

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



An SVM-Based Classification and Stability Analysis of Synthetic Emulsions Co-Stabilized by a Nonionic Surfactant and Laponite Clay

Abubakar A. Umar, Ismail M. Saaid and
Aliyu A. Sulaimon

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75707>

Abstract

Emulsions are metastable systems typically formed in the presence of surfactant molecules, amphiphilic polymers, or solid particles, as a mixture of two mutually immiscible liquids, one of which is dispersed as very small droplets in the other. These dispersions are unwanted occurrences in some areas, like those formed during crude oil production, but are also put into many other useful applications in the oil and gas industry, food industry, and construction industry, among others. These emulsions form when two immiscible liquids come together in the presence of an emulsifying agent and sufficient agitation strong enough to disperse one of the liquids in the other. Thermodynamically, these emulsions are unstable and thus would separate into their individual phases when left alone. To be stabilized, surface-active agents (surfactants) or solids (that act in so many ways like surfactants) ought to be used. Like many commercially available products, several pharmaceutical products are usually supplied in the form of emulsions that must be stabilized before they are being administered. Pharmaceutical emulsions used for oral administration either as medications themselves or as carriers come in form of stable emulsions. Either water-in-oil (w/o) or oil-in-water (o/w), these emulsions after formulation must be classified, majorly as stable or unstable. Only formulations that give stable emulsions are used, and the unstable ones reformulated or discarded. Classifying such emulsions using results obtained by visual observation in most cases can be very tedious and inaccurate. This necessitates the use of a more scientific and intelligent method of classification. The objective of this study is to employ support vector machine (SVM) as a new technique to classify synthetic emulsions. The study will assess the effects of nonionic surfactant (sodium monooleate) and Laponite clay (LC) on the stability of synthetic emulsions prepared using a response surface methodology (RSM) based on a Box-Behnken design. The stability of the emulsions was measured using batch test and TurbiScan, and the SVM was used to classify the emulsions into stable, moderately stable and unstable emulsions.

The study showed that an increase in surfactant concentration in the presence of moderate to high concentrations of LC can provide a stable emulsion. Also, a clear classification of the emulsion samples was provided by the SVM, with high accuracy and reduced misclassifications due to human error. A higher accuracy in classification would reduce the risk of using the wrong formulation for any pharmaceutical product.

Keywords: water-in-oil emulsions, oil-in-water emulsions, multiple emulsions, Pickering emulsions, support vector machine, response surface methodology

1. Introduction

Several processes that involve choice of materials and operational decisions are required in the formation of an emulsified system. One is tasked with selecting the type of emulsion that is needed (whether a water-in-oil or an oil-in-water), what types of surface-active-agents and how much of them are required to form the emulsion, what level of stability is required and so on. These are, no doubt, some of the most important operational decisions required when handling an emulsified system.

The process via which emulsion droplets are stabilized is achievable either by small molecular weight surfactants that reduce the interfacial tension between the two fluids or amphiphilic macromolecules (like proteins and polysaccharides) by the formation of steric elastic film as well as the reduction of interfacial tension. The process by which these surfactants stabilize emulsions has been widely studied. Surfactants form a unique class of chemical compounds, and their widespread applications in emulsification and other industrial processes have advanced a wealth of published literature [1]. Lecithin from egg yolk and various proteins from milk are some naturally occurring surfactants used in the food industry for the preparation of food products like mayonnaise, salad creams and so on [2]. Most of these compounds, like short-chain fatty acids, have a part of them with affinity for the nonpolar hydrocarbon chain and one the other part that has its affinity for polar group such as water. These compounds are referred to as amphiphilic or amphipathic. The most satisfactory orientation these molecules can assume is at the interface, so that each part of the molecule can stay in the phase (polar or non-polar) which it has the greatest affinity [3].

Apart from natural or artificial surfactants, dispersed colloidal particles were discovered to function as emulsion stabilizers in a fundamentally unique way, and this concept was formally recognized since the publication of Pickering. The knowledge that fine solid powders can stabilize emulsions dates to centuries ago. Clayton [4] reported that emulsions of oil and water were prepared with North-African argillaceous sand in 1898, and in 1903, Ramsden [5] concluded that the stability of many emulsions could be attributed, in part, to “the presence of solid or highly viscous matter at the interfaces of the two liquids.” Pickering performed the first extensive experimental study in 1907 [6] in connection with plant sprays [7].

The word “emulsion” is used very frequently in identifying both microemulsions and macroemulsions. The term has been defined severally by experts in different areas of its application. Emulsion according to Manning and Thompson [8] was defined as a quasi-stable suspension

of fine drops of one liquid in another liquid. However, Roberts [9] defined emulsion as a system containing two liquid phases, one of which is dispersed as globules in the other. Other researchers [10–12] defined emulsion as a mixture of two mutually immiscible liquids, one of which is dispersed as very small droplets in the other, and is stabilized by an emulsifying agent. According to Leal-Calderon and Schmitt [13], emulsions are metastable systems typically formed in the presence of surfactant molecules, amphiphilic polymers or solid particles. The relative balance of the hydrophilic and lipophilic properties of these emulsifiers is known to be the most important parameter dictating the emulsion type: oil-in-water (O/W) emulsions are preferentially obtained with molecules which are rather hydrophilic whereas water-in-oil (W/O) emulsions are produced in the presence of hydrophobic molecules. A close look at all the definitions provided above indicates that there is a common understanding and belief that emulsions are thermodynamically unstable [11]. By thermodynamic instability, it means that the contact between the oil and water molecules is unfavorable, and so they will always break down over time, leading to decrease in free energy. Based on the sizes of their dispersed droplets, these dispersions can either fall under macroemulsions or microemulsions. Macroemulsions are liquid-in-liquid dispersions with droplet size ranging usually from 1 to 100 μm (and can sometimes be extended down to 0.5 or up to 500 μm). This range of droplet sizes is in general large enough to allow settling due to gravity influence. Microemulsions are single-phase systems that are thermodynamically stable. According to Nielloud [14], many microemulsions would not qualify to be called dispersions of very small droplets, but reasonably as percolated or bicontinuous structures in which there is no dispersed nor continuous phase, and no probability of dilution as in normal emulsions. According to Lambert et al. [15], microemulsions can be defined as thermodynamically stable, isotropically clear dispersion of two immiscible liquids, like oil and water, stabilized by an interfacial film of surfactant molecules. The microemulsion has a mean droplet diameter of less than 200 nm, in general between 10 and 50 nm.

For emulsions to form, three conditions must be satisfied [17]: (1) the two liquids forming the emulsion must be immiscible; (2) there must be sufficient agitation to disperse one liquid as droplets in the other, and (3) the presence of an emulsifying agent. Lecithin from egg yolks or soybeans is a commonly used surfactant [15]. Pharmaceutical products can be stabilized by the addition of various amphiphilic molecules, including anionic, nonionic, cationic, and zwitterionic surfactants. The amphiphilic molecules in addition comprise of surfactants such as ascorbyl-6-palmitate, stearylamine, sucrose fatty acid esters, various vitamin E derivatives, and so on. One or a combination of these surfactants can be used to stabilize the emulsion, and excipients are added to render the emulsion more biocompatible, stable and less toxic [15]. In addition to surfactants, solids have been widely used as stabilizers of emulsions, a process called Pickering Stabilization [5].

