



Using Machine Learning to Discover Shape Descriptors for Predicting Emulsion Stability in a Microfluidic Channel

Journal:	<i>Soft Matter</i>
Manuscript ID	SM-ART-10-2018-002054.R2
Article Type:	Paper
Date Submitted by the Author:	10-Dec-2018
Complete List of Authors:	Khor, Jian Wei; Stanford University, Mechanical Engineering Jean, Neal; Stanford University, Computer Science Luxenberg, Eric; Stanford University, Mathematics Ermon, Stefano; Stanford University, Computer Science Tang, Sindy; Stanford University, Mechanical Engineering

Using Machine Learning to Discover Shape Descriptors for Predicting Emulsion Stability in a Microfluidic Channel

Jian Wei Khor¹, Neal Jean^{2,3}, Eric S. Luxenberg⁴, Stefano Ermon³, and Sindy K.Y. Tang^{1*}

¹*Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA*

²*Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA*

³*Department of Computer Science, Stanford University, Stanford, CA 94305, USA*

⁴*Department of Mathematics, Stanford University, Stanford, CA 94305, USA*

[*sindy@stanford.edu](mailto:sindy@stanford.edu)

ABSTRACT

In soft matter consisting of many deformable objects, object shapes often carry important information about local forces and their interactions with the local environment, and can be tightly coupled to the bulk properties and subsequent functions. In a concentrated emulsion, for example, the shapes of individual droplets are directly related to the local stress arising from interactions with neighboring drops, which in turn determine their stability and the resulting rheological properties. Shape descriptors used in prior work on single drops and dilute emulsions, where droplet-droplet interactions are largely negligible and the drop shapes are simple, are insufficient to fully capture the broad range of droplet shapes in a concentrated system. This paper describes the application of a machine learning method, specifically a convolutional autoencoder model, that learns to: 1) discover a low-dimensional code (8-dimensional) to describe droplet shapes within a concentrated emulsion, and 2) predict whether the drop will become unstable and undergo break-up. The input consists of images ($N = 500,002$)

of two-dimensional droplet boundaries extracted from movies of a concentrated emulsion flowing through a confined microfluidic channel as a monolayer. The model is able to faithfully reconstruct droplet shapes, as well as to achieve a classification accuracy of 91.7% in the prediction of droplet break-up, compared with $\sim 60\%$ using conventional scalar descriptors based on droplet elongation. It is observed that 4 out of the 8 dimensions of the code are interpretable, corresponding to drop skewness, elongation, throat size, and surface curvature, respectively. Furthermore, the results show that drop elongation, throat size, and curvature are dominant factors in predicting droplet break-up for the flow conditions tested. The method presented is expected to facilitate follow-on work to identify the relationship between drop shapes and the interactions with other drops, and to identify potentially new modes of break-up mechanisms in a concentrated system. Finally, the method developed here should also apply to other soft materials such as foams, gels, and cells and tissues.

1. INTRODUCTION

Object shapes carry important information in many areas of soft matter, which often consists of deformable objects or components that can adopt a range of different shapes. For example, in foams and emulsions, the shapes of individual bubbles and droplets are directly related to the local stress and strain they experience, which in turn govern their tendency to undergo instability such as break-up or coalescence, and can in turn determine their bulk rheological properties.¹⁻⁵ As another example, the shapes of cells in biology often indicate their physiological stage, and are tightly coupled to many biological functions such as their mobility and disease state.⁶⁻⁸

In this paper, we focus on shapes that describe the instability of drops within a concentrated emulsion flowing as a two-dimensional monolayer in a tapered microfluidic channel leading to a constriction that fits one drop at a time. We choose this channel geometry because it is commonly used in the serial interrogation of droplet content in droplet microfluidics applications, where monodisperse drops serve as individual biochemical reactors in high-throughput screening applications.⁹⁻¹¹ Unlike their solid well counterparts, droplets are metastable and can undergo instability, which leads to the loss of integrity of the biochemical reactors and decreases the accuracy of the assay. Previously, we have shown that the onset of the instability of the drops, specifically the break-up of the drops, sets the upper limit in the throughput of the interrogation process.¹²⁻¹⁵ The break-up process occurs primarily between two drops attempting to enter the constriction at the same time. Whether break-up occurs depends strongly on the relative position between the two drops at the entrance of the constriction.¹⁵ If these two drops are separated by a small offset in the streamwise direction (i.e., they are more synchronized in their entry into the constriction), break-up is more likely to occur. We have found that there exists a critical offset between the two drops below which break-up always

occurs, and another critical offset above which no break-up occurs. A bistable region also exists: when the offset is in this region, both break-up and non-break-up events occur. The physical parameters determining break-up in this bistable region are currently unknown.

