Journal of Pharmaceutical Sciences 000 (2022) 1–13

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Pharmaceutical Biotechnology

A Collaborative Study on the Classification of Silicone Oil Droplets and Protein Particles Using Flow Imaging Method

Hiroko Shibata[a,*], Masahiro Terabe[b], Yuriko Shibano[c], Satoshi Saitoh[b], Tomohiro Takasugi[d], Yu Hayashi[d], Shinji Okabe[e], Yuka Yamaguchi[e], Hidehito Yasukawa[e], Hiroyuki Suetomo[f], Kazuhiro Miyanabe[g], Naomi Ohbayashi[h], Michiko Akimaru[i], Shuntaro Saito[i], Daisuke Ito[j], Atsushi Nakano[j], Shota Kojima[k], Yuya Miyahara[l], Kenji Sasaki[l], Takahiro Maruno[m], Masanori Noda[m], Masato Kiyoshi[a], Akira Harazono[a], Tetsuo Torisu[c], Susumu Uchiyama[c], Akiko Ishii-Watabe[a]

[a] *Division of Biological Chemistry and Biologicals, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan*
[b] *Pharmaceutical Technology Division, Analytical Development Department, Chugai Pharmaceutical Co. Ltd., 5-1 Ukima, 5-chome, Kita-ku, Tokyo 115-8543 Japan*
[c] *Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*
[d] *Analytical Research Laboratories, Pharmaceutical Technology, Astellas Pharma. Inc., 5-2-3 Tokodai, Tsukuba, Ibaraki, 300-2698, Japan*
[e] *Research Division, CMC Development Research, Formulation Research Unit, Formulation Development, JCR Pharmaceuticals Co., Ltd., 2-2-9 Murotani, Nishi-ku, Kobe, Hyogo 651-2241, Japan*
[f] *Bio Process Research and Development Laboratories, Production Division, Kyowa Kirin Co., Ltd., 100-1, Hagiwara-machi, Takasaki, Gunma 370-0013, Japan*
[g] *CMC Regulatory and Analytical R&D., Ono Pharmaceutical Co., Ltd., 1-1, Sakurai 3-chome, Shimamoto-cho, Mishima-gun, Osaka, 618-8585, Japan*
[h] *Pharmaceutical Research Center, Formulation Research Lab., Meiji Seika Pharma Co., Ltd., 788 Kayama, Odawara, Kanagawa, 250-0852, Japan*
[i] *Analytical & Quality Evaluation Research Laboratories, Daiichi Sankyo Co., Ltd., 1-12-1, Shinomiya, Hiratsuka, Kanagawa, 254-0014, Japan*
[j] *Japan Blood Products Organization, 1007-31 Izumisawa, Chitose, Hokkaido, 066-8610, Japan*
[k] *Pharmaceutical Laboratory, Mochida Pharmaceutical Co., Ltd. 342 Gensuke, Fujieda, Shizuoka, 426-8640, Japan*
[l] *CMC Modality Technology Laboratories, Production Technology & Supply Chain Management Division, Mitsubishi Tanabe Pharma Corporation, 7473-2, Onoda, Sanyoonoda-shi, Yamaguchi, 756-0054 Japan*
[m] *U-Medico Inc., 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan*

ARTICLE INFO

Article history:
Received 1 March 2022
Revised 7 July 2022
Accepted 7 July 2022
Available online xxx

Keywords:
Flow imaging
Silicone oil
Protein particles
Machine learning

ABSTRACT

In this study, we conducted a collaborative study on the classification between silicone oil droplets and protein particles detected using the flow imaging (FI) method toward proposing a standardized classifier/model. We compared four approaches, including a classification filter composed of particle characteristic parameters, principal component analysis, decision tree, and convolutional neural network in the performance of the developed classifier/model. Finally, the points to be considered were summarized for measurement using the FI method, and for establishing the classifier/model using machine learning to differentiate silicone oil droplets and protein particles.

Introduction

The number of therapeutic protein injections, including monoclonal antibodies and Fc-fusions, has increased significantly over the last few decades, becoming essential therapeutic agents for various poorly treated diseases including autoimmune diseases and cancer.[1,2] Protein molecules can inherently self-associate to form aggregates ranging in size from nanometer-sized dimers to those detected as insoluble

* Correspondence author at: Division of Biological Chemistry and Biologicals, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa, 210-9501, Japan.
E-mail address: h-shibata@nihs.go.jp (H. Shibata).

https://doi.org/10.1016/j.xphs.2022.07.006

particles due to various physicochemical stresses.[3] Numerous studies have suggested that protein aggregates including insoluble particles can elicit immune responses.[4, 5] However it has not been directly proven that protein particles induce adverse events in humans. Therefore, protein aggregates and insoluble particles are thought of as impurities to be adequately characterized and controlled to ensure the efficacy and safety of therapeutic protein injections.[6-8]

The development of prefilled syringe formulations has been demanded because of the increase in therapeutic protein injections intended to treat chronic diseases, including diabetes, rheumatoid arthritis, and inflammatory bowel disease. The advantages of prefilled syringes include ease of use, accurate administration of required doses, and elimination of the risk of contamination, providing easy and safe self-administration of the dose for patients.[9] Although cyclo-olefin polymer silicone-oil (SO) free syringes have been commercialized and approved,[10] most pre-filled syringes used for therapeutic proteins still adopt glass syringes coated with SO. SO is used as a lubricant in prefilled glass syringes, and part of them migrate as droplets into drug products from the syringes. Even though SO is not proven to be harmful to the patient,[11] it has been described to contribute to the generation of protein particles and the formation of protein-SO complexes that might potentially be immunogenic.[12-15] From a quality control perspective, elevated concentrations of SO droplets between lots can indicate problems during the production process, for example, improper siliconization of syringes. Therefore, differentiation between protein particles and SO droplets, as well as counting of each particle, is required.

The flow imaging (FI) method is a powerful tool for characterizing insoluble particles in therapeutic protein injections.[16] We recently conducted a collaborative study toward standardizing the FI method, and concluded that setting FI as a release test would be feasible to precisely analyze the counts and shapes of insoluble particles for therapeutic protein injections.[17] Since the FI method can measure characteristic parameters as well as the size and number of particles by capturing and analyzing images of particles, it is believed to allow the classification of the particles based on morphological features. In fact, several reports have demonstrated that the FI method successfully classifies SO droplets and protein particles mainly by using filters composed of characteristic parameters calculated using the built-in software of the FI instruments.[18-21] Applying such classifiers/models is believed to be a more efficient approach than visually classifying and counting particles based on particle image data. However, there are no standardized or widely usable methods of analysis to differentiate between SO droplets and protein particles. Therefore, it should be investigated how accurately SO droplets can be classified from protein particles using the FI method, and how different in capability to classify these particles are among instruments. Then, the feasibility of establishing such a standardized method for the classification should be examined.

Recently, utilizing machine learning has significantly expanded.[22] In the pharmaceutical science field, machine leaning has been used to develop automated analysis of images captured by several techniques such as X-ray computed tomography.[23] It has already been shown that machine learning provides automatic classifiers that can differentiate between the protein particles produced under various manufacturing conditions or physical stresses, and can identify SO droplets and protein particles.[24-26] In these studies, convolutional neural network (CNN) analysis and/or random forest were used to develop an automatic classifier to identify of intended particles. CNN analysis enables highly accurate classification, but it also has the aspect of low interpretability, which might be an obstacle to become a widespread approach and standardization of the analysis. Conversely, a decision tree, a machine learning algorithm, could also provide a classifier/model with higher interpretability, although it may be less accurate compared to CNN.

To assess the feasibility of the standardization of the classifier/model, we conducted a collaborative study on the classification between SO droplets and protein particles detected using the FI method. Fig. 1 shows the scheme of this study. As in study 1, we collected particle data of SO droplets and protein particles measured using the FI method from 11 participating laboratories to prepare a training data set for developing the classifier/model. We compared several approaches, including a traditional classifier using a filter composed of characteristic parameters, decision trees developed by statistical tools or machine learning, and CNN in their performance as classifier/model. We also discussed the difference between the instrument types and the difference among laboratories on the classification results by CNN. In study 2, to assess the feasibility and adequacy of the standardization of the classifier/model, two types of decision trees for each instrument were developed. Particle data of SO droplets and protein particles obtained by measuring arbitrary samples were also collected from the participating laboratories to prepare a test (unlearned) data source. Then, the performance of developed decision trees was verified using the test data. Finally, the points to be considered are summarized for establishing the classifier/model using machine learning to differentiate SO droplets and protein particles measured using the FI method.

