An Integrated Histogram-Based Vision and Machine-Learning Classification Model for Industrial Emulsion Processing

Saritha Unnikrishnan, John Donovan, *Member, IEEE*, Russell Macpherson, and David Tormey

***Abstract*—Existing techniques in emulsion quality evaluation are found to be highly subjective, time-consuming, and prone to overprocessing. Other conventional droplet analysis techniques such as laser diffraction, which require dilution of samples, introduce an additional complexity to industrial processes. The possibility of developing a fully automated technique for droplet characterization during emulsification holds remarkable potential for overcoming the existing challenges. In this article, a histogram-based image segmentation technique detects droplets from emulsion micrographs. The evolution of droplet characteristics and their significance are studied by performing statistical analysis, and the significant characteristics are selected. The principal component analysis is applied to obtain a reduced set of uncorrelated components from the selected characteristics. The linear discriminant analysis classifies the micrographs into a set of quality categories called target, acceptable, marginal, and unacceptable. The model accuracy is validated using stratified five-fold cross-validation and is successful in classifying the micrographs obtained from two different manufacturing facilities with high accuracy up to 100%. The histogram-based technique is successful in detecting smaller droplets than previously reflected in the literature. The current approach is fully automated and is implemented as a soft-sensor, which supports its real-time deployment into an industrial environment. The entire approach has promising potential in the in-line prediction of emulsion quality leading to more efficient and sustainable manufacturing.**

***Index Terms*—Emulsion manufacturing, image processing, linear discriminant analysis (LDA), machine learning, machine vision, principal component analysis (PCA), soft-sensor.**

Manuscript received July 31, 2019; revised October 7, 2019 and October 30, 2019; accepted November 25, 2019. Date of publication January 24, 2020; date of current version May 26, 2020. This work was supported by the Institute of Technology Sligo's President's Bursary award. This work was carried out as part of the North West Centre for Advanced Manufacturing project funded by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body. Paper no. TII-19-3451. *(Corresponding author: Saritha Unnikrishnan.)*

S. Unnikrishnan, J. Donovan, and D. Tormey are with the Department of Engineering and Design and Centre for Precision Engineering, Materials and Manufacturing Research, Institute of Technology Sligo, F91 YW50, Sligo, Ireland (e-mail: unnikrishnan.saritha@itsligo.ie; donovan.john@itsligo.ie; tormey.david@itsligo.ie).

R. Macpherson is with GlaxoSmithKline Limited, F91 P773 Sligo, Ireland (e-mail: russell.i.macpherson@gsk.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2019.2959021

I. Introduction

RECENT advances in industrial automation have imposed tremendous pressure on pharmaceutical industries to rapidly improve the techniques applied for the evaluation of the emulsification process [1]–[3]. Emulsions have a wide range of applications in food, pharmaceutical, chemical, and biomedical industries. A better understanding of droplet size has been identified as the key factor in the control and optimization of industrial emulsification processes [1].

Manual evaluation of emulsion samples is one of the techniques currently employed in industries to identify the optimum processing time of emulsification [1], [4]–[8]. Such evaluation techniques were found to be significantly biased. 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 [7], [9]. However, those techniques were found incapable of delivering reliable droplet size measurements and require time-consuming sample preparation [8], [10], [11]. Such limitations have made the automation of these conventional techniques difficult. Nuclear magnetic resonance (NMR) is another common technique applied for emulsion droplet sizing. NMR is a well-established technique used to find droplet size distributions in a noninvasive manner. However, the technique requires approximately 5–20 min to provide a single-droplet size distribution, which is a very time-consuming procedure [12], [13]. Recent studies have found image processing of micrographs (image taken through a microscope) as a potential approach to rectify the existing challenges in emulsion droplet size evaluation [1], [7], [10].

Emulsion stability studies have been performed using light microscopy in conjunction with image processing and statistical analysis from the early 21st century [14]–[17]. Jorin's ViPA and J.M. Canty provide commercially available image analysis systems for the measurement of oil droplets from diluted samples [18]. A few studies have reported inline droplet size monitoring in emulsions using automated image analysis [1], [19], [20]. The existing image processing techniques applied for droplet characterization and emulsion quality evaluation have also identified certain challenges. These include droplet detection from highly concentrated emulsions and the detection of smaller droplets from production systems [1], [8]. Studies have also identified difficulty in using optical microscopy on emulsion samples due to its opaque nature [21].

ImageJ has been widely applied for image processing of emulsion micrographs [13], [15]–[17]. The utilities of ImageJ have been extended to Fiji, which is a "batteries-included" distribution of ImageJ [22]. Image processing techniques integrated with machine-learning classifiers have a long history of success in product quality evaluation in industries such as automotive, electronics, medical, pharmaceutical, and food. However, such techniques are currently underexplored in the area of industrial emulsification. It has been identified, in the literature, that developing inferential models using droplet data could benefit the optimization of the emulsification process [23], [24]. A very few studies have been reported in the literature, which have developed classification or prediction models using droplet data for the quality evaluation of emulsions [7], [25]. Wang *et al.* [25] have provided a detailed review about the use of various supervised and unsupervised machine learning models in the quality evaluation of olive oil using spectroscopic data.

Principal component analysis (PCA) and linear discriminant analysis (LDA) were described as the most widely used unsupervised and supervised methods for exploratory data analysis and classification, respectively [25]. PCA projects the variation in the original multivariate data set across an equal number of uncorrelated components onto an orthogonal subspace [26]. It aids in reducing the dimensionality of a multivariate dataset into a set of principal components (PCs) ordered hierarchically based on their explained variance. PCA has been extensively used as an efficient method for dimensionality reduction and pattern recognition in the field of computer vision and image classification [27], [28].

LDA is a supervised machine-learning method used for multilevel classification of a categorical response [29]. LDA is trained using a set of observations, taken from a multivariate dataset, in order to build a classification model. The model is then used to predict the category of unknown (new) data based on *a priori* knowledge.

