Establishing predictive model for the spreadability of cosmetic formulations by Large Amplitude Oscillatory Shear (LAOS) and machine learning

Lee, Su Hyun[1]; Kim, Sung Ryul[2]; Lee, Hyo Jeong [3]; Oh, Heemuk[3]; Lee, Jun Bae[3]; Park, Kyunghye[4]; Park, Kyunghye[4]; Yi, Yoon Ju[4]; Park, Chun Ho[4]; **Park, Jun Dong**[1*]

[1] Department of Chemical and Biological Engineering, Sookmyung Women's University, Seoul, Republic of Korea; [2] Department of Computer Software Engineering, Kumoh National Institute of Technology, Gyeongbuk, Republic of Korea; [3] Research and Innovation Center, Cosmax Inc., Gyeonggi-do, Republic of Korea; [4] CAI Laboratory, Cosmax Inc., Gyeonggi-do, Republic of Korea

* Jun Dong Park, 100 Cheongpa-ro 47gil, Yongsan-Gu, Seoul, 04310, Korea, +82-22077-7842, jdpark@sookmyung.ac.kr

Abstract (Maximum of 250 words)

Inspired by an analogy between the application process of cosmetics and large amplitude oscillatory shear (LAOS), we suggest a novel predictive model for the spreadability of cosmetic formulations via LAOS analysis and machine learning techniques. Rheological measurements of cosmetics formulations including the transient elastic and viscous modulus from the sequence of physical process (SPP) analysis are selected as features for the predictive models, and the spreadability of each formulation, which is quantitatively rated by trained panels, is set up as the target variable. Firstly, multiple linear regression prediction models are derived, and it is shown that the LAOS-SPP parameters are more effective feature than other rheological parameters that have been conventionally related to spreadability of cosmetics. Additionally, non-linear prediction model is built based on the random forest regressor algorithm with consideration for the possibility of the nonlinear correlation between rheological measurements and spreadability. Random forest regressor model shows better performance than linear regression model, and the LAOS-SPP parameters are found to be more effective features for random forest regressor model as in the multiple linear regression model. Correlation between the LAOS-SPP parameters and the spreadability is interpreted in terms of rheological transition during rubbing process of cosmetics. Our findings indicate the importance of the nonlinear rheological behavior in texture perception mechanism of cosmetics, and how rheological measurements can be combined with machine learning techniques to solve a complicated question.

Keywords: Spreadability; LAOS; SPP analysis; Multiple linear regression; Random forest regression

Introduction

The textural properties are important factors that affects consumer satisfaction on a cosmetic formulation. The most common and direct approach that has been employed to investigate the texture of cosmetics is the panel test, such as Quantitative Descriptive Analysis (QDA) and the Spectrum Descriptive Analysis (SDA) methods. While these panel test-based sensory evaluation methods have served as descent tools for analyzing the texture of cosmetics, they are time-consuming, expensive, laborious and easily influenced by irrelevant factors, which has limited their use. As an alternative to the panel-based approach, there have been several attempts to correlate the texture of cosmetics to instrumental measurements that are economical and objective. The skin feeling is a direct outcome of the cosmetic's stress response against flow and deformation, hence the texture of cosmetics has been most linked to the rheological properties.

Spreadability is one of the most important sensory attributes that determines the feeling of cosmetics, as it is a main texture felt under rubbing process which occupies largest time in cosmetics use. There have been many attempts to understand the spreadability in terms of rheology, however, most of them end in just reporting simple correlations between spreadability and rheological properties. In a recent study on sensory properties of cosmetic formulations, correlation between various rheological parameters and spreadability is investigated with the principal component analysis [1]. Although rheological properties that are related to spreadability are identified, quantitative relationship between them is not suggested. In some research, quantitative prediction models for spreadability are suggested as functions of the work of friction and viscosity [2]. Such predictive models show decent predictive performance, however, there still exists something to be desired. Rheological parameters selected as feature of prediction model are too simple to effectively represent the complex rheological transitions of cosmetics during rubbing process. Another problem is that sample size is rather insufficient. Even though spreadability is well predicted with high coefficient of determination (R^2 or $R^2_{adjusted}$) over 0.9 in previous studies, less than 20 samples are examined, which is insufficient to ensure performance of the predictive model. Moreover, small sample set cannot represent wide range of cosmetics with diverse rheological properties, and consequently restricts model's operation range. Another potential problem lies in prediction model construction. Previous studies have developed predictive models under the assumption of linear, or semi-logarithmic or logarithmic relationship between spreadability and rheological properties, which follows common assumption in psychophysics. Even though such assumptions have shown to work effectively in the studies on olfaction and food texture, it is still doubtful that complex relationship between rheological properties and spreadability, which might be nonlinear, can be effectively captured by the empirical simple regression model. Additionally, impact of the prediction model is dulled by the lack of elaborated rheological interpretation.

In this study, we aim to establish an effective predictive model for the spreadability of cosmetics formulations with rheological measurements. First step of prediction model establishment is gathering material information so called feature or descriptor for prediction model, which is one of the keys to build an accurate prediction model. Although there is no universal rule for choosing descriptors, a good set of features should be physically meaningful and remain as low dimensional as possible. Cosmetics go through a variety of rheological transitions according to the flow conditions during the application procedure on the skin, which determines the texture experienced by consumers. Therefore, a feature set should include rheological parameters that effectively reflect flow condition of actual cosmetics application process and resulted rheological transition. Considering an analogy between the rubbing and Large Amplitude Oscillatory Shear (LAOS), both of which are repeated application of large deformation, we select rheological parameters from the LAOS analysis as feature for prediction model, impact of which is compared to that of conventional rheological parameters such as elastic modulus and viscous modulus. While textural properties of cosmetics formulation have been studied with LAOS analysis [3,4] in a few studies, this study distinguishes itself with introduction of the Sequence of the Physical Process (SPP) technique that can provide temporally resolved information on the intra-cycle rheological transition.