Materials and Methods

Materials

Two different types of model protein particles, NPAS 90 and NPAS 70, were prepared as previously described[17] and used as shared samples. Briefly, diluted intravenous immunoglobulin (IVIG, Sanglopor, CSL Behring, 50 mg/ml) at 25 mg/mL were heated at 90°C or 70°C (corresponding to NPAS 90 or NPAS 70) with shaking at 2,000 rpm for one or two minutes, respectively. Then, the samples were diluted with 0.5% hydroxypropyl methylcellulose (HPMC, Alfa Aesar, USA) solution. NPAS 70 and 90 were stored at -80°C until they were distributed to the laboratories. The ETFE particles were kindly provided by Dr. Ripple from NIST and handled according to the instruction in our previous report.[17] The ETFE particles were also used as model protein particles in this study.[27]

Silicone oil droplets (SO) were produced at each laboratory by dropping 1 mL SO-coated syringes filled with 0.01% tween 20 PBS from a height of 30 cm ten times, and collected in a tube. Subsequently, the solutions were vacuumed for five minutes to remove bubbles.

Flow Imaging

FlowCam 8100, VS-1 (Yokogawa Fluid imaging technologies, Inc., USA), and MFI DPA 4100, 4200, and 5200 (Protein Simple, Canada) were used. A 0.5 mL of each sample was measured. Detailed operating procedures, such as flushing and cleaning of flow cells, and assessment of instrument qualification, were the same as previously described.[17] In the case of FlowCam, the flow rate, sample volume, autoimage rate and sampling efficiency were set to 0.1 or 0.05 mL/min, 0.5 mL, 16-18 frames/s, and 65%-70% respectively. The focus was adjusted by following the manufacturer's instructions for each instrument using reference particles. The threshold parameter was set as "dark and light, 10 and 20, respectively" based on the result of a prior experiment (Supplemental Figure 1a).

Model Data Set

Data including about 300 particles for each sample (SO, NPAS70, NPAS90, and ETFE) were collected from seven laboratories for FlowCam and four laboratories for MFI, and used as a training data set for

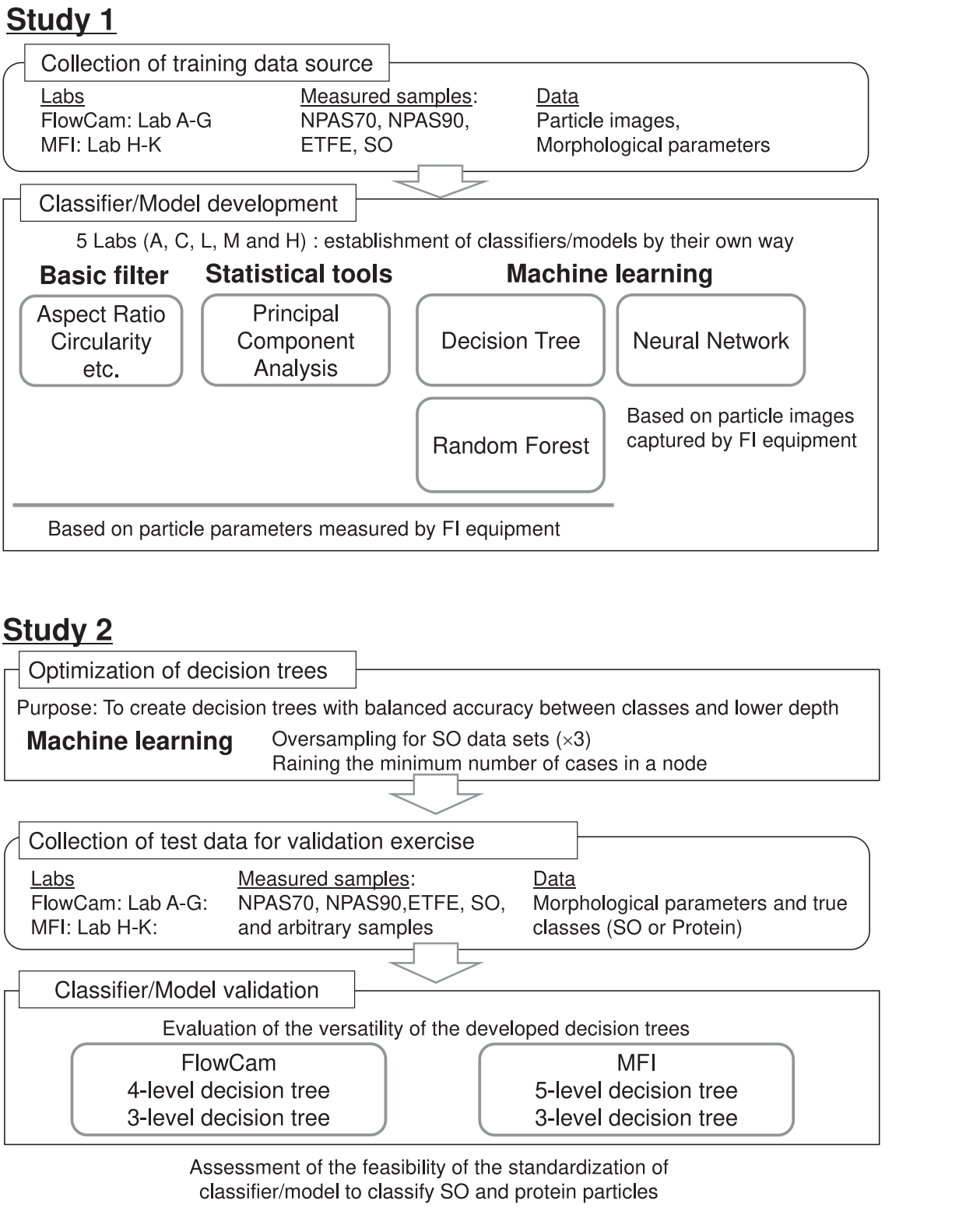


Figure 1. Scheme of this study.

study 1 (Table 1, Supplemental Figure 1b-c). The number of particle data used for each classifier/model development was listed in Supplemental Table 2-1, 3-1, 4-1, 5-1, and 6-2. Also, data were separately collected as test data for validation exercise, which were not used to develop the classifiers/models, from seven laboratories for FlowCam and four laboratories for MFI. The participating laboratories arbitrarily selected the measurement samples. The test data included each particle's class (SO or protein particle) visually determined by each laboratory. The characteristic parameters and images of all model data sets are shown in the Supplemental Material. To determine the error rate of classifier/model, the particles equal or greater than 10 μm whose origin can be easily determined visually were targeted in this study. Thus, all particles less than 10 μm were excluded, both during training and testing/classification.

Table 1
Number of particles collected from participating laboratories as training data source.

Target instrument		Type of particles*			
	Laboratory	SO	NPAS90	NPAS70	ETFE
FlowCam	A	189	300	300	300
	B	119	280	165	172
	C	36	99	37	76
	D	418	500	500	–
	E	331	271	293	292
	F	134	300	91	123
	G	50	454	79	81
	total	1277	2204	1465	1044
MFI	H	266	290	317	311
	I	100	300	300	–
	J	146	141	133	127
	K	108	110	108	294
	total	620	841	858	732

* Particles equal or greater than 10 μm.

Principal Component Analysis

Principal component analysis was employed when classifying SO droplets and protein aggregates (protein particles) to identify which particle characteristic parameters contributed. Principal component analysis was performed using JMP® version 15 (SAS Institute Inc., Cary, NC) multiple times on FlowCam data to narrow down the particle characteristic parameters for each sample that determine the features of the SO droplets (Supplemental Figure 2-1, 2-2). The classification was predicted based on the positive or negative sign of each principal component composed of the identified particle characteristic parameters, and the error rate was assessed for each sample. Furthermore, a decision tree was prepared using the parameters with a high proportion of variance in the calculated principal component. The particle classification was also performed on the decision tree, and the predicted results were compared to the actual results to assess the error rate (Supplemental Figure 2-3).

Classification Filter

To create a filter for classifying SO droplets and protein aggregates on FlowCam data (Supplemental Table 3-1), we focused on particle circularity, a large differences between SO and protein aggregate circularity. Aspect ratio, circle fit, and circularity were chosen as the three particle characteristic parameters representing how close the particles are to a circle. We plotted scatter diagrams for each sample from the FlowCam data to set the filter that produced the best error rate (Supplemental Figure 3-2).

Statistical Analysis-Decision Tree

Scatter diagrams were plotted for each sample with the particle characteristic parameters obtained from MFI, to check for bias in the data and identify particle characteristic parameters, which appear to be useful for classification.

Of the images, 90% (557 silicon oil droplets and 2188 protein aggregates) were assigned for decision tree creation, and 10% (63 silicon oil droplets and 243 protein aggregates) were assigned for decision tree validation. Statistical analysis software (JMP®, Version: 14.2.0) was used to optimize the decision tree (setting the sequence of nodes and threshold values for particle characteristic parameters). The validation data were used to check the accuracy of the optimized decision tree (Supplemental Figure 4).