Bertani *et al.* [30] achieved a partly automatic multivariate classification of hyperspectral micrographs of living cells using PCA followed by LDA. Application of computer vision integrated with a LDA classification model was investigated for the automated evaluation of durum wheat quality and their study has achieved 96% accuracy in classifying the wheat kernels into four categories [31]. A similar classification study of barley milk samples, obtained by blending barley grain, was conducted by Kljusuric *et al.* [7] to find the optimum processing time in barley milk production. Their study concluded with 45 s as the optimal blending time because the droplets appeared to form aggregates after that point according to the analyses conducted. However, their study concluded with a limited set of samples and lacked detailed analysis of the findings using independent barley milk samples. In addition to that, their PCA results were not extended to any supervised learning techniques to develop a predictive classification model or to optimize the process on an industrial scale.

In addition, there have been no previous studies reported in the supervised classification of in-process emulsion samples through fully automated image processing and droplet characterization. In this article, we have used a histogram-based image segmentation technique and investigated its potential in droplet characterization from micrographs acquired at fixed intervals from an emulsification process. A supervised machine-learning model is developed and integrated with the automated vision technique to classify the micrographs into four quality categories. This is a novel approach in emulsion manufacturing.

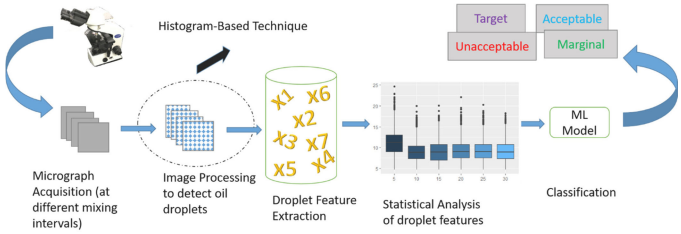

Fig. 1. Schematic representation of the TAMU classification of in-process micrographs.

II. Methodology

This article involved the industrial production of a topical skincare emulsion. The emulsion was continuously mixed for 30 min using a homogenizer at a tip speed of 25 m/s for the first 15 min and at a tip speed of 15 m/s for the last 15 min, respectively. Micrographs were acquired from the emulsion samples at 5-min intervals. A schematic representation of the applied methodology is shown in Fig. 1.

Automated image processing of the micrographs was performed to detect the oil droplets and their corresponding characteristics using a histogram-based technique (HBT). The droplet characteristics were then statistically analyzed to investigate their variation throughout the emulsification process and their impact on emulsion quality. Based on the analysis, the characteristics suitable for the classification of micrographs were identified. The micrographs were categorized into a set of quality categories ranging from unacceptable to target, based on the droplet characteristics. PCA of the selected set of characteristics was performed and a PC based LDA (PC-LDA) model was developed to classify the micrographs into four groups named target, acceptable, marginal, and unacceptable (TAMU).

A. Micrograph Acquisition

A Zeiss Microscope Axio imager A2m was used to obtain the micrographs. Samples were acquired from the production process at 5-min intervals. Ten bright field (BF) 40x micrographs were acquired from each sample and were saved as tagged file format files. The emulsification process was stopped after 30 min, as any further processing was considered as over processing. This represented a total of 60 micrographs. A sample micrograph obtained at every 5-min interval is shown in Fig. 2.

B. Image Processing and Droplet Detection

The image processing of the micrographs was performed in Fiji version 1.51 h, which is an extended distribution of ImageJ with additional functionalities. An automated image segmentation procedure named HBT was developed in Fiji to detect

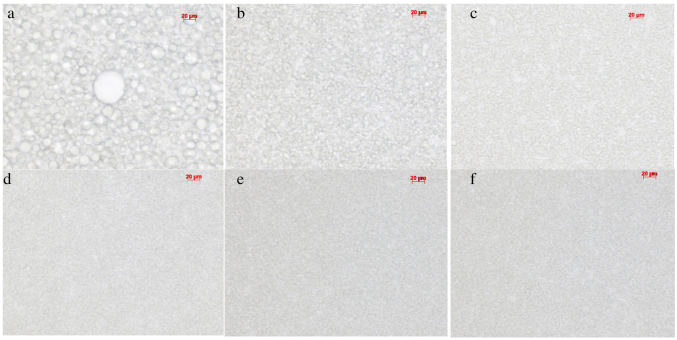

Fig. 2. BF 40x micrographs of the emulsion after (a) 5 min, (b) 10 min, (c) 15 min, (d) 20 min, (e) 25 min, and (f) 30 min of processing.

the droplets in the micrographs by Unnikrishnan *et al*. [32]. The existing image segmentation techniques, as described in the literature, for droplet detection from emulsion micrographs are mainly border-based edge detection methods. Such techniques have demonstrated potential only in the case of high-quality images, droplets with pronounced borders, less overlap, and in emulsions with low dispersed phase fraction $\leq$15%. The HBT technique is based on computing the histogram of the pixel intensity values (grey values) in the image, observing the number of peaks, and thresholding the image using the peak intensity value. The HBT technique also detects various size and shape characteristics of the detected droplets. The steps involved in the HBT technique are programmed as a macro and are presented schematically in Fig. 3.

The image processing and droplet detection steps given in Fig. 3 can be explained as follows.

1) The micrographs were calibrated to a scale of 4 pixels/μm.
2) These images were then converted to 8-bit greyscale, which provides 2^8 = 256 levels of intensity values for each pixel.
3) A histogram of the pixel intensity values was computed for each image.
4) The number of peaks of the histogram was observed and the peak intensity value was stored in a variable.
5) Each image was then thresholded using the calculated peak intensity value.
6) The images were then converted to binary, and watershed segmentation was applied to separate the droplets that touch/overlap each other.
7) Finally, each micrograph was analyzed for a droplet area range $\geq$1 μm^2 and a circularity range of 0.00–1.00 to extract the droplet characteristics.