Using two separate machine learning algorithms of the multiple linear regression and the decision tree regressor, prediction model is trained with 77 datasets of cosmetic formulations which is the largest in spreadability study of cosmetic formulations to best of our knowledge. Each dataset contains rheological measurements and the spreadability. Here, the spreadability is quantitatively rated by 10 professionally trained panels and set as target variable. Rheological measurements are divided into two groups and used as features for the prediction model. The LAOS-SPP parameters comprise one feature set, while linear rheological parameters and shear stress at high shear rate comprise the other feature set. Firstly, multiple linear regression models for the two different feature sets are developed with all possible combination of features, performance of which are evaluated by Root Mean Squared Error (RMSE). Additionally, feature importance is analyzed. We conduct same analysis with decision tree regressor and compare result to multiple linear regression model case. It is shown that the LAOS-SPP parameters are more effective feature for prediction of the spreadability, and the decision tree regressor that can consider nonlinear correlation between rheological measurements and the spreadability provides more accurate prediction. Moreover, importance of each feature is discussed in rheological perspective.

Materials and Methods

1) Materials

Ingredients	Type of Formulation			
	Solubilization			Emulsification
Polyol (wt%)	1~10	10~20	20~	2~37
Oil (wt%)	~1	~1	~1	0~15
Silicone (wt%)	0~1	0~0.5	0~10	0~33
Thickener (wt%)	0~0.85	0.04~1.93	0~0.8	0~1.6
Emulsifier (wt%)	-	-	-	0~3.5
Water (wt%)	60~93	57~85	33~75	27~82
Number of samples	15	21	18	22

Table 1. Ingredient information of cosmetic formulations.

77 cosmetic formulations from Cosmax Inc. were selected to study the correlation between their sensory attributes and rheological measurements. These include Solubilization type, Emulsification type formulations, as shown in table 1, that can give various skin sensations. The basic ingredients of those formulations were water, polyol, oil, silicone, emulsifier and thickener. Polyols and type of emulsion may play a significant role in determining the sensory characteristics of cosmetic formulations. All formulations were made using routine laboratory equipment such as agitator and homogenizer.

2) Spreadability evaluation

Highly trained ten panelists aged 25 to 37 years old were participated in the evaluation. For the panel training, QDA (Quantitative Descriptive Analysis) method was applied, and the evaluation was conducted in a room with controlled temperature and relative humidity and adequate light conditions according to ISO guideline (ISO 8589:2007). All testing samples were blinded with a random three-digit code and Latin square design was carried out for evaluation to avoid bias (panel effect and order effect). As previously introduced, our study focuses on the sensory “Spreadability”. Spreadability is defined as “Percepted degree (Amount) of the spread strength or the spread area within the test spot while the sample cover over the skin”. Spreadability was evaluated by panelists using micro pipette (Gilson MICROMAN M100E, France), deliver 50 μm of sample on inner forearm. Gently spread product using an index or middle finger, within a circle with 5 cm diameter, at a rate of 120 BPM (beats per minute) while 1 to 10 rubs, and rated from 0 to 150 (line scale) against the references samples.

3) Rheological measurements

All rheological measurements were performed using a HR-20 (TA instruments, USA) with 40mm cross-hatched geometry to prevent wall slip during the test. Temperature was maintained as 32℃ which corresponds to hand skin temperature in a room at 15-20℃. After sample loading, all samples were pre-sheared with large amplitude oscillatory shearing of strain amplitude $\gamma_0 = 10$ and frequency $\omega = 1\ rad/sec$ for 120 sec followed by 600 sec of rest time, ensuring that samples are in a consistent beginning condition.

The linear viscoelastic moduli (G',G'') and their ratio ($tan(\delta)$), which compose the conventional rheological feature set, were measured with oscillatory shearing of strain amplitude $\gamma_0 = 0.01$ and frequency $\omega = 1\ rad/sec$. Another component of the conventional rheological feature set, σ_{100} $^{-1}$ was defined as shear stress measured under steady shear of shear rate $\dot{\gamma} = 100$ sec. For the feature set of the LAOS-SPP analysis, large amplitude oscillatory shearing of strain amplitude $\gamma_0 = 10$ and frequency $\omega = 1\ rad/sec$ was employed. The stress response from the first 15 oscillation cycles were eliminated, and the stress response after that was examined to secure the steady oscillatory state, also known as the alternating state. The SPP analysis necessitates mathematically continuous and smooth stress response as it uses the first and second derivatives of the stress response. Therefore, stress responses are averaged over 16 different oscillation cycles and smoothed with the locally estimated scatterplot smoothing (LOESS) technique.