Machine Learning

We attempted to build a classification model by applying two machine learning methods, decision tree[27] and random forest,[28] to MFI and FlowCam data (Supplemental Table 5-1). R was used for model building, and J4.8 was used for the decision tree algorithm, implemented with "RWEKA", an R package WEKA.[28] For the random forest, we used the "randomForest" package. Training data were not sampled but used as-is. Unless noted otherwise, the hyper parameters for learning had the default values. The accuracy (error rate) of the created model was assessed using a 10-fold cross-validation.

Convolutional Neural Networks

A convolutional neural network (CNN) by the numerical analysis software MATLAB ® R2019 (Version: 9.6.0.1072779, The MathWorks, Natick, NJ) was used to classify particle images obtained using MFI and FlowCam. The convolutional neural network has 15 layers which are composed of a first image input layer, three blocks of four layers (convolution, batch normalization, rectified linear unit, and max pooling layers or fully connected layers), a fully connected layer, a softmax layer, and a classification output layer (Supplemental Figure 6-1).

The images obtained from FlowCam and MFI were cropped and resized to 32 pixels x 32 pixels x 3 channels and 32 pixels x 32 pixels x 1 channel, respectively. The identical number of the collected images was used to develop the classifier of each classification class, as shown in Supplemental Figure 6-2. 80% of images were randomly selected as training data and the other 20% were used as test data, and the process was repeated thrice. To evaluate the accuracy of the classification between the two devices, or among institutions, error rates of the classification between ETFE, NPAS70, NPAS90 and SO droplets (In laboratory D, the error rate when classifying between three types of particles excluding ETFE) were calculated for each device or laboratory.

Optimization of Decision Trees

We derived decision trees with multiple different condition settings to understand the overall error rate, the balance of error rate between SO droplets and protein aggregates, and the relationship to the decision tree depth. Specifically, for the training data for protein aggregates, decision trees were built while varying the two conditions of oversampling and the minimum number of cases to stop node branching. Finally, we proposed two types of optimal layers from the perspectives of balancing the error rate between classes and decision tree complexity, for both MFI and FlowCam (Figs. 5 and 6).

Evaluation of Classification Performance (Model Quality Indicators)

The classification performance of each filter or model was evaluated as the error rate (%), which was determined as the percentage of the number of false predictions for actual number of particles. It was also evaluated based on a confusion matrix, a tabulation of the predicted versus actual class for each particle. While ETFE was made of polymer and not protein, ETFE particles were classified as a protein class in this study. Counting errors were assessed as below:

Total error of classification [error rate (ALL)] = 1-accuracy=1-(TP+TN)/(TP+TN+FP+FN)

Misclassification rate of SO [error rate (SO)] = 1-specificity=1-TN/(TN+FP)

Misclassification rate of Protein [error rate (Protein)] = 1-precision=1-TP/(TP+FN)

	Actual Positive Protein	Actual Negative SO
Predicted Positive Protein	TP (True Positive)	FP (False Positive)
Predicted Negative SO	FN (False Negative)	TN (True Negative)

Results and Discussions

As shown in Fig. 1, in study 1, we developed the classifiers or models to classify SO droplets and protein particles, and compared them in terms of accuracy. Next, in study 2, focusing on decision trees developed using machine learning, we attempted to develop a standardized decision tree with balanced accuracy between classes and lower depth, to assess the feasibility of the standardization of classifier/model to classify SO and protein particles in biopharmaceuticals. Twelve laboratories contributed to this study. Some only provided model data, others were involved in the classifier/model development, while some contributed both.

Distribution of Morphology Parameters in Model Data Sets

Table 1 shows the model data set collected in this study, and Fig. 2 shows the representative particle images in each model data set. To grasp the characteristics of the model data set, we plotted the distribution of values of each particle characteristic parameter (e.g. aspect ratio, circularity, intensity, etc.) for each laboratory (Fig. 3, Supplemental Figure 1b). With MFI, the same values of each particle characteristic parameter were obtained among laboratories, while with FlowCam, large disparities of the values were observed among laboratories, depending on particle characteristic parameters, including compactness and perimeter. Also, looking at the actual particle images, similar images were obtained in the different laboratories with MFI, while different considerably appearances of images were confirmed in FlowCam among laboratories, possibly due to different resolution and focus settings. This tendency of difference in appearance was particularly pronounced in silicon oil droplets (Fig. 2).

Classification Performance of Each Filter or Model

Table 2 shows the particle characteristic parameters that appear useful for classification, as identified by each analysis method. The performance of the filter and classifier/model created, in terms of error rate, was shown in Table 3. The study results for each analysis method are stated below.

Principal Component Analysis

The classification of protein aggregates and SO droplets was performed using component 1, derived from principal component analysis. When classification was performed using the positive or negative

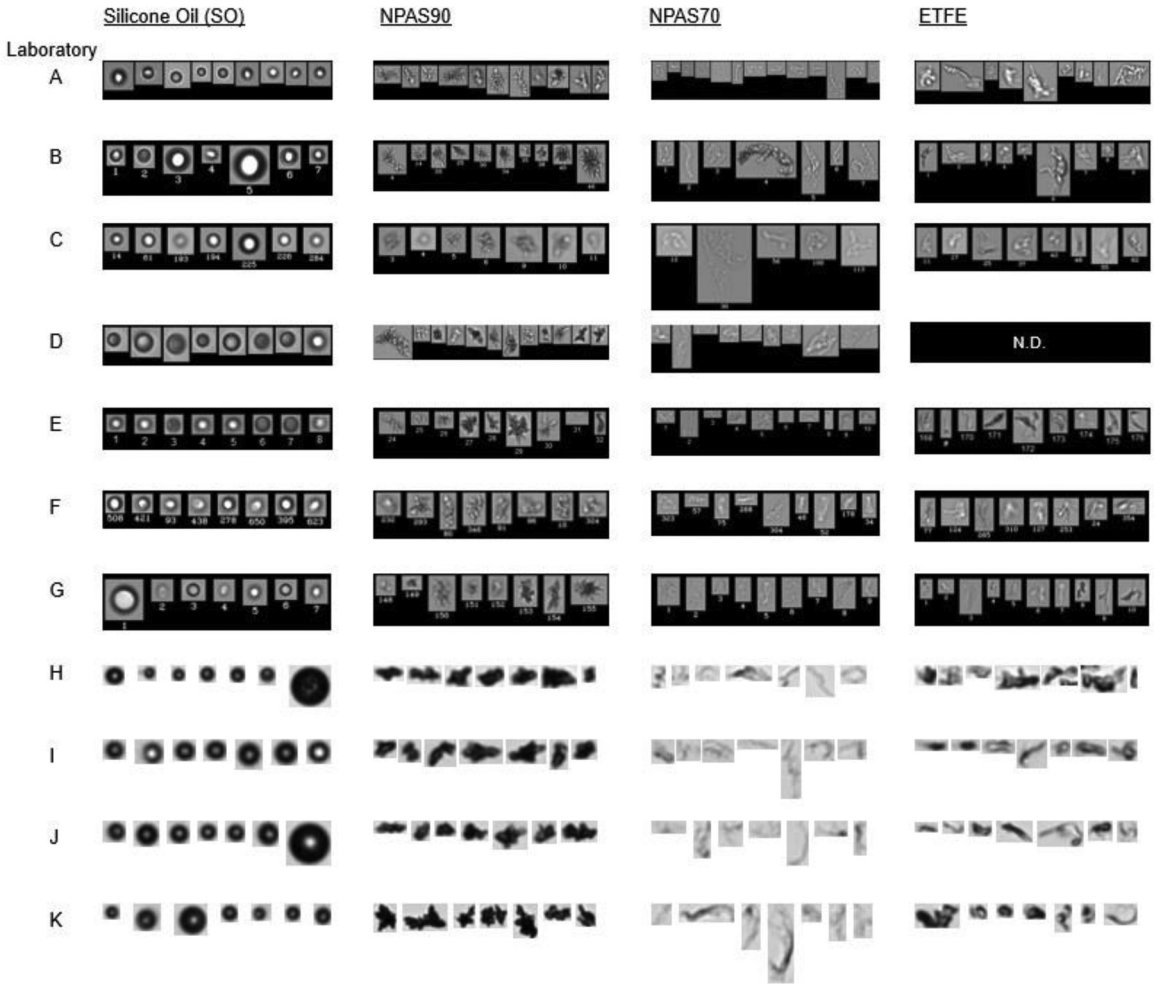


Figure 2. Representative images of particles detected in each sample. N.D.; not determined.

