measurement system with a Kappa statistic between 0.90 and 1.00 is considered excellent and if  $\kappa \leq 0$ , it is considered poor or no agreement (Marques *et al.* 2018).

AAA is widely performed on various operator evaluation systems such as defect classification of images, classification of machine parts into categories based on specification *etc.* (Marques *et al.* 2018, Al-Refaie and Bata 2010, Browne *et al.* 2010, Murphy *et al.* 2009, Vago and Kemeny 2012, Chen and Lyu 2011, Lyu and Chen 2008, de Mast and van Wieringen 2010).

## 6.2 Classification Study and Attribute Agreement Analysis (AAA)

A set of 20 random micrographs, five from each TAMU category, were selected for assessing the classification accuracy of the PC-LDA model against the manual classification performed by four micrograph analysts. The TAMU labels, used in this chapter, correspond to the Process-1 micrographs discussed in chapters 3 and 4. These are presented in Table 6.1.

**Table 6-1. Micrographs from Process-1 labelled as TAMU.**

Label	Process Interval (minutes)	Number of micrographs
U	5	5
M	10	5
A	15	5
T	20, 25, 30	5

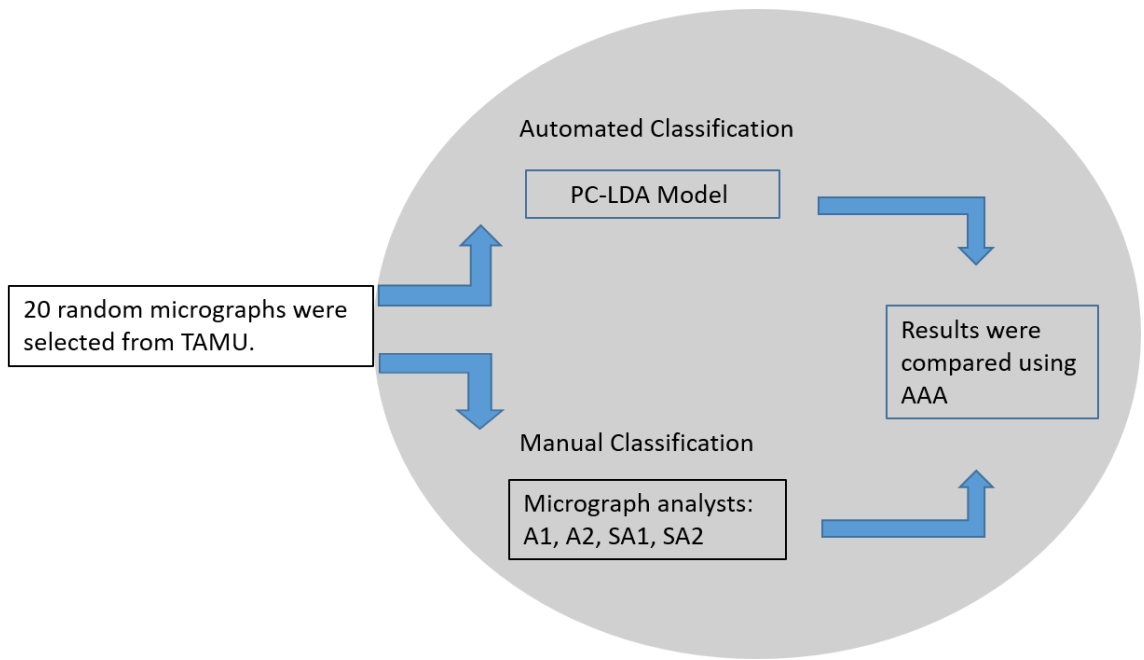
---

Two Analysts (A1 and A2) and two Senior Analysts (SA1 and SA2) were selected for the manual classification of the 20 micrographs. An independent set of 12 micrographs, three from each category, were used to train the four micrograph analysts. The training set was accessible to the analysts at all stages of the assessment. After the training phase, the validation set of 20 micrographs were classified automatically using the PC-LDA model and manually by the four micrograph analysts. For the manual classification, the micrographs were presented as A4 laminated printouts in a random order. The automated and manual classification results were analysed and compared in Minitab 17, using the AAA technique to assess the accuracy, *i.e.*, agreement with the known classification and repeatability and reproducibility between the assessments. The results of the AAA are discussed in the following section.

### **6.3 Results and Discussion**

A schematic representation of the comparative classification study is presented in Figure 6.1. The classification results of the Analysts (A1 and A2) and the PC-LDA model are presented in Table 6.2 and those of the Senior Analysts (SA1 and SA2) and the model are presented in Table 6.3. The Assessment1 and Assessment2 columns, of Tables 6.2 and 6.3, represent the two assessments performed by each analyst for the 20 micrographs. The model assessment column represents the classification results of the PC-LDA model. The standard column represents the ‘known classification’ category of the micrographs.





**Figure 6.1. The comparative study between manual and automated micrograph classification using Attribute Agreement Analysis (AAA).**

**Table 6-2. Classification results of the Analysts (A1 and A2) and the PC-LDA model. The orange cells represent the misclassifications.**

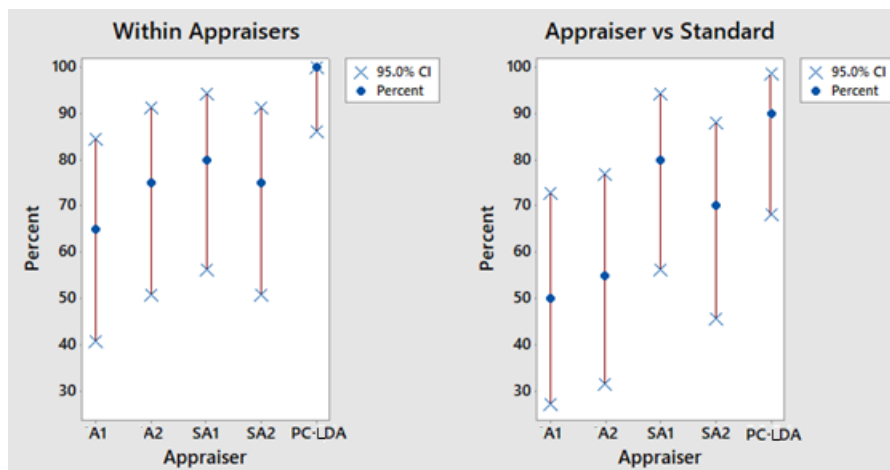
Micro-graphs	A1 Assessment1	A1 Assessment2	A2 Assessment1	A2 Assessment2	PC-LDA Model Assessment	Standard
1	Acceptable	Marginal	Acceptable	Acceptable	Marginal	Marginal
2	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
3	Target	Target	Acceptable	Acceptable	Target	Target
4	Target	Target	Target	Target	Unacceptable	Acceptable
5	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
6	Acceptable	Target	Target	Target	Target	Target
7	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
8	Acceptable	Marginal	Acceptable	Acceptable	Acceptable	Acceptable
9	Target	Target	Target	Acceptable	Acceptable	Acceptable
10	Target	Target	Marginal	Acceptable	Target	Target
11	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
12	Marginal	Acceptable	Acceptable	Target	Marginal	Marginal
13	Marginal	Acceptable	Acceptable	Marginal	Acceptable	Acceptable
14	Target	Target	Target	Target	Target	Target
15	Target	Acceptable	Target	Target	Target	Target
16	Acceptable	Acceptable	Acceptable	Acceptable	Marginal	Marginal
17	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
18	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
19	Acceptable	Target	Target	Acceptable	Unacceptable	Acceptable
20	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable

**Table 6-3. Classification results of the Senior Analysts (SA1 and SA2) and the PC-LDA model. The orange cells represent the misclassifications.**

Micro-graphs	SA1 Assessment1	SA1 Assessment2	SA2 Assessment1	SA2 Assessment2	PC-DA Model Assessment	Standard
1	Marginal	Marginal	Acceptable	Marginal	Marginal	Marginal
2	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
3	Target	Acceptable	Target	Target	Target	Target
4	Acceptable	Acceptable	Acceptable	Acceptable	Unacceptable	Acceptable
5	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
6	Target	Acceptable	Target	Target	Target	Target
7	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
8	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
9	Acceptable	Marginal	Marginal	Acceptable	Acceptable	Acceptable
10	Target	Target	Target	Target	Target	Target
11	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
12	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
13	Acceptable	Acceptable	Acceptable	Marginal	Acceptable	Acceptable
14	Target	Target	Acceptable	Target	Target	Target
15	Target	Target	Acceptable	Acceptable	Target	Target
16	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
17	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable
18	Marginal	Marginal	Marginal	Marginal	Marginal	Marginal
19	Marginal	Acceptable	Marginal	Acceptable	Unacceptable	Acceptable
20	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable	Unacceptable

The classification assessments, obtained from Tables 6.2 and 6.3, were statistically analysed using AAA. The corresponding repeatability and accuracy results of the classification are graphically presented in Figure 6.2. The left panel (Within Appraisers) explains the repeatability measure of each analyst and the model between their first and second assessments while the right panel (Appraiser vs Standard) represents the classification accuracy, i.e., the percentage of agreement of each analyst and the model with the standard (the known category). In each panel, the first two lines represent A1 & A2 respectively, while the next two lines represent SA1 & SA2 respectively. The fifth line represents the PC-LDA model.

The left panel of Figure 6.2 shows that A1 & A2 have a repeatability of 70% (average of 65 and 75) indicating that they have only 70% agreement between their first and second assessments. SA1 & SA2 give an average of 77.5% repeatability (80 and 75), which is higher compared to A1 & A2. The machine learning model gives 100% repeatability, as expected. The right panel of Figure 6.2 shows the accuracy, *i.e.*, percentage of agreement of the assessments with the standard. The values indicate that A1 & A2 are in agreement only 52.5% (average of 50 and 55) of the time with the standard and SA1 & SA2 are in agreement 75% of the time with the standard. The PC-LDA machine learning model shows 90% agreement with the standard.



**Figure 6.2. Attribute agreement analysis (AAA) results from Minitab.**

The percentage of agreement values of each analyst and the model along with their confidence intervals (CI) and Kappa statistics for repeatability and accuracy are provided in Tables 6.4 and 6.5 respectively. Fleiss's Kappa Statistic is calculated, for each category of classification, as a measure of the evaluation system and an overall Kappa statistic is also calculated for each analyst and the model. The higher the Kappa value, the better the agreement is. A Kappa statistic of 0.90 to 1.00 represents the best assessment system,

0.89 to 0.70 reflects a capable assessment system that can be improved and a value less than 0.70 indicates that the assessment method requires improvement.

**Table 6-4. Repeatability agreement within appraisers.**

<b>Assessment Agreement</b>				
Appraiser	Inspected	Matched #	Percent	95% CI
A1	20	13	65	(40.78, 84.61)
A2	20	15	75	(50.90, 91.34)
SA1	20	16	80	(56.34, 94.27)
SA2	20	15	75	(50.90, 91.34)
PC-LDA Model	20	20	100	(86.09, 100.00)
# Matched: Appraiser agrees with themselves across trials.				
<b>Fleiss' Kappa Statistics</b>				
Appraiser	Response	Kappa		
A1	Unacceptable	1.000		
	Marginal	0.375		
	Acceptable	-0.003		
	Target	0.658		
	<b>Overall</b>	<b>0.528</b>		
A2	Unacceptable	1.000		
	Marginal	0.608		
	Acceptable	0.430		
	Target	0.624		
	<b>Overall</b>	<b>0.659</b>		
SA1	Unacceptable	1.000		
	Marginal	0.762		
	Acceptable	0.467		
	Target	0.688		
	<b>Overall</b>	<b>0.732</b>		
SA2	Unacceptable	1.000		
	Marginal	0.524		
	Acceptable	0.373		
	Target	0.827		
	<b>Overall</b>	<b>0.663</b>		
PC-LDA Model	Unacceptable	1.000		
	Marginal	1.000		
	Acceptable	1.000		
	Target	1.000		
	<b>Overall</b>	<b>1.000</b>		

**Table 6-5. Accuracy Agreement of each Appraiser vs Standard.**

<b>Assessment Agreement</b>				
Appraiser	Inspected	Matched #	Percent	95% CI
A1	20	10	50	(27.20, 72.80)
A2	20	11	55	(31.53, 76.94)
SA1	20	16	80	(56.34, 94.27)
SA2	20	14	70	(45.72, 88.11)
PC-LDA Model	20	18	90	(68.30, 98.77)
# Matched: Appraiser agrees with the standard.				
<b>Fleiss' Kappa Statistics</b>				
Appraiser	Response	Kappa		
A1	Unacceptable	1.000		
	Marginal	0.570		
	Acceptable	0.098		
	Target	0.574		
	<b>Overall</b>	<b>0.565</b>		
A2	Unacceptable	1.000		
	Marginal	0.375		
	Acceptable	0.204		
	Target	0.420		
	<b>Overall</b>	<b>0.497</b>		
SA1	Unacceptable	1.000		
	Marginal	0.875		
	Acceptable	0.740		
	Target	0.844		
	<b>Overall</b>	<b>0.866</b>		
SA2	Unacceptable	1.000		
	Marginal	0.749		
	Acceptable	0.553		
	Target	0.772		
	<b>Overall</b>	<b>0.766</b>		
PC-LDA Model	Unacceptable	0.762		
	Marginal	1.000		
	Acceptable	0.688		
	Target	1.000		
	<b>Overall</b>	<b>0.866</b>		

Table 6.4 shows that the 95% CI for the true percentage of repeatability of the analysts varies from 40.78% to 94.27% ("Minitab 18 Support, "). The corresponding 95% CI for the PC-LDA model varies from 86% to 100%. The matched column, in Table 6.4, shows the number of first and second assessments of each analyst in agreement. The individual

---

Kappa statistic for each category shows that the repeatability of the analysts reduced because of primarily failing to classify the micrographs of the “Acceptable” category. These were incorrectly classified as either “Marginal” or “Target”. The overall Kappa statistic of all the four analysts for their repeatability measures is  $\leq 0.73$ , while the overall Kappa statistic for the model is 1.0. From this result, it can be inferred that the machine learning classification model performed much more effectively compared to the manual evaluation of micrographs.

The matched column in Table 6.5 shows the number of assessments of each analyst in agreement with the standard across all the trials. The 95% CI shows that the accuracy agreement of the model with the standard varies from 68% to 99%, while the CIs of the individual analysts vary from 27% to 94%. The overall Kappa statistic of the individual analysts for their accuracy agreement with the standard, varies from 0.50 to 0.87. The analysts have their overall kappa statistic less than 0.7, with the lack of accuracy agreement attributed to the ‘Target’, ‘Acceptable’ and ‘Marginal’ categories. The senior analysts demonstrate a better overall classification performance, than the analysts, with a Kappa statistic  $> 0.7$ . The overall kappa statistic of the model was 0.87.

However, the PC-LDA model presented two misclassifications in the ‘Acceptable’ category by classifying those micrographs into ‘Unacceptable’ and therefore, the kappa statistic presented by the model for the Unacceptable category was 0.762 as seen from Table 6.5.

This is mainly due to the following reasons:

1. The TAMU classification models presented in chapter 5 used HBT detection, which was found superior to the EST detection on which the AAA was conducted.

2. The EST, to a certain extent, had difficulty in detecting smaller droplets. This could have contributed to reduced efficiency in detecting the actual size parameters of smaller droplets. Details of the EST analysis are presented in Chapter 4.
3. Some of the micrographs in the ‘Acceptable’ category contained air bubbles. This resulted in a decrease in the number of droplets detected and has reduced the droplet count value measured by the EST algorithm.

Due to the above stated reasons, the model has misinterpreted those two ‘Acceptable’ micrographs as ‘Unacceptable’ with a reduced droplet count and not enough size accuracy. The PC-LDA model performed well in the classification of the ‘Target’ and the ‘Marginal’ categories, which presented a kappa statistic equal to 1.00. The overall Kappa statistic shows a comparable accuracy between the senior analysts and the model, however the performance of the senior analysts in the classification of borderline categories is inferior compared to the PC-LDA model.

The reproducibility, otherwise known as inter-observer agreement, results between the four analysts are presented in Table 6.6. The results indicate that the analysts are in agreement only 35% of the time with a 95% CI varying from only 15.39% to 59.22%, which is considered very poor.

**Table 6-6. Reproducibility agreement between Appraisers.**

<b>Assessment Agreement</b>			
Inspected	# Matched	Percent	95% CI
20	7	35.00	(15.39, 59.22)
# Matched: All appraisers’ assessments agree with each other			
<b>Fleiss’ Kappa Statistics</b>			
Response	Kappa		
Unacceptable	1.000		
Marginal	0.507		
Acceptable	0.232		
Target	0.462		
Overall	0.547		



---

The overall results of this study demonstrated that the PC-LDA model performed much more effectively in the repeatability and accuracy of micrograph classification, compared to manual evaluation. The reproducibility of the analysts was also found to be limited, which is mainly due to the subjectivity associated with identifying the correct micrograph category. The majority of the classification errors in the manual evaluation occurred in the border line cases between the 'T','A' and 'M' categories, which agreed with a previous study conducted by Ulery *et al.* (Ulery et al., 2012) in the case of forensic fingerprint examiners. In addition, the micrograph classification performed by each analyst for the 20 micrographs took approximately 15 minutes to complete, while the automated classification approach using the PC-LDA model was completed in five seconds.