4) Sequence of the Physical Processes (SPP) analysis and feature selection

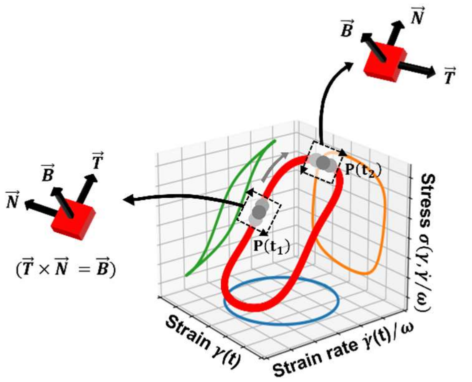


Figure 1. Description of stress response as a trajectory in three-dimensional space.

Stress response of each cosmetics formulation in LAOS is quantitively analyzed by the SPP techniques, results from which form a set of features for the prediction model. As details are available in many previous studies [5,6], we give a brief account of the SPP analysis. In the SPP analysis, stress response from the cosmetics is regarded as a function of strain and strain rate, $\sigma(\gamma, \dot{\gamma}/\omega)$. Rheological transition under oscillatory shear strain is represented by a trajectory in a three-dimensional space ($\mathbb{R}^3$) that consists of strain (γ)-strain rate ($\dot{\gamma}/\omega$)-stress (σ) axes as demonstrated in Fig.2. Thus, rheological state at a certain state is denoted by a point *P* on the trajectory that has a position vector $\vec{\boldsymbol{P}}(t) = (\gamma(t), \dot{\gamma}(t)/\omega, \sigma(t))$. At each point $\vec{\boldsymbol{P}}(t)$, the Frenet-Serret frame (tangent vector $\vec{\boldsymbol{T}}(t)$, normal vector $\vec{\boldsymbol{N}}(t)$, and binormal vector $\vec{\boldsymbol{B}}(t)$) is defined as follows

$$\vec{\boldsymbol{T}}(t) = \frac{\vec{\boldsymbol{P}}'(t)}{\left|\vec{\boldsymbol{P}}'(t)\right|}, \vec{\boldsymbol{N}}(t) = \frac{\vec{\boldsymbol{T}}'(t)}{\left|\vec{\boldsymbol{T}}'(t)\right|}, \vec{\boldsymbol{B}}(t) = \vec{\boldsymbol{T}}(t) \times \vec{\boldsymbol{N}}(t) \tag{1}$$

where $\vec{\boldsymbol{P}}'(t)$ and $\vec{\boldsymbol{T}}'(t)$ are time derivative of $\vec{\boldsymbol{P}}(t)$ and $\vec{\boldsymbol{T}}(t)$. The tangent ($\vec{\boldsymbol{T}}$) vector and normal ($\vec{\boldsymbol{N}}$) vector span a plane that is tangent to the trajectory at a point $\vec{\boldsymbol{P}}(t)$. Such plane is referred to as osculating plane that is normal to $\vec{\boldsymbol{B}}$. The osculating plane at an arbitrary $\vec{\boldsymbol{P}}(t^*)$ is

$$\vec{\boldsymbol{B}}(t^*) \cdot \left(\vec{\boldsymbol{P}}(t) - \vec{\boldsymbol{P}}(t^*)\right) = 0 \text{ or} \tag{2}$$

$$B_{\boldsymbol{\gamma}}(t^*)\big(\gamma(t) - \gamma(t^*)\big) + B_{\frac{\dot{\gamma}}{\omega}}(t^*)\left(\frac{\dot{\gamma}(t)}{\omega} - \frac{\dot{\gamma}(t^*)}{\omega}\right) + B_{\sigma}(t^*)\big(\sigma(t) - \sigma(t^*)\big) = 0 \tag{3}$$

With $\Delta t \to 0$, three consecutive points $\vec{\boldsymbol{P}}(t^* - \Delta t), \vec{\boldsymbol{P}}(t^*), \vec{\boldsymbol{P}}(t^* + \Delta t)$ sit within the osculating plane, and Eq.(2) can be written in the form of a differential change

$$\begin{aligned} &B_{\boldsymbol{\gamma}}(t^*)\big(\gamma(t^* \pm \Delta t) - \gamma(t^*)\big) + B_{\frac{\dot{\gamma}}{\omega}}(t^*)\left(\frac{\dot{\gamma}(t^* \pm \Delta t)}{\omega} - \frac{\dot{\gamma}(t^*)}{\omega}\right) \\ &\qquad + B_{\sigma}(t^*)\big(\sigma(t^* \pm \Delta t) - \sigma(t^*)\big) = B_{\boldsymbol{\gamma}}(t^*)d\gamma + B_{\frac{\dot{\gamma}}{\omega}}(t^*)d\left(\frac{\dot{\gamma}}{\omega}\right) + B_{\sigma}(t^*)d\sigma \\ &\qquad = 0 \end{aligned} \tag{4}$$

Eq.(4) can be rewritten in terms of $d\sigma$

$$d\sigma = -\frac{B_{\gamma}(t^*)}{B_{\sigma}(t^*)}d\gamma - \frac{B_{\frac{\dot{\gamma}}{\omega}}(t^*)}{B_{\sigma}(t^*)}d\left(\frac{\dot{\gamma}}{\omega}\right), \tag{5}$$

and compared to the total derivative of stress $\sigma(\gamma, \dot{\gamma}/\omega)$

$$d\sigma = \frac{\partial\sigma}{\partial\gamma}d\gamma + \frac{\partial\sigma}{\partial\left(\frac{\dot{\gamma}}{\omega}\right)}d\left(\frac{\dot{\gamma}}{\omega}\right). \tag{6}$$

$-\frac{B_\gamma(t^*)}{B_\sigma(t^*)}$ and $-\frac{B_{\frac{\dot\gamma}{\omega}}(t^*)}{B_\sigma(t^*)}$ are equivalent to $\frac{\partial\sigma}{\partial\gamma}$ and $\frac{\partial\sigma}{\partial\left(\frac{\dot\gamma}{\omega}\right)}$ that are referred to as transient elastic modulus $G_t'(t)$ and transient viscous modulus $G_t''(t)$. The transient moduli can be considered as time-dependent analogues of the dynamic moduli (G', G''), although the transient moduli distinguish itself by providing temporally resolved information on the intra-cycle rheological transition.