Here, we aim to explore factors that determine break-up in this bistable region. We focus on the shape of the drops since it is a direct measure of the local interactions among the drop, its neighbors, and the channel walls, and should be a good predictor of droplet outcome (to undergo break-up or stay intact). Indeed, in the extensive studies performed on single drops, explicit relations between droplet shape, the local stress and strain (typically expressed in terms of capillary number), and its outcome have been identified.¹⁶⁻²⁶ In these studies on single drops, the imposed physical constraints were relatively simple. For example, a simple shear or extension flow was applied to an isolated, single drop. As such, the shapes the drops could adopt were relatively simple, and varied primarily between spherical and elliptical shapes. It was thus sufficient to use simple scalar descriptors, such as the ratio of the major axis to the minor axis, to describe the shapes of these drops. In a concentrated emulsion confined in a microfluidic system, however, the drops are always in contact with their neighbors and the channel walls. Due to these interactions which are heterogeneous and time-dependent, individual drops adopt a much wider range of shapes than unconfined isolated drops do. Approximating these drops as ellipses and representing their shapes with major and minor axes alone is inadequate to capture the local physics each drop experiences. To our knowledge, no work thus far has reported shape descriptors to capture such a wide range of droplet shapes, nor to identify the relation of these shapes to droplet instability.

On the other hand, a suite of machine learning methods have been developed for computer vision applications involving the classification of complex shapes.²⁷⁻³³ Many of these

techniques are widely used for tasks like object recognition, but are not yet common in soft matter, especially in emulsion physics. Learning useful object features is one of the central goals of unsupervised machine learning. In recent years, many deep learning models based on a general autoencoding approach have been proposed and evaluated.³⁰⁻³³ Briefly, an autoencoder is a model that encodes inputs, such as images, into a lower-dimensional feature representation or “code” that can be decoded to reconstruct the input image.³² After the autoencoder is trained on the data, the learned code can be used for downstream tasks such as classification of the input data.

In this paper, we describe an autoencoder model that learns to: 1) discover a low-dimensional code (8-dimensional) to describe droplet shapes within a concentrated emulsion, and 2) predict whether the drop will undergo instability, i.e., break-up into small drops. Our input consists of images ($N = 500,002$) of two-dimensional droplet boundaries extracted from movies of the concentrated emulsion flowing through the microfluidic tapered channel as a monolayer; each image has an associated binary label specifying whether the drop breaks or stays intact. The input images are fed into an autoencoder. We choose to use a convolutional autoencoder architecture that borrows from the design principles presented in Turchenko *et al.*³³ Important architectural choices include: 1) the use of hyperbolic tangents as activation function rather than the more common rectified linear unit (ReLU), allowing for stable gradient propagation and optimization, and 2) the decision not to use unpooling layers, which would allow the model to “cheat” by passing shape information in a way that bypasses the low-dimensional encoding (see Turchenko *et al.* for further details). This second condition is essential for learning shape descriptors that capture the physical properties of the inputs. Our autoencoder consists of an encoder with 4 convolutional layers and 2 max pooling layers, and a decoder with 8 layers of

transpose convolutional layers. The convolutional, transpose convolutional, and max pooling layers contain trainable weights that the machine learning model modifies to learn features describing the drop shapes during the training process. We will not describe the details of these layers here — we refer readers to Goodfellow *et al.* for a review of necessary concepts instead.³² The code produced by the encoder of the autoencoder is then fed into a prediction model consisting of two fully connected layers that predicts the outcome of the drop, i.e., whether the droplet undergoes break-up or stays intact. Our model produces two outputs. The first output is a reconstructed droplet image from the autoencoder. The second output is a prediction of droplet outcome.