Automated processing, analysis and classification of microscopic images based on machine learning algorithms have been identified as a powerful technique in many areas such as molecular biotechnology, bioinformatics, microsystems technology *etc.* (Pfeil et al., 2018; Schulze et al., 2013). An automated micrograph classification approach using neural networks has recently been investigated by Pfeil et al., (2018) in the analysis of brewing yeasts for cell counting and classification. Their study concluded that the automated approach was ten times faster than manual classification, more cost effective, objective and, free from human errors. Another study previously conducted on the measurement and calibration of droplet size distributions in emulsions have reported an average difference of 5.1% in the accuracy of evaluation between two analysts (Boxall *et al.* 2010). A similar study conducted by Maaß *et al.* (Maass et al., 2012) found  $\pm 5\%$  deviation in the manual evaluation of average droplet size and  $\pm 15\%$  deviation in the evaluation of droplet count, despite the micrographs being of good quality with a smaller

---

number of droplets. The deviation was almost double when repeated with the same analysts.

The manual micrograph classification, presented in the current study, shows 5% to 10% variation in the accuracy within the analyst and senior analyst groups respectively and 30% variation between the groups. The automated machine learning approach is 40% more accurate than the least performing analyst, 10% more accurate than the best performing analyst and 100% repeatable. The results clearly demonstrate that the automated classification is superior to the manual classification of micrographs with respect to greater speed (180 times faster), accuracy and repeatability. The machine learning approach reported in this study is significantly faster than manual evaluation when compared to the results from similar studies reported previously (Pfeil et al., 2018).

#### **6.4 Detailed Analysis of the study**

A comparative study between manual micrograph assessment and a machine learning based assessment is presented in this chapter. This study has presented a detailed insight into the subjectivity associated with the manual technique, which is currently practised in emulsion manufacturing industries. This is a proof-of-concept study conducted using the in-process micrographs of 10x magnification obtained from a 30 minutes long emulsification process (Process 1). A detailed review of the important aspects, novelty and limitations of this study are presented below.

A set of 20 micrographs from the TAMU categories were used to train the machine learning model and the analysts, which included both junior and senior. An independent set of TAMU micrographs (3 from each category) were used for validation. Various aspects of the validation process such as accuracy, repeatability and reproducibility were evaluated using the AAA technique. The machine learning model presented an improved

---

overall accuracy of 10% to 40% compared to the manual validation performed by the analysts. The model also proved 180 times faster, by finishing the classification of 12 micrographs in 5 seconds, compared to the manual validation, which took 15 minutes to finish. The repeatability of the model is 100%. These results formed a significant contribution to the existing literature which has studied how the subjectivity in visual inspection techniques can affect product validation in industry.

The limitations encountered in this study can be attributed to the Edge and Symmetry Technique (EST) used for droplet detection. The machine learning model used for the Attribute Agreement Analysis is a PC-LDA model developed using the droplet count and average droplet size (area, perimeter, Feret and minFeret) variables obtained from the EST. In addition, the presence of air bubbles in two of the 'Acceptable' micrographs also affected the correct interpretation of the micrograph category by the model. The limitations presented by the EST in the detection of smaller droplets throughout the emulsification process is rectified by the new Histogram Based Technique (HBT). The PC-LDA model developed using the fully automated and improved HBT segmentation, presented in Chapter 5, is expected to overcome the limitations presented in this study and to provide improved accuracy compared to the EST approach. A future extension of this comparative study using the HBT integrated PC-LDA model is proposed in the future work section of this thesis.

## **6.5 Conclusions**

The manual versus machine learning model comparison, presented in this chapter, is a novel approach in the area of emulsion micrograph classification. The machine learning approach demonstrated higher accuracy (90%) and repeatability (100%) in the classification of in-process emulsion micrographs. The accuracy of the manual evaluation

---

varied from 55% to 80% and the repeatability measure varied from 65% to 80% between the analysts. The Kappa statistic for the repeatability of manual classification fell below 0.7, which shows that manual evaluation is not a satisfactory method and requires improvement. The accuracy of the manual classification was also limited in the case of borderline categories, which were much more effectively classified by the PC-LDA model. The automated evaluation approach presented enhanced speed, accuracy and repeatability, when compared to the manual classification performed by the analysts.

---

## Chapter 7

### Predicting the Emulsification Time

#### 7.1 Introduction

This chapter discusses the regression models developed, for Process-2, to predict the Remaining Processing Time (RPT) required to attain the target droplet characteristics, at all stages of emulsification. The major objective of the prediction model is to enable the process analysts in evaluating ‘how much more time is required to complete the processing’ when presented with micrographs obtained from in-process emulsion samples. This will potentially avoid over-processing in emulsion manufacturing and reduce wastage of resources such as time, energy and manpower.

The optimum processing time, of Process-2, was previously identified as 65 minutes according to the statistical analysis and the TAMU classification performed in Chapters 4 and 5 respectively. The machine learning techniques that demonstrated high efficiency in the TAMU classification were the PC-LDA, the Random Forest (RF) and the Multinomial Logistic Regression (MLR) in that order. Considering these techniques, a Linear Regression model and a Random Forest regression model were developed for the RPT prediction. The regression model’s performance, on unseen data, was evaluated using the independent samples/micrographs acquired from a different batch of the emulsion product. The micrograph acquisition from this independent batch is explained in the methodology chapter (section 3.3.2). The optimum processing time of the independent batch was identified as approximately 60 minutes, which was five minutes earlier than the first batch. This is presented later in this chapter. The intention of the

---

regression model is to predict the remaining emulsification time when presented with an unseen sample micrograph.

## 7.2 Regression Modelling: Results and Discussion

The results obtained from the regression models are discussed in this section. The methodology was explained previously in Chapter 3.

### 7.2.1 Linear Regression Model

A linear regression (LR) model was developed using the same set of five predictor variables,  $x_1$  to  $x_5$ , which were used to develop the MLR and RF classification models (Table 7.1). The response (dependant) variable of the model was the Remaining Processing Time (RPT), which was the time required by the emulsion to attain the target characteristics. The Specified RPT of the emulsion, at various stages of emulsification, was calculated as the difference between the current process time and the optimum processing time, which was identified as 65 minutes for the first batch of Process-2 (Figure 4.12).

**Table 7-1. Predictor variables of the LR model.**

$x_1$	Mean droplet area
$x_2$	Mean Feret
$x_3$	Mean perimeter
$x_4$	Mean minFeret
$x_5$	Droplet count

The total dataset of 150 micrographs, from the first batch, were split into training (70%) and test (30%) sets. The createDataPartition function from the R caret package was used for splitting the data. The linear model function, lm, was used to build the prediction model. Table 7.2 shows the model summary for the training set.

**Table 7-2. LR Model Summary for the 70% training data.**

lm(formula = RPT ~ x1 + x2 + x3 + x4 + x5, data = training)				
<b>Residuals:</b>				
Min	1Q	Median	3Q	Max
-12.210	-3.452	0.352	3.477	11.242
<b>Coefficients:</b>				
	Estimate	Std. Error	t value	Pr (> t )
Intercept	159.342	36.132	4.410	2.63e <sup>-05</sup> ***
x1	1.084	0.862	1.258	0.211
x2	-57.346	17.968	-3.192	0.002 **
x3	1.866	3.872	0.482	0.631
x4	53.186	22.011	2.416	0.018 *
x5	-0.016	0.002	-7.816	5.94e <sup>-12</sup> ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 5.9 on 99 degrees of freedom (DF)				
Multiple R-squared: 0.929		Adjusted R-squared: 0.926		
F-statistic: 260.6 on 5 and 99 DF			p-value: < 2.2e-16	

The adjusted  $R^2$  value of the model was obtained as 93%, which means that the five predictor variables explained 93% of the total variance in the response variable, the RPT. The Standard Error or Root Mean Square Error (RMSE) given by the model was 5.9, which is a reasonably low value. RMSE is the square root of the ‘mean of squared residuals’ of the model. The  $p$ -values of the predictor variables in the summary table explains their significance in the model prediction. These are calculated using  $t$ -tests, which estimates the statistical significance of each variable in the presence of the other variables. Droplet count ( $x_5$ ) and Feret diameters ( $x_2$  and  $x_4$ ) were found highly significant compared to droplet area and perimeter. The F-statistic of 260.6 with a  $p$ -value < 0.01 indicates that the null hypothesis could be rejected, and the model performs reasonably well with the training data. The linear regression model is represented by Equation 7.1. Table 7.3 presents the ANOVA results of the model.

$$y = 159.34 + 1.08x_1 - 57.35x_2 + 1.87x_3 + 53.19x_4 - 0.02x_5 \quad \text{Equation 7-1}$$

Where  $y$  is the response variable, the RPT.  $x_1$  to  $x_5$  represent the predictor variables. The model coefficients are taken from the output table (Table 7.2).

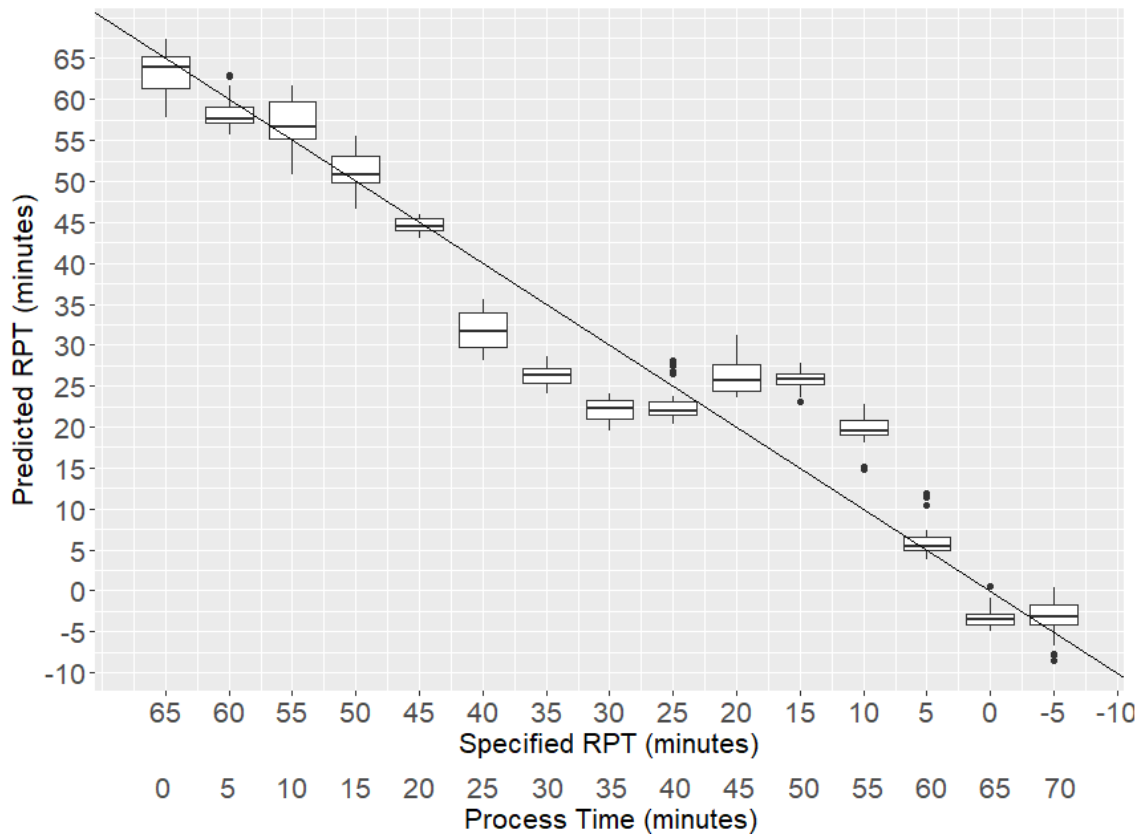
**Table 7-3. ANOVA results from the LR model.**

Analysis of Variance Table					
Response: RPT					
	DF	Sum Sq.	Mean Sq.	F value	Pr(>F)
$x_1$	1	37480	37480	1076.70	$< 2.2e^{-16} ***$
$x_2$	1	4072	4072	116.98	$< 2.2e^{-16} ***$
$x_3$	1	1027	1027	29.52	$3.98 e^{-07} ***$
$x_4$	1	647	647	18.60	$3.82 e^{-05} ***$
$x_5$	1	2127	2127	61.09	$5.94 e^{-12} ***$
Residuals:	99	3446	35		

The ANOVA test calculates the significance of each predictor variable, in the model prediction, independent of the other predictor variables. Therefore, ANOVA is a more precise and powerful way of testing the variable significance. The low  $p$ -values obtained for the F-statistic of each predictor variable, from the ANOVA results, implies that all the five independent variables are identified as statistically significant in the model prediction.

Ten random partitions of the training and test data sets were made from the 150 micrographs. Each partition consisted of a 70/30 split of the training and test samples. Ten LR models were built from the training sets. The average RMSE of the 10 training models was 6.08. The Predicted RPT, of the 10 models, on their respective test sets was recorded and was plotted against the Specified RPT. A total number of 450 observations (45x10) were predicted. The box plots are presented in Figure 7.1.





**Figure 7.1. Predicted RPT of the test data (from the 10 LR models) plotted against the Specified RPT. The additional  $x$ -axis shows the current process time. The diagonal line represents the ideal case with zero error. Each box plot represents the test micrographs selected from that process interval.**

The  $x$ -axis of the above plot represents the specified RPT, for e.g. 65 represents the Specified RPT of the emulsion at the start of the process, while 60 represents the Specified RPT after five minutes of processing. The  $y$ -axis represents the Predicted RPT of the emulsion at each process interval. The values below zero represents over-processed samples. The diagonal line represents the ideal case with zero error. Each box plot of Figure 7.1 represents the test micrographs selected from a particular process time.

The box plots presented a good linear fit between the Predicted and the Specified RPT, during the initial 20 minutes and the final 15 minutes of processing, with the data points aligned close to the diagonal line, which represents an ideal case with zero error. This region of Figure 7.1 corresponds to the ‘Unacceptable’, ‘Acceptable’ and ‘Target’

---

categories of Process-2. The Predicted RPT of the samples obtained from 25 to 55 minutes of the industrial processing are represented by the box plots that do not intersect the diagonal line (Figure 7.1). These micrographs belonged to the period of Process-2, which presented minimal variation in the droplet size and count and required extra time to break down the droplets further (section 4.3). This explains the slightly non-linear predictions observed in the RPT values of those samples. The RMSE of the 450 test predictions, from the 10 models, was calculated using Equation 7.2 and was obtained as 6.03. This was close to the average RMSE achieved from the summary of the 10 training models (6.08). The overall variance explained by the test predictions was obtained by calculating the  $R^2$  and the adjusted  $R^2$  values using Equation 7.3 and Equation 7.4 respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \hat{y}_{ij})^2}{(n \times m)}} \quad \text{Equation 7-2}$$

Where  $y$  represents the Specified RPT and  $\hat{y}$  is the Predicted RPT.  $n$  represents the number of models, *i.e.*, 10 and  $m$  represents the number of test micrographs assessed by each model, *i.e.*, 45.

$$R^2 = 1 - \left( \frac{SS_{RES}}{SS_{TOT}} \right) \quad \text{Equation 7-3}$$

Where  $SS_{RES}$  is the residual sum of squares and  $SS_{TOT}$  is the total sum of squares (variance of the specified RPT).

$$R^2_{adj} = 1 - \left[ \frac{(1-R^2)(n-1)}{(n-k-1)} \right] \quad \text{Equation 7-4}$$

Where  $n$  is the number of observations and  $k$  is the number of predictor variables ( $n = 450$  and  $k = 5$  in the current case).

The adjusted  $R^2$  value for the test predictions was obtained as 92%. This demonstrated that the test predictions were successful in explaining 92% of the variance in the observed data. The RMSE and the adjusted  $R^2$  values obtained for the training and the test data

demonstrated that the regression model is not over fitting or under fitting the data. A detailed summary of the RPT predictions from the test data is given in Table 7.4.