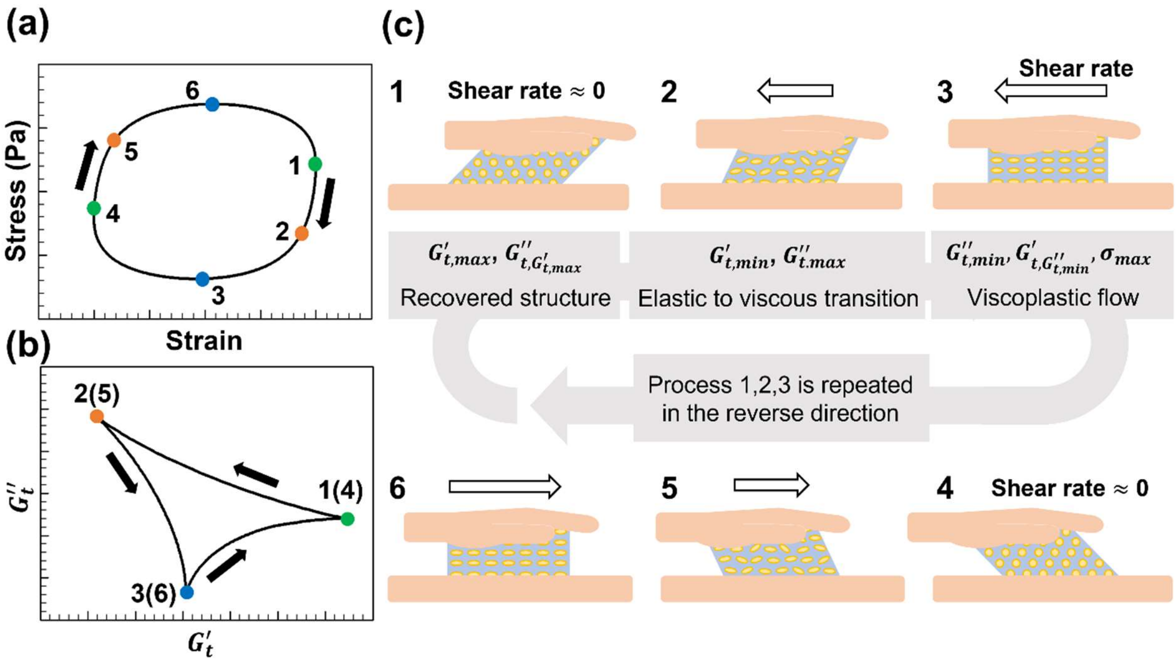


Figure 2. Feature variable selection inspired by an analogy between LAOS and rubbing out process of cosmetics. (a) Elastic Lissajous curve. (b) Cole-Cole plot. (c) Simplified rubbing out process corresponding to LAOS, and seven LAOS-SPP parameters defined at each point.

Variable Set	Feature variables		Target variable
	Description	Feature	
LAOS-SPP	Recovered structure	$G'_{t,max}$, $G''_{t,G'_{t,max}}$	Spreadability
	Elastic to viscous transition	$G'_{t,min}$, $G''_{t,max}$	
	Viscoplastic flow	$G''_{t,min}$, $G'_{t,G''_{t,min}}$, σ_{max}	
Conventional	Linear rheological property	$G', G'', \tan(\delta)$	

parameters	Spreading stress	$\sigma_{100s^{-1}}$	

Table 2. Feature selection for the prediction model

In the SPP analysis, a rheological state is described by the transient moduli $G_t'(t)$ and $G_t''(t)$, and rheological transition is represented by a change in the transient moduli. Typically, the rheological transition is interpreted in the transient Cole-Cole plot form whose abscissa and ordinate are given as $G_t'(t)$ and $G_t''(t)$. Shown in Fig 3 are (a) elastic Lissajous curve and (b) Cole-Cole plot of a cosmetic sample studied in this study. It should be noted that all 77 samples have similar the elastic Lissajous curve and Cole-Cole plot that do not remarkably deviate from Fig 2(a) and (b). Looking at the corresponding trace in the Cole-Cole plot can reveal the rheological transition in a region of interest. For example, rheological transition between point 1 and 2 in the elastic Lissajous curve in Fig 3(a) is identified as the elastic to viscous transition, because the trace marked with corresponding points shows decrease of $G_t'(t)$ and increase of $G_t''(t)$ in the Cole-Cole plot in Fig 2(b).