To our knowledge, this work is the first to show that it is possible to combine an autoencoder with a droplet outcome classifier to discover descriptors to represent a wide range of drop shapes, and to identify droplet shapes that predict break-up within a confined concentrated emulsion. We demonstrate that the new shape descriptor is capable of: 1) Successful reconstruction of droplet shapes capable of capturing complex local droplet curvatures. 2) Achieving a classification accuracy up to 91.7% in the prediction of droplet break-up.

2. METHODS

2.1 Device fabrication and droplet generation

We used methods in soft lithography to fabricate microchannels using poly(dimethylsiloxane) (PDMS).³⁴ The microchannels were rendered hydrophobic using Aquapel (Pittsburgh, PA) to prevent drops from wetting the wall.³⁵ 50 pL monodisperse drops were generated using flow-focusing nozzles (with drop volume dispersity less than 3%).³⁶ The disperse phase of the emulsion was deionized water, and the continuous phase was a hydrofluoroether HFE-7500 (3M, St. Paul, MN). The viscosity of HFE-7500 was approximately 1.24 cP at 20°C, the temperature at which all experiments were performed. The continuous phase contained an ammonium salt of Krytox (2% w/w) as surfactant to stabilize the drops against coalescence.³⁷ The generated drops were collected from the flow-focusing nozzles into a syringe. As water has a lower density than HFE-7500 ($\rho = 1.614 \text{ g/mL}$), the drops creamed to the top of the syringe to form an emulsion with volume fraction $\phi \sim 85\%$ after 6 hours of storage at room temperature. The drop's volume remained unchanged after this storage time.

2.2 Emulsion injection

Figure 1A shows the geometry of the tapered microchannel with a linearly decreasing width leading to a constriction downstream. The constriction had a width of 30 μm . The height of the channel was 30 μm . The half-angle of the taper was 30°. A syringe pump drove the emulsion flow at a constant volumetric flow rate, and an inverted optical microscope combined with a high-speed camera (Phantom v7.3, Vision Research) recorded videos of the emulsion flowing in the microchannel. Figure 1B shows an example of the break-up process. The applied flow rate was fixed at 400 $\mu\text{L/hr}$ where the proportion of drops that underwent break-up was 8.4%. The

corresponding capillary number calculated at the constriction was 7.00×10^{-3} . For section 3.4, we also performed experiments at a flow rate of $1000 \mu\text{L/hr}$ where the proportion of break-up was 33.2%. The corresponding capillary number calculated at the constriction was 1.75×10^{-2} .

To analyze the video data, we used custom *MATLAB* scripts to track the drops in each frame. The tracked drops were analyzed, and only those having an offset within the bistable region were retained.¹⁵ Detailed explanation of the custom *MATLAB* scripts can be found in our previous work.¹²⁻¹⁵ As all break-up events occurred only after the drop reached the entrance of the constriction ($x = 0$, Figure 1C), we used drop shapes when their leading edge reached $x = 0$ only.¹⁵ We identified the time point (or frame in the movie) when the droplet leading edge was at $x = 0$, and the boundary of the drop at this point was extracted. This drop boundary, along with its outcome (subsequent break-up or not), was stored into a dataset. The procedure was repeated for all drops in the bistable region. The data was then used to train and test the machine learning model.

2.3 Machine learning methods

All code for the machine learning model and data visualization was written in Python 3.0 and used the PyTorch library. The machine learning model has two main components: the autoencoder and the droplet outcome classifier. Figure 1C shows a scheme of the process flow of the model.

2.3.1 Input images and datasets

Our dataset contained the drop boundary (represented as a binary $210 \text{ pixels} \times 210 \text{ pixels}$ image) and the drop outcome with an equal number of drops that break and those that stay intact. The

datasets at 400 $\mu\text{L/hr}$ and 1000 $\mu\text{L/hr}$ each contained information about $N = 500,002$ droplets.

For all sections except 3.3 and 3.4, the training and testing were all performed on a single dataset from the same flow rate at 400 $\mu\text{L/hr}$.