**Table 7-4. Summary of the median Predicted RPT compared with the Specified RPT values of the test predictions. The green cells represent reasonably good predictions and the white cells represent poor predictions.**

Process Time (minutes)	Specified RPT (minutes)	Median of Predicted RPT (minutes)	Delta (minutes)
0	65	63.85	1.15
5	60	57.6	2.4
10	55	56.7	-1.7
15	50	50.8	-0.8
20	45	44.5	0.5
25	40	31.6	8.4
30	35	26.25	8.75
35	30	22.35	7.65
40	25	21.9	3.1
45	20	25.65	-5.65
50	15	25.9	-10.9
55	10	19.45	-9.45
60	5	5.5	-0.5
65	0	-3.5	3.5
70	-5	-3.1	-1.9

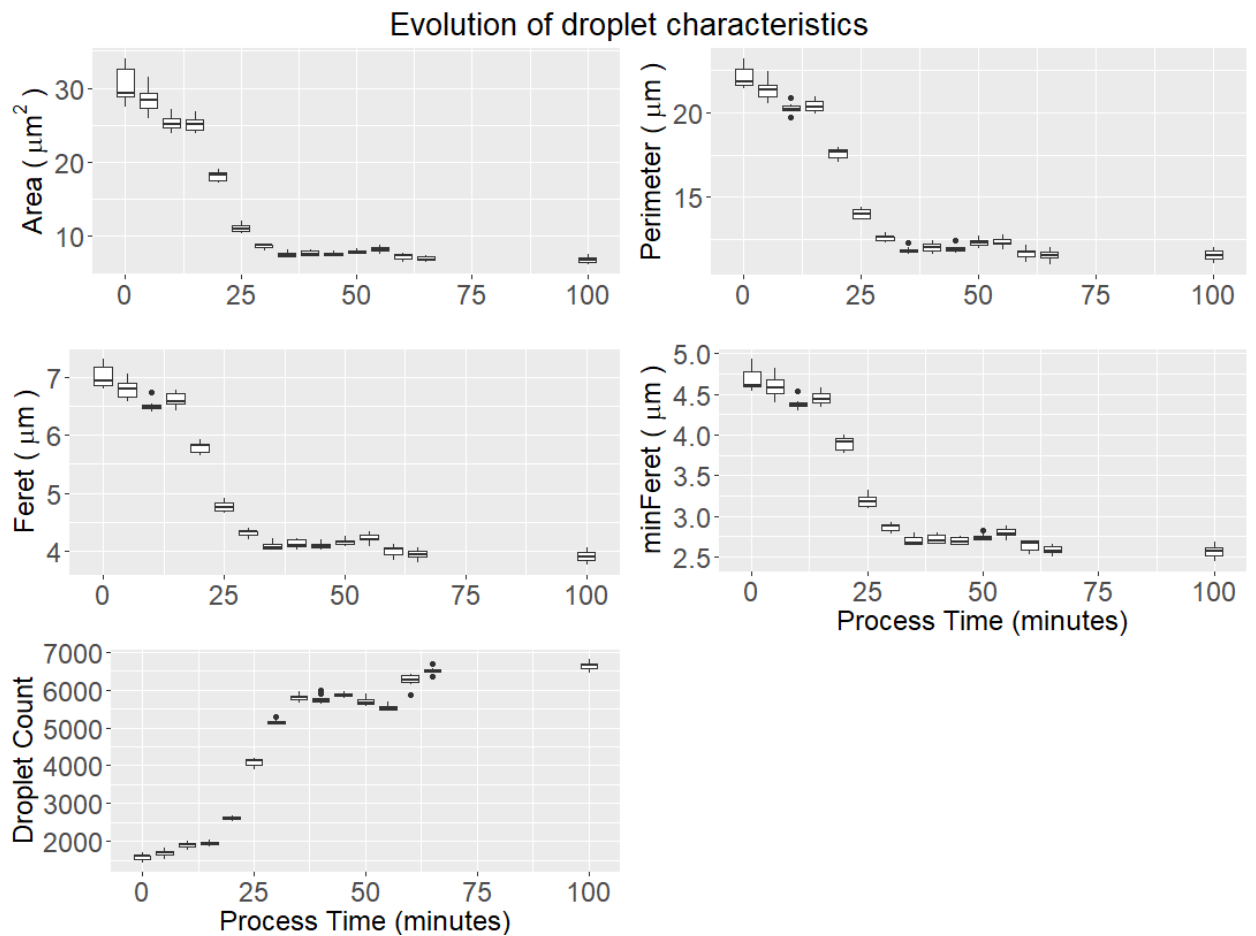
Column 4, Delta in Table 7.4 presents the difference between the Specified RPT and the median of the Predicted RPT at each process interval. Delta values  $\leq 5$  minutes are considered as reasonably good predictions and those  $> 5$  minutes are considered as poor predictions. The delta values  $\leq 5$  minutes were considered reasonably good because the variation in the droplet characteristics, between the TAMU categories observed in Chapter 5, was minimal within a 5-minute duration. In addition,  $\leq 5$  minutes was considered reasonably good relative to the wider range of variation observed in the overall prediction. The median values of the Predicted RPT presented minimal deviation, of 0.5 to 3.5 minutes, from the Specified RPT in the initial 20 minutes of processing. Higher deviations ranging from 3.1 to 10.9 minutes were observed afterwards. This is due to the

---

difficulty the model has in interpreting the droplet characteristics, which presented negligible variation, during the period of emulsification from 25 to 55 minutes.

The reasonably good predictions presented by the model during the initial 20 minutes of emulsification can benefit the process analysts in avoiding sampling during the middle of the process. For example., after 20 minutes of processing, the model predicted 44.5 minutes as the RPT required by the emulsion to achieve the Target. This value was in close agreement with the Specified RPT (45 minutes) of the emulsion at that point. This indicated that subsequent sampling can be performed after 44.5 minutes or slightly later. This information, provided by the model, can save resources and avoid repeated sampling, reduce the wastage of samples and, reduce the frequent risk of introducing impurities into the emulsion mixer.

The performance of the LR model was also evaluated, on unseen data, using the micrographs obtained from the independent batch of Process-2. Ten micrographs were obtained at every five-minute interval up to 65 minutes of processing and the final discharge samples were obtained after a further 35 minutes of cooling (section 3.3.2). The evolution of mean droplet size (area, Feret, minFeret and perimeter) and droplet count of this batch are presented as box plots in Figure 7.2.

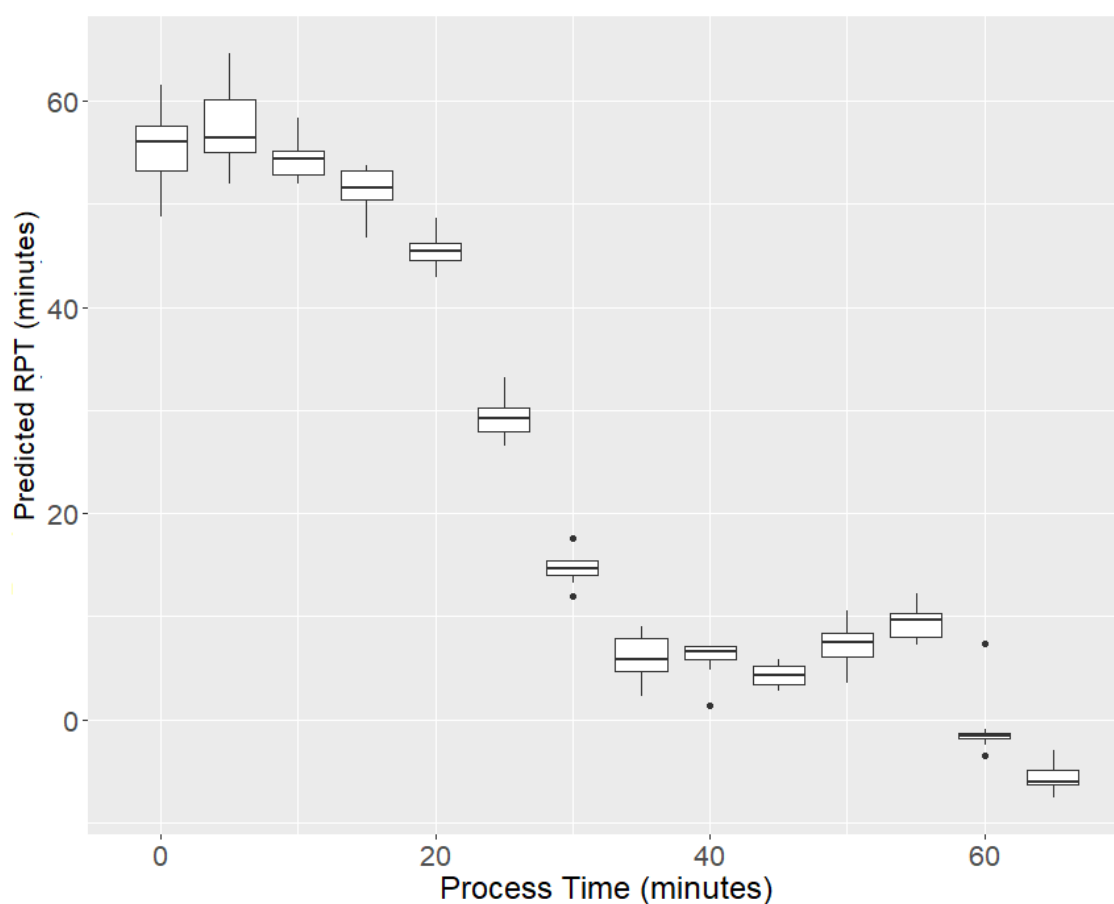


**Figure 7.2. Evolution of mean droplet characteristics a) area, b) perimeter, c) Feret, d) minFeret and e) droplet count from the emulsification process of the independent batch. The last box plot represents the 10 micrographs obtained from the discharge sample.**

The droplet size and count characteristics attained acceptable values quite early, which was after 30 minutes of processing. In addition, minimal variation was observed in the droplet characteristics between 35 and 55 minutes. After a further five minutes of processing, the characteristics attained their target values and remained steady. Therefore, this batch was observed to reach the optimum state of emulsification after approximately 60 minutes of processing, which was five minutes earlier than the first batch. The micrographs of the discharge sample were not considered for the prediction modelling, as there was no processing after 65 minutes and the emulsion was allowed to cool down in the mixer until the discharge sample was obtained.

---

The RPT of the unseen micrographs were predicted using the LR model, which was trained with the 150 micrographs from the first batch of Process-2. The Predicted RPT obtained for the 140 independent micrographs were plotted against their current process time. The box plots are shown in Figure 7.3. Each box represents 10 micrographs.



**Figure 7.3. Box plots showing the Predicted RPT (minutes) of the micrographs plotted against their current process time. Each box represents 10 micrographs.**

The RPT predicted for the independent micrographs after the initial five minutes (the first box plot) was approximately 56 to 60 minutes. The Predicted RPT was approximately 45 minutes for the micrographs obtained after 20 minutes. These predictions can provide a better insight, for the process analysts, into the emulsification process and confirm the completion time at an early stage. The model predicted an RPT value of approximately -

2 minutes for the micrographs obtained after 60 minutes. This indicated that the emulsion was over-processed for 2 minutes at that point.

A comparison between the median of the Predicted RPT and the Specified RPT values at each processing interval is presented in Table 7.5. The Specified RPT values, for the independent batch, were calculated in the same way as the Specified RPT was calculated for the first batch.

**Table 7-5. Comparison of the Predicted RPT values of the independent set of micrographs with the Specified RPT values. The reasonably good predictions are highlighted in green. The white cells represent poor predictions.**

Process Time (minutes)	Specified RPT (minutes).	Median of Predicted RPT (minutes)	Delta (minutes)
0	65	56.65	8.35
5	60	56.7	3.3
10	55	53.9	1.1
15	50	51.55	-1.55
20	45	44.75	0.25
25	40	28.2	11.8
30	35	13.95	21.05
35	30	5.2	24.8
40	25	6.15	18.85
45	20	3.8	16.2
50	15	6.55	8.45
55	10	8.85	1.15
60	5	-2.05	7.05
65	0	-6.45	6.45

The predictions showed higher deviations for the micrographs acquired from 25 to 50 minutes of processing. This is the period of the emulsification process, when there was little variation observed in the droplet size and count (Figure 7.2). This introduced difficulty for the model in interpreting the influence of the processing time on droplet characteristics during that period. The independent batch of Process-2 was observed to have reached the optimum emulsification time five minutes earlier than the first batch.

---

This has introduced a deviation of  $\approx -5$  minutes in the predictions. The predicted RPT after 60 minutes of processing was in close agreement with the optimum processing time identified for this batch, which was 60 minutes. In summary, the model demonstrated reasonably good potential in identifying over-processing at early stages of emulsification and as one approaches the optimum emulsification time.

### 7.2.2 Random Forest Regression Model

The Random Forest (RF) regression model was developed using the same set of five predictor variables,  $x_1$  to  $x_5$ , used for the LR model (Table 7.1). The response variable of the model was the RPT (Remaining Processing Time). This was calculated in the same way as for the LR model.

The 150 micrographs acquired from the industrial process were split into training (70%) and test (30%) samples. The model was built with the training data using R's `randomForest` package (Andy & Matthew, 2002). The number of random trees (*ntree*) was selected by default as 500 for the initial model. The number of variable splits at each node (*mtry*) was calculated as  $p/3$ , where  $p$  is the number of predictor variables. In the current case,  $p=5$  and therefore, the value of *mtry* was rounded to 1 by the algorithm. The model output is presented in Table 7.6 and the variation in the model error as *ntree* grows from 0 to 500 is shown in Figure 7.4.

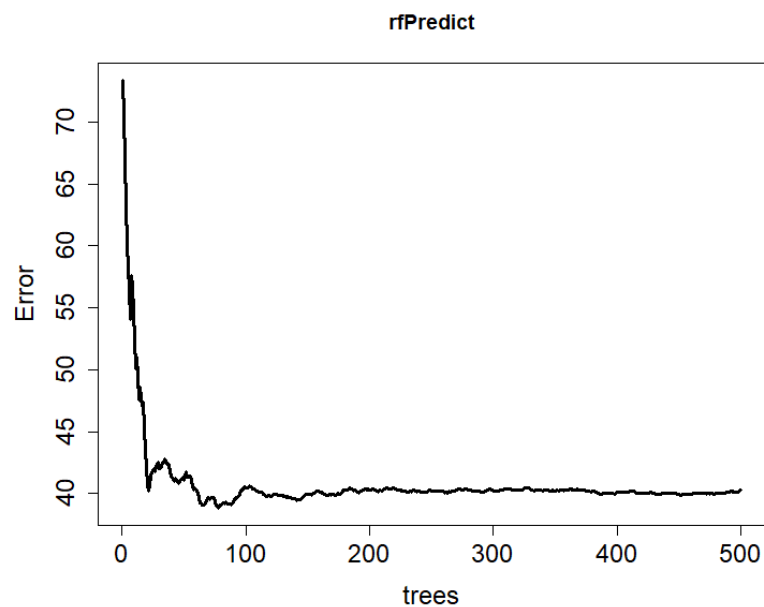


**Table 7-6. Random Forest Regression Model with default *n*tree and *m*try values.**

```
randomForest(formula = RPT ~ x1 + x2 + x3 + x4 + x5, data = training,
importance = TRUE)
      Type of random forest: regression
      Number of trees: 500
No. of variables tried at each split: 1

      Mean of squared residuals: 40.3
      % Var explained: 91.3
```

Table 7.6 reflects that the out-of-bag (OOB) predictions of the model explained 91.3% of the total variance of the training set. The mean of squared residuals represents the mean square error of the OOB predictions, which was recorded as 40.3. The square root of this value gives the RMSE of the model, which is 6.35.



**Figure 7.4. OOB error of the RF regression model plotted against the number of trees (*n*tree)**

Figure 7.4 shows that the OOB error of the model appears steady after an *n*tree value of approximately 200. Therefore, the model was tuned for a set of *n*tree values less than or equal to 200 and *m*try values from 1 to 5. The optimum values of the model parameters were obtained as 141 and 1 for the *n*tree and *m*try respectively. These values were used

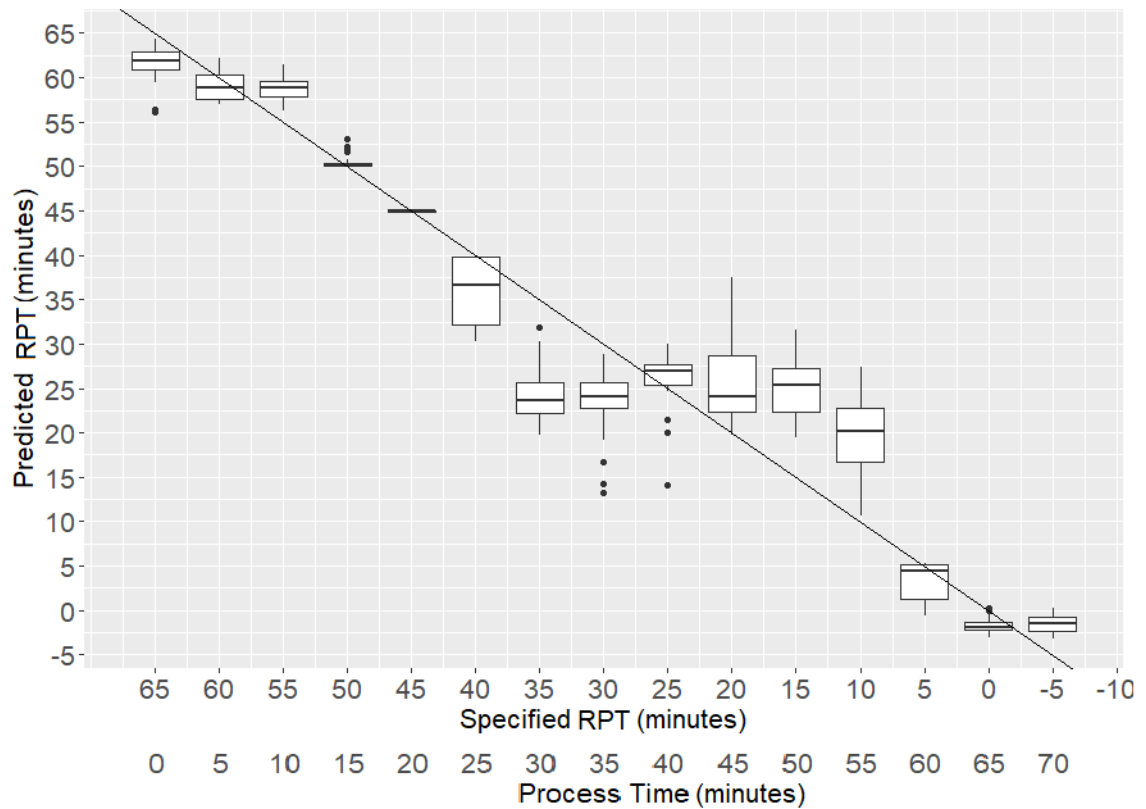
---

to rebuild the model. This enhanced the model with a marginal increase of 0.2% in the  $R^2$  value to 91.5% and a decrease in the mean of the squared residuals from 40.3 to 39.6. These results are presented in Table 7.7.