Thirteen characteristics were extracted for each droplet from the 60 micrographs. These are the following:

1) Size features: area (μm^2), perimeter (μm), maximum Feret diameter (Feret in μm), minimum Feret diameter (MinFeret in μm);
2) Centroid coordinates: *X* and *Y*;
3) Starting coordinates of Feret diameter: FeretX and FeretY;
4) Orientation characteristic: Feret angle;

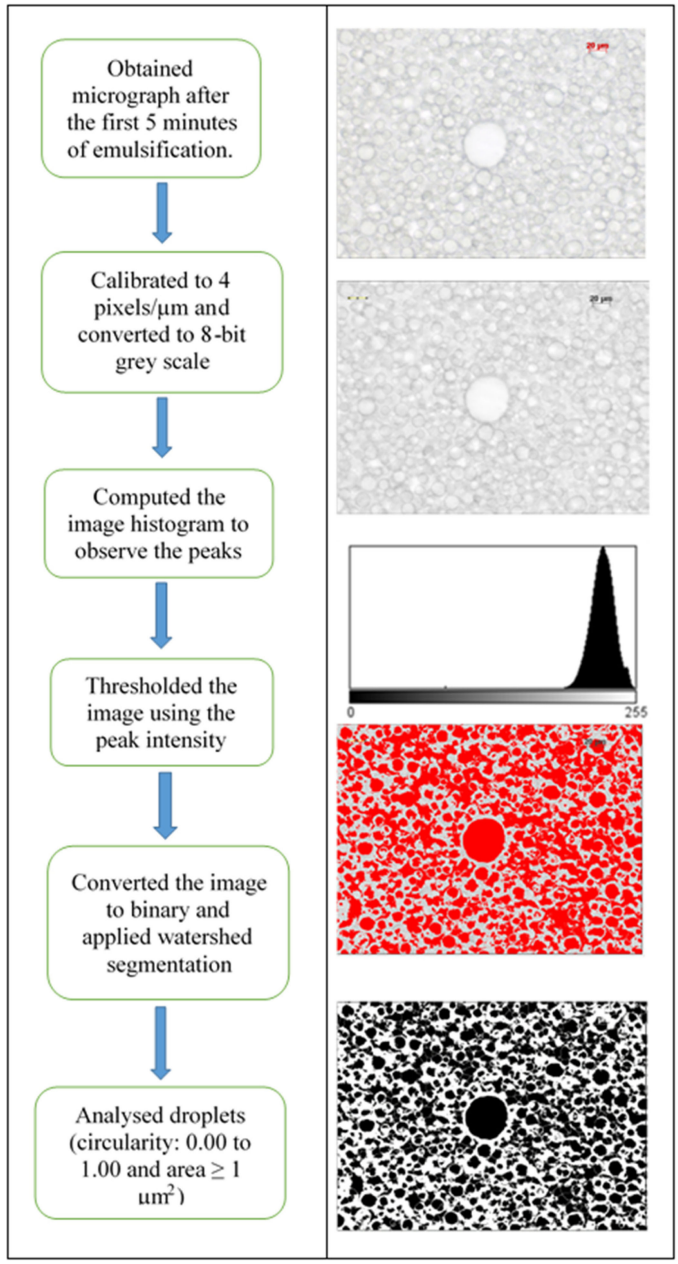

Fig. 3. Schematic showing the detection of oil droplets and their characteristics from an emulsion micrograph using the HBT image segmentation macro in Fiji.

5) Shape features: Circularity, roundness, aspect ratio, and solidity.

The droplet characteristics were automatically exported, by the macro, into a CSV file in the user-specified directory.

C. Statistical Analysis

The droplet characteristics were statistically analyzed to investigate their variation at every 5-min interval throughout the 30-min emulsification process. The analysis was performed in RStudio, which is an integrated development environment for programming using the R language. R 3.4 is the version used in this article. The mean values of the droplet characteristics

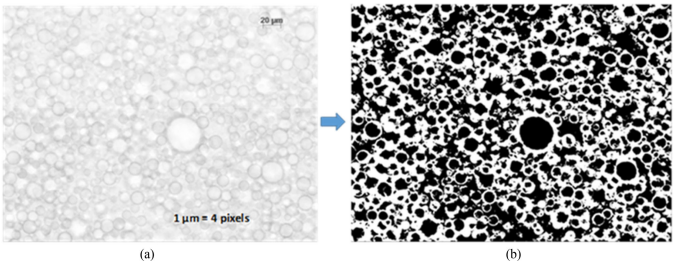

Fig. 4. Droplet detection. (a) Micrograph obtained from an emulsion sample processed for 5 min. (b) Output image with the droplets detected using the HBT.

were calculated, and box plots were generated to identify the characteristics that varied significantly over the emulsification period. The *R* data visualization package ggplot2 was used to create the plots. The variation in the mean and median of each characteristic and the correlation between them were also investigated. A reduced set of droplet characteristics, deemed suitable, was finalized as the input feature space for the classification of micrographs. This analysis and feature selection are presented in detail in Section III.

D. Multivariate Classification

The total set of 60 micrographs were categorized manually into four groups named unacceptable (U), marginal (M), acceptable (A), and target (T) by micrograph analysts from industry. An unsupervised cluster analysis of the droplet characteristics was performed using PCA to observe the patterns in the data. The PCA technique also helped to reduce the dimensionality of the data and to obtain a set of uncorrelated components. A scree plot was used to select the most significant PCs. Scree plot shows the percentage of total variance in the data as explained by each PC. The score plots of the first three PCs were also graphed. The PCs, which explained a significant percentage of the variance in the data and were also identified as the most relevant for classification, were selected as the predictor variables to build the supervised classification model using LDA. The PC-LDA model accuracy was evaluated using stratified five-fold cross-validation. The model results are presented in Section III.

III. Results and Discussion

The results of droplet detection and characterization are discussed in this section.