These colloidal particles perform in several ways like surfactant molecules, mostly if adsorbed to a fluid-fluid interface [16]. The same way a surfactant's oil or water-liking tendency is defined by the hydrophilic-lipophilic balance (HLB) [18], so are spherical particles defined with respect to their wettability via contact angle, as shown in **Figure 1**.

There exists some important dissimilarities between these two types of surface-active solids, partially as a result of how they are held at the oil-water interface [4, 18, 19]. Hydrophilic particles

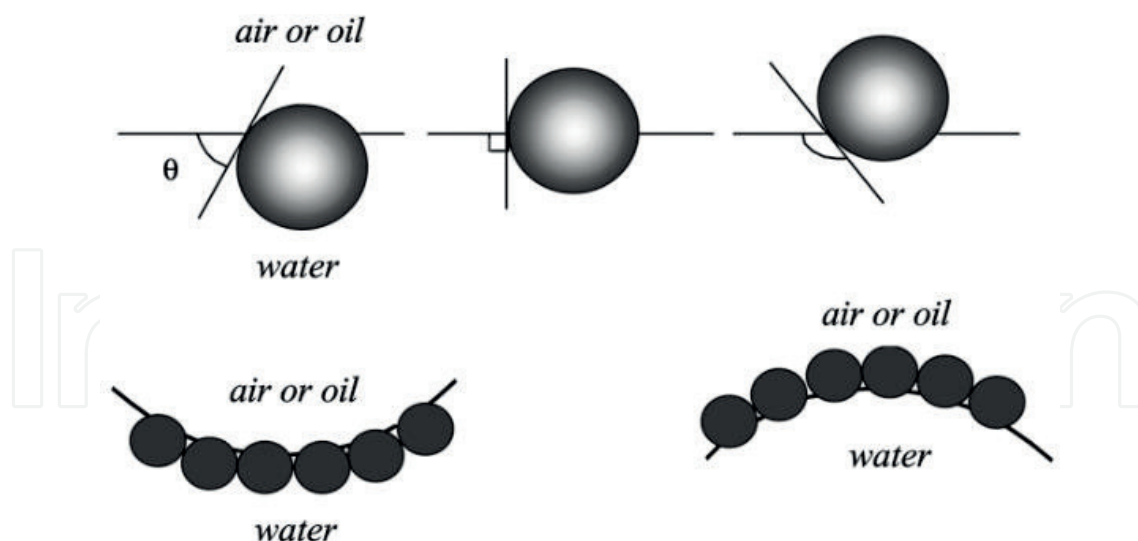


Figure 1. Position of a small spherical particle at w/o interface [16].

have the tendency of forming oil-in-water (o/w) emulsions while hydrophobic particles form water-in-oil (w/o) emulsions. Many of the properties can be attributed to the very large free energy of adsorption for particles of intermediate wettability (contact angle at the oil-water interface, say, between 50 and 130°) [19]. This adsorption of solids at the fluids interface is efficiently irreversible and leads to extreme stability for certain emulsions.

In this study, the effects of Laponite clay (a colloidal particle) and a nonionic surfactant (sorbitan monooleate) on the stability of synthetic emulsions were investigated. The synthetic emulsions were further classified using a kind of typical machine learning method, support vector machine (SVM), designed and developed to classify the synthetic emulsions formulated based on the response surface methodology (RSM). According to Hu et al. [20], the novel SVM algorithm has its origin from Vapnik [21]. Together with Cortes, Vapnik suggested the modified maximum margin idea that allows for mislabeled examples [22]. Support vector machine (SVM) is a supervised machine learning algorithm used both in classification and regression-related challenges, though it has higher applications in classification problems. In an SVM algorithm, each data item is plotted as a point in n -dimensional space (n -representing the number of features to be classified) with the value of each feature being the value of a particular coordinate. Then, classification is performed by finding the hyper-plane that differentiates the two classes very well [23]. The primary focus while drawing the hyperplane is on maximizing the distance from hyperplane to the nearest data point of either class. The drawn hyperplane called as a maximum-margin hyperplane [21, 24]. For linearly separable data, two parallel hyperplanes that separate the two classes of data are chosen, so that distance between both the lines is the maximum. The region between these two hyperplanes is known as “margin” and the maximum margin hyperplane is the one that lies in the middle of them, as shown in **Figure 2**. Details of the algorithm and derivations are beyond the scope of this paper. For those interested in understanding the details, the following references would suffice [21, 22, 25, 26].

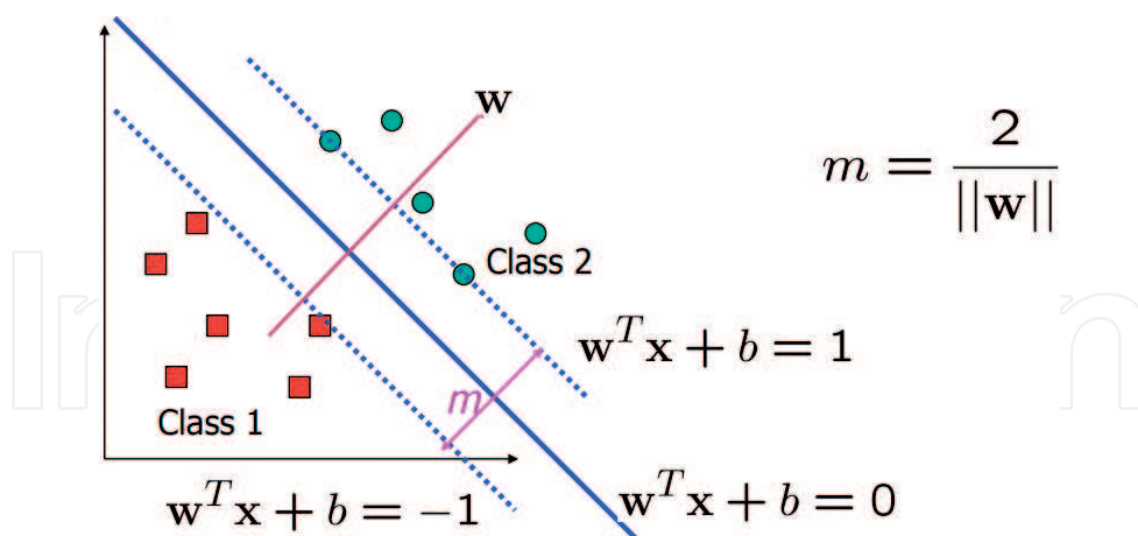


Figure 2. A depiction of the decision boundary showing the maximum margin m and other parameters [27].

The support vectors defined as $w^T + w_0 = \pm 1$ are the training vectors that are the closest to the linear classifier, and they constitute the critical elements of the training set.

Support vector machines, Bayes point machines, kernel principal component analysis and other Kernel-based techniques represent a key advancement in machine learning algorithms. Support vector machines (SVM) are a group of supervised learning methods that can be used in classification or regression [28]. Industrial processes that require classifications have utilized the emergence of rapid development in artificial intelligence like neural networks to solve classification problems [20]. As in many industries, pharmaceutical emulsions require accurate classification in order to avoid waste of materials during formulation or using an unstable emulsion where a stable one is required.