**Table 7-7. Random Forest Regression model summary with tuned *ntree* and *mtry* values**

<pre>randomForest(formula = RPT ~ x1 + x2 + x3 + x4 + x5, data = training, mtry = 1, ntree = 141, importance = TRUE)       Type of random forest: regression       Number of trees: 141 No. of variables tried at each split: 1        Mean of squared residuals: 39.6       % Var explained: 91.5</pre>
--

Ten random partitions of the training data (70%) and test data (30%) sets were created and 10 RF models were trained using these training sets. The average RMSE of the 10 training models was 6.23. These 10 models were tested using the corresponding test sets. The Predicted RPT values of the 10 models were recorded. The overall goodness of fit of the models was examined by plotting the Predicted RPT values, obtained from the 450 observations of the 10 test data sets, against their Specified RPT values. This is presented in Figure 7.5.



**Figure 7.5.** Predicted RPT of the test data (from the 10 RF models) on the y-axis and Specified RPT on the x-axis. The additional x-axis shows the corresponding process time. The diagonal line represents the ideal case with zero error. Each box plot represents the test micrographs selected from that process time.

The Predicted RPT of the test micrographs, in Figure 7.5, appears closer to their Specified RPT values during the initial 25 minutes and the final 15 minutes of the emulsification process. The predictions for the samples processed between 30 and 55 minutes lie off the diagonal line. This is similar to what was observed in the LR model predictions. This can be explained by the minimal variation observed in the model’s predictor variables during the 30 to 55-minute period of Process-2.

The RMSE of the test predictions was calculated as 6.25, which was close to the average RMSE obtained from the summary of the 10 training models (6.23). The adjusted  $R^2$  of the test predictions was calculated as  $0.9166 \approx 91.7\%$ . A summary of the predicted RPT is provided in Table 7.8.

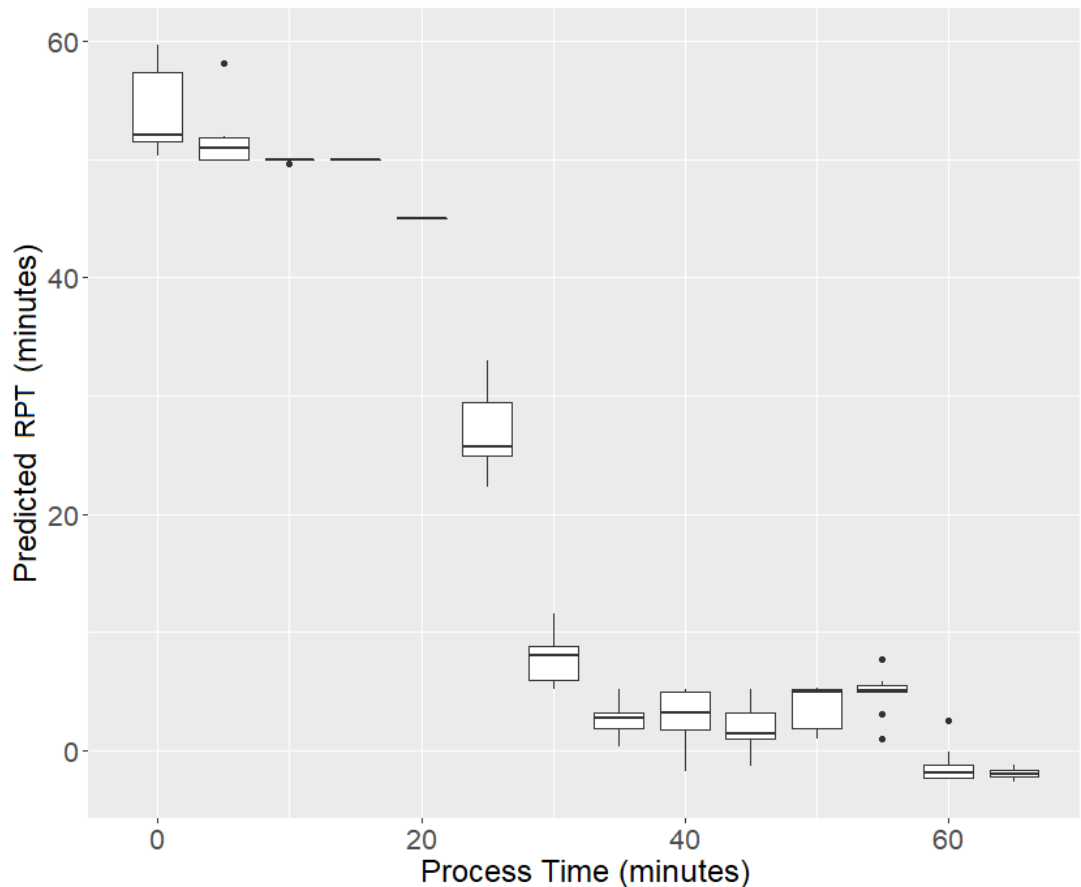
**Table 7-8. Summary statistics of the Predicted RPT compared with the Specified RPT of the test data from the 10 RF models. The green cells represent reasonably good predictions and the white cells represent poor predictions. The values in bold shows the best predictions.**

Process Time (minutes)	Specified RPT (minutes)	Median of Predicted RPT (minutes)	Delta (minutes)
0	65	61.85	3.15
5	60	58.9	1.1
10	55	58.8	-3.8
15	50	50	0
20	45	45	0
25	40	36.6	3.4
30	35	23.7	11.3
35	30	24	6
40	25	27	-2
45	20	24.1	-4.1
50	15	25.3	-10.3
55	10	20.1	-10.1
60	5	4.4	0.6
65	0	-1.9	1.9
70	-5	-1.55	-3.45

Overall, the RF model presented reasonably good predictions for the test micrographs. The median of the Predicted RPT, in Table 7.8, presented minimal deviation from their Specified RPT values for most of the micrographs except for those samples obtained between 30 and 55 minutes of emulsification. The median predictions obtained at the early stage of emulsification, samples obtained after 15 to 20 minutes, presented zero deviation from their Specified RPT. This result is very useful for the identification of the completion point of emulsification at an early stage. The median predictions obtained towards the end of the process, between 60 and 65 minutes, also presented minimal deviation (0.6 to 1.9 minutes) from their Specified RPT.

---

The RF regression model was also evaluated with the independent set of 140 micrographs. The RPT values predicted by the model are presented using box plots in Figure 7.6. Each box plot represents 10 micrographs.



**Figure 7.6. Predicted RPT plotted against current process time for the independent batch of 140 micrographs. Each box plot represents 10 micrographs.**

The RPT, predicted by the RF model, during the initial 15 minutes of emulsification is close to 50 minutes. These predictions closely aligned with the optimum processing time of the emulsion. These early stage predictions are quite useful to make assumptions on the total processing time required and reduce repeated sampling performed in the middle of the emulsification process. The RF model, similar to the LR predictions, predicted lower than expected RPT values for the samples obtained between 30 and 55 minutes of

processing. However, the Predicted RPT was closer to zero for the samples obtained after 60 minutes, which reaffirmed that the process has reached the target.

Additional comparison was performed between the Predicted RPT of the independent samples and their Specified RPT. This is presented in Table 7.9.

**Table 7-9. Comparison of the Predicted RPT values of the independent set of micrographs with their Specified RPT values. The reasonably good predictions are highlighted in green.**

Process Time (minutes)	Specified RPT (minutes)	Median of Predicted RPT (minutes)	Delta (minutes)
0	65	52.95	12.05
5	60	51.6	8.4
10	55	50	5
15	50	50	0
20	45	45	0
25	40	28	12
30	35	6.75	28.25
35	30	2.9	27.1
40	25	3.4	21.6
45	20	1.65	18.35
50	15	4.9	10.1
55	10	5	5
60	5	-2.1	7.1
65	0	-2.8	2.8

The higher deviations, as given by Delta, in the Predicted RPT during the middle of emulsification was consistently observed in both the LR and RF models. This is due to the absence of any linear variation in the droplet characteristics, which are the model predictors, during that period of emulsification. The predictions aligned closely (100%) with the Specified RPT for the 30 micrographs acquired during the initial 10 to 20 minutes of emulsification. The Predicted RPT of the 10 micrographs acquired after 65 minutes indicated that the emulsion was over-processed for 2.8 minutes, which was in close agreement with the optimum processing time identified for the independent batch (60 minutes).

---

### 7.2.3 Comparison between the LR and RF models

Both regression models presented reasonably good predictions for the early and latter stages of the emulsification processes. The *RMSE* and the adjusted  $R^2$  values obtained for the test predictions of the first batch are presented in Table 7.10.

**Table 7-10. RMSE and  $R^2$  values obtained for the LR and RF models for the test data predictions.**

Regression Model	RMSE	$R^2_{adj}$
LR	6.03	92%
RF	6.25	91.7%

Both LR and RF models presented approximately similar values for *RMSE* and adjusted  $R^2$ . With respect to individual predictions, the models presented similar results at the early stages of emulsification as well as towards the completion point. Both the models presented comparatively higher deviations for the samples processed between 25 and 55 minutes, although the RF model offered slightly better predictions between 40 and 45 minutes for the first batch. The end point predictions, after 60 minutes, of both the models were similar.

A comparison of the prediction results obtained for the independent batch, which attained the target characteristics five minutes earlier than the first batch, showed that both the models were reasonably good in predicting over-processing. In addition, the LR and the RF models presented reasonable potential, at the earlier stages of processing, in predicting the completion time of the emulsification process well in advance.

---

### 7.3 Conclusions

Two regression models, Linear Regression and Random Forest, were investigated to predict the Remaining Processing Time required by the emulsion to attain the target droplet characteristics. Five predictor variables, mean droplet area, mean Feret, mean minFeret, mean perimeter and droplet count, were used for the models. The 150 micrographs acquired from the first batch of Process-2 were initially used to train and test the models. Both the regression models explained the variance in the training and the test data reasonably well (approximately 92%). These trained models were also evaluated for their performance on unseen data using an independent set of 140 micrographs acquired from a different batch of the emulsion. This batch was observed to attain the target characteristics 5 minutes earlier than the first batch.

Both the regression models demonstrated promising potential for predicting the remaining emulsification time when presented with samples from the initial 20 minutes of processing. This information is extremely useful, for process analysts, to reduce the frequency of sampling performed during processing. This could save the extra resources required for sampling, reduce the wastage of samples and, minimise the risk of introducing impurities into the emulsion mixer. The models' performance was limited when it came to predict the RPT during the middle of the emulsification process. This can be attributed to the absence of any linear variation in the droplet characteristics during that period. However, the models demonstrated reasonably good accuracy for predicting over-processing.

The Remaining Processing Time, predicted by these regression models, is highly beneficial for the industry to identify the completion time of emulsification, when



---

presented with micrographs from any stage of the process. These prediction models could be, in the future, integrated in line with the industrial process for real-time prediction of the APT. This approach will have the potential to reduce the current demand for personnel resources, time and energy in emulsion manufacturing, thus, leading to more efficient and sustainable manufacturing practices.

---

## Chapter 8

### Conclusions and Future Work

This chapter discusses the conclusions of the current work and the recommendations for future research.

#### 8.1 Conclusions

The aim of this research was to develop an automated quality evaluation tool for industrial emulsion processing, which could potentially overcome the current challenges faced by the manual and other conventional techniques. Each section in the conclusions is related to a corresponding objective identified in Chapter 1.

##### 8.1.1 A novel Image segmentation for automated droplet characterisation

This study has investigated the competence of two image segmentation techniques, EST and HBT, to develop a novel method for droplet detection and characterisation from in-process emulsion micrographs. The EST is a discontinuity detection-based approach, while the HBT is a similarity detection approach. The techniques were applied to a trial process as a proof of concept and the best technique was used to evaluate an industrial emulsification process. Droplet size characteristics such as area, perimeter, maximum Feret diameter, minimum Feret diameter as well as droplet count were identified, through statistical analysis, as the emulsion quality indicators. It was a requirement of the EST algorithm to recalibrate the radius parameter when droplet sizes changed. The EST failed to detect varying sizes of droplets, especially smaller ones, due to the difficulty for the user in identifying accurate radii values. This challenge made it difficult to automate the technique. Alternatively, the histogram-based approach (HBT) is completely automated, as the technique does not demand any user input parameters and, was able to detect

---

droplets of various sizes in a single pass. The HBT has demonstrated significant potential in the detection of droplets and their characteristics, throughout the emulsification process, presenting a progressive evolution of droplet size and count. In addition, the technique was successful in detecting droplets with radii smaller than that reflected in the literature. The oil concentration detected in the micrographs, using the HBT approach, was also consistent and, in close agreement with the industrial formulation of the studied emulsion.

### 8.1.2 Machine Learning Classification

This study has also investigated the integration of microscopic image analysis with machine learning models to classify in-process emulsion samples into Target, Acceptable, Marginal and Unacceptable categories. The Histogram-Based droplet detection, integrated with the micrograph classification approach, indicated that the time of completion of an emulsification process can be automatically determined from an emulsion micrograph.

The Principal Component based Linear Discriminant Analysis and the Random Forest classification approaches presented 96% to 100% accuracy in the TAMU classification of the studied emulsion. This demonstrated the potential of applying HBT integrated with machine learning, to classify unknown micrographs at various stages of emulsification into the TAMU categories. The PC-LDA approach presented 100% accuracy in the classification of the unseen data set. This may be due to the approach used in the PC-LDA model. The characteristics of the individual droplets in a micrograph were used to train the PC-LDA model, while the MLR and the RF models used the summary characteristics of each micrograph such as the mean droplet area, mean Feret, mean minFeret, mean perimeter and the droplet count. The PC-LDA approach was successful in achieving a low bias in the model predictions and aided in identifying the underlying

---

pattern of the training data. In addition, the dimensionality reduction of the input feature space also helped to achieve low variance in the model predictions.

The vanilla neural network classification model, developed in this study, with a single hidden layer and five units presented only 85% accuracy in the classification of unseen data. The reduced accuracy of the VNN model compared to the PC-LDA, MLR and the RF models can be attributed to the limited set of training micrographs available in this study. The CNN classification model, which used a pixel-based approach, presented much poorer accuracy of 44%. The poor performance of the CNN model can also be attributed to the limited data size compared to the scale of data required to train a multilayer perceptron. In addition to this, the inter-class variation between the TAMU categories presented by the pixel intensity data was minimal.

### 8.1.3 Processing Time Prediction

Two regression models, Linear Regression and Random Forest, were investigated to predict the Remaining Processing Time (RPT), i.e., the time required by an unknown emulsion samples (micrographs) to attain the target characteristics. The optimum processing time identified for the studied emulsion was 65 minutes. Both the regression models explained 92% ( $R^2_{\text{adj}} \approx 0.92$ ) of the total variance in the training and the test data from the first batch of Process-2. These models were also evaluated using the independent batch of the product. The models presented high accuracy in the prediction of the RPT during the initial 20 minutes of the emulsification process for both the batches. This information is highly beneficial for the process analysts to identify, at an early stage, the total time and the amount of resources required to complete the process. Moreover, it helps to eliminate the huge amount of extra sampling currently done during the middle of the process to assess the emulsion quality, which adds a potential risk of adding impurities

---

into the emulsion mixer. The models demonstrated difficulty in predicting the RPT during the middle of emulsification, which is due to the absence of any linear variation in the model predictors during that period. However, both the prediction models confirmed reasonable potential in identifying over-processing when presented with unknown micrographs of the studied emulsion.

#### 8.1.4 Knowledge Contribution and Impact

A precise understanding of droplet size distribution has been identified, in the literature, as the key factor to control and optimise industrial emulsification processes (Panckow *et al.* 2017). Manual evaluation of in-process samples is one of the techniques currently employed in the industrial quality evaluation of emulsions. Such evaluation techniques have been found highly subjective, erroneous and time consuming in previous studies (Gwyn *et al.* 1965, Boxall *et al.* 2010, Maaß *et al.* 2012). Other conventional techniques such as laser diffraction and spectroscopy have also been applied to study the droplet size distribution and determine the optimum processing time in food emulsions. Such techniques require time-consuming sample preparation, sample dilution and were found incapable of delivering reliable droplet size measurements in the past (Abidin *et al.* 2013, Vankeirsbilck *et al.* 2002)).