A. Selection of Feature Space

Approximately 1500–2000 droplets were detected from a typical micrograph after the first 5 min of emulsification processing, using HBT segmentation, and their characteristics were obtained. The droplets detected from a sample micrograph acquired after the initial 5 min of processing are shown in Fig. 4. Similarly, the droplet characteristics of all the 60 micrographs were extracted automatically, in an iterative loop, by the macro.

The HBT method presented good detection of both large and small droplets. An *R* function was written to read the droplet characteristics of each micrograph from the CSV file (saved by

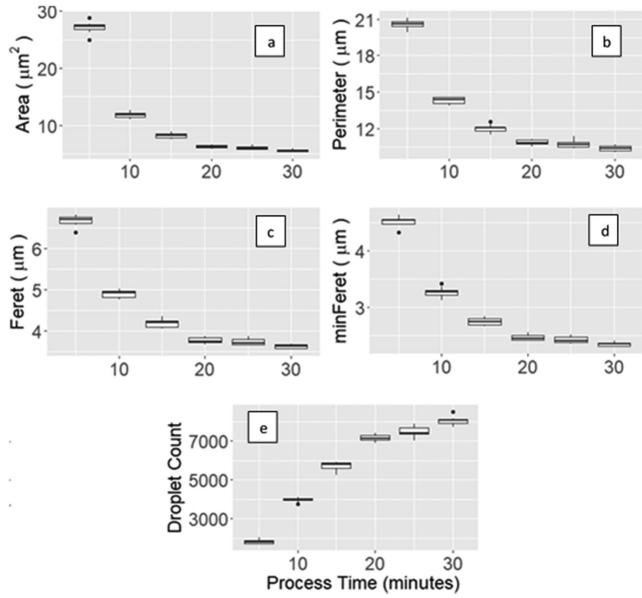

Fig. 5. Box plots of mean droplet size characteristics and droplet count obtained from the HBT: (a) area, (b) perimeter, (c) Feret, (d) MinFeret, and (e) droplet count. Each box plot represents ten micrographs.

the macro) in a sequential loop and create box plots showing their evolution over the emulsion processing time. A plugin was developed in Java language, as a soft sensor, to integrate the Fiji macro with the *R* function.

The variation in the droplet characteristics throughout the process was statistically analyzed using the box plots presented in Fig. 5. Among the 13 characteristics obtained from each droplet, only the size features such as area, perimeter, Feret, and MinFeret as well as droplet count showed significant variation during the emulsification process (see Fig. 5). The remaining droplet characteristics such as orientation, shape, and centroid were not considered relevant for this article, as they showed no variation throughout the emulsification process.

Each box plot, from Fig. 5(a) to (d), represents the mean droplet size obtained from the ten micrographs at every 5-min interval. A sharp decrease was observed in the mean droplet size during the first 10 min followed by a progressive decrease throughout the remaining emulsification process. In the last 15 min, i.e., from 20 to 30 min on the *x*-axis of Fig. 5(a)–(d), minimal variation was observed in the droplet size, which indicated the process approaching a steady state. The droplet count presented in Fig. 5(e) shows a sharp increase initially followed by minimal variation during the last 10 min of the emulsification process. The overall variation in the mean droplet characteristics during the whole 30-min process can be summarized as follows.

1) Area decreased from 27.1 to 5.6 μm^2.
2) Perimeter decreased from 20.6 to 10.3 μm.
3) Maximum Feret diameter (Feret) dropped from 6.7 to 3.6 μm.
4) Minimum Feret diameter (MinFeret) dropped from 4.5 to 2.3 μm.
5) Droplet count increased from 1500 to 8500.

The 60 micrographs were manually grouped into four quality-based categories referred to as TAMU based on the variation

TABLE I
MICROGRAPH CATEGORIES NAMED FROM 5 TO 30 MIN BASED ON THE VARIATION IN THE DROPLET CHARACTERISTICS

Process time	Category
5 minutes	Unacceptable
10 minutes	Marginal
15 minutes	Acceptable
20, 25, 30 minutes	Target

TABLE II
CORRELATION MATRIX OF THE DROPLET SIZE CHARACTERISTICS SELECTED AS THE INPUT FEATURE SPACE FOR THE CLASSIFICATION MODEL

	Area	Perimeter	Feret	MinFeret
Area	1.00	0.85	0.84	0.87
Perimeter	0.85	1.00	0.98	0.96
Feret	0.84	0.98	1.00	0.94
MinFeret	0.87	0.96	0.94	1.00

presented by the mean droplet size as well as the droplet count characteristics, including the guidance obtained from the industrial collaborator. The processing intervals to which the categories belonged to are shown in Table I.

The ten sample micrographs obtained from the first 5 min were categorized as "U," another ten micrographs obtained after ten min of emulsification were labeled as "M," the next 10 obtained after 15 min, in total, were categorized as "A," and the 30 micrographs acquired after 20 min until the end of the process were grouped as "T."

The five droplet characteristics—area, perimeter, Feret, MinFeret, and count—were selected as the input features for building the classification model. A correlation matrix of the droplet characteristics was obtained and is presented in Table II.

The droplet size characteristics were found to be highly correlated ($r = 0.84$–0.98). The next major step in the analysis was to covert the correlated feature space into a reduced set of uncorrelated PCs. This was followed by an unsupervised cluster analysis of the components to distinguish the TAMU patterns identified by the micrograph analysts of the industrial collaborator.