2.3.2 Autoencoder

We used a convolutional autoencoder architecture that borrows from the design principles presented in Turchenko *et al.* with the following differences. 1) To accommodate our increased input image sizes, our model had 6 layers in the encoder and 8 layers in the decoder. 2) We found that using only deconvolutional layers in the decoder, instead of both deconvolutional layers combined with centralized unpooling layers, achieved minimal reconstruction loss.³³ Compared with the more rigid unpooling functions, deconvolutional layers have more flexibility in learning effective filters for reconstruction. In addition, this architecture allows us to output the descriptors separately before feeding them into the classifier. Outputting the descriptors separately achieves two objectives. First, we are interested in exploring unsupervised feature learning in our emulsion system. Unlike a supervised classifier, the autoencoder does not need labeled data (i.e., whether each drop breaks or not). We can thus take advantage of large quantities of unlabeled data that are easy to collect. It also opens up the possibility of transfer learning approaches (e.g., training an autoencoder at one flow rate and using the learned shape descriptors on another flow rate). Second, we are interested in learning a representation that could be considered a "shape descriptor" capable of capturing all information about drop shape. A supervised neural network could learn features that are useful for achieving high predictive accuracy, but potentially not useful for preserving shape information (i.e., for reconstruction),

especially if that information has no bearing on the downstream classification task (whether the drop breaks or not).

Figure S1 shows the details of the architecture of our model. The autoencoder has two components, the encoder and the decoder. The encoder consists of the following layers: *conv-16* \rightarrow *max* \rightarrow *conv-8* \rightarrow *max* \rightarrow *conv-4* \rightarrow *conv-2*, where *conv-N* is a convolutional layer with *N* channels and *max* is a max pooling layer. The decoder consists of: *uconv-4* \rightarrow *uconv-8* \rightarrow *uconv-16* \rightarrow *uconv-16* \rightarrow *uconv-8* \rightarrow *uconv-4* \rightarrow *uconv-1* \rightarrow *uconv-1*, where *uconv-N* is a transpose convolutional layer with *N* channels. In both the encoder and decoder, all transpose convolutional and convolution layers are followed by a *tanh* activation function. The convolutional and transpose convolutional layers contain trainable weights that the autoencoder modifies while learning to represent and reconstruct drop shapes during training. The loss function for the autoencoder is a mean squared error (MSE) loss defined in Eq. (1).

$$Loss_{MSE} = \frac{1}{N} \sum (X(i,j) - X_r(i,j))^2 \quad (1)$$

where $X(i,j)$ is the pixel intensity at location (i,j) for the input image, $X_r(i,j)$ is the pixel intensity at location (i,j) for the reconstructed image, and N is the total number of pixels in the image. The MSE loss directly compares the reconstructed image with the original input image by computing the mean of the squared differences between each pixel (also see Figure S2). We chose to use MSE loss here because it performed better in reconstruction than using a cross entropy loss.

The autoencoder learns to compress each input into a lower-dimensional continuous representation that can be used for reconstruction. The encoder network learns a mapping from the input space (i.e., images of drops) to the 8-dimensional shape descriptor space. In other words, the encoder learns to compress data, where the goal is to represent a high-dimensional array as a low-dimensional vector. The decoder network learns the inverse mapping from shape

descriptor back to image. We jointly train the encoder and the decoder to minimize the reconstruction loss, using gradient descent to update the parameters of both models at each iteration.

In our setting, we interpret the output of the encoder as a shape descriptor, the dimensionality of which is a hyperparameter that can be tuned. In our experiments, we found that an 8-dimensional shape descriptor was the smallest representation that allowed for accurate reconstruction of the original drop shapes (see details in the main text). The low-dimensional shape descriptor is then fed into the droplet outcome classifier.