2. Methodology

2.1. Materials and methods

To accomplish the set objectives of this study, Castor oil supplied by Merck Malaysia in its pure form (99%) was used as the oleic phase, while deionized water obtained from a PureLab Flex 3 purifier as the internal phase. Laponite clay and Sorbitan monooleate were used as emulsifiers, in the form they were supplied by Avantis Chemicals Bhd Malaysia. The batch of the Laponite clay used in this study is the 9/4156 batch, with chemical compositions expressed as wt%: Li_2O ; 0.8, Na_2O ; 28, SiO_2 ; 59.6 and MgO ; 27.4. A cross polarized microscopy was used to measure the droplet size of the emulsions prepared. The CPM provides us a unique window into the internal structure of crystals and at the same time is esthetically pleasing due to the colors and shapes of the crystals.

2.2. Surface response methodology based on box: Behnken design

Response surface methodology is a collection of statistical and mathematical methods that are useful for modeling and analyzing engineering problems. In this technique, the main objective is to optimize the response surface that is influenced by various process parameters [29]. It is indispensable that experimental design methodology is a cost-effective way for mining the maximum amount of multifaceted data, a weighty experimental time saving factor and likewise, it saves the material used for analyses and personal costs as well [29–31]. An experiment is a series of tests in which the input variables are changed according to a given rule in order to identify the reasons for the changes in the output response [32]. Such tasks that require investigating the effects of different variables on one or more outputs can be tedious and may be accompanied with different kinds of errors.

This experiment was designed using a Statgraphic Centurion XVII a flagship data analysis and visualization software. It encloses 32 statistical procedures and significant upgrades to 20 other existing procedures. As shown in **Tables 1** and **2**, 15 castor oil synthetic emulsions were prepared using different compositions of Laponite clay (0.1–0.3% w/w), Span80 (0.5–1.5% v/v), and deionized water. The emulsifier compositions were prepared based on Box-Behnken design for the estimation of emulsion stability (as amount of water released from the emulsions). Analysis of variance (ANOVA) and regression surface analysis were conducted to determine the statistical significance of model terms and fit a regression relationship relating the experimental data to the independent variable.

Runs	Clay% (w/w)	Span80% (v/v)
1	0.2	1.0
2	0.3	1.0
3	0.2	0.5
4	0.1	1.0
5	0.2	1.5
6	0.3	0.5
7	0.2	1.5
8	0.1	1.5
9	0.2	1.0
10	0.3	1.0
11	0.2	1.0
12	0.3	1.5
13	0.2	0.5
14	0.1	0.5
15	0.1	1.0

Table 1. Box-Behnken design with actual values for three size fractions and results.

Variables	Symbol	Actual variable levels		
		Low	Center	High
Weight concentration of Laponite clay (%)	X1	0.1	0.0	0.3
Volume concentration of Span80 (v/v)	X2	0.5	0.0	1.5

Table 2. The level of variables chosen for the Box-Behnken design.

Assuming that the variation of Y (STABILITY) obeys an eight parameter, second-order equation of the following type:

$$STABILITY = \beta_0 + \sum \beta_i x_i + \sum \beta_{ii} x_i^2 + \sum \beta_{ij} x_i x_j \quad (1)$$

where Y (STABILITY) is the response value predicted by the model; β_0 is an offset value; $\beta_i, \beta_{ii}, \beta_{ij}$ are linear, quadratic and interaction regression coefficients, respectively. The competence of the models was determined using model analysis; lack-of fit test and coefficient of determination (R^2) analysis. Joglekar [31] recommended that R^2 should be at least 0.80 for a good fitness of a response model. The corresponding variables will be more significant ($p < 0.05$), if the absolute t value becomes larger and the p-value becomes smaller [33]. For all terms statistically found non-significant ($p > 0.05$), they would be dropped from the initial models and the experimental data refitted only to significant ($p < 0.05$) independent variable effects in order to obtain the final reduced model.

2.3. Preparations of synthetic emulsions

Emulsions used in this study were prepared using a w/o ratio of 30/70 (v/v). Deionized water was used as the aqueous phase and castor oil was used as the oleic phase. For every 28 mL of the oil phase, an equivalent 12 mL of the aqueous phase was added, into a 50 mL plastic centrifuge bottle with a Wheaton adjustable pipet. Before the two phases were mixed, different wt% and v/v% concentrations of the Laponite clay (LC) and sorbitan monooleate (Span80) were dispersed in the oil phase with a homogenizer for 1 min. The oil-phase which now contains the LC and Span80, and the deionized water was then mixed with a Virtis Virtishear Cyclone IQ Homogenizer. The homogenization was performed at 15,000 rpm for 5 min. All the 15 emulsion samples were prepared based on the RSM design.

2.4. Support vector machine classification

The emulsions prepared were investigated of the percent water released, and based on that were preliminarily classified into stable, moderate stable and unstable emulsions. This scheme is applied for the optimization of the sparse coefficient matrix. In each case, there was a repetition of the analysis over all the training set where one sample was left out for testing. The training group was used in building the SVM classifier while the performance of the classifier was calculated by the testing group. The important terms used to evaluate the performance of the classifier are accuracy, sensitivity and specificity. A confusion matrix, which comprises

of actual and predicted classifications, is usually used in the performance evaluation. The different components of a confusion matrix that are used in the performance evaluation are: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Another important parameter used in the performance evaluation and also to select the optimum model is the area under the curve (AUC), also called the receiver operating characteristics (ROC). It is obtained by plotting the TPR against the FPR at different settings. Another commonly used metric is the rate of detection (RoD) also known as the Sensitivity or Recall. Equal error rate (EER) or crossover rate (COR) is the percentage of misclassified frames when the acceptance and rejection errors are equal, that is, $FPR = FNR$.

3. Results and discussions

3.1. Emulsion stability measurements: bottle test

Emulsion stability measurement by bottle test is the most popular technique employed in determining the stability of emulsions. It measures the amount of water resolved from the emulsion over time. In this study, it is employed to assess the stability of the prepared emulsions at 60°C. This phenomenon is controlled by the gravity separation, where the amount of water released is being observed with time, and used as a measure of the stability. All prepared emulsions were kept in a graduated plastic centrifuge bottle in a water bath of 60°C, and aged for 60 h. The percentage of water separated was calculated, and plotted against time (Tables 1 and 2). The results are as shown in Figures 3 and 4.

From Figures 3–7, the trend observed is that of varying water separation percent over the entire period of this study. Figure 3 indicates that Emulsion B with 0.3 wt% of LC and 1.0 v/v% of Span80 is the most stable, having released barely 10% of its emulsified water over the entire study period. As the Span80 and LC concentrations increase, the extent of emulsified water release decreases and the dispersed droplet size correspondingly decreases. Although sorbitan monooleate (Span80), a nonionic oil soluble surfactant has the ability to stabilize emulsions on its own, a synergy between the two emulsifiers has shown a higher stability being achieved.