Emulsion droplet detection studies have been performed previously using image processing techniques (Hosseini *et al.* 2015, Silva *et al.* 2010, Freire *et al.* 2005). The existing image processing techniques have focused on identifying the droplet border, using edge detection, at the phase change between oil and water (Hosseini *et al.* 2015). Such discontinuity-based segmentation techniques have identified certain challenges, which include droplet detection from highly concentrated emulsions and, the detection of smaller droplets from production systems (Khalil *et al.* 2010, Mickler *et al.* 2011, Maaß

---

*et al.* 2012, Panckow *et al.* 2017). The HBT, presented in this study, was successful in detecting droplets with diameter as low as *ca.* 1  $\mu\text{m}$  from emulsion samples having dispersed phase fraction  $\approx 50\%$  (Unnikrishnan *et al.* 2018).

In addition, the application of machine vision in the in-process quality evaluation of emulsions is found to be an area that requires further investigation, particularly in respect to the use of automation within pharmaceutical industries. The current study has investigated the application of microscopic image analysis combined with machine learning algorithms in the classification of in-process emulsion samples. The classification is based on the variation in the droplet characteristics observed during the emulsification process. Computer vision analysis, integrated with machine learning, has gained wide attention in a range of industries to achieve fast, accurate and objective evaluation of product quality (Pfeil *et al.* 2018, Caballero *et al.* 2018, Manak *et al.* 2018, Bertani *et al.* 2017). However, there have been very few studies reported previously for the quality evaluation of emulsions using various imaging techniques (Amokrane *et al.* 2016, Maaß *et al.* 2012, Khalil *et al.* 2010, Panckow *et al.* 2017). The existing literature has focused only on the inline extraction of droplet characteristics from emulsions. The extension of such techniques using machine learning algorithms to develop an automated evaluation tool is currently unexplored in emulsion manufacturing. The TAMU classification models, using RF and PC-LDA, developed in the current study presented promising potential in the in-process classification of emulsion samples.

Moreover, the current study has demonstrated the effectiveness of machine learning models for the automated quality evaluation of emulsion systems compared to manual evaluation. The TAMU classification, using machine learning, was 150 times faster and

---

10% to 40% more accurate than the manual approach currently practiced in industry. The results prove that there is significant difference, in terms of speed, accuracy, repeatability and reproducibility, between micrograph classifications performed manually and using automated machine learning.

The HBT is currently implemented as an offline soft sensor, which integrates the droplet detection macro in Fiji, the statistical analysis and the machine learning models in R to present automated classification and prediction. Moreover, the successful detection of smaller droplets, with  $d_{32}$  as low as  $3.5\ \mu\text{m}$  and equivalent circle diameter  $\approx 1\ \mu\text{m}$ , enables the integration of the technique into real-time production systems.

## **8.2 Future Work**

The future potential of this research is to extend the application of the image segmentation integrated machine learning approach to food and chemical industries, where emulsification is an integral part of manufacturing. Extension of the technique to detect both particles and droplets simultaneously, as separate clusters, from the micrographs of medicinal creams and ointments, which uses Active Pharmaceutical Ingredients (API) is also under consideration. Fractal analysis, which is a technique used to extract and assess the fractal dimensions of patterns in an image, will be considered for the future analysis of particles (Shahidi-Noghabi *et al.* 2019).

Another future objective is to increase the amount of training data used for the machine learning models to eliminate any chances of overfitting and to improve the accuracy of the models. The reason for the limited set of micrographs in the current study was due to the physical sampling performed by the micrograph analysts in the industry. There are several

---

methods, which could be adapted, to potentially overcome this data limitation. The usage of generative models, droplet data simulation and real-time imaging are among them. The future work is also focused on extending the current offline technique to an inline soft sensor.

### 8.2.1 Generative Models

Generative models could be a solution to expand the limited dataset in the current research. These are classification models which learn the probability of  $X$  given  $Y$ , i.e.,  $P(X|Y)$  and it also learns the probability distribution of  $Y$ , the response classes, from the training data. The model then generates  $X$ , the features, for an unknown set of  $Y$ 's by predicting  $P(X|Y)$ . Generative models can be used for fast simulation of input features. Bayesian networks are the most commonly used generative models in classification problems with a limited set of training data (Sarkar *et al.* 2017). These are probabilistic graphical models also known as directed acyclic graphs, which generates a joint distribution of statistically dependent random variables (Friedman *et al.* 2001, Sarkar *et al.* 2017). Bayesian networks are suitable for both supervised and unsupervised classification.

Generative Adversarial Networks (GANs) are multilayer neural networks applied for deep machine learning (Ian *et al.* 2014). These are widely used in medical image classification problems, where there are insufficient labelled data sets. GANs are a combination of a generative model and a discriminative model. The former generates synthetic images from the limited training set given to the model and the latter classifies them from the original (Bowles *et al.* 2018, Goodfellow 2016). These GAN derived images can be introduced into the training set of the discriminative model to improve the classification accuracy. Moreover, the chances of overfitting can be avoided by improving the generalisation ability of deep learning models using more training data. GANs can learn dense



---

distributions over images, audio or any form of data, which are potentially difficult to model with an explicit likelihood (Saatci and Wilson 2017). Bayesian Generative Adversarial Networks (BGANs) are recommended as a powerful addition to the family of GANs to introduce more robust classification with minimal human intervention (Saatci and Wilson 2017).

### 8.2.2 Ready-to-use object detection networks

The minimal inter-class variation presented by the pixel data in the micrographs was identified as one of the reasons for the poor performance of the CNN model reported in the current study. This could be attributed to the lack of image clarity caused by varying illumination and poor resolution. There are ready-to use object detection CNNs, such as MobileNet and Inception-v2, available for the detection of objects from greyscale images even with low clarity. A recent study by Soldati *et al.* (2018) has used these object detection CNNs for the classification of microfluidic droplets from the images of tumour cells, which resulted in improved accuracy compared to conventional CNNs.

### 8.2.3 Real-Time Integration

A future extension of this work is planned for the inline quality evaluation and optimisation of industrial emulsification processes. The offline soft sensor, developed in this research, can be integrated with real-time imaging of emulsification processes (Khalil *et al.* 2010, Maaß *et al.* 2012). This will generate a large number of images for training the machine learning models used in image classification as well as process time prediction. The models developed in this study could be retrained using these real-time images to avoid any possibility of overfitting. Moreover, the accuracy of the neural network models could potentially be improved using these larger datasets generated in real-time. The inline implementation of the technique would contribute towards evaluating emulsion quality,

---

facilitating instantaneous process feedback and identifying the optimum time to achieve the target droplet characteristics. The proposed approach developed by this research work has demonstrated promising potential in avoiding the over-processing of emulsions, leading to smart and sustainable manufacturing practices.

#### 8.2.4 Extension of the Manual Versus Machine Learning Classification

The work published from this research on the comparison of manual microscopic evaluation with machine learning classification is based on the droplet characteristics obtained using the EST (Unnikrishnan *et al.* 2019). It is planned to carry out additional work using the machine learning models derived from the HBT droplet data. This will be achieved by repeating the micrograph classification study using the well-performed machine learning models such as the PC-LDA and the Random Forest and comparing the results with manual classification. The micrograph analysts from the industrial partner will be involved in the manual classification.

---

## Bibliography

- Abidin, M. I. I. Z., Raman, A. A. A. and Nor, M. I. M. (2013) 'Review on measurement techniques for drop size distribution in a stirred vessel', *Industrial & Engineering Chemistry Research*, 52 (46), pp. 16085-16094.
- Al-Refaie, A. and Bata, N. (2010) 'Evaluating measurement and process capabilities by GR&R with four quality measures', *Measurement*, 43 (6), pp. 842-851.
- Aly, A. A., Deris, S. B. and Zaki, N. (2011) 'Research review for digital image segmentation techniques', *International Journal of Computer Science & Information Technology*, 3 (5), p. 99.
- Amokrane, A., Maass, S., Lamadie, F., Puel, F. and Charton, S. (2016) 'On droplets size distribution in a pulsed column. Part I: In-situ measurements and corresponding CFD-PBE simulations', *Chemical Engineering Journal*, 296, pp. 366-376.
- Andy, L., & Matthew, W. (2002) 'Classification and Regression by randomForest', *R News*, pp. 18-22.
- Anjna, E. A. and Er, R. K. (2017) 'Review of image segmentation technique', *International Journal of Advanced Research in Computer Science*, 8 (4).
- Bakry, A. M., Abbas, S., Ali, B., Majeed, H., Abouelwafa, M. Y., Mousa, A. and Liang, L. (2016) 'Microencapsulation of oils: a comprehensive review of benefits, techniques, and applications', *Comprehensive Reviews in Food Science and Food Safety*, 15 (1), pp. 143-182.
- Berrueta, L., Alonso-Salces, R. and Heberger, K. (2007) 'Supervised pattern recognition in food analysis', *Journal of Chromatography a*, 1158 (1-2), pp. 196-214.
- Bertani, F. R., Mozetic, P., Fioramonti, M., Iuliani, M., Ribelli, G., Pantano, F., Santini, D., Tonini, G., Trombetta, M., Businaro, L., Selci, S. and Rainer, A. (2017) 'Classification of M1/M2-polarized human macrophages by label-free hyperspectral reflectance confocal microscopy and multivariate analysis', *Scientific Reports*, 7 (1), article number. 8965.
- Biau, G. and Scornet, E. (2016) 'A random forest guided tour', *Test*, 25 (2), pp. 197-227.
- Blonk, J. and Van Aalst, H. (1993) 'Confocal scanning light microscopy in food research', *Food Research International*, 26 (4), pp. 297-311.
- Bora, D. J. and Gupta, A. K. (2014) 'A new approach towards clustering based color image segmentation', *International Journal of Computer Applications*, 107 (12).
- Borooh, V. K. (2002) *Logit and probit: Ordered and multinomial models*. Sage.
- Bosque-Sendra, J. M., Cuadros-Rodríguez, L., Ruiz-Samblás, C. and De La Mata, A. P. (2012) 'Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data—A review', *Analytica Chimica Acta*, 724 pp. 1-11.
- Boudier, T.  
[http://imagejdocu.tudor.lu/doku.php?id=plugin:filter:3d\\_edge\\_and\\_symmetry\\_filter:st\\_art](http://imagejdocu.tudor.lu/doku.php?id=plugin:filter:3d_edge_and_symmetry_filter:st_art) [Online]. [Accessed 10/03/2017].

- 
- Boutros, P. and Okey, A. (2005) 'Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data', *Briefings in Bioinformatics*, 6 (4), pp. 331-343.
- Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D. A., Hernández, M. V., Wardlaw, J. and Rueckert, D. (2018) 'GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks', *arXiv preprint arXiv:1810.10863*.
- Boxall, J. A., Koh, C. A., Sloan, E. D., Sum, A. K. and Wu, D. T. (2010) 'Measurement and Calibration of Droplet Size Distributions in Water-in-Oil Emulsions by Particle Video Microscope and a Focused Beam Reflectance Method', *Industrial & Engineering Chemistry Research*, 49 (3), pp. 1412-1418.
- Breiman, L. (1996a) 'Bagging predictors', *Machine Learning*, 24 (2), pp. 123-140.
- Breiman, L. (1996b) Out-of-bag estimation. Citeseer.
- Breiman, L. (2001) 'Random forests', *Machine learning*, 45 (1), pp. 5-32.
- Breiman, L. and Cutler, A. (2003) 'Setting up, using, and understanding random forests V4. 0', *University of California, Department of Statistics*.
- Brás, L. M. R., Gomes, E. F., Ribeiro, M. M. M. and Guimarães, M. M. L. (2009) 'Drop distribution determination in a liquid-liquid dispersion by image processing', *International Journal of Chemical Engineering*, 2009, article number. 746439.
- Browne, R., Mackay, J. and Steiner, S. (2010) 'Leveraged Gauge R&R Studies', *Technometrics*, 52 (3), pp. 294-302.
- Caballero, D., Caro, A., Dahl, A. B., Ersbøll, B. K., Amigo, J. M., Pérez-Palacios, T. and Antequera, T. (2018) 'Comparison of different image analysis algorithms on MRI to predict physico-chemical and sensory attributes of loin', *Chemometrics and Intelligent Laboratory Systems*, 180 pp. 54-63.
- Calderon, C. P., Daniels, A. L. and Randolph, T. W. (2017) 'Using Deep Convolutional Neural Networks to Circumvent Morphological Feature Specification when Classifying Subvisible Protein Aggregates from Micro-Flow Images', *arXiv preprint arXiv:1709.00152*.
- Canny, J. (1987) 'A computational approach to edge detection'. *Readings in computer vision*. Elsevier.
- Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V. and Kraemer, D. J. M. (2019) 'Decoding individual differences in STEM learning from functional MRI data', *Nature Communications*, 10 (1), article number. 2027.
- Chen, C., Liaw, A. and Breiman, L. (2004) 'Using random forest to learn imbalanced data', *University of California, Berkeley*, 110, pp. 1-12.
- Chen, W., Xie, X. S., Wang, J. L., Pradhan, B., Hong, H. Y., Bui, D. T., Duan, Z. and Ma, J. Q. (2017) 'A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility', *Catena*, 151 pp. 147-160.
- Chen, M. and Lyu, J. (2011) 'Enhancement of measurement capability for precision manufacturing processes using an attribute gauge system', *Proceedings of the*

---

*Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture*, 225 (B10), pp. 1912-1924.

- Cho, H.-Y., Choi, J.-H., Oh, I.-J. and Lee, Y.-B. (2015) 'Self-emulsifying drug delivery system for enhancing bioavailability and lymphatic delivery of tacrolimus', *Journal of nanoscience and nanotechnology*, 15 (2), pp. 1831-1841.
- Crawley, G. and Malcolmson, A. (2004) 'Online particle sizing as a route to process optimization', *Chemical Engineering*, 111 (9), pp. 37-41.
- Cumbaa, C. and Jurisica, I. (2005) 'Automatic Classification and Pattern Discovery in High-throughput Protein Crystallization Trials', *Journal of Structural and Functional Genomics*, 6 (2), pp. 195-202.
- Cumbaa, C. A. and Jurisica, I. (2010) 'Protein crystallization analysis on the World Community Grid', *Journal of Structural and Functional Genomics*, 11 (1), pp. 61-69.
- Dagher, I. and Nachar, R. (2006) 'Face recognition using IPCA-ICA algorithm', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (6), pp. 996-1000.
- Dao, H., Lakhani, P., Police, A., Kallakunta, V., Ajjarapu, S. S., Wu, K.-W., Ponkshe, P., Repka, M. A. and Murthy, S. N. (2018) 'Microbial stability of pharmaceutical and cosmetic products', *AAPS PharmSciTech*, 19 (1), pp. 60-78.
- Das, R., Piciuccio, E., Maiorana, E. and Campisi, P. (2019) 'Convolutional neural network for finger-vein-based biometric identification', *IEEE Transactions on Information Forensics and Security*, 14 (2), pp. 360-373.
- Dehghani, M. and Kamps, J. (2018) 'Learning to Rank from Samples of Variable Quality', *arXiv preprint arXiv:1806.08694*.
- De Mast, J. and Van Wieringen, W. N. (2010) 'Modeling and Evaluating Repeatability and Reproducibility of Ordinal Classifications', *Technometrics*, 52 (1), pp. 94-106.
- Dokania, S. and Joshi, A. K. (2015) 'Self-microemulsifying drug delivery system (SMEDDS)–challenges and road ahead', *Drug delivery*, 22 (6), pp. 675-690.
- Doyle Iii, F. J., Harrison, C. A. and Crowley, T. J. (2003) 'Hybrid model-based approach to batch-to-batch control of particle size distribution in emulsion polymerization', *Computers & Chemical Engineering*, 27 (8-9), pp. 1153-1163.
- Dreiseitl, S. and Ohno-Machado, L. (2002) 'Logistic regression and artificial neural network classification models: a methodology review', *Journal of Biomedical Informatics*, 35 (5-6), pp. 352-359.
- Du, C.-J. and Sun, D.-W. (2006) 'Learning techniques used in computer vision for food quality evaluation: a review', *Journal of Food Engineering*, 72 (1), pp. 39-55.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2012) 'Pattern classification', *John Wiley & Sons*, pp. 216-217.
- Díaz-Uriarte, R. and De Andres, S. A. (2006) 'Gene selection and classification of microarray data using random forest', *BMC bioinformatics*, 7 (1), p. 3.
- Emmerich, J., Tang, Q., Wang, Y., Neubauer, P., Junne, S. and Maaß, S. (2018) 'Optical inline analysis and monitoring of Particle Size and shape distributions for multiple applications: Scientific and industrial relevance', *Chinese Journal of Chemical Engineering*.