An analogy between the LAOS and rubbing-out process of cosmetics served as an insight in selecting transient moduli as features for the spreadabilty prediction model. By projecting the LAOS onto cosmetics application process, we employ transient moduli at three characteristic points in the Cole-Cole plot as features for the prediction model. Generally, the Cole-Cole plot under LAOS has deltoid shape due to the dominance of the third harmonic in Fourier spectrum, and the transient moduli at three vertices of the deltoid can characterize rheological behavior of cosmetics during rubbing out process of cosmetics as shown in Fig2 (c). To begin with, as strain peaks and flow is reversed near point 1 (or 4) in Fig 2, the deformation rate approaches zero allowing more time for structural relaxation in cosmetics. Point 1 (or 4), where the transient elastic modulus reaches its highest ($G_{t,max}'$), represents the most structure-recovered state during oscillation in rheological perspective. In terms of cosmetics application, the transient elastic modulus ($G_{t,max}'$) and the transient viscous modulus ($G_{t,G_{t,ma}'}''$) at point 1 (or 4) can be linked to the spreadability of cosmetics perceived when a user changes their application direction and cosmetics instantaneously recover their structure. Secondly, the minimum transient elastic modulus $G_{t,min}'$ and the maximum transient viscous modulus $G_{t,max}''$ at point 2 (or 5) are utilized to characterize the elastic to viscous transition. As structure recovered from point 6 (or 3) to point 1 (or 4) starts to be deformed and ruptured by flow reversal, cosmetics undergo elastic to viscous transition from point 1 (or 4) to point 2 (or 5). The minimum transient elastic modulus $G_{t,min}'$ and the maximum transient viscous modulus $G_{t,max}''$ indicate that the elastic to viscous transition occurs most rapidly at point 2 (or 5). The minimum transient elastic modulus $G_{t,min}'$ and the maximum transient viscous modulus $G_{t.max}''$ are included in feature set as rheological parameters to consider the elastic to viscous transition process during cosmetics application. After going through the structure recovery and elastic to viscous transition processes sequentially, cosmetics flow with the least structured state in the vicinity of point 3 (or 6) where strain is almost 0 and strain rate reaches its highest. Here the least structured state is ascribed to large shear rate. The transient elastic modulus is near to 0 at point 3 (or 6), while the transient viscous modulus G_t'' is minimum, indicating viscoplastic flow. The minimum G_t'' and the transient elastic modulus at the same point are defined as $G_{t,min}''$ and $G_{t,G_{t,min}''}'$ respectively, and they will be used to characterize viscoplastic flow during rubbing out of cosmetics. $G_{t,min}''$ and $G_{t,G_{t,min}''}'$ are predicted to have a significant influence on the spreadability of cosmetics since viscoplastic flow accounts for the majority of the cosmetics application procedure. Additionally, the stress maximum (σ_{max}) observed around point 3 (or 6) is selected as feature for the prediction model according to previous studies that reported close correlation between shear stress and spreadability [1]. The LAOS-SPP parameters stated above constitute a feature set for the spreadability prediction model, which will be compared to another feature set of conventional rheological parameters that has been closely correlated to the spreadability. Table 2 summarizes the two feature sets for the spreadability prediction model.

5) Machine learning algorithms

A. Multiple linear Regression model

In this work, prediction models were built with *scikit-learn* modules of *RidgeCV*, *LASSOCV*, and *ElasticNetCV*. Figure 4 shows schematic diagram of multiple linear regression model built in this work. For each variable set, we tried every possible combination of variables to find the best feature set. For example, 127 ($2^7 - 1$) combinations of variables were tried in the case of LAOS-SPP set. For each feature set, multiple linear regression models in the form

$$\hat{y}(i) = \beta_0 + \beta_1 x_1(i) + \beta_2 x_2(i) \cdots \tag{7}$$

are created, where $\hat{y}(i)$, $x_k(i)$, β_k indicate predicted spreadability of the *i*-th sample, *k*-th feature of the *i*-th sample, and regression weight for *k*-th feature, respectively. Here, $x_k(i)$ is scaled with min-max normalization for 1) feature importance analysis and 2) utilization of regularization techniques to prevent overfitting problem. Three different regularization methods of Ridge, and Lasso, elastic net are introduced, and each regularization method employ regularized loss of the following forms in sequence

$$\sum_{i=1}^{n}\left(y(i)-\hat{y}(i)\right)^{2}+\alpha\sum_{j=1}^{k}\beta_j^2 \tag{8}$$

$$\sum_{i=1}^{n}\left(y(i)-\hat{y}(i)\right)^{2}+\alpha\sum_{j=1}^{k}\left|\beta_j\right| \tag{9}$$

$$\sum_{i=1}^{n}\left(y(i)-\hat{y}(i)\right)^{2}+\alpha\lambda\sum_{j=1}^{k}\left|\beta_j\right|+\frac{\alpha(1-\lambda)}{2}\times\sum_{j=1}^{k}\beta_j^2 \tag{10}$$

where hyper parameters α and λ were optimized with grid search approach. Details on the regularized regression are available elsewhere [7]. As we have limited data set, 5-fold cross validation technique was adopted to resample and evaluate machine learning models on a limited data sample. For each of the 5-folds, the Root Mean Squared Error (RMSE), defined as

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y(i)-\hat{y}(i)\right)^{2}} \tag{11}$$

is calculated and their average is utilized as model evaluation metric.