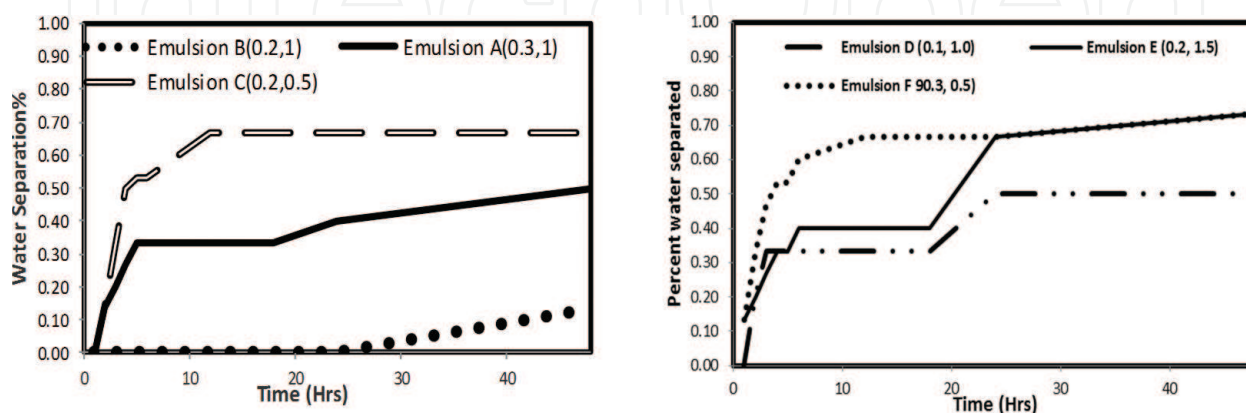


Figure 3. Water separation percent for emulsions A to F observed at 60°C for 48 h.

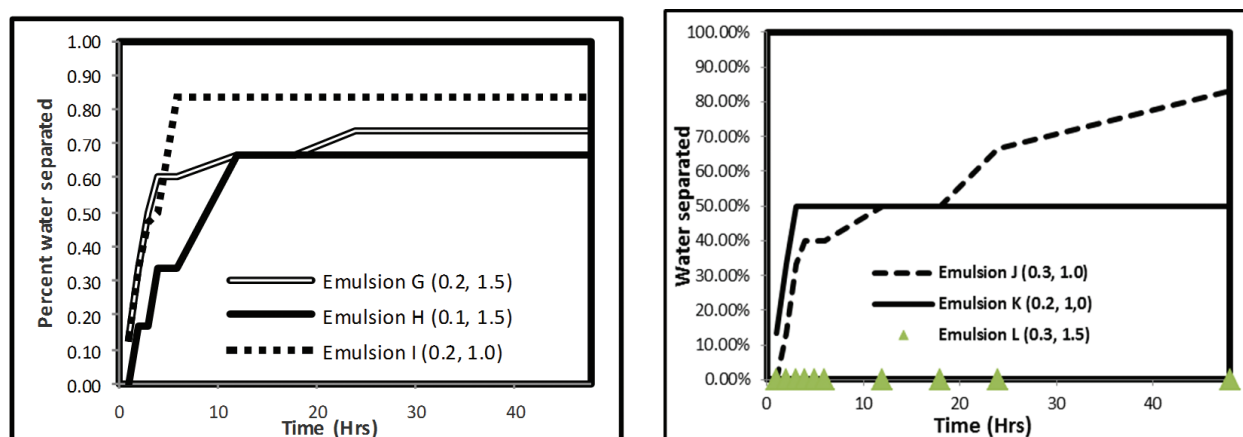


Figure 4. Water separation percent for emulsions G to L observed at 60°C for 48 h.

This is evident from Emulsion B which though has the same concentration of Span80, but a lower LC concentration of 0.2 wt% indicated a wide range of difference between water released by the two emulsions under comparison. This trend also applies to Emulsion C, which has equal concentration of LC as Emulsion B, but lower Span80 concentration, and subsequently lower stability with the highest water release of almost 70% within the time under study.

The Laponite clay discs are believed to play a critical role in the nucleation stage of the Pickering emulsion polymerization process. The use of increasing amounts leads to smaller average particle sizes but inflicts longer nucleation periods [34]. This same trend observed in **Figure 1** applies through **Figure 4**. Emulsion L in **Figure 4** indicates the most stable emulsion, having released 0% throughout the study time. This is due to the high concentrations of both Span80 and LC used in its preparation. Emulsions H and K exhibit somewhat similar behavior. As observed, Emulsion H released around 70% of its emulsified water in 12 h, and there was no visible increase till the end of this study. This is similar to what obtains in Emulsion K; where it released around 50% of its emulsified water within 5 h with no further increase in release. In both cases, higher concentrations of Span80 were used with relatively lower concentrations of LC. This could be as a result of the irreversible nature of particle adsorption at oil/water interfaces, where it is rapid at the beginning and ceases completely over a long period of time [35].

3.2. Design data analysis

In this section, we present the estimated regression coefficients for the response variable (Y-Stability) together with the corresponding R^2 , F-value and p-value of lack of fit. The response, Y was evaluated as a function of main, cubic and interaction effects of Laponite clay (X1 or A), Span80 (X2 or B) and time taken for separation to occur (X3 or C). The individual significance F-value and p-value of independent variables are presented in **Table 3**. The ANOVA table segregates the variability in STABILITY into separate pieces for each of the effects. It further tests the statistical significance of each effect by comparing the mean square against an estimate of the experimental error. In this case, three effects have P-values less than 0.05, indicating that they are significantly different from zero at the 95.0% confidence level. The R-Squared statistic

Source	Sum of squares	Df	Mean square	F-Ratio	P-Value
A:Laponite clay (wt%)	0.216234	1	0.216234	5.63	0.0190
B:Span80 (v/v%)	0.350053	1	0.350053	9.12	0.0030
C:Time (Hours)	1.49966	1	1.49966	39.07	0.0000
AA	0.0707332	1	0.0707332	1.84	0.1768
AB	0.330028	1	0.330028	8.60	0.0039
AC	0.0155063	1	0.0155063	0.40	0.5261
BB	0.097644	1	0.097644	2.54	0.1130
BC	0.00746267	1	0.00746267	0.19	0.6599
CC	0.867603	1	0.867603	22.60	0.0000
Total error	5.33539	139	0.0383841		
Total (corr.)	9.53095	148			

Table 3. Analysis of variance for stability index.

indicates that the model as fitted demonstrates 44.0204% of the variability in STABILITY INDEX. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 40.3958%. The standard error of the estimate shows the standard deviation of the residuals to be 0.195919. The mean absolute error (MAE) of 0.147747 is the average value of the residuals.

The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 5.0%, there is an indication of possible serial correlation at the 5.0% significance level. Plot the residuals versus row order to see if there is any pattern that can be seen. **Figure 5** shows the Pareto chart displaying all the variables of the general model. Pareto analysis is a statistical procedure that seeks to discover from an analysis of defect reports or customer complaints which “critical few” causes are responsible for most of the reported problems. The old adage states that 80% of reported problems can usually be traced to 20% of the various underlying causes [29, 31, 33].

From **Figure 5**, it shows that all the variables in the study have significance on emulsion stability index, albeit at different level. The time it takes the emulsions to release their emulsified water has the highest effect on the stability index, then followed by the square of its value, Span80, a mixture of span80 and Laponite clay and finally Laponite clay. Those parameters with standardized effect below 2 are AB, AA, AC and BC. These are statistically ineffective on this model and therefore are removed and the pareto chart recomputed, as shown in **Figure 6**.