B. Principal Component Analysis

PCA was performed to transform the five variable feature spaces into an equal set of uncorrelated PCs. The first three PCs (PC1, PC2, and PC3 explained as 77%, 17.4%, and 4.1%, respectively) together explained a cumulative variance of 98.5%, which represented a significant proportion of the total variance in the original feature space. A graphical representation of the variance explained by the PCs is shown by the scree plot presented in Fig. 6. An unsupervised clustering pattern of the four response categories (TAMU) was also explored

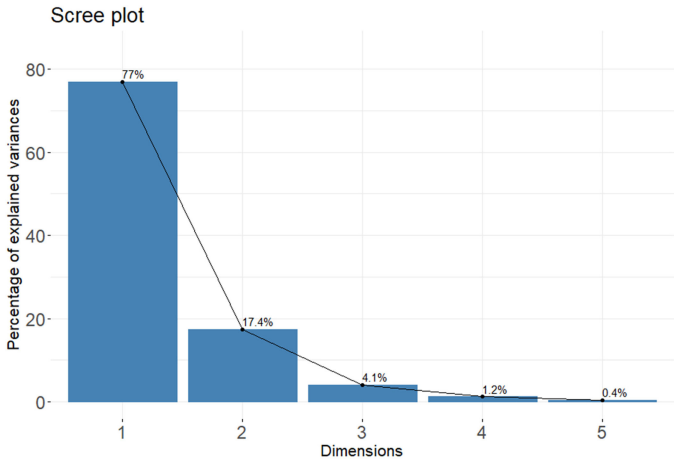

Fig. 6. Scree plot of the five PCs.

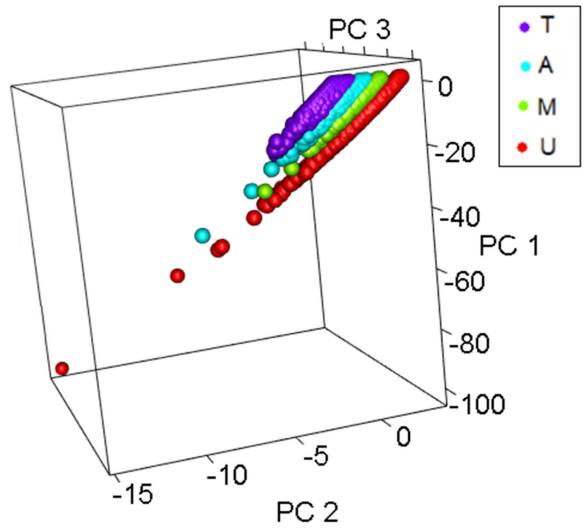

Fig. 7. PCA score plots using the first three PCs. "U" is represented by red circles, "M" by green, "A" by blue, and "T" by purple, respectively.

using the score plots of the first three PCs. This is presented in Fig. 7.

The droplets obtained from the TAMU categories, represented by the four different colors, were separated into four clusters along the PC1-PC2 plane as shown in Fig. 7. PCA also helped to reduce the dimensionality and the correlation of the original feature space from five correlated variables down to three uncorrelated components. PC1, PC2, and PC3 were selected as the predictor variables for developing the LDA classification model.

C. Supervised Classification Model

The PC-LDA model was developed as a linear combination of the first three PCs, PC1, PC2, and PC3. 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 values of p and G-1 are equal and, therefore, the PC-LDA model resulted in three discriminant functions, LD1, LD2, and LD3, as given by (1)–(3). The model

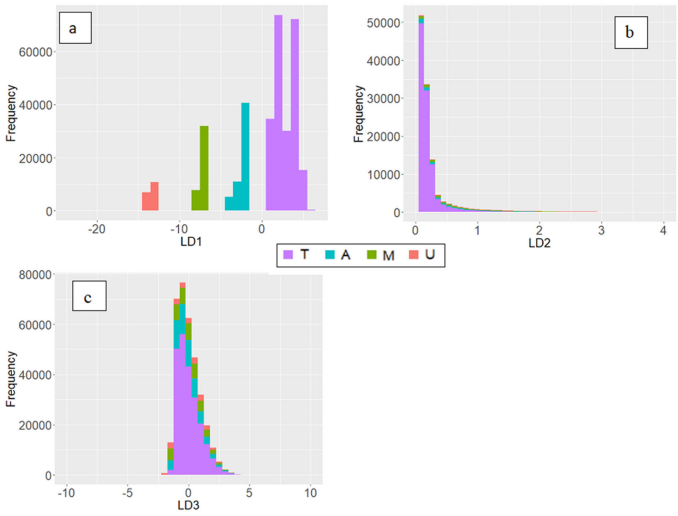

Fig. 8. LDA and the classification presented by the three discriminant functions. (a) LD1. (b) LD2. (c) LD3.

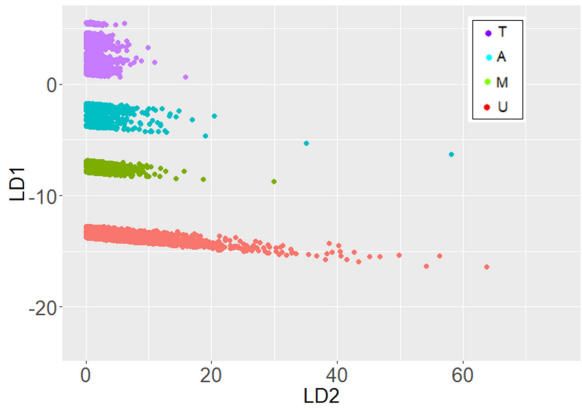

Fig. 9. TAMU classification presented by LD2 vs LD1 scatter plot.

formula is represented by (4)

$$\mathbf{LD1} = \mathbf{1.03PC} - \mathbf{0.31PC2} - \mathbf{0.35PC3} \quad (1)$$

$$\mathbf{LD2} = -\mathbf{4.68PC1} - \mathbf{0.29PC2} - \mathbf{0.35PC3} \quad (2)$$

$$\mathbf{LD3} = \mathbf{0.13PC1} + \mathbf{1.70PC2} - \mathbf{1.46PC3} \quad (3)$$

$$\mathbf{Category} \sim \mathbf{LD1} + \mathbf{LD2} + \mathbf{LD3}. \quad (4)$$

The classification presented by each discriminant function was observed by plotting the histograms shown in Fig. 8(a)–(c). The first discriminant function, LD1, was found to best separate the four TAMU categories. The percentage of classification achieved by each discriminant function is explained by the proportion of trace given by the model, which was 99.86% for LD1. A two-dimensional scatter plot of LD2 vs LD1 was also plotted to see the overall separation between the four categories (see Fig. 9).