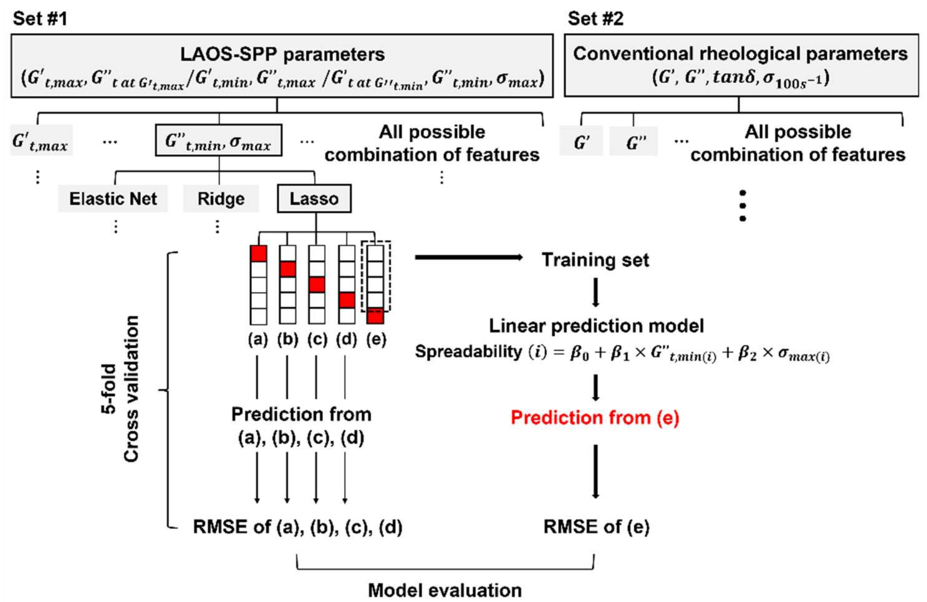


Figure 3. Schematic diagram of multiple linear regression model

B. Random forest

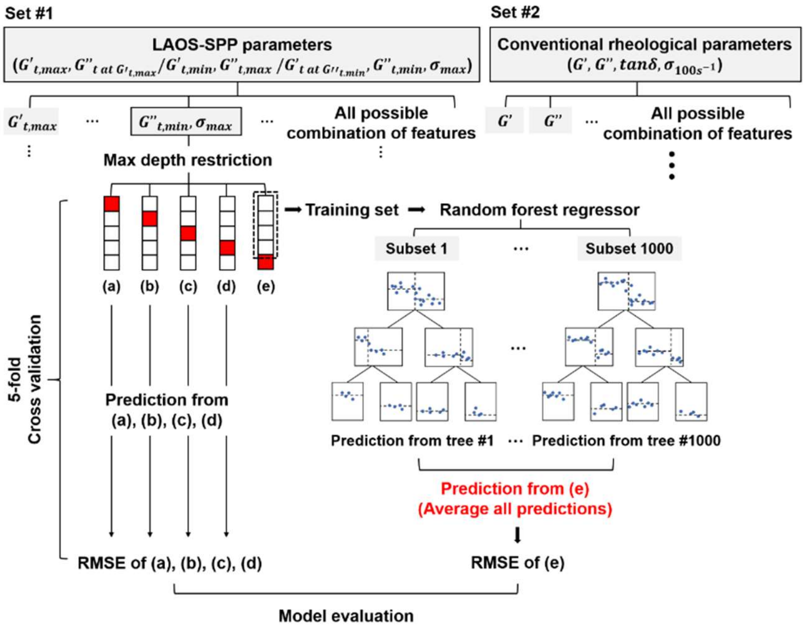


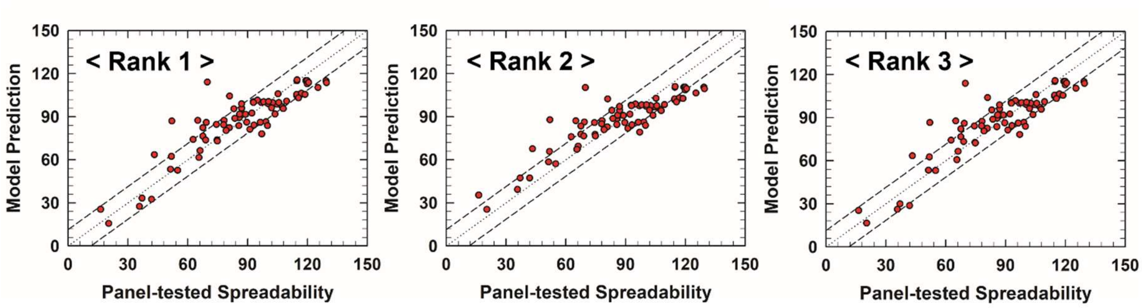
Figure 4. Schematic diagram of random forest regressior model

While a multiple linear regression model may adequately capture the linear correlation between features and spreadability, it cannot account for nonlinea relationships. As an alternative, we employ the random forest regressor that is a supervised learning algorithm with an ensemble of decision trees. The random forest regressor operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the mean predcition (or regression) of the idividual trees. Random forest regressor model has advantage in that it provides accuarate but not overfitted preidction. Details on the random forest regressor is given in elsewhere [8]. As in the case multiple linear regression model, every possible combination of features are attempted to find the best feature set, and 5-fold cross validation technique is used. Herein, maximum depth of trees is restricted between 1 and 10 to ascertain overfitting possibility. Each of the 5-folds consists of 1000 trees (or estimators), and model is evaluated by the RMSE as before.