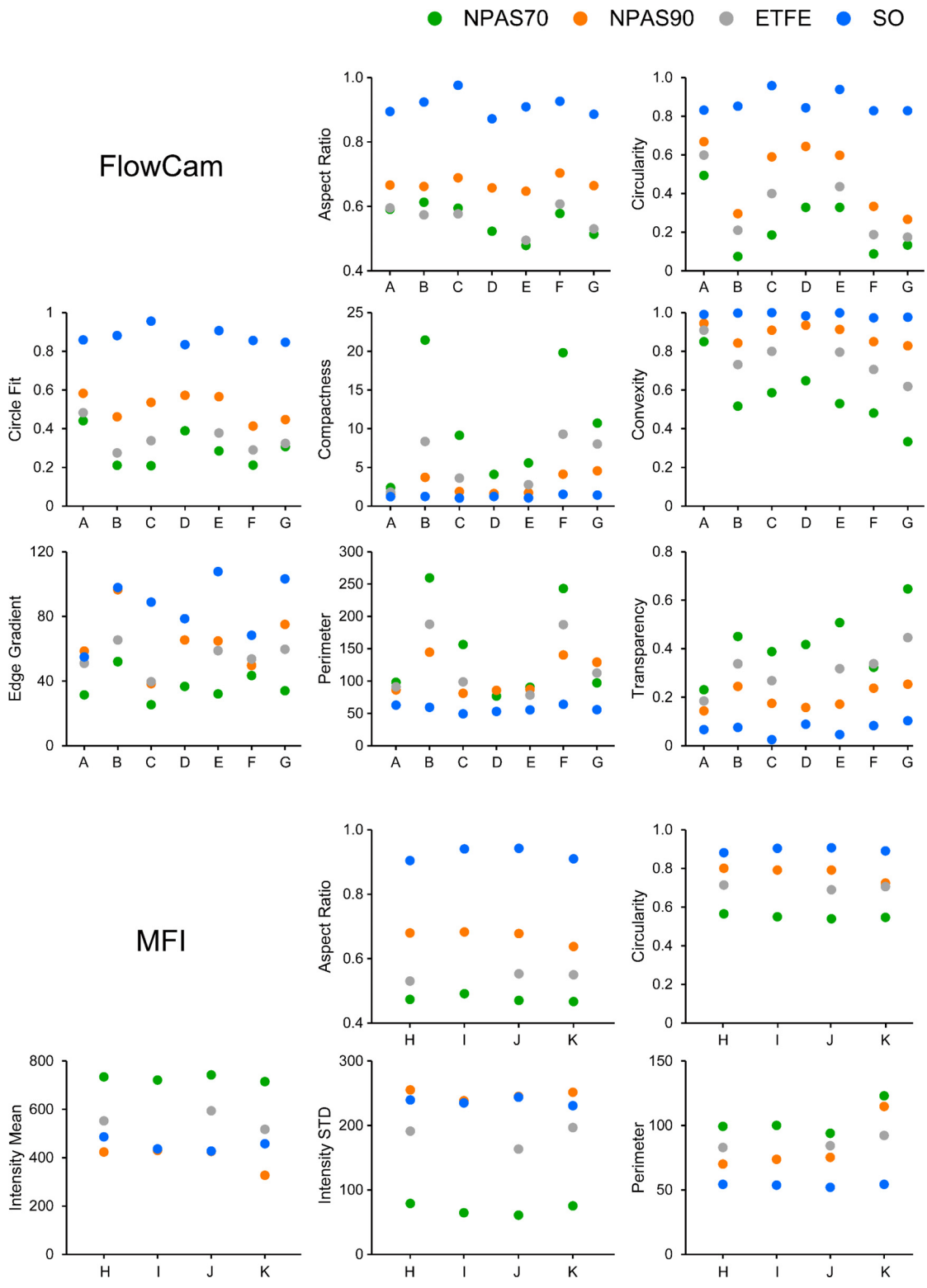


Figure 3. Representative particle characteristic parameters measured by each laboratory on model samples (SO, NPAS70, NPAS90 and ETFE). For FlowCam, measured values of aspect ratio, circularity, circle fit, compactness, convexity, edge gradient, perimeter and transparency are plotted on laboratory A-G. For MFI, measured values of aspect ratio, circularity, intensity mean, intensity STD and perimeter are plotted on laboratory H-K. All show mean values. The distribution of each parameter value is shown in Supplemental Fig. 1a.

value of this component 1, the accuracy was extraordinarily good for silicon oil droplets, with an error rate of approximately 2%, but the error rate for protein aggregates was approximately 14% (Table 3). This result indicates that the principal component analysis is a highly accurate analysis method for recognizing SO as SO. From component 1, we identified the useful parameters that determine the elements of the silicon oil droplets shown in Table 2. Because the error rate varied in the range of 13.6%-25.4% as the type of protein aggregate changed, it was found that the accuracy changed with the subjects being classified (Table 3). Since we found convexity to have a high

Table 2
List of extracted particle characteristic parameters to classify SO and protein partices in each analysis.

Laboratory	A	C	L	H	
Approach	PCA	Filter	Statistical analysis-Decision tree	Machine learning-Decision tree	
Instrument	FlowCam	FlowCam	MFI	FlowCam	MFI
Particle characteristic parameters	Convexity Circle fit Circularity (Hu) Symmetry	Aspect ratio Cricle fit	Aspect ratio Intensity Min Intensity STD	Circularity (Hu) Circle Fit Length Sigma intensity Intensity Width Edge gradient Geodesic aspect ratio Geodesic thickness Geodesic length Aspect ratio Roughness Length Area Area (filled) Volume (ESD) Perimeter	Circularity Aspect ratio Perimeter Intensity Min Intensity Max Intensity Mean Intensity STD Max Feral Diameter ECD

contribution to component 1, we created a decision tree in which the first node is set using convexity (Supplemental Figure 2-3). It was possible to create a decision tree that enables the classification of SO droplets with a low error rate of approximately 3.0%, depending on the type of protein aggregate. Therefore, it appears that, for classification using positive/negative for the principal component, the accuracy of the decision tree depends on the subjects being classified (Supplement Figure 2-3).

Classification Filter

The scatter diagrams of the parameters from the FlowCam data were created for the combinations of circle fit with aspect ratio as well as that of circularity with aspect ratio on each sample (Supplemental Figure 3_2a). The SO data distribution was concentrated in the area where the parameter values for circle fit and aspect ratio are close to 1. There was a similar trend for circularity, but with a larger variation than for circle fit and aspect ratio. While there were broad ranges of particle distributions in NPAS70, NPAS90, and ETFE, almost no distributions of particles with circle fit, aspect ratio, and circularity close to 1 were confirmed. Based on these results, SO and particles in each sample could be classified by setting a filter consisting of circle fit and aspect ratio. The optimum values for filter settings were examined. If a circle fit was 0.80 or more and aspect ratio was 0.78 or more, the error rates were 10.7% for SO, 0.7% for NPAS70, 2.1% for NPAS90, and 1.6% for ETFE (Table 3, Supplemental Figure 3_2b). Factors involved in the high SO error rate originated from the incorrect recognition of multiple SO particles as a single particle because of the inappropriate reorganization of the edges for a part of the SO images due to the out of focus. This incorrect recognition also caused a failure in the morphological evaluation of SO particles as circular, causing a high error rate. The same phenomenon might happen in the case of proteinous particles, while the raising circularity is an infrequent occurrence, resulting in a lower error rate.

Statistical Analysis-Decision Tree

From the scatter diagrams of particle characteristic parameters for silicon oil droplets and protein aggregates for MFI data in the model data set, aspect ratio, intensity Min, and intensity STD are the potential parameters useful for the classification (Table 2). Using the above parameters, statistical analysis software was used to optimize the decision tree. The condition for identifying protein particles is that 1, 2, or 3 below are satisfied, as shown in Supplemental Figure 4-2:

1. Aspect ratio <0.82
2. Aspect ratio ≥0.82 and Intensity Min ≥237
3. Aspect ratio ≥0.82 and Intensity Min <237 and Intensity STD (≥257.06)

The error rate for the above decision tree using learning data was 6.7 %, and the error rates were 13.3% for silicon oil droplets and 5.1% for protein particles. A decision tree with an error rate ≤10 % can be created. The error rate obtained using learning data was also confirmed using validation data. However, considering the particle characteristic parameters used, the optimum values of the parameters for silicon oil droplets do not change significantly among products, but

Table 3
Clasification performance of filter or models.

	Error rate (%)								
Laboratory	A	C	L	H				M	
Approach	PCA	Filter	Statistical analysis-Decision tree	Machine learning-Decision tree		Machine learning-Random forest		CNN	
Instrument	FlowCam	FlowCam	MFI	FlowCam	MFI	FlowCam	MFI	FlowCam	MFI
ALL	-	-	6.7	2.2	3.5	3.2	2.2	2.8	4.4
SO	2.1 – 2.6	10.7	13.3	4.2	8.9	9.9	6.5	-	-
Protein*	-	-	5.1	1.7	2.1	1.3	1.4	-	-
NPAS 70	13.6	0.7	-	0.3	0.3	0.3	0.0	-	-
NPAS 90	25.4	2.1	-	2.6	4.4	1.7	3.1	-	-
ETFE	23.1	1.6	-	1.6	1.6	1.8	1.2	-	-

* Protein: all of NPAS70, NPAS90 and ETFE.