2.3.3 Droplet outcome classifier

The droplet outcome (break-up or stay intact) classifier has two fully connected layers $8\text{-FC-}32 \rightarrow 32\text{-FC-}2$, where $M\text{-FC-}N$ is a fully connected layer with an input size of $M \times 1$ and an output size of $N \times 1$. A fully connected layer is given in Eq. (2).

$$y = wx + b \quad (2)$$

where y is the output, w is the weight, x is the input, and b is the bias. A fully connected layer is always followed by a rectified linear unit (ReLU) activation function in our model. A ReLU activation function f implements the rectifier $f(x) = \max(0, x)$. The loss function for the classifier is the cross entropy (CE) loss defined in Eq. (3).

$$Loss_{CE} = - (y \log p + (1 - y) \log(1 - p)) \quad (3)$$

where y is the class label (0=break, 1=intact) and p is the probability of the drop staying intact as predicted by the trained model. A threshold is determined by the machine learning model to classify drop outcome into break or intact.

2.3.4 Training paradigm

Unless stated otherwise, our learning algorithm used a training method where training was separated into two phases (Eq. (4)).

$$Loss = \begin{cases} L_1, & t \leq k \\ L_2, & t > k \end{cases} \quad (4)$$

where L_1 is the loss function at phase 1, L_2 is the loss function at phase 2, t is the training epoch and k is the epoch switching threshold where the loss computation was switched from L_1 to L_2 . The first phase of the training only involved minimizing the image reconstruction loss (MSE loss). After a predetermined epoch (see below for its identification), the training transitioned to the second phase, which involved minimizing the combined loss of both the image reconstruction loss (MSE loss) and the cross entropy loss for the droplet outcome classifier (CE loss). The combined loss was the sum of the MSE loss and the CE loss multiplied by an empirical scaling factor α to ensure they were of the same order of magnitude (Eq. (5)).

$$Loss_{combined} = Loss_{MSE} + \alpha \cdot Loss_{CE} \quad (5)$$

The value of α was determined through experimentation. We observed that if the empirical scaling factor was too high, the model overemphasized training the droplet outcome classifier component of the model to reach a minimal combined loss, but the image reconstruction did not improve while training. On the other hand, if the empirical scaling factor was too low, the model overemphasized training the image reconstruction component of the model to reach a minimal combined loss, but the prediction accuracy did not improve. We found that the image reconstruction did not improve qualitatively past 40 or 50 epochs (Figure S3A). As such, we chose 50 epochs to initiate phase 2 of the training. The model was trained using the ADAM optimizer, with learning rate determined via grid search and set at 0.001 (Figure S3B).³⁸ Briefly,

ADAM is a first-order gradient-based optimizer with adaptive learning rate widely used for training neural networks.³⁸

We initially attempted to optimize the combined loss in a single-phase training, but found that this approach resulted in a lower quality of image reconstruction. In contrast, the two-phase training paradigm achieved better reconstruction and better classification accuracy (91.7%) compared with the model optimized solely with the combined loss (89.6%) (also see Figure S2 and Table S1).

For section 3.3, we compared the accuracy of the model when we trained the autoencoder and the classifier with datasets using different number of dimensions of the shape descriptor. For these comparisons, we used training paradigms IIa – III as listed in Table 1. For section 3.4, we compared the accuracy of the model when we trained the autoencoder and the classifier with datasets from different flow rates. For this comparison, we used training paradigms III and IV as listed in Table 1. For all training paradigms, the number of data points used for training was always $N = 498,002$. The number of data points used for testing was always $N = 2,000$.

2.3.5 t-SNE plots

t-SNE is a data visualization method that projects high-dimensional data into two or three dimensions to allow for interpretable visualization.³⁹ In our case, we used t-SNE to project the 8-dimensional shape descriptor onto a 2-dimensional plane to visualize the distribution of drop shapes that break and stay intact at 400 $\mu\text{L/hr}$ and 1000 $\mu\text{L/hr}$ respectively. In this experiment, we trained an autoencoder on a dataset collected at the high flow rate of 1000 $\mu\text{L/hr}$, and then used it to encode both the dataset that it was trained on (i.e., 1000 $\mu\text{L/hr}$) and the dataset at a low flow rate of 400 $\mu\text{L/hr}$. We hypothesized that the diversity of drop shapes observed at the lower

flow rate would be less than that observed at the higher flow rate (see details in section 3.3). To generate the t-SNE plot, 1500 data points with equal number of drops that break and stay intact were sampled from each dataset through random permutation of the list of identification numbers assigned to individual images.