- 
- Efron, B. and Tibshirani, R. J. (1994) 'An introduction to the bootstrap', *CRC press*.
- Freire, M. G., Dias, A. M. A., Coelho, M. A. Z., Coutinho, J. A. P. and Marrucho, I. M. (2005) 'Aging mechanisms of perfluorocarbon emulsions using image analysis', *Journal of Colloid and Interface Science*, 286 (1), pp. 224-232.
- Friedman, J., Hastie, T. and Tibshirani, R. (2001) *The elements of statistical learning*. Springer series in statistics New York, NY, USA:.
- Garcia, H. and Filzmoser, P. (2015) *Multivariate Statistical Analysis using the R package chemometrics* [Online]. Available: <https://cran.r-project.org/web/packages/chemometrics/vignettes/chemometrics-vignette.pdf> [Accessed March 29 2016].
- Geladi, P. (2003) 'Chemometrics in spectroscopy. Part 1. Classical chemometrics', *Spectrochimica Acta Part B: Atomic Spectroscopy*, 58 (5), pp. 767-782.
- Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2010) 'Variable selection using random forests', *Pattern Recognition Letters*, 31 (14), pp. 2225-2236.
- George, A., Shah, P. A. and Shrivastav, P. S. (2018) 'Guar gum: Versatile natural polymer for drug delivery applications', *European Polymer Journal*.
- Geuther, B. Q., Deats, S. P., Fox, K. J., Murray, S. A., Braun, R. E., White, J. K., Chesler, E. J., Lutz, C. M. and Kumar, V. (2019) 'Robust mouse tracking in complex environments using neural networks', *Communications Biology*, 2 (1), p. 124.
- Goodfellow, I. (2016) 'NIPS 2016 tutorial: Generative adversarial networks', *arXiv preprint arXiv:1701.00160*.
- Gosselin, R., Vachon Lachance, E., Cournoyer, A. and Clarke, F. (2016) 'Classifying Pharmaceutical Capsules Through X-Ray Image Analysis Based on the Agglomeration of Their Contents', *Journal of Pharmaceutical Innovation*, 11 (1), pp. 92-101.
- Grané, A. and Jach, A. (2014) 'Applications of principal component analysis (PCA) in food science and technology', *Mathematical and Statistical Methods in Food Science and Technology*, pp. 55-86.
- Greaves, D., Boxall, J., Mulligan, J., Montesi, A., Creek, J., Sloan, E. D. and Koh, C. A. (2008) 'Measuring the particle size of a known distribution using the focused beam reflectance measurement technique', *Chemical Engineering Science*, 63 (22), pp. 5410-5419.
- Gwyn, J., Crosby, E. and Marshall Jr, W. (1965) 'Bias in Particle-Size Analyses by the Count Method', *Industrial & Engineering Chemistry Fundamentals*, 4 (2), pp. 204-208..
- Günther, O. P., Gardy, J. L., Stafford, P., Fluge, Ø., Mella, O., Tang, P., Miller, R. R., Parker, S. M., Johnston, S. A. and Patrick, D. M. (2018) 'Immunosignature Analysis of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)', *Molecular neurobiology*, pp. 1-9.
- Haralick, R. M. and Shapiro, L. G. (1985) 'Image segmentation techniques', *Computer Vision, Graphics, and Image Processing*, 29 (1), pp. 100-132.
- Heffels, C., Polke, R., Rädle, M., Sachweh, B., Schäfer, M. and Scholz, N. (1998) 'Control of Particulate Processes by Optical Measurement Techniques', *Particle & Particle Systems Characterization*, 15 (5), pp. 211-218.

- 
- Heo, Y. J., Lee, D., Kang, J., Lee, K. and Chung, W. K. (2017) 'Real-time Image Processing for Microscopy-based Label-free Imaging Flow Cytometry in a Microfluidic Chip', *Scientific Reports*, 7 (1), p. 11651.
- Heung, B., Hodul, M. and Schmidt, M. G. (2017) 'Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes', *Geoderma*, 290 pp. 51-68.
- Hong, H., Xiaoling, G. and Hua, Y. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. (2016) *7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 26-28 Aug. p. 219-224.
- Honkanen, M., Eloranta, H. and Saarenrinne, P. (2010) 'Digital imaging measurement of dense multiphase flows in industrial processes', *Flow Measurement and Instrumentation*, 21 (1), pp. 25-32.
- Hosseini, A., Jafari, S., Mirzaei, H., Asghari, A. and Akhavan, S. (2015) 'Application of image processing to assess emulsion stability and emulsification properties of Arabic gum', *Carbohydrate Polymers*, 126, pp. 1-8.
- <https://imagej.net/Welcome> [Online]. [Accessed 15 March 2017].
- <https://www.rstudio.com/> [Online]. [Accessed 22 February 2017].
- Ian, J. G., Jean, P.-A., Mirza, M., Xu, B. and David, W.-F. (2014) 'Generative adversarial networks', *Advances in neural information processing systems*, 3 pp. 2672-2680.
- Isaksson, A. J., Harjunkoski, I. and Sand, G. (2018) 'The impact of digitalization on the future of control and operations', *Computers & Chemical Engineering*, 114, pp. 122-129.
- Isaza, C., Anaya, K., De Paz, J. Z., Vasco-Leal, J. F., Hernandez-Rios, I. and Mosquera-Artamonov, J. D. (2018) 'Image analysis and data mining techniques for classification of morphological and color features for seeds of the wild castor oil plant (*Ricinus communis* L.)', *Multimedia Tools and Applications*, 77 (2), pp. 2593-2610.
- Islam, M. J., Basalamah, S. M., Ahmadi, M. and Sid-Ahmed, M. A. (2012) 'Computer vision-based quality inspection system of transparent gelatin capsules in pharmaceutical applications', *American Journal of Intelligent Systems*, 2 (1), pp. 14-22.
- Jackson, J. E. (1980) 'Components and Factor Analysis: Part I - Principal Components', *Journal of Quality Technology*, 12 (4), pp. 201-213.
- Jackson, J. E. (2003) *A User's Guide to Principal Components*. Hoboken, New Jersey & Canada: John Wiley & Sons.
- Jafari, S. M., He, Y. and Bhandari, B. (2007) 'Role of Powder Particle Size on the Encapsulation Efficiency of Oils during Spray Drying', *Drying Technology*, 25 (6), pp. 1081-1089.
- Jain, D. K., Shamsolmoali, P. and Sehdev, P. (2019) 'Extended Deep Neural Network for Facial Emotion Recognition', *Pattern Recognition Letters*.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An introduction to statistical learning*. Springer.
- Jian, Y., Zhang, D., Frangi, A. F. and Jing-Yu, Y. (2004) 'Two-dimensional PCA: a new approach to appearance-based face representation and recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (1), pp. 131-137.

- 
- Joly, A., Geurts, P. and Wehenkel, L. Random forests with random projections of the output space for high dimensional multi-label classification. *Joint European conference on machine learning and knowledge discovery in databases*. 2014. Springer, pp. 607-622.
- Jousse, F. (2008) 'Modeling to improve the efficiency of product and process development', *Comprehensive Reviews in Food Science and Food Safety*, 7 (1), pp. 175-181.
- JJ, A., & François, C. (2019). keras: R Interface to 'Keras', <https://CRAN.R-project.org/package=keras>.
- JJ, A., & Yuan, T. (2019). tensorflow: R Interface to 'TensorFlow', <https://CRAN.R-project.org/package=tensorflow>
- Khalil, A., Puel, F., Chevalier, Y., Galvan, J. M., Rivoire, A. and Klein, J. P. (2010) 'Study of droplet size distribution during an emulsification process using in situ video probe coupled with an automatic image analysis', *Chemical Engineering Journal*, 165 (3), pp. 946-957.
- Kirby, M. and Sirovich, L. (1990) 'Application of the Karhunen-Loeve procedure for the characterization of human faces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (1), pp. 103-108.
- Kljusuric, J., Benkovic, M. and Bauman, I. (2015) 'Classification and Processing Optimization of Barley Milk Production Using NIR Spectroscopy, Particle Size, and Total Dissolved Solids Analysis', *Journal of Chemistry*.
- Kourti, T., Lee, J. and Macgregor, J. F. (1996) 'Experiences with industrial applications of projection methods for multivariate statistical process control', *Computers & Chemical Engineering*, 20, pp. S745-S750.
- Kourti, T. and Macgregor, J. F. (1995) 'Process Analysis, Monitoring and Diagnosis, using Multivariate Projection Methods', *Chemometrics and Intelligent Laboratory Systems*, 28 (1), pp. 3-21.
- Kourti, T. and Macgregor, J. F. (1996) 'Multivariate SPC Methods for Process and Product Monitoring', *Journal of Quality Technology*, 28 (4), pp. 409-428.
- Krizhevsky, A. and Hinton, G. (2009) *Learning multiple layers of features from tiny images*. Citeseer.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- Kuhn, M. and Johnson, K. (2013) *Applied predictive modeling*. Springer.
- Kumar, R., Arthanari, M. and Sivakumar, M. (2012) 'Image Segmentation using Discontinuity-Based Approach', *International Journal Multimedia and Image Processing (IJMIP)*, 2 (1/2), pp. 72-78.
- Lenth, R. V. (2009) 'Response-Surface Methods in R, Using rsm', *Journal of Statistical Software*, 32 (7), pp. 1-17.
- Lyu, J. and Chen, M.-N. (2008) 'Gauge capability studies for attribute data', *Quality and Reliability Engineering International*, 24 (1), pp. 71-82.
- Maaß, S., Rojahn, J., Haensch, R. and Kraume, M. (2012) 'Automated drop detection using image analysis for online particle size monitoring in multiphase systems', *Computers & Chemical Engineering*, 45, pp. 27-37.



- 
- Manak, M. S., Varsanik, J. S., Hogan, B. J., Whitfield, M. J., Su, W. R., Joshi, N., Steinke, N., Min, A., Berger, D., Saphirstein, R. J., Dixit, G., Meyyappan, T., Chu, H.-M., Knopf, K. B., Albala, D. M., Sant, G. R. and Chander, A. C. (2018) 'Live-cell phenotypic-biomarker microfluidic assay for the risk stratification of cancer patients via machine learning', *Nature Biomedical Engineering*, 2 (10), pp. 761-772.
- Marques, C., Lopes, N., Santos, G., Delgado, I. and Delgado, P. (2018) 'Improving operator evaluation skills for defect classification using training strategy supported by attribute agreement analysis', *Measurement*, 119 pp. 129-141.
- Martinez-Bazan, C., Montanes, J. L. and Lasheras, J. C. (1999) 'On the breakup of an air bubble injected into a fully developed turbulent flow. Part 1. Breakup frequency', *Journal of Fluid Mechanics*, 401 pp. 157-182.
- Max, K. (2020). caret: Classification and Regression Training, R package version 6.0-86, <https://CRAN.R-project.org/package=caret>
- Menard, S. (2002) *Applied logistic regression analysis*. Sage.
- Mickler, M., Didas, S., Jaradat, M., Attarakih, M. and Bart, H.-J. (2011) 'Tropfenschwarmanalytik mittels Bildverarbeitung zur Simulation von Extraktionskolonnen mit Populationsbilanzen', *Chemie Ingenieur Technik*, 83 (3), pp. 227-236.
- Minitab 18 Support [Online]. Available: <https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/measurement-system-analysis/how-to/attribute-agreement-analysis/attribute-agreement-analysis/methods-and-formulas/assessment-agreement/#confidence-intervals-for-percent-agreement> [Accessed 01/03/2019].
- Minitab, Attribute Agreement Analysis, n.d. [Online]. Available: [http://support.minitab.com/en-us/minitab/17/Assistant\\_Attribute\\_Agreement\\_Analysis.pdf](http://support.minitab.com/en-us/minitab/17/Assistant_Attribute_Agreement_Analysis.pdf) [Accessed 02/05/2018].
- Murphy, S. A., Moeller, S. E., Page, J. R., Cerqua, J. and Boarman, M. (2009) 'Leveraging Measurement System Analysis (MSA) to Improve Library Assessment: The Attribute Gage R&R', *College & Research Libraries*, 70 (6), pp. 568-577.
- Nair, V. and Hinton, G. E., 'Rectified linear units improve restricted boltzmann machines', *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010. 807-814.
- Nichols, T. R., Wisner, P. M., Cripe, G., & Gulabchand, L. (2010), Putting the Kappa Statistic to Use, *The Quality Assurance Journal*, 13, 57-61.
- Nielsen, M. A. (2015) *Neural Networks and Deep Learning*, Determination Press.
- Niknafs, N., Spyropoulos, F. and Norton, I. T. (2011) 'Development of a new reflectance technique to investigate the mechanism of emulsification', *Journal of Food Engineering*, 104 (4), pp. 603-611.
- Ollion, J., Cochenec, J., Loll, F., Escude, C. and Boudier, T. (2013) 'TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization', *Bioinformatics*, 29 (14), pp. 1840-1841.
- Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. and Johnson, G. R. (2018) 'Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy', *Nature Methods*, 15 (11), pp. 917-920.

- 
- Oviedo, F., Ren, Z., Sun, S., Settens, C., Liu, Z., Hartono, N. T. P., Ramasamy, S., Decost, B. L., Tian, S. I. P., Romano, G., Gilad Kusne, A. and Buonassisi, T. (2019) 'Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks', *npj Computational Materials*, 5 (1), p. 60.
- Pacek, A. W., Moore, I. P. T., Nienow, A. W. and Calabrese, R. V. (1994a) 'VIDEO TECHNIQUE FOR MEASURING DYNAMICS OF LIQUID-LIQUID DISPERSION DURING PHASE INVERSION', *Aiche Journal*, 40 (12), pp. 1940-1949.
- Pacek, A. W., Nienow, A. W. and Moore, I. P. T. (1994b) 'ON THE STRUCTURE OF TURBULENT LIQUID-LIQUID DISPERSED FLOWS IN AN AGITATED VESSEL', *Chemical Engineering Science*, 49 (20), pp. 3485-3498.
- Panckow, R., Reinecke, L., Cuellar, M. and Maaß, S. (2017) 'Photo-Optical In-Situ Measurement of Drop Size Distributions: Applications in Research and Industry', *Oil & Gas Science and Technology*, 72(3), article number. 14.
- Patel, K. K., Kar, A., Jha, S. N. and Khan, M. A. (2012) 'Machine vision system: a tool for quality inspection of food and agricultural products', *Journal of Food Science and Technology*, 49 (2), pp. 123-141.
- Pearson, K. (1901) 'Principal components analysis', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6 (2), p. 559.
- Pfeil, J., Frohme, M. and Schulze, K. (2018) 'Mobile Microscopy and Automated Image Analysis: The ease of cell counting and classification', *Optik & Photonik*, 13 (1), pp. 36-39.
- Puntanen, S. (2013) 'Regression Analysis by Example, by Samprit Chatterjee, Ali S. Hadi', *International Statistical Review*, 81 (2), p. 347.
- Qu, H., Wang, J., Wu, Y., Zheng, J., Krishnaiah, Y. S., Absar, M., Choi, S., Ashraf, M., Cruz, C. N. and Xu, X. (2018) 'Asymmetric flow field flow fractionation for the characterization of globule size distribution in complex formulations: A cyclosporine ophthalmic emulsion case', *International journal of pharmaceutics*, 538 (1-2), pp. 215-222.
- R Core Team. (2019) R: A Language and Environment for Statistical Computing.
- Refaeilzadeh, P., Tang, L. and Liu, H. (2009) 'Cross-Validation'. In: L. Liu & M. T. Özsu (eds.) *Encyclopedia of Database Systems*. Boston, MA: Springer US.
- Ripley, W. N. V. a. B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.
- Rodriguez, J., Martín, M. J., Ruiz, M. A. and Clares, B. (2016) 'Current encapsulation strategies for bioactive oils: From alimentary to pharmaceutical perspectives', *Food Research International*, 83 pp. 41-59.
- Rossner, M. and O'donnell, R. (2004) 'The JCB will let your data shine in RGB', *Journal of Cell Biology*, 164 (1), pp. 11-13.
- Saatci, Y. and Wilson, A. G. Bayesian gan. *Advances in neural information processing systems*. 2017. 3622-3631.
- Sachweh, B., Heffels, C., Polke, R. and Rädle, M. Light scattering sensor for in-line measurements of mean particle sizes in suspensions. *Proceedings of 7th Europ. Symp. Part. Char., Nuremberg*. 1998, p. 644.

- 
- Sarkar, S., Jiang, Z., Akintayo, A., Krishnamurthy, S. and Tewari, A. (2017) 'Probabilistic graphical modeling of distributed cyber-physical systems'. *Cyber-Physical Systems*. Elsevier.
- Scherze, I., Knofel, R. and Muschiolik, G. (2005) 'Automated image analysis as a control tool for multiple emulsions', *Food Hydrocolloids*, 19 (3), pp. 617-624.
- Schindelin, J. Fiji Is Just ImageJ (batteries included). *ImageJ User and Developer Conference 2008*.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012) 'Fiji: an open-source platform for biological-image analysis', *Nat Meth*, 9 (7), pp. 676-682.
- Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9 p. 671.
- Schulze, K., Tillich, U. M., Dandekar, T. and Frohme, M. (2013) 'PlanktoVision - an automated analysis system for the identification of phytoplankton', *BMC Bioinformatics*, 14 (1), p. 115.
- Schuster, S., Bernewitz, R., Guthausen, G., Zapp, J., Greiner, A. M., Köhler, K. and Schuchmann, H. P. (2012) 'Analysis of W1/O/W2 double emulsions with CLSM: Statistical image processing for droplet size distribution', *Chemical Engineering Science*, 81 (Supplement C), pp. 84-90.
- Sendra, G. H., Hoerth, C. H., Wunder, C. and Lorenz, H. (2015) '2D map projections for visualization and quantitative analysis of 3D fluorescence micrographs', *Scientific Reports*, 5 p. 12457.
- Shahidi-Noghabi, M., Naji-Tabasi, S. and Sarraf, M. (2019) 'Effect of emulsifier on rheological, textural and microstructure properties of walnut butter', *Journal of Food Measurement and Characterization*, 13 (1), pp. 785-792.
- Shapiro, L. (2001) 'Computer Vision/Linda G', *Shapiro, George C. Stockman-Pearson*.
- Shi, D., El-Farra, N. H., Li, M., Mhaskar, P. and Christofides, P. D. (2006) 'Predictive control of particle size distribution in particulate processes', *Chemical Engineering Science*, 61 (1), pp. 268-281.
- Silva, K. A., Rocha-Leão, M. H. and Coelho, M. A. Z. (2010) 'Evaluation of aging mechanisms of olive oil-lemon juice emulsion through digital image analysis', *Journal of Food Engineering*, 97 (3), pp. 335-340.
- Singh, A. K., Chaurasiya, A., Awasthi, A., Mishra, G., Asati, D., Khar, R. K. and Mukherjee, R. (2009) 'Oral bioavailability enhancement of exemestane from self-microemulsifying drug delivery system (SMEDDS)', *Aaps Pharmscitech*, 10 (3), pp. 906-916.
- Singh, P. and Chadha, R. S. (2013) 'A novel approach to image segmentation', *International Journal of Advanced Research in Computer Science and Software Engineering*, 3 (4).
- Slanzi, G., Balazs, J. A. and Velasquez, J. D. (2017) 'Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention', *Information Fusion*, 35 pp. 51-57.
- Soldati, G., Del Ben, F., Brisotto, G., Biscontin, E., Bulfoni, M., Piruska, A., Steffan, A., Turetta, M. and Della Mea, V. (2018) 'Microfluidic droplets content classification and analysis through convolutional neural networks in a liquid biopsy workflow', *American journal of translational research*, 10 (12), p. 4004.