Table 4
Clasification performance of CNN-based classifier between protein particles and SO on the each data set.

Instrument	FlowCam							MFI			
Laboratory	A	B	C	D	E	F	G	H	I	J	K
Error rate (%)	1.3	2.9	0.0	3.8	0.8	5.6	3.3	5.3	2.5	1.1	1.5

that the optimum values of the parameters for protein particles may differ for each product. Therefore, practically, it may be necessary to prepare learning data for each product to optimize (particle characteristic parameter selection, sequence of node, and threshold value setting) the decision tree for each product.

Machine Learning-Decision Tree and Random Forest

When classification models were built using the decision tree and random forest, the accuracy of these models was high enough, with error rates ≤5% for both MFI and FlowCam (Table 3).The bias trend between classes in accuracy was confirmed, where the error rate for SO droplets was ≤10% but higher than that for protein particles. There was also a tendency for the overall error rate for MFI to be lower with random forest, and for FlowCam to be lower with the decision tree. The decision tree depth was ten levels for MFI and eight levels for FlowCam (Supplemental Figure 5-2).

Convolutional Neural Networks: Classification Between Protein Particles and SO

The CNN classifiers were developed for the classification of SO droplet and protein particle (ETFE, NPAS70 and NPAS90) images obtained using MFI and FlowCam (Supplemental Figure 6-2a), and the error rates were used as indicators of the accuracy of the classification (Table 4). The error rate calculated for each laboratory was <6% for almost all laboratories despite the small number of data like tens to hundreds of images. This result shows the high accuracy of the CNN classification.

Generally, CNNs require hundreds to tens of thousands of data, and the classification accuracy becomes low when the number of data is insufficient. To investigate the effect of the number of images on CNN accuracy, the error rate was calculated for data sets composed of different numbers of particle images obtained using FlowCam at laboratory M (Supplemental Figure 6-3). The error rate was consistently ≤ 3% in the entire range from 20 to 6,000 images (10 to 3,000 images per class), suggesting that tens of images were enough to develop a classifier by CNN for SO droplets and protein aggregates ≥10 μm.

For the classification of SO and the other particles in this study, lower image sharpness does not necessarily cause the lower classification accuracy. This is also suggested by the result that classification accuracy of all MFI is better than that of all FlowCam. While SO images from laboratory C were blurry because of the previous model (VS-1) with lower resolution than the other models like 8100, the SO images look more uniform than those from other organizations that used FlowCam. The uniformity could cause the best classification accuracy of laboratory C.

Convolutional Neural Networks: Classification of Between ETFE, NPAS70, NPAS90 and SO

The classifier was developed to discriminate between each sample type: ETFE, NPAS70, NPAS90, and SO droplets (Supplemental Figure 6-2b). Although the relationship between the morphological characteristics of protein particles and immune responses remains unclear, if it is revealed in the future, classifying and monitoring the certain particles would bring more efficient quality and safety control. Additionally, characterizing particles contained in drug substances and products may be useful to establish a quality control strategy and evaluate manufacturing consistency. Consequently, the error rates calculated for each laboratory were between 9% and 31%, except for laboratory D. Regarding the type of instrument, the error rates were 18% for FlowCam, and 14% for MFI, resulting in a higher error rate compared to the classification between protein particles and SO (Table 5). As can be seen from the breakdown of the classification results, SO droplets were classified with high accuracy, and ETFE, NPAS70, and NPAS90 tended to be misclassified (Fig. 4).Particularly, it seemed that the classification of ETFE and NPAS70, as well as ETFE and NPAS90 were difficult. The distances between these classes based on extracted texture features from three types of particles (ETFE, NPAS70 and NPAS90) indicated that ETFE has texture features similar to NPAS70 and NPAS90 (Supplemental Figure 6-4). Additionally, in the PCA, subset of ETFE data were overlapped with NPAS70 and NPAS 90 (Supplemental Figure 6-5). These results indicate that the similarity between ETFE and NPAS70/90 resulted in the high error rates. The lower error rate (5%) was obtained for the data set of laboratory D, which did not contain the ETFE class. This also indicated that the similarity of ETFE to other particles causes lower classification accuracy.

To evaluate the variation in the error rates between laboratories that used FlowCam, the error rate of each laboratory was calculated for a varying number of images ranging from 140 to 1,000 (35 to 250 images per class). As an overall trend, the error rate varied depending on the laboratory even when the number of images used for the CNN analysis was exact. Especially, when 200 images or more were used, the gap attributed to the difference in laboratories rather than the difference in the number of images was observed. These results suggest that a small difference in images between laboratories affects the classification accuracy. However, as the number of images decreased, the error range of the error rate tended to increase, so it was difficult to accurately compare the classification accuracy between each laboratory.

Because the number of images used in this study was lower than that generally required for classification by CNN, transfer learning which was expected to be effective in improving the accuracy in such a case was conducted. When the images of laboratories A to C and E to G were classified using the pre-trained AlexNet, the classification accuracy for all institutions was improved (Supplemental Figure 6-6).

Table 5
Clasification performance of CNN-based classifier between ETFE, NPAS70, NPAS90, and SO on the each data set.

Instrument	FlowCam								MFI				
Laboratory	A	B	C	D	E	F	G	ALL	H	I	J	K	ALL
Error rate (%)	19.5	17.0	29.8	4.9	17.3	31.0	30.8	17.7	16.7	15.0	13.0	9.5	13.8

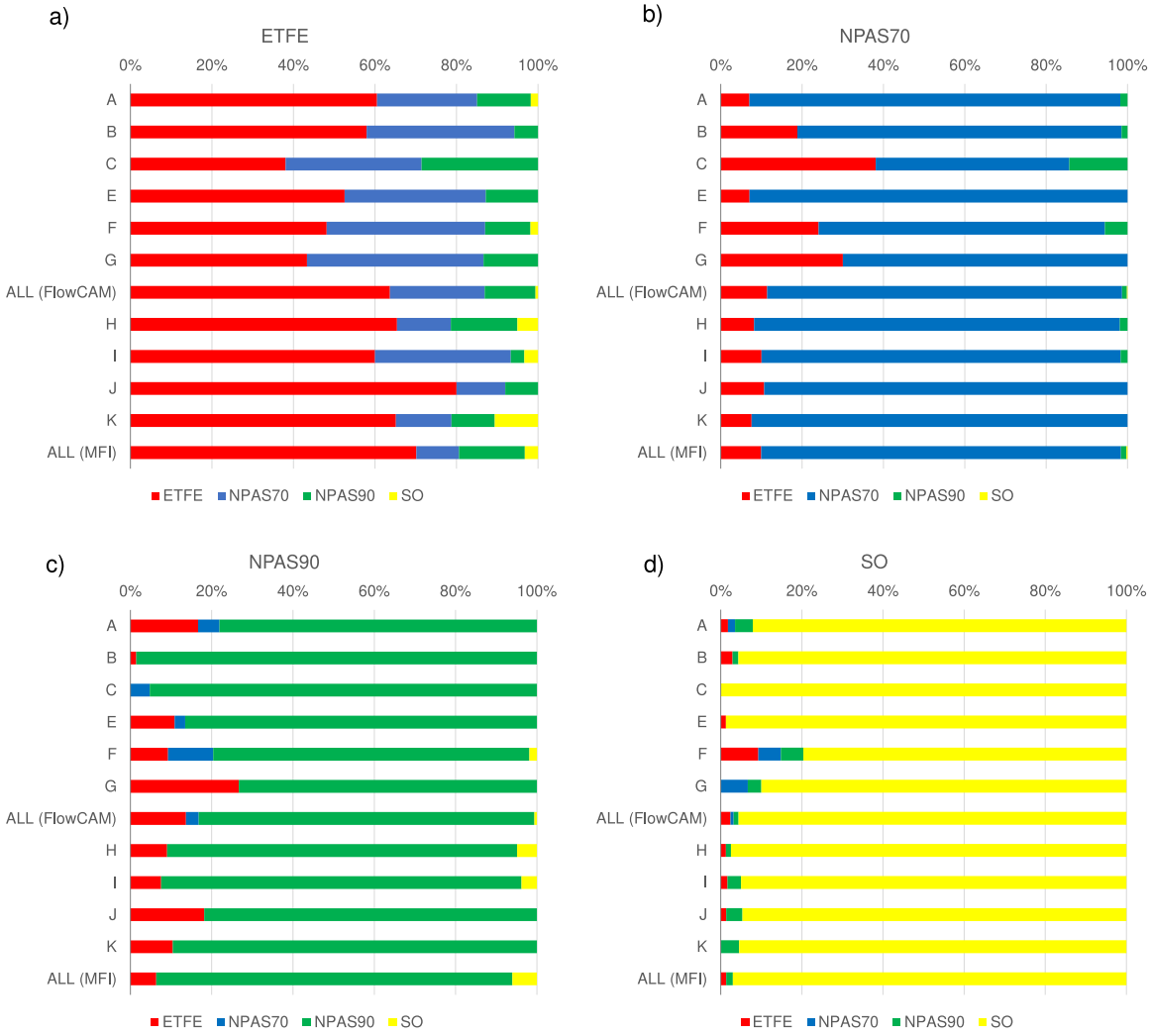


Figure 4. Ratio of predicted classes in each true class by the CNN-based classifiers; a) ETFE, b) NPAS70, c) NPAS90 and d) SO. (red: ETFE, blue: NPAS70, green: NPAS90, yellow: SO).