From the modified pareto chart, only those variables with significant effect are shown. These are the variables that are statistically significant on emulsion stability index studied. The longer

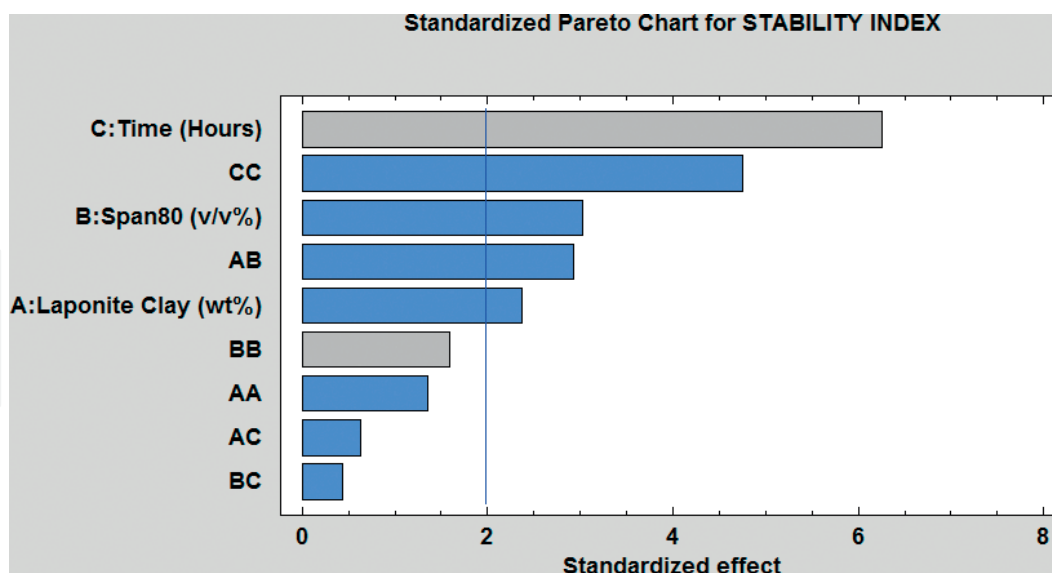


Figure 5. Pareto chart showing all the factors in the initial model.

the horizontal bars on the pareto Chart, the higher the significance. This is again confirmed in **Table 3**. For all terms statistically found significant, their p-values should be less than 0.05. All terms with $p > 0.05$ are statistically non-significant.

From **7A**, an increase in the concentrations of span80 and LC leads to an increase in the stability index. The effects of both LC and Span80 with time on stability index were shown on **Figure 7B** and **C**. Both response plots indicate that there is a concentration value for both emulsifiers that may tend to decrease the stability index, indicated by the curved surface on the plot. Also, the longer the time the studied emulsions stay, the more water release from them, also indicated by the curved surface on the response plot.

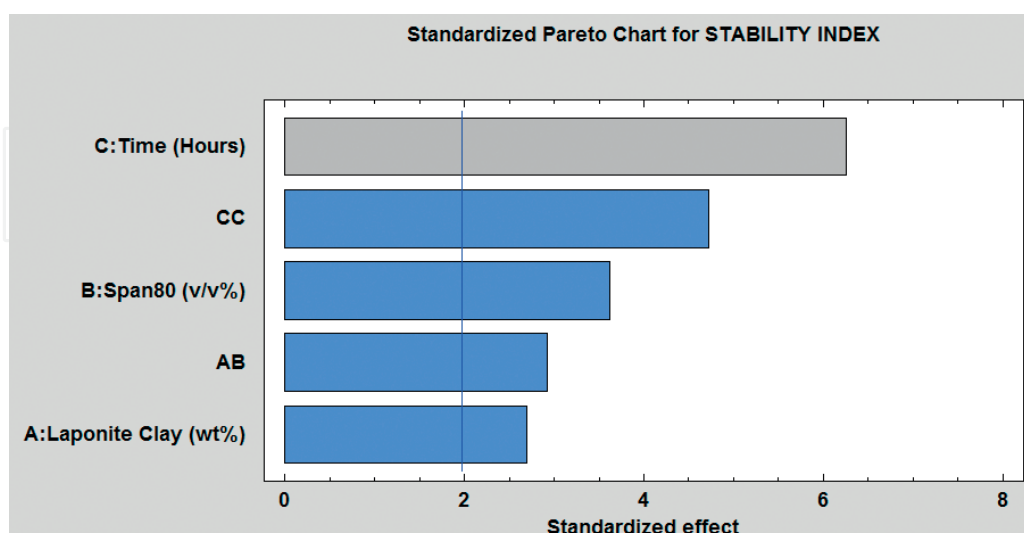


Figure 6. Pareto chart showing all the factors in the modified model.

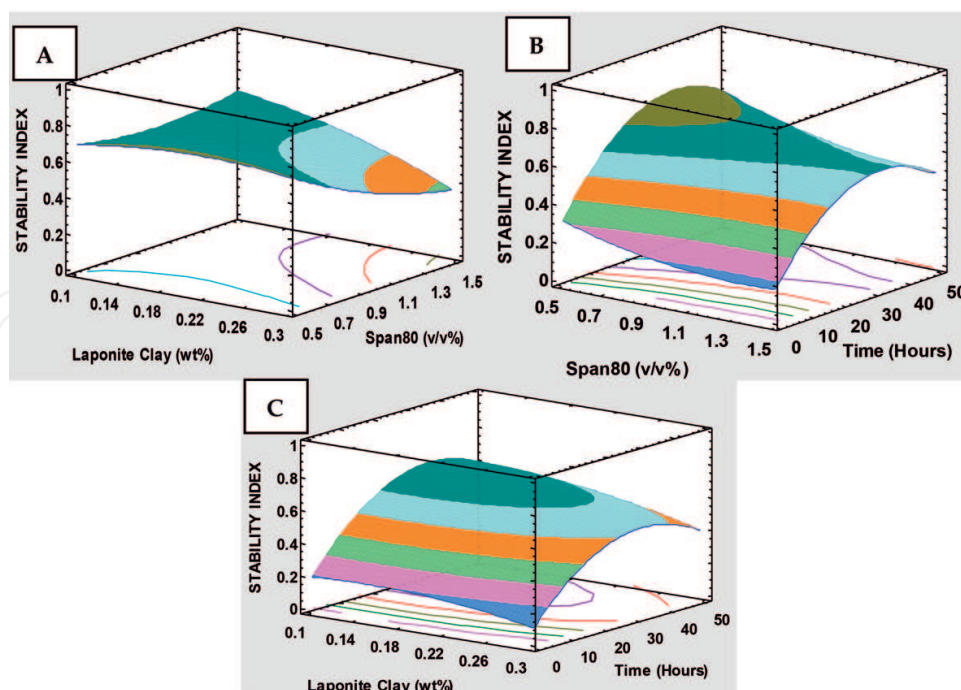


Figure 7. Response surface plots showing the effects of (A) LC and Span80 on stability index, (B) Span80 and time on stability index and (C) LC and time on stability index.

For both pareto charts shown in **Figures 5** and **6**, the equations of the fitted models are given by Eqs. (2) and (3) where the values of the variables are specified in their original units. In the equations, Laponite clay is represented by (X_1), Span80 is represented by (X_2) and Time is represented by (X_3)

$$\begin{aligned} \text{StabilityIndex} = & 0.0925421 + 3.10227 * X_1 - 0.190561 * X_2 \\ & + 0.0305891 * X_3 + 0.205924 * X_2^2 \\ & - 0.00139185 * X_2 X_3 - 0.000399451 * X_3^2 \end{aligned} \quad (2)$$

Equation (2) contains variables that are not significant statistically in the model. Therefore, these variables were excluded from the model and refitted, and the outcome is Eq. (3)

$$\begin{aligned} \text{StabilityIndex} = & 0.455345 - 0.595982 * X_1 - 0.142047 * X_2 \\ & + 0.0284712 * X_3 - 0.00139185 * X_2 X_3 - 0.000398865 * X_3^2 \end{aligned} \quad (3)$$

These two equations can be used to predict the emulsion stability index of the emulsions studied.