- 
- Spier, N., Nekolla, S., Rupprecht, C., Mustafa, M., Navab, N. and Baust, M. (2019) 'Classification of Polar Maps from Cardiac Perfusion Imaging with Graph-Convolutional Neural Networks', *Scientific Reports*, 9 (1), p. 7569.
- Stone, M. (1974) 'CROSS-VALIDATORY CHOICE AND ASSESSMENT OF STATISTICAL PREDICTIONS', *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 36 (2), pp. 111-147.
- Sun, K., Liu, J., Kang, J.-L., Jang, S.-S., Wong, D. S.-H., & Chen, D.-S. (2014). Soft Sensor Development with Nonlinear Variable Selection Using Nonnegative Garrote and Artificial Neural Network. In J. J. Klemeš, P. S. Varbanov & P. Y. Liew (Eds.), *Computer Aided Chemical Engineering*, 33, pp. 883-888.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. and Feuston, B. P. (2003) 'Random forest: a classification and regression tool for compound classification and QSAR modeling', *Journal of chemical information and computer sciences*, 43 (6), pp. 1947-1958.
- Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818-2826.
- Tasker, A., Sainsbury, F. and Puttick, S. (2018) 'Particle-Stabilized Fluid-Fluid Interfaces: The Impact of Core Composition on Interfacial Structure', *Frontiers in Chemistry*, 6 p. 7.
- Tontul, I. and Topuz, A. (2015) 'Chapter 13 - Microencapsulation of Plant Oils Rich in Alpha-Linolenic Acid: Effect of Processing Parameters'. In: L. M. C. Sagis (ed.) *Microencapsulation and Microspheres for Food Applications*. San Diego: Academic Press.
- Turk, M. and Pentland, A. (1991) 'Eigenfaces for Recognition', *Journal of Cognitive Neuroscience*, 3 (1), pp. 71-86.
- Ulery, B., Hicklin, R., Buscaglia, J. and Roberts, M. (2012) 'Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners', *Plos One*, 7 (3).
- Unnikrishnan, S., Donovan, J., Macpherson, R. and Tormey, D. (2017) 'Multi-response Optimisation of Image Processing Parameters using CCD and Desirability Function ', *International Manufacturing Conference (IMC)*. Sligo, Ireland.
- Unnikrishnan, S., Donovan, J., Macpherson, R. and Tormey, D. Machine vision for the quality assessment of emulsions in pharmaceutical processing. *4th International Conference on Universal Village (UV)*. Boston, MA, USA, 21-24 Oct. 2018 2018. IEEE, pp. 1-6.
- Unnikrishnan, S., Donovan, J., Macpherson, R. and Tormey, D. (2019) 'Machine Learning for Automated Quality Evaluation in Pharmaceutical Manufacturing of Emulsions', *Journal of Pharmaceutical Innovation*, pp. 1-12.
- Vago, E. and Kemeny, S. (2012) 'Random Effects Model for Attribute Gauge R&R', *Quality and Reliability Engineering International*, 28 (8), pp. 807-823.
- Van Dalen, G. (2002) 'Determination of the water droplet size distribution of fat spreads using confocal scanning laser microscopy', *Journal of Microscopy*, 208 (2), pp. 116-133.

- 
- Vankeirsbilck, T., Vercauteren, A., Baeyens, W., Van Der Weken, G., Verpoort, F., Vergote, G. and Remon, J. P. (2002) 'Applications of Raman spectroscopy in pharmaceutical analysis', *TrAC trends in analytical chemistry*, 21 (12), pp. 869-877.
- Vasconcelos, T., Marques, S. and Sarmento, B. (2018) 'Measuring the emulsification dynamics and stability of self-emulsifying drug delivery systems', *European Journal of Pharmaceutics and Biopharmaceutics*, 123 pp. 1-8.
- Vaswani, N. and Chellappa, R. (2006) 'Principal components null space analysis for image and video classification', *IEEE Transactions on Image Processing*, 15 (7), pp. 1816-1830.
- Venora, G., Grillo, O. and Saccone, R. (2009) 'Quality assessment of durum wheat storage centres in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis system', *Journal of Cereal Science*, 49 (3), pp. 429-440.
- Wang, P., Sun, J. B., Zhang, T. T. and Liu, W. J. (2016) 'Vibrational spectroscopic approaches for the quality evaluation and authentication of virgin olive oil', *Applied Spectroscopy Reviews*, 51 (10), pp. 763-790.
- Wang, Y., Li, Z., Zhang, Z. and An, D. (2004) 'Applications of Raman spectroscopy in pharmaceutical analysis', *Acta Pharmaceutica Sinica*, 39 (9), pp. 764-768.
- Wang, C. and Xi, Y. 'Convolutional Neural Network for Image Classification', *Johns Hopkins University Baltimore, MD*, 21218.
- Webster, D. D. and Breckon, T. P. Improved raindrop detection using combined shape and saliency descriptors with scene context isolation. *2015 IEEE International Conference on Image Processing (ICIP)*. 27-30 Sept. 2015, pp. 4376-4380.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.
- Wold, S., Kim, E. and Geladi, P. (1987) 'Principal Component Analysis', *Chemometrics and Intelligent Laboratory Systems*, 2 (1-3), pp. 37-52.
- Xanthopoulos, P., Pardalos, P. M. and Trafalis, T. B. (2013) 'Linear discriminant analysis'. *Robust data mining*. Springer.
- Xudong, X. and Kin-Man, L. (2006) 'Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image', *IEEE Transactions on Image Processing*, 15 (9), pp. 2481-2492.
- Yang, M.-S. and Wu, K.-L. (2006) 'Unsupervised possibilistic clustering', *Pattern Recognition*, 39 (1), pp. 5-21.
- Yu, Y., Wang, J., Ng, C. W., Ma, Y., Mo, S., Fong, E. L. S., Xing, J., Song, Z., Xie, Y. and Si, K. (2018) 'Deep learning enables automated scoring of liver fibrosis stages', *Scientific reports*, 8 (1), article number. 16016.
- Zeaiter, J., Romagnoli, J. A. and Gomes, V. G. (2006) 'Online control of molar mass and particle-size distributions in emulsion polymerization', *Aiche Journal*, 52 (5), pp. 1770-1779.
- Zhang, K., Chen, X. and Wang, H. (2018) 'Research on External Quality Inspection Technology of Tropical Fruits Based on Computer Vision', *Springer International Publishing*, pp. 165-174.
- Zhao, K., He, H., Gao, P., Liu, S., Zhang, X., Li, G. and Wang, Y. (2019) 'A new approach for vaginal microbial micrograph classification using convolutional neural network combined with decision-making tree (CNN-DMT)', In Tenth International Conference on Graphics and

---

Image Processing (ICGIP 2018) (Vol. 11069, p. 110692U). *International Society for Optics and Photonics*.

## Appendix A: Formulae for the droplet size characteristics used in this study

**Table A1. Droplet size characteristics extracted in Fiji and their formulae (Venora et al., 2009).**

Droplet Feature	Formula <sup>a</sup>	Description
Area ( $\mu\text{m}^2$ )	$\frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$ $(x, y) \in \text{ROI}$	Dimension
Perimeter ( $\mu\text{m}$ )	$\sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$	Dimension
Maximum Feret Diameter, Feret ( $\mu\text{m}$ )	$\left[ \frac{[\mu_{2,0} + \mu_{0,2} - \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2}]}{\mu_{0,0}} \right]^{1/2}$	The maximum distance between any two points along the selected droplet boundary.
Minimum Feret Diameter, minFeret ( $\mu\text{m}$ )	$\left[ \frac{[\mu_{2,0} + \mu_{0,2} + \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2}]}{\mu_{0,0}} \right]^{1/2}$	The minimum distance between any two points along the selected droplet boundary.

<sup>a</sup> Where: ROI is the region of interest,  $\mu_i$  is the  $i$ th central moment of the pixel intensity distribution.

---

## Appendix B: Multi-response Optimisation of Image Processing Parameters using CCD and Desirability Function

This section summarises the Design of Experiments, conducted for the parameter optimisation of the Edge and Symmetry technique (EST) used for droplet detection, presented in the IMC Conference 34 (Unnikrishnan *et al.* 2017). The abstract of the conference paper is presented here.

### Abstract

This paper focuses on the achievement of an optimum set of image processing parameters leading to robust detection of the maximum objects in an image with minimum error. Bright Field micrographs of oil-in-water (o/w) emulsions have been processed in Fiji, a scientific image processing package, for this study.

The image processing algorithm, Edge and symmetry filter (ESF), used for the detection of oil droplets in the micrographs is based on a two-step process. The first step of the algorithm performs a Gaussian convolution to smoothen the image for the given value of alpha Canny and subsequently detects the edges of the objects in the image using Canny edge detection criteria. The second phase of the algorithm applies a symmetry filter to the objects based on the supplied radius parameter value. The optimum values of three selected parameters of the algorithm such as alpha Canny, radius and normalisation are determined.

The design of experiments is performed using a Central Composite Design (CCD), one of the most popular designs of Response Surface Methodology (RSM). Regression models are created for multiple responses such as droplet count, average droplet area, integrated pixel density of droplets and % of area occupied by the droplets.



---

A desirability function approach is applied to achieve simultaneous optimisation of the multiple responses. The aim was to identify the input parameters that can detect maximum droplet count and integrated density, while simultaneously achieving set target values for average droplet area and % of area occupied by the droplets. Finally, an overall desirability level is achieved for the responses.

The predicted responses of the models showed good correlation with the corresponding experimental outputs. The desirability function approach offered an overall desirability of 72%. The optimum input parameter setting of  $\alpha = 0.8$ , radius = 40 pixels and normalisation = 9.55 were achieved for the desired responses. The study concludes that an RSM approach together with multi-response optimisation can be proposed as a potential technique to achieve an optimum set of input parameters of software algorithms, which can improve the accuracy of computing and minimise error rate, leading to robust results.

---

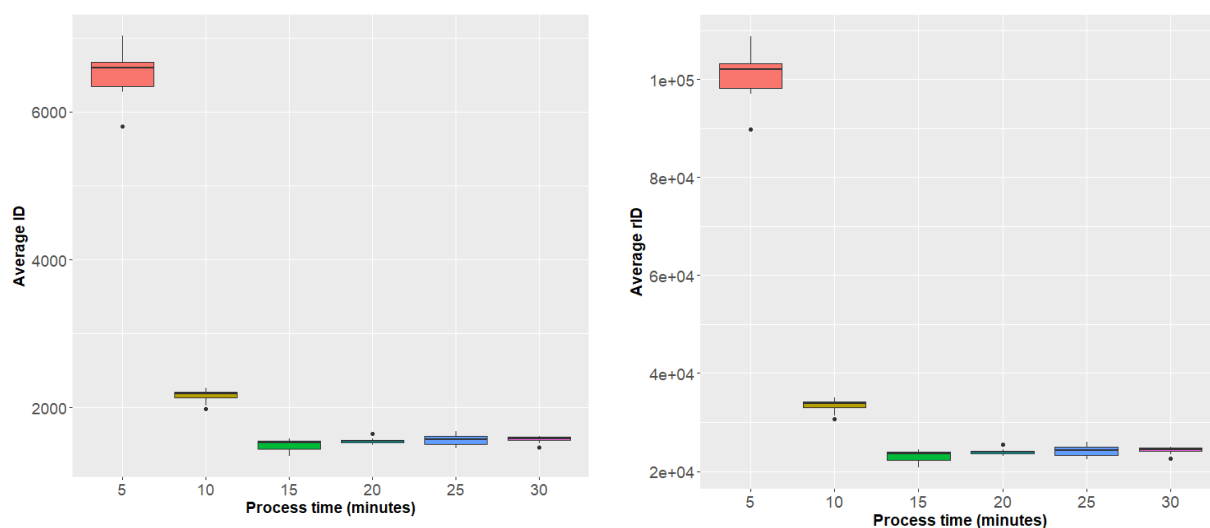
## Appendix C: Plots of Additional Droplet Characteristics

Distribution plots of the droplet characteristics, obtained from EST and HBT, which demonstrated zero/minimal variation during Process-1 and therefore, were found irrelevant for this study are presented in this section.

### Droplet Characteristics obtained from Edge and Symmetry Technique (EST)

Box plots of the average pixel intensity characteristics of the droplets in each micrograph at five-minute process intervals are presented in Figure C.1. Pixel Intensity measures obtained are integrated density (ID) and raw integrated density (rID) respectively. rID is measured as the sum of the pixel values in the ROI (droplets) of the image. ID is calculated as the product of droplet area in calibrated units ( $\mu\text{m}^2$ ) and the mean pixel intensity of the droplet.

#### 1. Pixel Intensity Characteristics

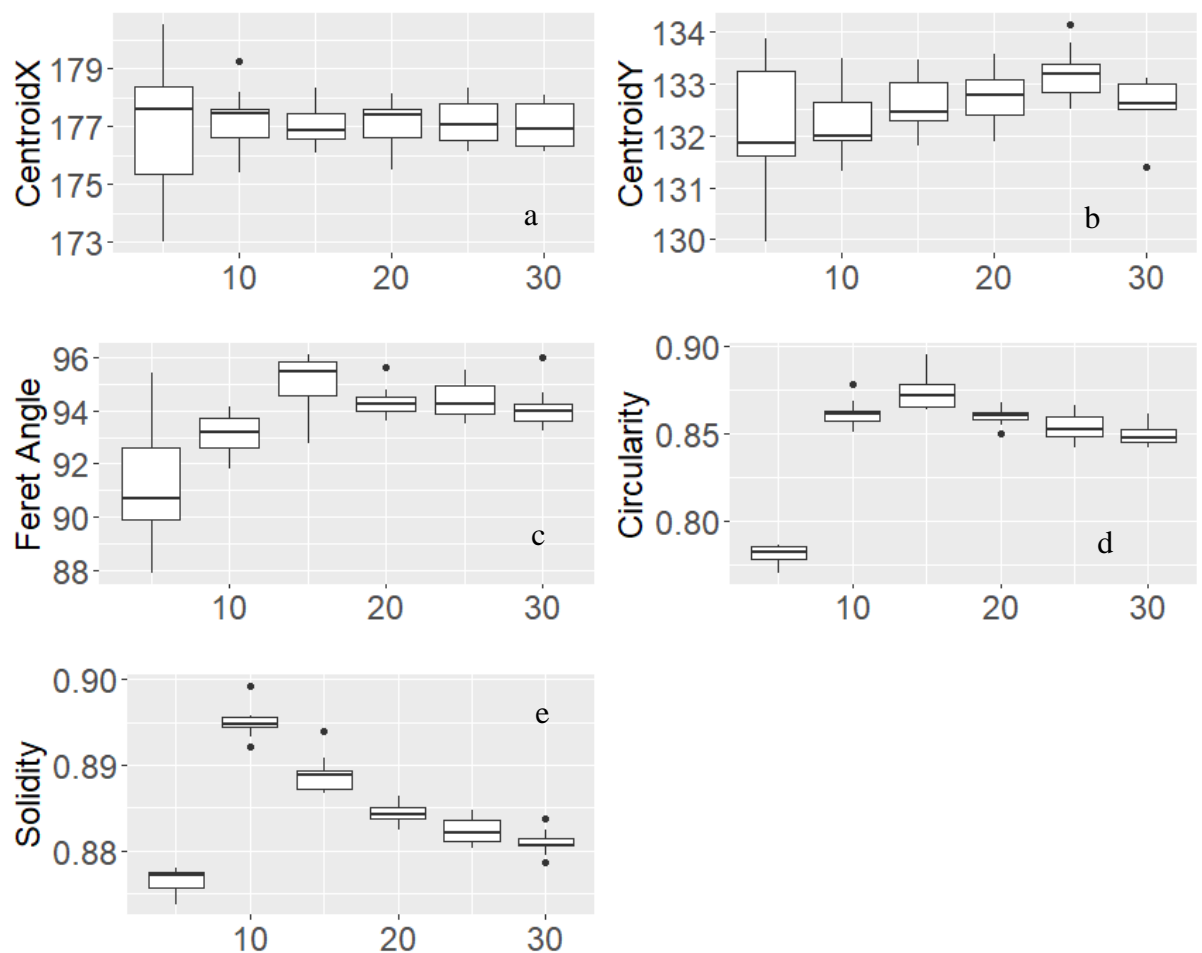


**Figure C.1. Box plots of average pixel intensity characteristics of droplets. a) integrated density b) raw integrated density. Each box plot represents the average of 10 micrographs.**

The integrated density (ID) and the raw integrated density (rID) presented a variation similar to that of droplet area as seen from Figure 4.3a. This is because pixel intensity characteristics are a function of droplet area. Therefore, these characteristics were redundant.