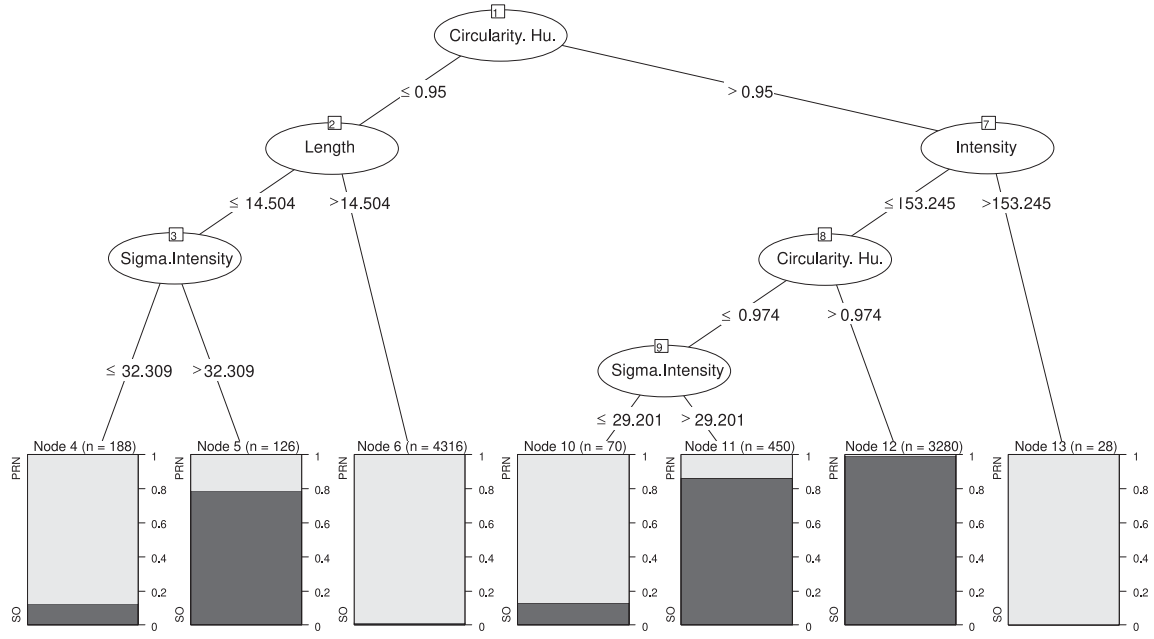
Additionally, the error ranges due to the change in the dataset combination decreased. Therefore, transfer learning is helpful for appropriate classification when classifying with limited data.

Building Streamlined Decision Trees

With the filter using two types of characteristic parameters, classification was possible with low error rates for protein particles, but the error rate for silicon oil droplets was high, at 10%. The same trend was observed even in decision trees optimized using statistical software. Conversely, it was confirmed that the overall error rate could be kept low by applying machine learning, and protein particles as well as SO droplets can be classified with higher accuracy, particularly for the FlowCam data set. Furthermore, we found that a high-accurate classifier/model could be built using CNN. Although CNN is an extremely powerful approach for classifying between SO droplets and protein particles, it would be difficult to share the classifier/model between laboratories due to its lower interpretability. Therefore, creating a universal decision tree using machine learning, was attempted to provide a standard analysis method. Discussion was divided on whether accuracy should be prioritized for SO droplets or protein particles. While protein aggregates are generally believed to be related to the immunogenicity of biopharmaceuticals, it was also reported that soluble and/or insoluble proteins accumulation around SO droplets may elicit undesired immune responses.[14,28,29] Continuously monitoring lot-to-lot consistency in protein particles and SO droplets would be crucial to ensure the quality and safety of prefilled syringe products. Therefore, in study 2, creating a decision tree with the balance of accuracy adjusted between the two was attempted from the perspective of quality control and safety assurance. A decision tree with numerous layers complicates analytical operations and interpretation. Therefore, the relationship between depth and accuracy was analyzed to examine whether depth could be reduced.

Concerning the balance of accuracy, it is generally known that, when the number of data is imbalanced between training data classes, accuracy is worse in the classes with fewer data [30]. Various resampling methods have been proposed as methods for adjusting the accuracy balance independent of the machine learning algorithm [31]. In this study, adjustment by relatively simple oversampling [32] was adopted, emphasizing on sharing and reproducibility between laboratories. Specifically, oversampling was performed for SO droplets, in which the number of training data was relatively low, to adjust by increasing the number of training data and evening the balance between classes in the number of data. Consequently, it was learned that the bias in accuracy can be solved by increasing the default for oversampling from x1 to x3 (Supplemental Figure 7). Concerning the depth of the decision tree, we assessed the balance between accuracy and layer depth by increasing the minimum number of cases (the branch is cut if the number of cases belonging to a

a) 4-level decision tree for FlowCam data



b) 3-level decision tree for FlowCam data

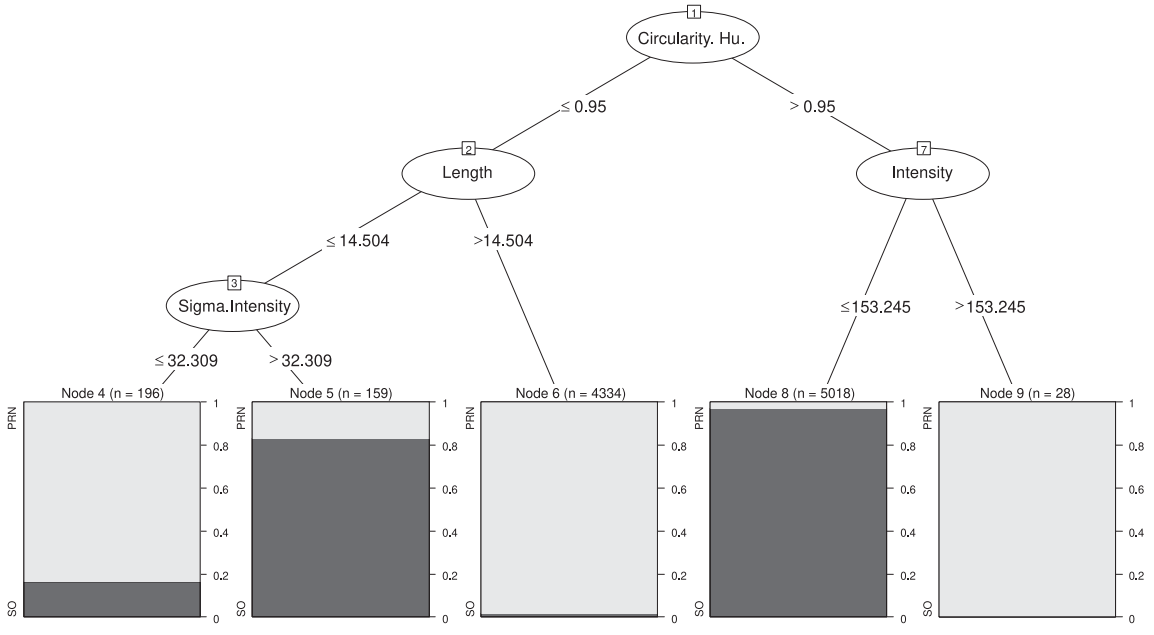
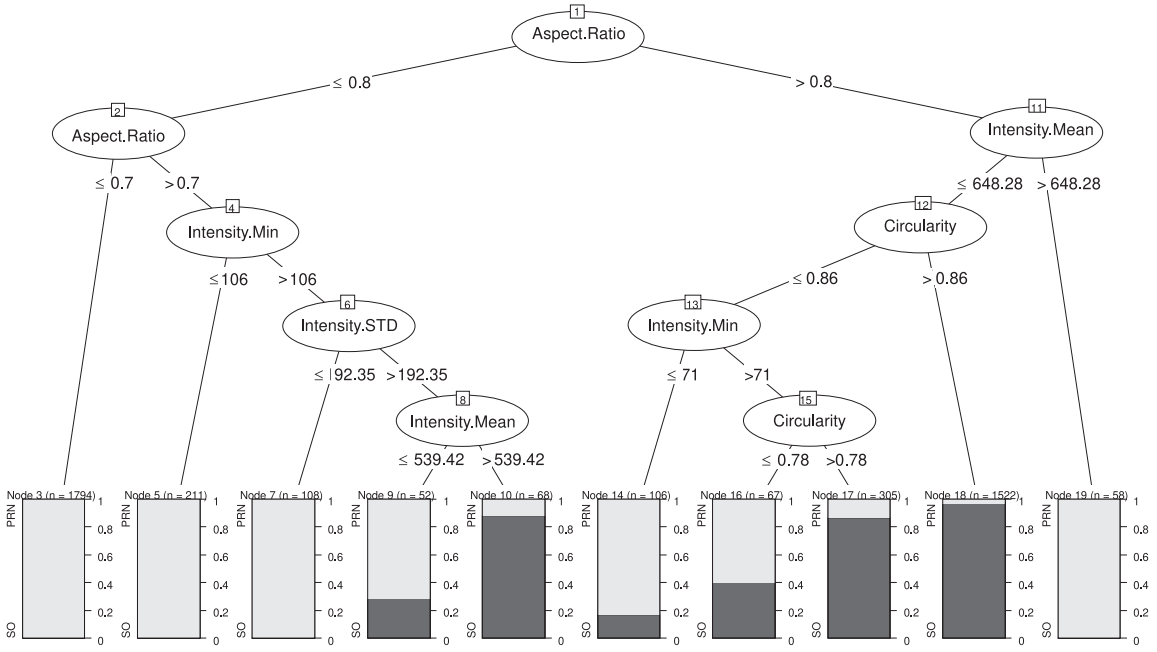