Results

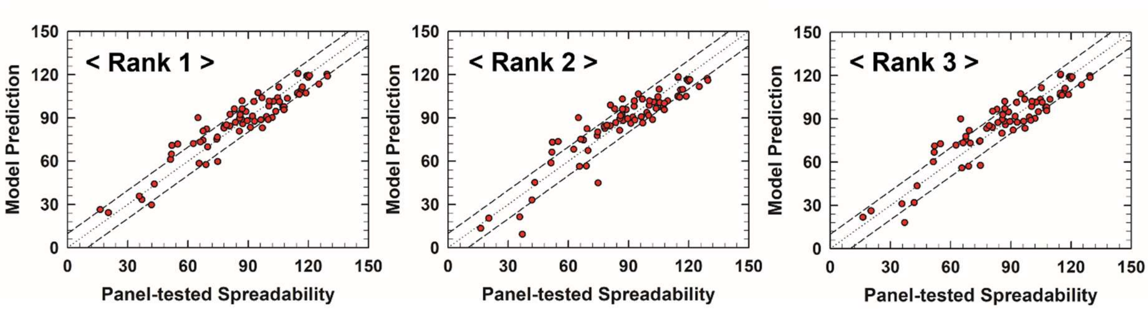
1) Multiple linear Regression model

Figure 5 shows the three best multiple linear regression models with the conventional rheological paramters of $G', G'', \tan(\delta), \sigma_{100s^{-1}}$. The best model with elastic net regularization shows RMSE of 11.38. It is remarkable that feature important analysis indicates that $\sigma_{100s^{-1}}$ is the most important rheological property in determining the spreadability.



Rank	Model (α)	Feature combination (importance)	RMSE
1	Elastic net (10.00)	$\sigma_{100s^{-1}}(2.60), G''(0.70)$	11.38
2	Lasso (3.00)	$\sigma_{100s^{-1}}(17.97)$	11.39
3	Ridge (0.05)	$\sigma_{100s^{-1}}(20.36), G''(2.49)$	11.4

Figure 5. Spreadability prediction from the multiple linear regression model with conventional rheological parameters as features.

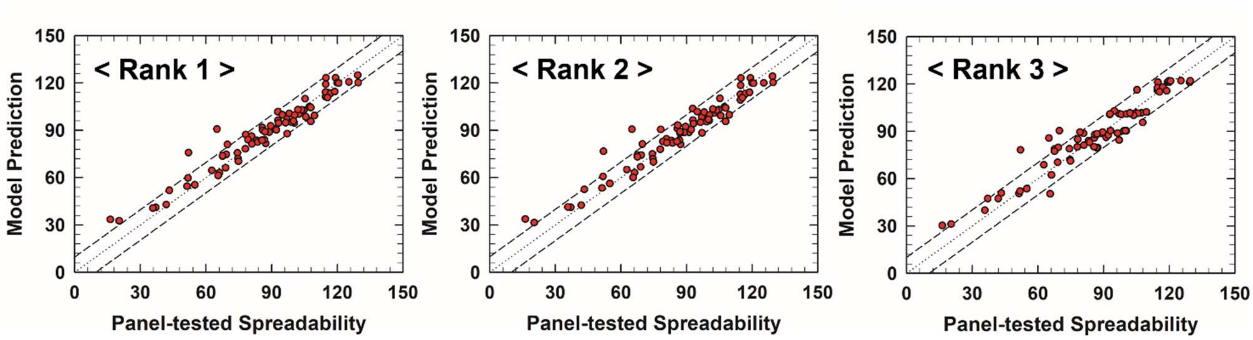


Rank	Model (α)	Feature combination (importance)	RMSE
1	Ridge (0.05)	$G''_{t,max}(30.28), \sigma_{max}(29.07), G'_{t,min}(19.50),$ $G''_{t,min}(9.55), G'_{t,G''_{t,min}}(3.21)$	9.85
2	Ridge (0.05)	$\sigma_{max}(28.86), G''_{t,max}(26.91), G'_{t,min}(24.02),$ $G''_{t,min}(10.62), G'_{t,min}(7.89), G'_{t,G''_{t,min}}(2.84)$	9.97
3	Ridge (0.05)	$G''_{t,max}(26.18), G'_{t,min}(15.59), \sigma_{max}(15.13),$ $G''_{t,G'_{t,max}}(13.92), G''_{t,min}(8.91), G'_{t,G''_{t,min}}(3.19)$	9.99

Figure 6. Spreadability prediction from the multiple linear regression model with the LAOS-SPP parameters as features.

Figure 6 demonstrates the three best multiple linear regression models with the LAOS-SPP paramters. The best model with Ridge regularization shows RMSE of 9.85, which indicates that the LAOS-SPP parameter based regression model can give more accurate prediction. Contrary to the multiple linear regression model with the conventional rheological parameters where $\sigma_{100\,s^{-1}}$ is dominantly important feature for spreadability prediction, multiple linear regression model with the LAOS-SPP parameters has several important features.