The scatter plot in Fig. 9 presented very good classification between the four TAMU categories along the LD1 axis. In summary, the number of canonical variables selected for the PC-LDA model was reduced to one (LD1).

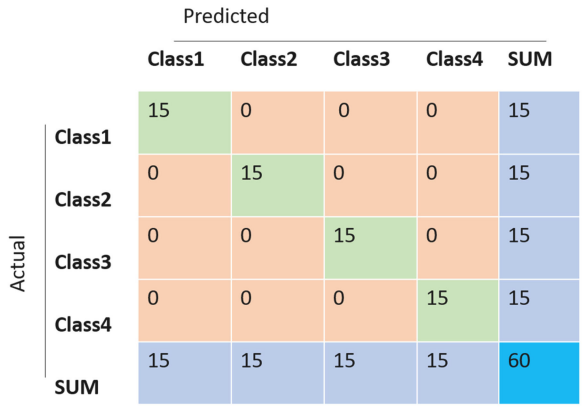

Fig. 10. Confusion Matrix of PC-LDA model from five-fold cross-validation. The sum of five confusion matrices from the five models is represented. The green cells represent the correct classification from each category. The blue cell at the bottom right-hand corner represents the total number of correct classifications.

A stratified five-fold cross-validation was performed to evaluate the classification accuracy of the PC-LDA model. Ten micrographs from each category were selected for the cross-validation. In each fold, the micrographs were randomly split into 70% for training and the remaining 30% for testing. Five models were created in such a way that each model consisted of a training set of 28 micrographs (seven from each category) and a test set of 12 micrographs (three from each category), respectively. For each model, the classification score was recorded, and a confusion matrix was created. A summation of all the five confusion matrices is shown in Fig. 10.

Each micrograph was classified into a particular category, by the model, based on the highest percentage of droplets classified from that micrograph. In the five-fold cross-validation, the test sets of micrographs (12 in each model) in all the five folds (models) achieved 100% correct classification of their corresponding droplets. The green cells, in Fig. 10, represent the sum of the correct classification in each category obtained from the five models ($3 \times 5 = 15$). The blue cell at the bottom right-hand corner represents the overall sum of the correct classification ($12 \times 5 = 60$).

The models were further validated using a set of six micrographs (two from "U" category, two from "M," one from "A," and one from "T," respectively). These micrographs were obtained from a laboratory, in a different country, under varying illumination settings. All the five models successfully classified five out of the six micrographs into the correct category except the T one, which was classified as A. This gives an overall accuracy of 83%.

This article has investigated the application of microscopic image analysis, combined with a supervised machine-learning algorithm, for the classification of in-process emulsion samples. The HBT image segmentation approach demonstrated significant potential in the detection of droplets and their corresponding characteristics throughout the emulsification process. The HBT has provided a progressive evolution of decreasing droplet size and increasing droplet count as the emulsification process progressed. This was a very promising result compared to the

edge-based detection techniques reported in previous studies [33]. The majority of the existing conventional droplet analysis techniques such as laser diffraction and spectroscopy require time-consuming sample preparation of emulsion samples, which introduce additional time and complexity to industrial processes [10], [11]. Such techniques have also presented inconsistency in droplet size measurement. The current HBT approach provided consistent droplet size measurements and is entirely automated. It has also identified the completion of the emulsification process, when the droplets have attained their target characteristics. Any further processing was identified as over-processing.

The supervised classification model, in the current research, has demonstrated significant potential in classifying the micrographs at various stages of the emulsification process from unacceptable to target. The benefit of developing such supervised classification models using droplet characteristics has been identified as vital in food, pharmaceutical, and biomedical industries to meet the increasing demand for high-quality product [23]. There have been no systematic studies reported in the literature, which have investigated machine-learning models using in-process emulsion droplet data acquired through automated image segmentation.

In the current article, the validation of the machine-learning model presented 100% accuracy in the classification of 60 micrographs obtained from an emulsion manufacturing facility and 83% accuracy with an independent set of micrographs obtained from a different facility. The micrograph droplet characteristics and the classification were confirmed by the industrial collaborator using their existing validation technique. The approach reported in this article is developed as a soft-sensor, which supports the real-time deployment of the technique into an industrial environment.

IV. Conclusion

In this article, the histogram-based droplet detection and micrograph classification approach indicated that the time when the emulsification process is completed can be automatically determined from an emulsion micrograph. The HBT correctly characterized the droplet evolution throughout the emulsification process. Droplet characteristics such as count, area, perimeter, minimum, and maximum Feret diameters were identified as the emulsion quality indicators that vary with the process. The PC-LDA-based machine-learning approach presented 83% to 100% accuracy in the micrograph classification of the studied emulsion. Future work was planned to integrate the current methodology, implemented as a soft-sensor, with real-time image acquisition to execute in-line quality assessment in emulsion manufacturing. The proposed approach has significant potential in streamlining production, avoiding over-processing, and enabling efficient utilization of resources leading to efficient and sustainable emulsion manufacturing.

Acknowledgment

This research work was partially funded by the North West Centre for Advanced Manufacturing (NW CAM) project, which is supported by the European Union's INTERREG VA Programme and managed by the Special EU Programmes Body (SEUPB). The views and opinions in this document do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB). The authors wish to thank E. McHenry (GlaxoSmithKline, Sligo, Ireland) for providing the microscopic images for analysis. The support provided by GlaxoSmithKline is greatly acknowledged with appreciation.

References

[1] R. Panckow, L. Reinecke, M. Cuellar, and S. Maass, "Photo-optical In-situ measurement of drop size distributions: Applications in research and industry," *Oil Gas Sci. Technol.-Rev. IFP Energies Nouvelles*, vol. 72, no. 3, May/Jun. 2017, Art. no. ARTN 14. doi: 10.2516/ogst/2017009.