Figure 5. Two types of decision tree for FlowCam data constructed by machine learning.

node is lower than the minimum number of cases). Generally, raising the minimum number of cases is expected to make the decision tree more compact, but to degrade accuracy. Raising the minimum number of cases from the default setting of 2 to 150 reduced depth, and the overall error rate was observed to only increase by approximately 2%. At the beginning, the depth of the decision tree was 10-level and 8-level for MFI and FlowCam respectively. Ultimately, it was confirmed with MFI data that adopting oversampling = x3 and minimum number of cases = 50 or 150 could make the depth of the decision tree five or three layers, with an error rate of less than 10%. Fig. 5 (FlowCam) and Fig. 6 (MFI) show the decision trees with accuracy balanced and depth adjusted. For the FlowCam data, the condition

a) 5-level decision tree for MFI data



b) 3-level decision tree for MFI data

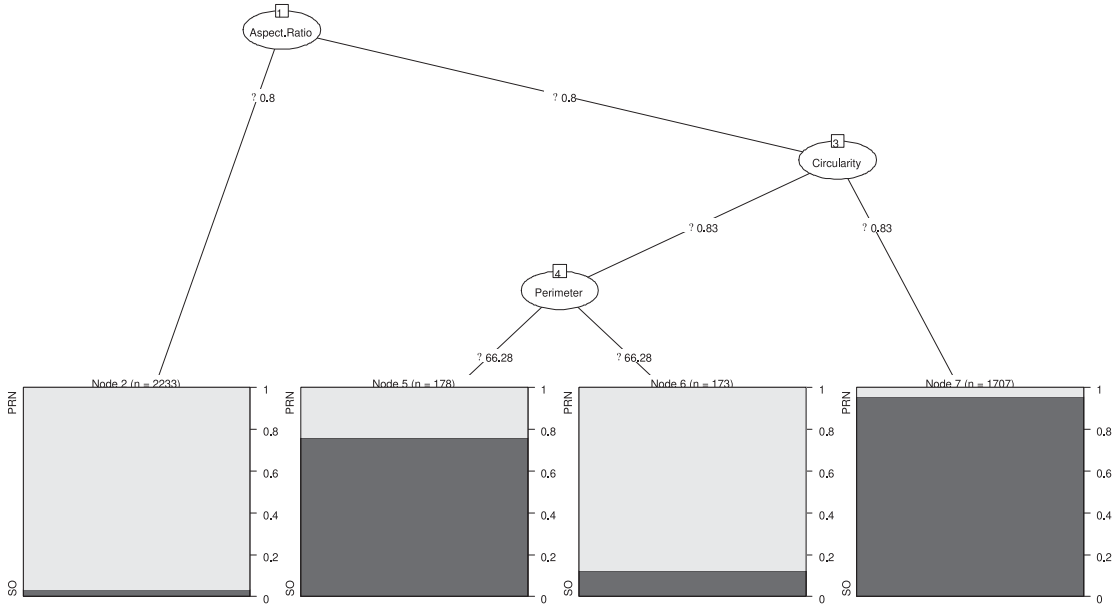


Figure 6. Two types of decision tree for MFI data constructed by machine learning.

was unsatisfied at five layers, so decision trees with three and four layers were derived. Tables 7 and 8 show the error rates during learning for each decision tree.

To validate the versatility of the created decision trees, each tree was used to classify (predict) test data measured at each laboratory (Table 6, Supplemental Table 7-2), and cross-referenced to actual results to assess predicted results (Tables 7 and 8). A trend was observed by which the error rate was mostly lower for the decision tree with more layers, for both FlowCam and MFI. When images of NPAS and ETFE particles that were subjects of training data, were used as arbitrary data, error rates were found to be kept down to ≤10% for both protein particles and silicon oil droplets, except for laboratories in which the characteristics of data differed from the other laboratories. For example, it appeared to be possible for shared

Table 6
Number of particles in each data set as test data.

Target instrument	Laboratory	Sample	Type of particles*	
			SO	Protein
FlowCam	A	NPAS, ETFE, SO	189	900
	B	NPAS, ETFE, SO	239	422
	C	Sample S (protein products)	9	205
	C	Sample L (protein products)	107	73
	D	NPAS, ETFE, SO	143	152
	E	NPAS, ETFE, SO	21	177
	E	Pre-filled syringe products	50	115
	F	NPAS, ETFE, SO	162	138
	G	Pre-filled syringe products	149	150
MFI	I	NPAS, ETFE, SO	75	225
	J	NPAS, ETFE, SO	80	240
	K	NPAS, ETFE, SO	107	200
	L	Protein products	79	619

* Particles equal or greater than 10 μm.

Table 7
Evaluation of 4-level decision tree on the FlowCam test data.

Laboratory Sample	Error rate (%) Traning data	A NPAS ETFE	B ETFE	C Protein products	C Protein products	D NPAS	E NPAS	E Pre-filled syringe products	F NPAS ETFE	G Pre-filled syringe products
4-level decision tree										
ALL	2.4	2.9	4.7	30.0	9.2	4.4	6.1	10.9	8.0	8.7
SO	2.4	3.2	1.2	12.3	9.5	6.6	6.8	7.8	2.9	1.3
Protein	2.3	1.6	10.9	42.1	0.0	2.1	0.0	18.0	12.3	16.1
3-level decision tree										
ALL	3.4	4.8	4.5	30.0	10.8	4.4	8.1	10.9	9.0	6.7
SO	3.7	5.6	1.2	13.7	11.1	6.6	9.0	10.4	5.1	1.3
Protein	2.0	1.1	10.5	41.1	0.0	2.1	0.0	12.0	12.3	12.1

decision trees to perform classification on the same samples at laboratories A, D, and E. Conversely, when validation was conducted using data for prefilled syringes that were unused in the training data, the error rates for overall and protein particles were found to be kept down to ≤10%, but the error rate for SO droplets tended to increase above 10%. Also, for other sample data, error rates were far higher than 10% in laboratory C and L, where validation used the data with characteristics very different from the training data. The factors for the high error rate at laboratory C were considered as follows: as there were many particles with smaller diameters in the silicon oil droplet data used as training data, “length” was used in the decision tree for classification between them and protein aggregates, but the data obtained at laboratory C included large silicon oil droplets, resulting in misclassification of those droplets as protein aggregates.

It was confirmed that the protein particles misclassified as silicon oil droplets had parameter values (e.g., circle fit or circularity) close to those of silicon oil droplets, and that, similarly, the silicon oil droplets misclassified as protein particles had values close to those of protein particles (Supplemental Figure 8a). The images of SO droplets misclassified as protein particles were observed like that circles were missing (Supplemental Figure 8b). After investigating the laboratory F data as data with different characteristics, it was found that changing node 3 in Fig. 5a from sigma intensity to sum intensity can improve the error rates for node 4 and 5, and that adding convex perimeter after node 6 can also slightly improve the error rate (Supplemental Figure 8c, d).

Points to Consider When Establishing the Model or Classifier Using Machine Learning to Differentiate SO Droplets and Protein Particles As Well As Their Maintenance Through the Lifecycle

It is tacitly assumed, not limited to decision trees, that the training data used in training for machine learning have the same distribution (independent and identically distributed: i.i.d) in their characteristics as the test data for classification (prediction) [33,34]. Within this validation experiment, it would be said that the assumption had broken down in the data for laboratory C and D. There are two main approaches to solve this problem. One is to prepare comprehensive and unbiased training data to match the distribution anticipated for the data that would be used in the actual classification, and then create a general-purpose decision tree. It is operationally superior to use the same decision tree regardless of the data to be classified, but the

Table 8
Evaluation of 5-level decision tree on the MFI test data.