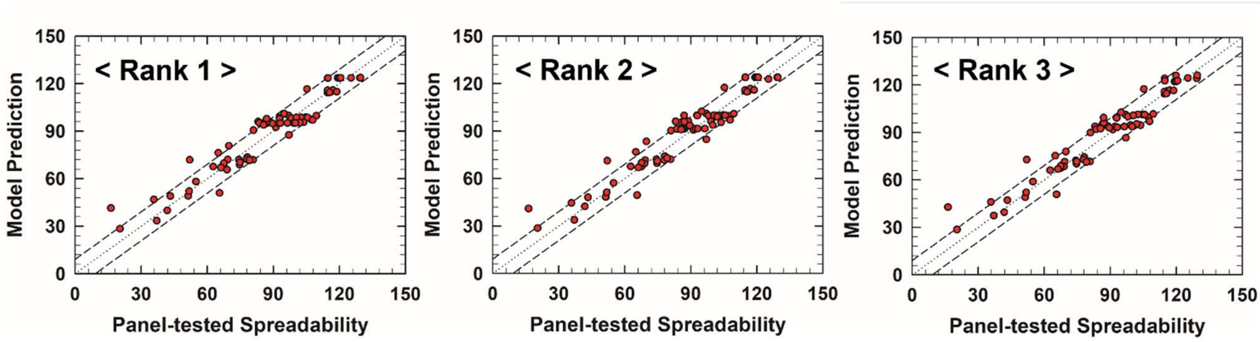
2) Random forest regression model



Rank	Feature combination (Importance)	Max_depth	RMSE
1	$\sigma_{100s^{-1}}(0.52), G''(0.43), \tan(\delta)\ (0.05)$	7	9.56
2	$\sigma_{100s^{-1}}(0.49), G''(0.38), G'(0.09), \tan(\delta)\ (0.04)$	8	9.90
3	$G''(0.53), \sigma_{100s^{-1}}(0.47)$	4	9.99

Figure 7. Spreadability prediction from the random forest regression model with conventional rheological parameters as features.

Shown in Figure 7 is the spreadability prediction from the random forest regression model with features of the convetnional rheological parameters. Compared to the prediction from the multiple linear regression model, the best result from the random forest regression model shows better performance with RMSE of 9.56, which indicates that there exists nonlinear correlation between rheological property and spreadability. It should be noticed that while the random forest regression model wih conventional rheological parameters seems to work well, maximum depth of the random forest regression model is large with values of 7,8 and 4. This indicates that there can be a overfitting problem in this model.



Rank	Feature combination (Importance)	Max_depth	RMSE
1	$G''_{t,min}$ (0.70), σ_{max}(0.30)	3	9.14
2	$G''_{t,min}$(0.68), σ_{max}(0.26), $G'_{t,min}$(0.06)	3	9.24
3	$G''_{t,min}$(0.68), σ_{max}(0.30), $G'_{t,\,G''_{t,min}}$ (0.02)	4	9.30

Figure 8. Spreadability prediction from the random forest regression model with the LAOS-SPP parameters as features.

Figure 8 demonstrates results from the random forest regression model with the LAOS-SPP parameters as features. This prediction model shows the best performance with the RMSE of 9.14. Maximum depth of trees in the best 3 model is less than or equal to 4, which means the trained model works effectively without overfitting problem. Feature importance analysis shows that $G''_{t,min}$ and σ_{max} play an important role in determining the spreadability.

Discussion.

It is shown that the best spreadability prediction model can be established by the combination of the random forest regression technique and the LAOS-SPP analysis. This signifies that nonlinear rheological behavior of cosmetic formulations is a key factor in determining the spreadability. Furthermore, our result emphasizes that the relationship between sensory texture (in this work spreadability) and rheological property can be nonlinear. Therefore, it seems essential to use nonlinear modeling approach in prediction of the sensory texture.

Feature importance analysis result reveals a new remarkable point. In the conventional study, shear stress or viscosity were considered to be the most important rheological parameter that determines the spreadability, which is well reflected in our study using the multiple linear regression model. However, it is shown that the spreadability of cosmetics is demonstrated to be closely related to a novel nonlinear rheological parameter $G''_{t,min}$, which measures the degree of the viscous transition during the rubbing out process.

Conclusion.

Our results demonstrates the importance of the nonlinear rheological behavior in texture perception mechanism of cosmetics. Additionally, it is shown that how rheological measurements can be combined with machine learning techniques to solve a complicated question of sensory texture.

Acknowledgments.

J. D. Park acknowledges support from the Cosmax, Inc. (Grant No.4-2103-0005)

Conflict of Interest Statement.

NONE

References.

1. Gilbert L, Savary G, Grisel M, Picard C, (2013) Predicting sensory texture properties of cosmetic emulsions by physical measurements. Chemometrics and Intelligent Laboratory Systems 124:21–31.
2. Cyriac F, Yi T, Chow P, Macbeath C, (2022) Tactile friction and rheological studies to objectify sensory properties of topical formulations. Journal of Rheology 66:305–311.
3. Ozkan S, Gillece T, Senak L, Moore D, (2012) Characterization of yield stress and slip behaviour of skin/hair care gels using steady flow and LAOS measurements and their correlation with sensorial attributes. International Journal of Cosmetic Science 34:193–201.
4. Pan S, Natalie G, (2020) The transient . Act Mech 231:3031–3042.
5. Park J.D , Rogers A. S, (2018) The transient behavior of soft glassy materials far from equilibrium. Journal of Rheology 62:869–888.
6. Park J.D , Rogers A. S, (2020) Rheological manifestation of microstructural change of colloidal gel under oscillatory shear flow. Physics of Fluids 32:063102-18.
7. Raschka S, Liu Y, Mirjalili V, Machine Learning with Pytorch and Scikit-Learn. Packt,

Birmingham, 2019.
8. Biau G, (2012) Analysis of a Random Forests Model. Journal of Mahcine Learning Research 13:1063-1095.