3.3. SVM classification

As earlier discussed, this study went further to classify the emulsions studied into three classes: stable, moderately stable and unstable. Those emulsions that after the period of study have released 0–20% of their emulsified water were labeled stable. Those emulsions that have released above 20% up to 40% were labeled moderately stable. The last class is those emulsions

that have released above 40% of their emulsified water up to 100%. The support vector machine (SVM) was used to make the classification, and some terms are used in evaluating the performance of the classifier. These terms are as follows: accuracy, sensitivity and specificity. A confusion matrix, which comprises of actual and predicted classifications, is usually used in the performance evaluation. In this study, five different kernels were used in this classification and the kernel that best classifies the emulsions was reported.

From **Table 4**, it can be seen that almost all the classifiers have good performance in terms of overall accuracy. Both cubic and medium Gaussian kernels provide up to 94.0% overall accuracy. In this study, we reported the cubic kernel. We will now define all the terms used in evaluating the performance of the SVM classifier.

The accuracy, denoted by ACC is given as the total number of correct predictions. Mathematically, it is written as:

$$ACC = \frac{TP + FN}{TP + FP + TN + FN} \quad (4)$$

The true positive rate (TPR) also known as “sensitivity” is the portion of positive classes that were correctly identified, as given below:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

The true negative rate (TNR) also called “Specificity” is the proportion of negative cases that were classified correctly, as given below:

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

Finally, the precision, also called positive predictive value (PPV), is the proportion of the predicted positive cases that were correctly classified, as given in Eq. (7):

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

S/No	Kernel type	Overall accuracy (%)	Overall error (%)
1	Linear SVM	92.0	8.0
2	Quadratic SVM	93.3	6.7
3	Cubic SVM	94.0	6.0
4	Fine Gaussian SVM	86.7	13.3
5	Medium Gaussian	94.0	6.0
6	Coarse Gaussian	84.0	16.0

Table 4. Comparison of different kernels used in the classification.

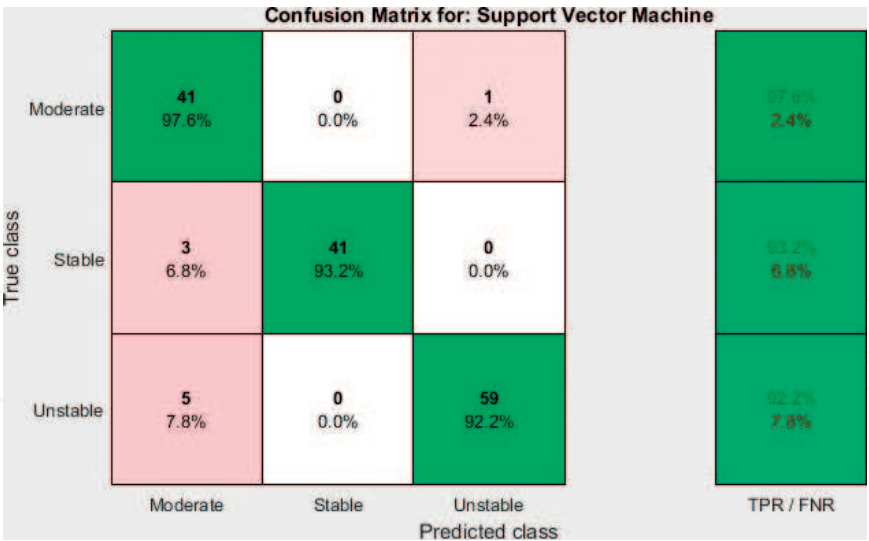


Figure 8. Confusion matrix for cubic SVM.

The next performance evaluation term used to assess the performance of the classifier is the receiver operating characteristics curve (ROC) (Figures 8 and 9). The most recurrently used performance measure removed from the ROC curve is the value of the area under the curve, normally symbolized as AUC. An AUC of 1 indicated that the classifier achieves perfect accuracy if the threshold is accurately chosen, and a classifier that predicts the class at random has an associated AUC of 0.5. Another remarkable point of the AUC is that it portrays a general behavior of the classifier since it is independent to the threshold used for obtaining a class label. Figures 10 and 11 present the AUC graphs for all the classes of emulsions. A quick look at the AUCs shown in Figure 10 indicates how close both areas are to unity, indicating a near

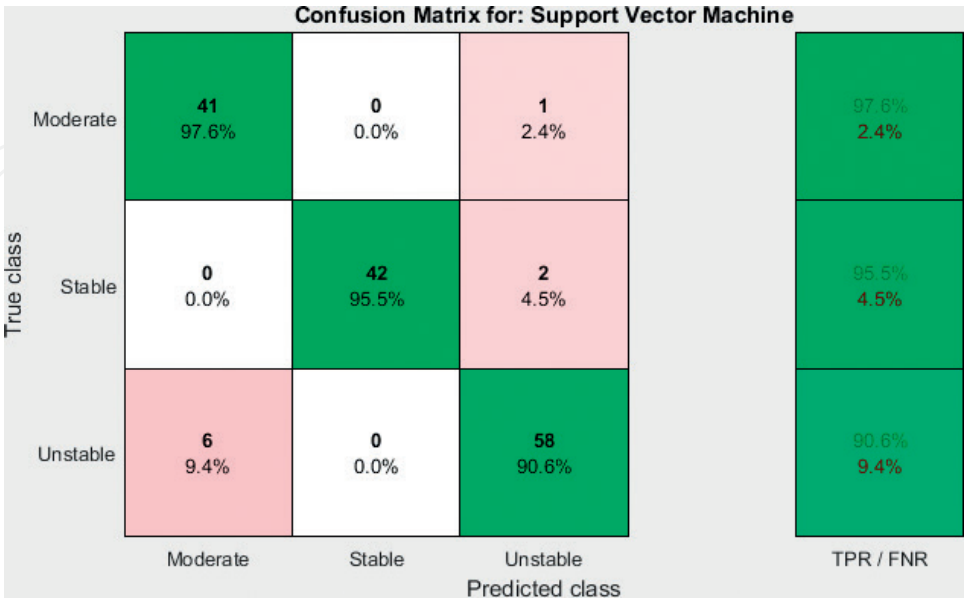


Figure 9. Confusion matrix for medium Gaussian SVM.

perfect classification of the stable emulsions by both cubic SVM and medium Gaussian SVM kernel. **Figures 11** and **12** also show the AUCs for the moderate and unstable emulsions from both kernels under comparison.