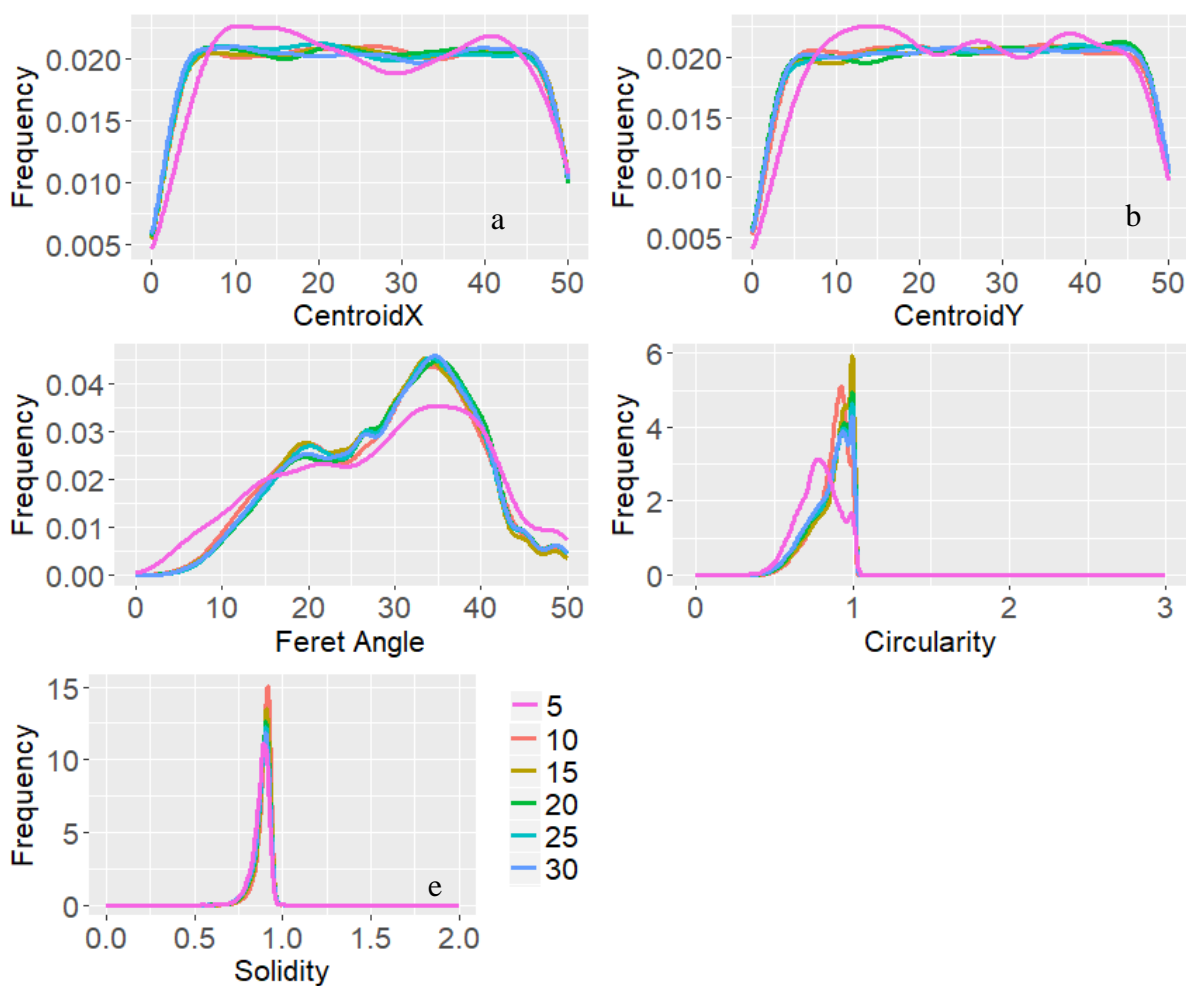
## 2. Centroid, Orientation and Shape characteristics of droplets

The X and Y coordinates of the centroid, Feret Angle, circularity and solidity of droplets are the other features presented in this section. Figure C.2 shows the box plots of the mean values of these characteristics in each micrograph obtained from Process-1.



**Figure C.2.** EST box plots of the mean: a) CentroidX, b) CentroidY, c) Feret Angle, d) Circularity and e) Solidity of the droplets in each micrograph. Each box plot represents the mean of 10 micrographs.

The mean droplet characteristics presented in Figure C.2 showed very minimal variation during the emulsification process. Frequency distribution plots of these droplet characteristics during the 30-minute process were also plotted and are presented in Figure C.3.



**Figure C.3.** EST Frequency distribution plots of: a) CentroidX, b) CentroidY, c) Feret Angle, d) Circularity and e) Solidity of the droplets at every five-minute interval of Process-1.

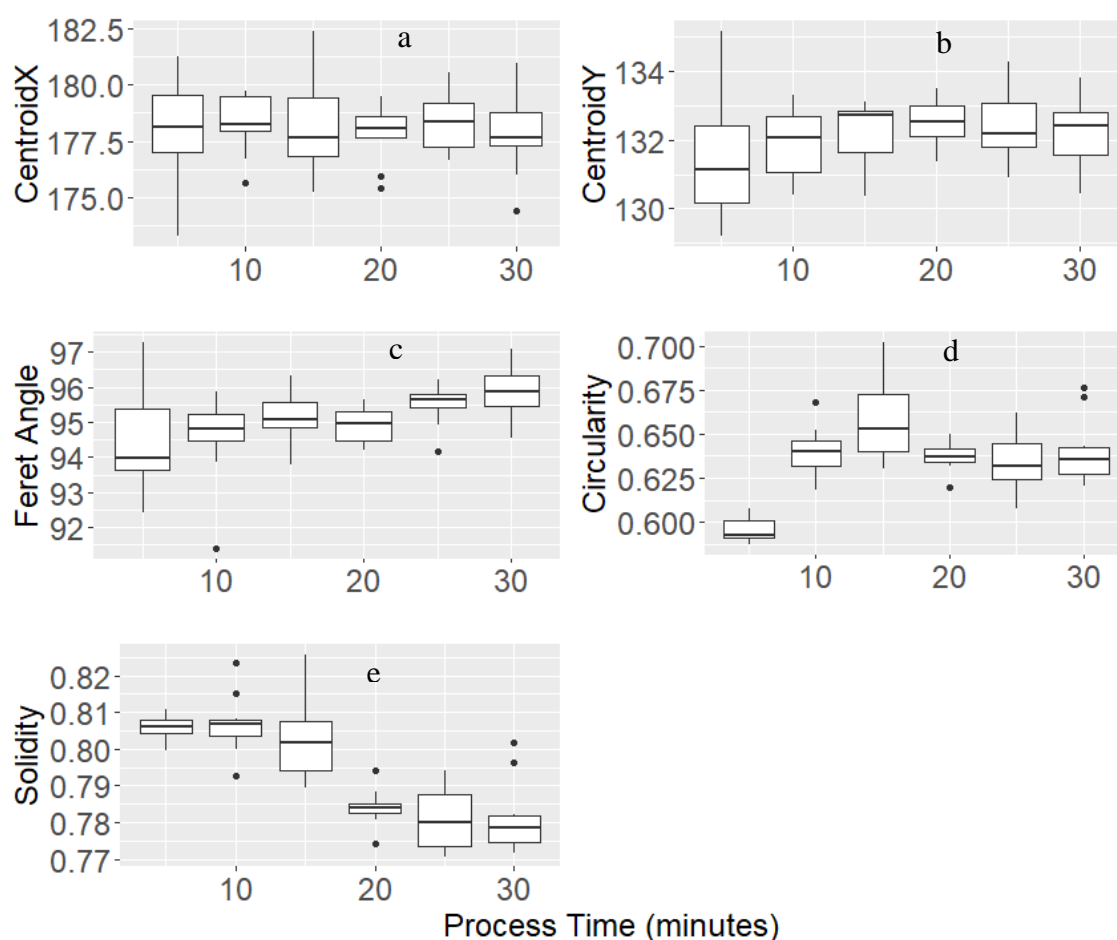
The frequency plots also presented very minimal variation of the above characteristics during the 30-minute emulsification process. Therefore, based on the analyses conducted, the centroid coordinates, Feret angle, circularity and solidity of the droplets were not significant predictors for the machine learning models developed in this study.

---

## Droplet Characteristics obtained from Histogram-Based Technique (HBT)

The pixel intensity characteristics of the droplets obtained using the HBT has a constant value. This is because, the technique sets the mean pixel intensity of each micrograph as the threshold value for droplet detection. For this reason, all the droplets detected with a pixel intensity value greater than the mean value will be assigned zero. These droplets are detected as black objects/regions. Therefore, pixel intensity characteristics are not considered for the analysis.

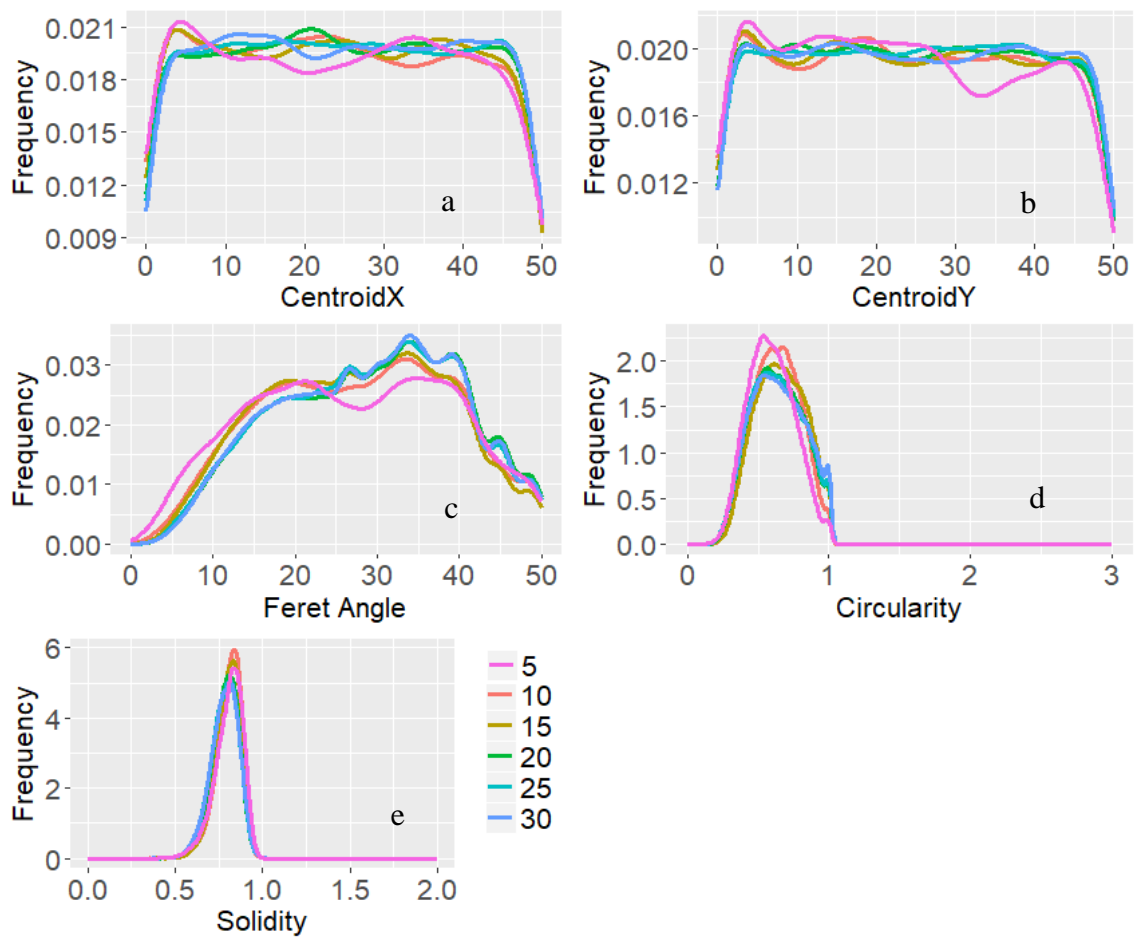
Centroid, Orientation and Shape characteristics of droplets



**Figure C.4.** HBT box plots of the mean: a) CentroidX, b) CentroidY, c) Feret Angle, d) Circularity and e) Solidity of the droplets in each micrograph. Each box plot represents the mean of 10 micrographs.

The mean droplet characteristics presented in Figure C.4 showed very minimal variation during the 30-minute emulsification process. In addition, frequency distribution plots of

these droplet characteristics, during the process, were also plotted and are presented in Figure C.5.



**Figure C.5. HBT frequency distribution plots of: a) CentroidX, b) CentroidY, c) Feret Angle, d) Circularity and e) Solidity of the droplets at every five-minute interval of Process-1.**

The frequency plots of the centroid coordinates, Feret angle, circularity and solidity of the droplets also presented very minimal variation during the 30-minute emulsification process. Therefore, based on the analyses conducted, these were not considered to be significant predictors for developing the machine learning models in this study.

---

## Appendix D: TAMU Micrographs

A sample micrograph from each of the TAMU categories are presented in this section.

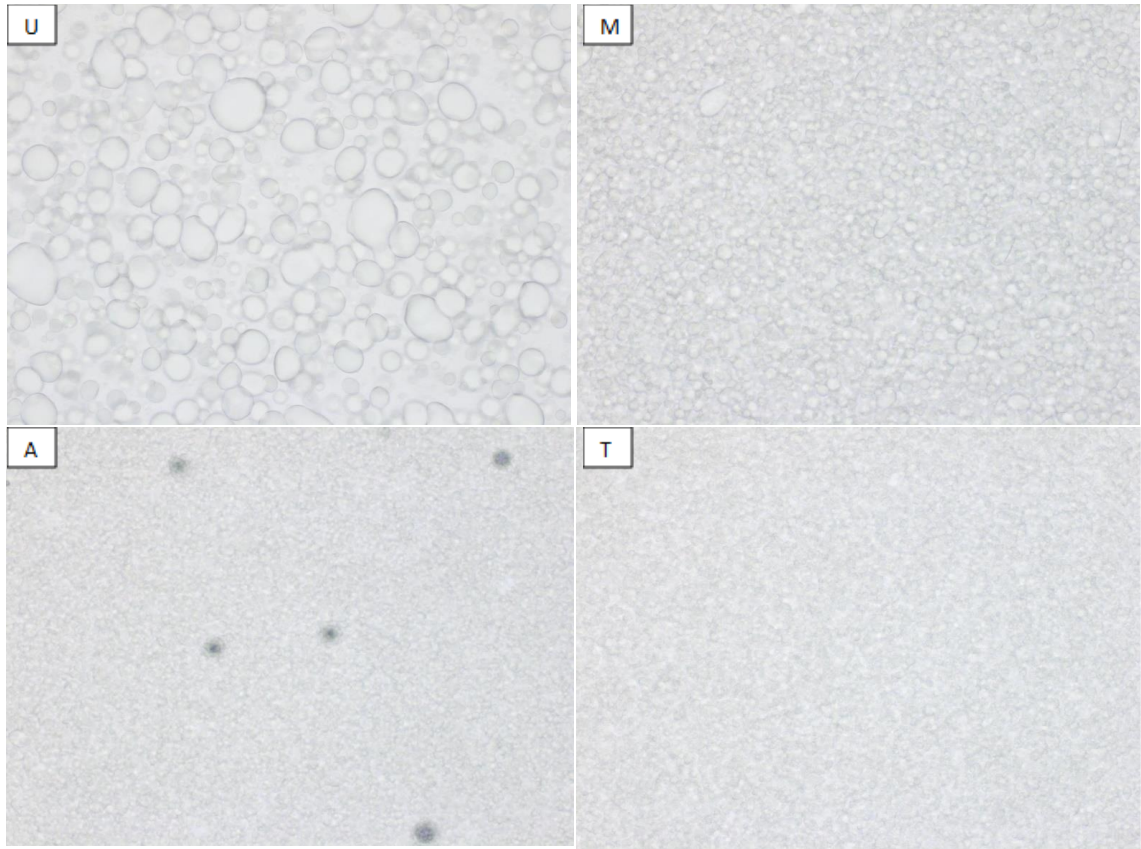


Figure D.1. TAMU micrographs. 'U' stands for Unacceptable, 'M' stands for Marginal, 'A' stands for Acceptable and 'T' stands for Target.