[2] A. J. Isaksson, I. Harjunkoski, and G. Sand, "The impact of digitalization on the future of control and operations," *Comput. Chem. Eng.*, vol. 114, pp. 122–129, 2018.

[3] M. J. Islam, S. M. Basalamah, M. Ahmadi, and M. A. Sid-Ahmed, "Computer vision-based quality inspection system of transparent gelatin capsules in pharmaceutical applications," *Amer. J. Intell. Syst.*, vol. 2, no. 1, pp. 14–22, 2012.

[4] N. Meijer, H. Abbes, and W. G. Hansen, "Particle size distribution and dispersion of oil-in-water emulsions: An application of light microscopy," *Amer. Lab.*, vol. 33, no. 8, pp. 28–31, Apr. 2001.

[5] B. Junker, "Measurement of bubble and pellet size distributions: Past and current image analysis technology," *Bioprocess Biosyst. Eng. Rev.*, vol. 29, no. 3, pp. 185–206, Aug. 2006.

[6] J. E. Gwyn, E. J. Crosby, and W. R. Marshall, "BIAS in particle-size analyses by count method," *Ind. Eng. Chem. Fundam.*, vol. 4, no. 2, pp. 204–208, 1965. doi: 10.1021/i160014a018.

[7] J. Kljusuric, M. Benkovic, and I. Bauman, "Classification and processing optimization of barley milk production using NIR spectroscopy, particle size, and total dissolved solids analysis," *J. Chemistry*, vol. 2015, 2015, Art. no. ARTN 896051, doi: 10.1155/2015/896051.

[8] S. Maaß, J. Rojahn, R. Haensch, and M. Kraume, "Automated drop detection using image analysis for online particle size monitoring in multiphase systems," *Comput. Chem. Eng.*, vol. 45, pp. 27–37, Oct. 2012, doi: 10.1016/j.compchemeng.2012.05.014.

[9] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, and S. E. J. Bell, "Recent applications of chemical imaging to pharmaceutical process monitoring and quality control," *Eur. J. Pharmaceutics Biopharmaceutics*, vol. 69, no. 1, pp. 10–22, 2008. doi: https://doi.org/10.1016/j.ejpb.2007.10.013.

[10] M. I. I. Z. Abidin, A. A. A. Raman, and M. I. M. Nor, "Review on measurement techniques for drop size distribution in a stirred vessel," *Ind. Eng. Chem. Res.*, vol. 52, no. 46, pp. 16085–16094, 2013.

[11] T. Vankeirsbilck *et al.*, "Applications of Raman spectroscopy in pharmaceutical analysis," *TrAC Trends Analytical Chem.*, vol. 21, no. 12, pp. 869–877, 2002.

[12] K. Hollingsworth, A. Sederman, C. Buckley, L. Gladden, and M. Johns, "Fast emulsion droplet sizing using NMR self-diffusion measurements," *J. Colloid Interface Sci.*, vol. 274, no. 1, pp. 244–250, 2004.

[13] S. Schuster *et al.*, "Analysis of W1/O/W2 double emulsions with CLSM: Statistical image processing for droplet size distribution," *Chem. Eng. Sci.*, vol. 81, pp. 84–90, 2012. doi: http://dx.doi.org/10.1016/j.ces.2012.06.059.

[14] I. Scherze, R. Knofel, and G. Muschiolik, "Automated image analysis as a control tool for multiple emulsions," *Food Hydrocolloids*, vol. 19, no. 3, pp. 617–624, May 2005. doi: 10.1016/j.foodhyd.2004.10.029.

[15] A. Hosseini, S. Jafari, H. Mirzaei, A. Asghari, and S. Akhavan, "Application of image processing to assess emulsion stability and emulsification properties of Arabic gum," *Carbohydrate Polymers*, vol. 126, pp. 1–8, 2015. doi: 10.1016/j.carbpol.2015.03.020.

[16] K. A. Silva, M. H. Rocha-Leão, and M. A. Z. Coelho, "Evaluation of aging mechanisms of olive oil–lemon juice emulsion through digital image analysis," *J. Food Eng.*, vol. 97, no. 3, pp. 335–340, 2010. doi: http://dx.doi.org/10.1016/j.jfoodeng.2009.10.026.

[17] M. G. Freire, A. M. A. Dias, M. A. Z. Coelho, J. A. P. Coutinho, and I. M. Marrucho, "Aging mechanisms of perfluorocarbon emulsions using image analysis," *J. Colloid Interface Sci.*, vol. 286, no. 1, pp. 224–232, Jun. 2005. doi: 10.1016/j.jcis.2004.12.036.

[18] M. Yang, "Measurement of oil in produced water," in *Produced Water: Environmental Risks and Advances in Mitigation Technologies*, K. Lee and J. Neff Eds., New York, NY, USA: Springer-Verlag, 2011, pp. 57–88.