Laboratory Sample	Error rate (%) Traning data	I NPAS, ETFE	J NPAS, ETFE	K NPAS	L Protein products
5-level decision tree					
ALL	3.8	5.0	3.1	4.2	59.0
SO	3.9	5.8	4.2	0.0	10.1
Protein	3.2	2.7	0.0	12.1	65.3
3-level decision tree					
ALL	5.0	5.7	5.3	5.5	62.0
SO	5.0	6.7	7.1	0.5	12.7
Protein	5.2	2.7	0.0	15.0	68.3

accuracy is generally inferior. The other approach is to prepare training data for each feature of data to be classified, and create a different decision tree for each, so that a decision tree based on training data matching the distribution of data to be classified can be used at the operation stage. That would raise accuracy but necessitate a decision tree selection for each data set. When operating a model based on either approach, it would be necessary to 1) visualize (if two dimensions are problematic, dimension reduction methods such as principal component analysis and multi-dimensional scaling are valuable) and compare the features of the training data and the data to be classified, or 2) assess the distance between the data distributions to periodically assess whether there has been any drift from the data used at the learning stage. Furthermore, when building any kind of model or classifier, it is necessary to explain background information such as the applicability domain of the model, the measurement instruments, conditions, etc., and to provide the model together with the training data employed to build it as it is assumed that the training data and the data to be classified have the same distribution. Also, it is critical to periodically assess and maintain the distributions of the training data and the data to be classified in actual operation.

Conclusion

In this study, we conducted a collaborative study among multiple laboratories using the shared model data set to standardize the analysis method for classifying silicon oil droplets and protein particles. As a result, the utility of classification using machine learning including CNN has been demonstrated. Since the data features of SO droplets and protein particles are likely to differ depending on the therapeutic protein product, it was considered difficult to create a universal decision tree for classifying SO droplets and protein particles. However, the optimized decision trees in this study (Figs. 5 and 6) exhibited good accuracy against the test data acquired at some laboratories with similar distribution to the training data. Therefore it seems possible to transfer a decision tree created by a certain laboratory to another laboratory in which the instruments of the same manufacture are used to classify SO and protein particles in shared samples. Comparing the CNN classifiers created by different laboratories using shared model data sets would be a next challenge. The particle characteristic parameters identified during the development of a classifier/model proposed in this study, would be useful information when building a classifier/model in each laboratory.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported in part by a Research Grant from the Japan Agency for Medical Research Development (AMED) by under Grant Number JP18ak0101074-21ak0101074. The authors would like to thank I. Takeuchi for technical assistance with the principal component analysis.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.xphs.2022.07.006.

References

1. Ecker DM, Jones SD, Levine HL. The therapeutic monoclonal antibody market. *mAbs*. 2015;7(1):9–14.
2. Kaplon H, Muralidharan M, Schneider Z, Reichert JM. Antibodies to watch in 2020. *mAbs*. 2020;12(1):1703531.
3. Mahler H-C, Friess W, Grauschopf U, Kiese S. Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci*. 2009;98(9):2909–2934.
4. Rosenberg AS, Verthelyi D, Cherney BW. Managing uncertainty: a perspective on risk pertaining to product quality attributes as they bear on immunogenicity of therapeutic proteins. *J Pharm Sci*. 2012;101(10):3560–3567.
5. Moussa EM, Panchal JP, Moorthy BS, et al. Immunogenicity of therapeutic protein aggregates. *J Pharm Sci*. 2016;105(2):417–430.
6. Fda US. *Guidance for Industry, Immunogenisity Assessment for Therapeutic Protein Products*. 2014.
7. Singh SK, Afonina N, Awwad M, et al. An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. *J Pharm Sci*. 2010;99(8):3302–3321.
8. Carpenter JF, Randolph TW, Jiskoot W, et al. Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. *J Pharm Sci*. 2009;98(4):1201–1205.
9. Makwana S, Basu B, Makasana Y, Dharamsi A. Prefilled syringes: an innovation in parenteral packaging. *Int J Pharm Investig*. 2011;1(4):200–206.
10. Yoshino K, Nakamura K, Yamashita A, et al. Functional evaluation and characterization of a newly developed silicone oil-free prefillable syringe system. *J Pharm Sci*. 2014;103(5):1520–1528.
11. Felsovalyi F, Janvier S, Jouffray S, Soukiassian H, Mangiagalli P. Silicone-oil-based subvisible particles: their detection, interactions, and regulation in prefilled container closure systems for biopharmaceuticals. *J Pharm Sci*. 2012;101(12):4569–4583.
12. Gerhardt A, Mcgraw NR, Schwartz DK, Bee JS, Carpenter JF, Randolph TW. Protein aggregation and particle formation in prefilled glass syringes. *J Pharm Sci*. 2014;103(6):1601–1612.
13. Kossovsky N, Heggers JP, Robson MC. Experimental demonstration of the immunogenicity of silicone-protein complexes. *J Biomed Mater Res*. 1987;21(9):1125–1133.
14. Krayukhina E, Yokoyama M, Hayashihara KK, et al. An Assessment of the ability of submicron- and micron-size silicone oil droplets in dropped prefillable syringes to invoke early- and late-stage immune responses. *J Pharm Sci*. 2019;108(7):2278–2287.
15. Ludwig DB, Carpenter JF, Hamel JB, Randolph TW. Protein adsorption and excipient effects on kinetic stability of silicone oil emulsions. *J Pharm Sci*. 2010;99(4):1721–1733.
16. Sharma DK, King D, Oma P, Merchant C. Micro-flow imaging: flow microscopy applied to sub-visible particulate analysis in protein formulations. *AAPS J*. 2010;12(3):455–464.
17. Kiyoshi M, Shibata H, Harazono A, et al. Collaborative study for analysis of subvisible particles using flow imaging and light obscuration: experiences in Japanese biopharmaceutical consortium. *J Pharm Sci*. 2019;108(2):832–841.
18. Strehl R, Rombach-Riegraf V, Diez M, et al. Discrimination between silicone oil droplets and protein aggregates in biopharmaceuticals: a novel multiparametric image filter for sub-visible particles in microflow imaging analysis. *Pharm Res*. 2012;29(2):594–602.
19. Zolls S, Weinbuch D, Wiggenhorn M, et al. Flow imaging microscopy for protein particle analysis–a comparative evaluation of four different analytical instruments. *AAPS J*. 2013;15(4):1200–1211.
20. Weinbuch D, Zolls S, Wiggenhorn M, et al. Micro-flow imaging and resonant mass measurement (Archimedes)–complementary methods to quantitatively differentiate protein particles and silicone oil droplets. *J Pharm Sci*. 2013;102(7):2152–2165.
21. Corvari V, Narhi LO, Spitznagel TM, et al. Subvisible (2-100 mum) particle analysis during biotherapeutic drug product development: Part 2, experience with the application of subvisible particle analysis. *Biologicals*. 2015;43(6):457–473.
22. Ekins S, Puhl AC, Zorn KM, et al. Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater*. 2019;18(5):435–441.
23. Ma X, Kittikunakorn N, Sorman B, et al. Application of deep learning convolutional neural networks for internal tablet defect detection: high accuracy, throughput, and adaptability. *J Pharm Sci*. 2020;109(4):1547–1557.
24. Calderon CP, Daniels AL, Randolph TW. Deep convolutional neural network analysis of flow imaging microscopy data to classify subvisible particles in protein formulations. *J Pharm Sci*. 2018;107(4):999–1008.
25. Gambe-Gilbuena A, Shibano Y, Krayukhina E, Torisu T, Uchiyama S. Automatic identification of the stress sources of protein aggregates using flow imaging microscopy images. *J Pharm Sci*. 2020;109(1):614–623.
26. Saggu M, Patel AR, Koulis T. A random forest approach for counting silicone oil droplets and protein particles in antibody formulations using flow microscopy. *Pharm Res*. 2017;34(2):479–491.
27. Ripple D, Telikepalli S, Steffens K, et al. Reference material 8634 ethylene tetrafluoroethylene for particle size distribution and morphology. *NIST Spec Publ*. 2019:193–260. https://doi.org/10.6028/NIST.SP.260-193.
28. Kiyoshi M, Tada M, Shibata H, Aoyama M, Ishii-Watabe A. Characterization of aggregated antibody-silicone oil complexes: from perspectives of morphology, 3D Image, and Fcgamma receptor activation. *J Pharm Sci*. 2021;110(3):1189–1196.
29. Nakajima T, Nagano K, Fukuda Y, et al. Subvisible particles derived by dropping stress enhance anti-PEG antibody production and clearance of PEGylated proteins in mice. *J Pharm Sci*. 2022;111(5):1363–1369.