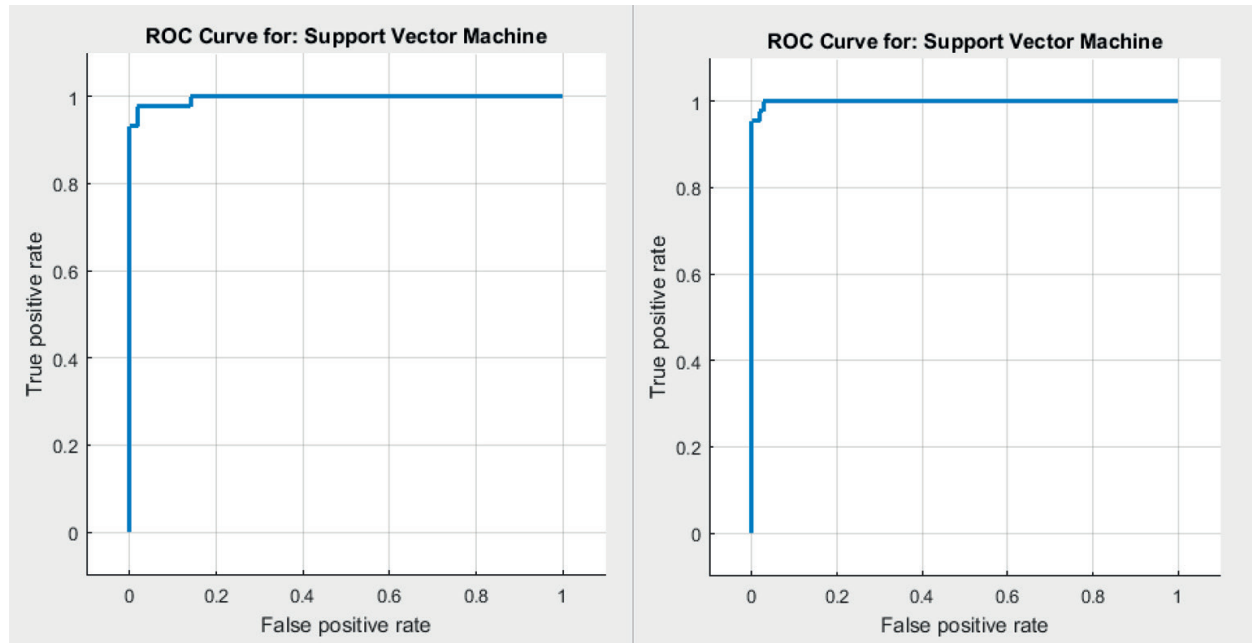


Figure 10. Area under the curve for a stable emulsions classified by (left) cubic SVM (99.59%) and (right) medium Gaussian SVM (99.89%).

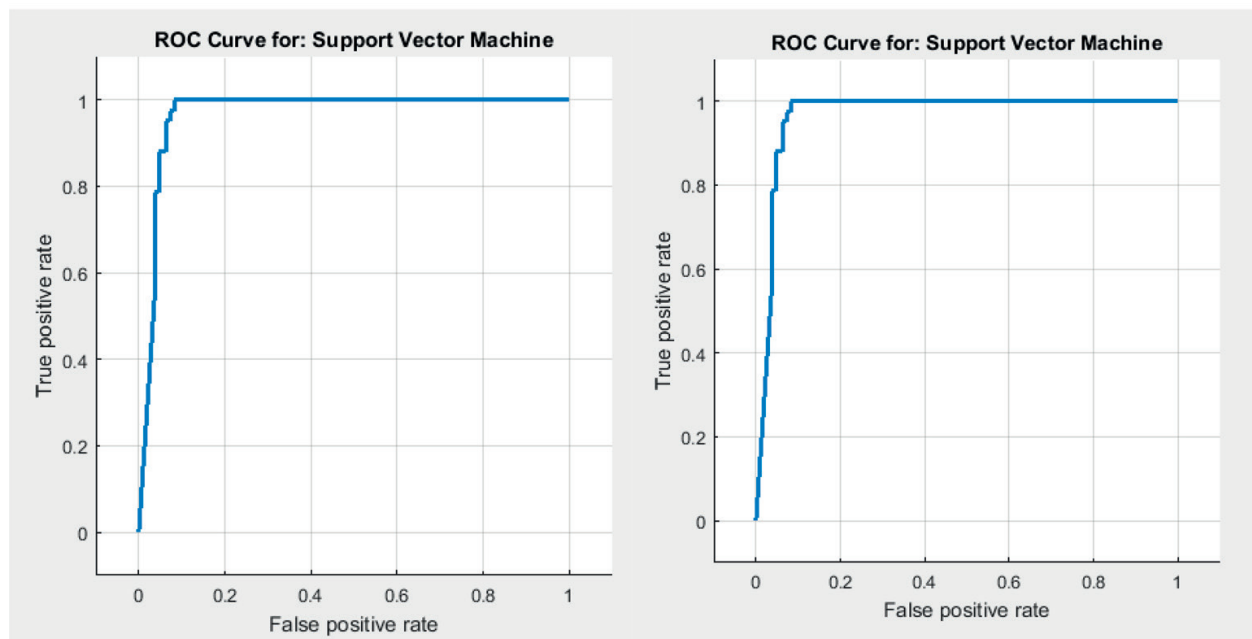


Figure 11. Area under the curve for a moderately stable emulsions classified by (left) cubic SVM (96.83%) and (right) medium Gaussian SVM (96.37%).

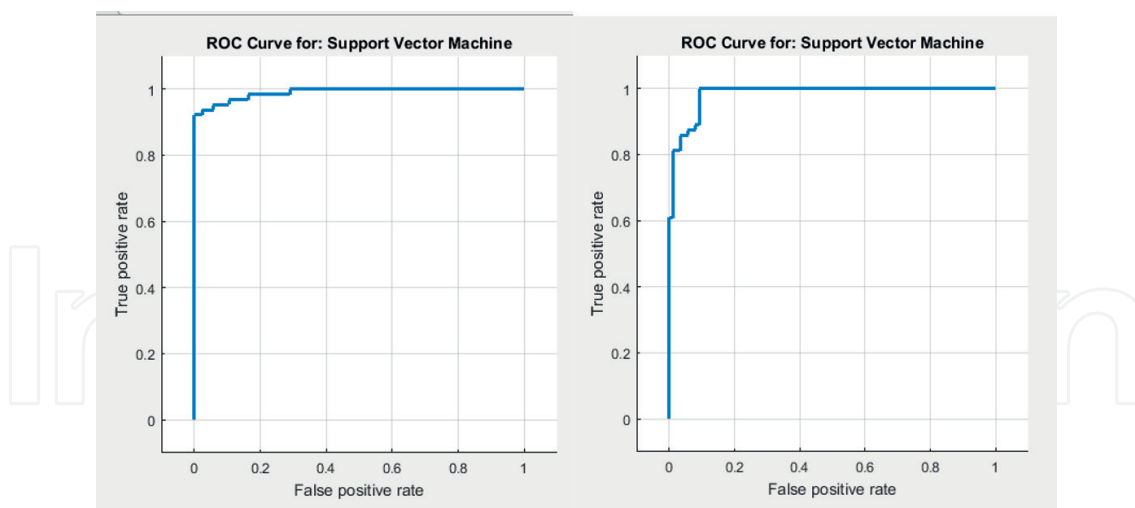


Figure 12. Area under the curve for an unstable emulsion classified by (left) cubic SVM (99.00%) and (right) medium Gaussian SVM (98.69%).