[19] A. Khalil, F. Puel, Y. Chevalier, J. M. Galvan, A. Rivoire, and J. P. Klein, "Study of droplet size distribution during an emulsification process using in situ video probe coupled with an automatic image analysis," *Chem. Eng. J.*, vol. 165, no. 3, pp. 946–957, Dec. 2010. doi: 10.1016/j.cej.2010.10.031.
[20] J. A. Boxall, C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu, "Measurement and calibration of droplet size distributions in water-in-oil emulsions by particle video microscope and a focused beam reflectance method," *Ind. Eng. Chem. Res.*, vol. 49, no. 3, pp. 1412–1418, Feb. 2010. doi: 10.1021/ie901228e.
[21] D. J. McClements, "Critical review of techniques and methodologies for characterization of emulsion stability," *Critical Rev. Food Sci. Nutrition*, vol. 47, no. 7, pp. 611–649, 2007. doi: 10.1080/10408390701289292.
[22] J. Schindelin *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nat. Meth.*, vol. 9, no. 7, pp. 676–682, 2012.
[23] F. Jousse, "Modeling to improve the efficiency of product and process development," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 7, no. 1, pp. 175–181, Jan. 2008. doi: 10.1111/j.1541-4337.2007.00033.x.
[24] D. Shi, N. H. El-Farra, M. Li, P. Mhaskar, and P. D. Christofides, "Predictive control of particle size distribution in particulate processes," *Chem. Eng. Sci.*, vol. 61, no. 1, pp. 268–281, 2006.
[25] P. Wang, J. B. Sun, T. T. Zhang, and W. J. Liu, "Vibrational spectroscopic approaches for the quality evaluation and authentication of virgin olive oil," *Appl. Spectrosc. Rev.*, vol. 51, no. 10, pp. 763–790, 2016. doi: 10.1080/05704928.2016.1176034.
[26] S. Wold, E. Kim, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, Aug. 1987. doi: 10.1016/0169-7439(87)80084-9.
[27] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digit. Signal Process.*, vol. 20, no. 2, pp. 433–441, 2010.
[28] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *Proc. Int. Conf. Syst. Sci. Eng. Des. Manuf. Inf.*, 2010, pp. 27–30, vol. 1.
[29] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," in *Robust Data Mining*. New York, NY, USA: Springer-Verlag, 2013, pp. 27–33.
[30] F. R. Bertani *et al.*, "Classification of M1/M2-polarized human macrophages by label-free hyperspectral reflectance confocal microscopy and multivariate analysis," *Sci. Rep.*, vol. 7, Aug. 2017, Art. no. 8965. doi: 10.1038/s41598-017-08121-8.
[31] G. Venora, O. Grillo, and R. Saccone, "Quality assessment of durum wheat storage centres in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis system," *J. Cereal Sci.*, vol. 49, no. 3, pp. 429–440, May 2009. doi: 10.1016/j.jcs.2008.12.006.
[32] S. Unnikrishnan, J. Donovan, R. Macpherson, and D. Tormey, "Machine vision for the quality assessment of emulsions in pharmaceutical processing," in *Proc. Int. Conf. Universal Village*, Oct. 2018, pp. 1–6. doi: 10.1109/UV.2018.8642158. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp = &arnumber = 8642158&isnumber = 8642106
[33] S. Unnikrishnan, J. Donovan, R. Macpherson, and D. Tormey, "Machine learning for automated quality evaluation in pharmaceutical manufacturing of emulsions," *J. Pharmaceutical Innov.*, pp. 1–12, Apr. 16 2019, doi: 10.1007/s12247-019-09390-8.

Saritha Unnikrishnan received the B. Tech. (Hons.) degree in electrical and electronics engineering from the College of Engineering Trivandrum, University of Kerala, Kerala, India, in 1999.

She is currently a Lecturer with the Department of Computing and Electronics, Institute of Technology Sligo, Sligo, Ireland. After finishing her degree, she has worked as a Software Engineer in various multinational companies in India and Ireland. She joined the Precision Engineering Materials and Manufacturing Research group in 2015 as a Ph.D. Researcher in the area of computer vision and machine learning with a focus on pharmaceutical manufacturing applications. Her research interests include discriminative and generative machine learning algorithms, image segmentation, and pattern recognition.

Dr. Unnikrishnan is working as a National Delegate and management committee member of the EU funded COST Action 18206 (Glioma MR Imaging 2.0).

John Donovan (Member, IEEE) received the Honors degree in electronic engineering and the Ph.D. degree in reliability engineering from the University of Limerick, Limerick, Ireland, in 1982 and 1998, respectively.

Since 1993, he has been a Lecturer with the Institute of Technology, Sligo, Ireland. Prior to lecturing, he worked with Digital Equipment Corporation in Clonmel and Galway and held the position of Design Assurance Manager in the Wide Area Networks telecommunications division. His current research interests include process optimization, simulation modeling, and applied industrial statistics including the applications of machine learning.

Dr. Donovan was a recipient of the National Teaching Expert Award in 2015 by the National Forum for the Enhancement of Teaching and Learning in Higher Education, Ireland.

Russell Macpherson received the M.Eng. degree in electrical and electronic engineering and the Ph.D. degree in applied artificial intelligence from the University of Aberdeen, Aberdeen, Scotland, U.K., in 1999 and 2004, respectively.

He previously was with Arjo Wiggins Fine Papers, Aberdeen, Scotland, U.K., and QinetiQ, Farnborough, U.K. Since 2004, he has been with GlaxoSmithKline where he has worked across the manufacturing and R&D areas. He is currently a manufacturing site Director, based in Ireland. His research interests include advancing the use of artificial intelligence and analytics in real-world manufacturing and supply chain applications.

David Tormey received the graduation degree in manufacturing engineering in 1998 from the Galway Mayo Institute of Technology (GMIT), and M.Eng.Sc. and Ph.D. degrees in industrial engineering from the Computer Integrated Manufacture Research Unit (CIMRU), National University of Ireland, Galway, in 2000 and 2006, respectively.

He is currently a Senior Lecturer with the Department of Mechanical and Manufacturing Engineering, Institute of Technology Sligo, Sligo, Ireland, and Academic Director with the Institute's Centre for Precision Engineering, Materials and Manufacturing Research (PEM Research Centre). His previous positions held in IT Sligo include Head of Department of Mechanical and Electronic Engineering and Manager of the Institute's Centre for Design Innovation. Following graduation, he worked as a Manufacturing Engineer with Boston Scientific in Galway before joining IT Sligo. His current research interests include additive manufacturing and manufacturing process optimization.

Dr. David is a Funded Investigator and executive committee member of the I-Form Advanced Manufacturing Research Centre funded by SFI and Principal Investigator with the Northwest Centre for Advanced Manufacturing funded by Interreg VA.