4. Conclusion(s)

The studies of the stability of emulsions, either where emulsion is an unwanted process occurrence (like during the production of crude oil), or where it is wanted (like many food or pharmaceutical companies), require understanding the parameters that determine its stability as well as the classes under which emulsions can fall is very important and critical. In this study, the effects of Laponite clay and sorbitan monooleate were investigated, and the resulting emulsions were classified using support vector machine. In this study, synthetic emulsions were prepared using castor oil as the oleic phase and deionized water as the dispersed aqueous phase. Fifteen synthetic emulsions were formulated, and their stabilities analyzed based on the Box-Behnken design of response surface methodology. Due to the inaccuracies of misclassifications normally encountered during conventional classifications, support vector machine (SVM) was used to classify the emulsions. This study has revealed that the RSM is a valuable tool that can be used to appraise the combined effect key variables in the formulation of stable emulsions. The stability of these emulsions is believed to be synergized by the steric effect originating from the Laponite clay particles adsorbed at the oil/water interface and the surfactant, which is form a thin film at the o/w interface. Although Span80 would normally stabilize emulsions as investigated by many researchers, the synergy in stabilization between the nonionic surfactant and Laponite clay has been elucidated in this study. An SVM classifier with different kernels was used to classify the emulsions studied, and two kernels (cubic and medium Gaussian SVM kernels) were presented having the highest overall accuracies. In order to fully maximize the benefit of SVM in the formulation of synthetic emulsions, future works should employ the technique in predicting the stability of emulsions. The SVM is a novel data mining methodology that has great potentialities in many areas, but yet to be fully utilized in the field of emulsion studies.

Acknowledgements

The authors wish to acknowledge the management of Universiti Teknologi PETRONAS (UTP), Malaysia for providing an enabling environment for the conduct of this study. Also, we wish to acknowledge the Yayasan UTP (YUTP) research grant 0153AA-H05 for supporting this study. Our sincere appreciations to Intan Khalidah Bint Saleh of PETRONAS Research Sdn Bhd and her team, for their support and suggestions throughout this study.

Author details

Abubakar A. Umar*, Ismail M. Saaïd and Aliyu A. Sulaimon

*Address all correspondence to: abubakar_g02934@utp.edu.my

Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia

References

- [1] Schramm LL, Stasiuk EN, Marangoni DG. 2 Surfactants and their applications. Annual Reports Section "C" (Physical Chemistry). 2003;**99**:3-48
- [2] Kralova I, Sjöblom J. Surfactants used in food industry: A review. Journal of Dispersion Science and Technology. 2009;**30**(9):1363-1383
- [3] Schramm LL. Emulsions, Foams, Suspensions, and Aerosols: Microscience and Applications. John Wiley & Sons; 2014
- [4] Clayton W. The Theory of Emulsions and Emulsification. J. & A. Churchill; 1923
- [5] Ramsden W. Separation of solids in the surface-layers of solutions and 'suspensions' (Observations on surface-membranes, bubbles, emulsions, and mechanical coagulation). Preliminary Account. Proceedings of the Royal Society of London. 1903;**72**:156-164
- [6] Pickering SU. CXCVI.—Pickering emulsions. Journal of the Chemical Society, Transactions. 1907;**91**:2001-2021
- [7] Gelot A, Friesen W, Hamza H. Emulsification of oil and water in the presence of finely divided solids and surface-active agents. Colloids and Surfaces. 1984;**12**:271-303
- [8] Manning FS, Thompson R. Oilfield Processing. 1991
- [9] Roberts CH. Use of treating compounds for oil field emulsions in the mid-continent field. Transactions of AIME. 1926;**1**:321-334

- [10] Aziz HMA, Darwish SF, Abdeen FM. Downhole emulsion problem, the causes and remedy, Ras Budran Field. In: SPE Asia Pacific Oil and Gas Conference and Exhibition; Society of Petroleum Engineers; 2002
- [11] Singh P et al. Flow properties of Alaskan heavy-oil emulsions. In: SPE Annual Technical Conference and Exhibition; Society of Petroleum Engineers; 2004
- [12] Kokal S. Crude Oil Emulsions: Everything you Wanted to Know but Were Afraid to Ask. 2008.
- [13] Leal-Calderon F, Schmitt V. Solid-stabilized emulsions. *Current Opinion in Colloid & Interface Science*. 2008;**13**(4):217-227
- [14] Nielloud F. *Pharmaceutical Emulsions and Suspensions: Revised and Expanded*. CRC Press; 2000
- [15] Lambert KJ, Constantinides PP, Quay SC. *Emulsion Vehicle for Poorly Soluble Drugs*. Google Patents; 2002
- [16] Binks BP. Particles as surfactants—Similarities and differences. *Current Opinion in Colloid & Interface Science*. 2002;**7**(1):21-41
- [17] Smith HV, Arnold KE. Crude Oil Emulsions. *Petroleum Engineering Handbook*. 3rd ed. Richardson: Society of Petroleum Engineers; 1992. pp. 19.1-19.34
- [18] Zafeiri I et al. Emulsions co-stabilised by edible Pickering particles and surfactants: The effect of HLB value. *Colloid and Interface Science Communications*. 2017;**17**:5-9
- [19] Aveyard R, Binks BP, Clint JH. Emulsions stabilised solely by colloidal particles. *Advances in Colloid and Interface Science*. 2003;**100**:503-546
- [20] Hu H et al. Classification of defects in steel strip surface based on multiclass support vector machine. *Multimedia Tools and Applications*. 2014;**69**(1):199-216
- [21] Vapnik V. *The Nature of Statistical Learning Theory*. Springer Science & Business Media; 2013
- [22] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;**20**(3):273-297
- [23] Wang L. *Support Vector Machines: Theory and Applications*. Vol. 177. Springer Science & Business Media; 2005
- [24] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*. 2004;**14**(3):199-222
- [25] Osuna E, Freund R, Girosi F. An improved training algorithm for support vector machines. In: *Proceedings of the 1997 IEEE workshop on Neural Networks for Signal Processing [1997] VII*; IEEE; 1997
- [26] Scholkopf B, Smola AJ. *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press; 2001

- [27] Law M. A simple introduction to support vector machines. Lecture for CSE. 2006;**802**
- [28] Ivanciuc O. Applications of support vector machines in chemistry. *Reviews in Computational Chemistry*. 2007;**23**:291
- [29] Aslan N, Cebeci Y. Application of Box-Behnken design and response surface methodology for modeling of some Turkish coals. *Fuel*. 2007;**86**(1):90-97
- [30] Box GE, Draper NR. A basis for the selection of a response surface design. *Journal of the American Statistical Association*. 1959;**54**(287):622-654
- [31] Joglekar AM et al. *Product Excellence Through Experimental Design*. Food Product and Development: From Concept to the Marketplace; 1987. pp. 211-230
- [32] Box GE, Hunter WG, Hunter JS. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Vol. 1. JSTOR; 1978
- [33] Atkinson AC, Haines LM. 14 designs for nonlinear and generalized linear models. *Handbook of Statistics*. 1996;**13**:437-475
- [34] Teixeira RF et al. Pickering emulsion polymerization using laponite clay as stabilizer to prepare armored “soft” polymer latexes. *Macromolecules*. 2011;**44**(18):7415-7422
- [35] Ashby N, Binks B. Pickering emulsions stabilised by Laponite clay particles. *Physical Chemistry Chemical Physics*. 2000;**2**(24):5640-5646