2.4 Scalar descriptors

To test the accuracy of simple scalar shape descriptors in predicting droplet break-up, we used a single dataset at a flow rate of 400 $\mu\text{L/hr}$ only. The scalar shape descriptors (SS_1 , SS_2 , and SS_3) were calculated for each drop using the following equations (Eq. (6)).

$$SS_1 = \frac{L - W}{L + W} \quad (6a)$$

$$SS_2 = \frac{L}{W} \quad (6b)$$

$$SS_3 = \frac{P}{\sqrt{4\pi A}} \quad (6c)$$

where L and W are the lengths of the major and minor axes of the drop assuming it is an ellipse; and P and A are the perimeter and area of the drop respectively. We split the data into two datasets, a training dataset (498,002 drops) and a test dataset (2,000 drops). Figure S4 shows the frequencies of occurrence of drops that underwent break-up and stayed intact as a function of the shape descriptor value. For each shape descriptor, we chose a threshold value that resulted in the lowest binary classification error on the training set. This classification threshold was then used to classify data in the test dataset. For example, drops with $SS_1 \geq SS_{1th}$ were predicted to break, and those with $SS_1 < SS_{1th}$ were predicted to stay intact. By comparing these predictions with the actual break-up results, we were able to calculate the test accuracies achieved using each shape descriptor.

3. RESULTS AND DISCUSSIONS

3.1 Reconstruction of droplet shapes

Figure 2A shows the MSE loss in reconstruction of droplet shapes after the two-phase training as a function of the number of dimensions for the shape descriptor (or the code). As can be seen, the MSE loss decreased with increasing number of dimensions used, but plateaued at 8 with a loss of $Loss_{MSE} \sim 0.0027$. As such, we chose to use a code with 8 dimensions for the rest of the paper.

While the 8-dimensional descriptor was able to represent the range of drop shapes in the concentrated emulsion with high fidelity, assigning physical meaning to each of the dimensions of the shape descriptor was difficult. In most object classification applications, high classification accuracy is the main goal, while physical interpretability of the learned representations is of secondary importance. Here, in order to derive physical interpretations of the shape descriptor, we perturbed the value of one dimension of the shape descriptor at a time while holding the other dimensions fixed, and reconstructed the images using the trained decoder to visualize the influence of each dimension on the drop shapes.

We were able to identify four dimensions of the shape descriptor that were interpretable. Figure 2B shows that they correspond to drop elongation, skewness, throat size, and surface curvature respectively. As we varied the value of the individual dimensions, the drop varied from a contracted form to an elongated form (Figure 2Bi); or from skewing to the left to skewing to the right (Figure 2Bii); or from one with a wide throat (a narrow region in the middle of the drop) to one with a narrow throat (Figure 2Biii); or from a more rounded shape to one with more sharp curvature or “corners” (Figure 2Biv). We note that training the autoencoder is a stochastic optimization procedure, and we would not expect to recover exactly the same shape descriptors

if we were to retrain the model. The same physical characteristics would likely still be captured by the new set of descriptors, though perhaps in different combinations among the 8 dimensions.

3.2. Prediction of droplet break-up: comparison between conventional scalar descriptor and 8D descriptor from machine learning

Even though the learned shape descriptor was not fully interpretable, it was much more effective than traditional scalar shape descriptors in predicting the break-up of the drops. Figure 3A and Table 2 compare the accuracy of prediction of droplet break-up using scalar descriptors based on droplet major and minor axes or normalized perimeter, versus that using the 8D descriptor identified from our machine learning model. We define accuracy in Eq. (7).

$$Accuracy = \frac{true\ intact + true\ break}{true\ intact + true\ break + false\ intact + false\ break} \quad (7)$$

As can be seen, the accuracies of the scalar descriptors were poor (57.6 – 59.5%), meaning they were only ~9% higher than a random guess. Figure 3B shows two arrays of 20 x 20 different drop shapes that are predicted to break or stay intact by the scalar descriptor SS_2 . The color of the drops represents whether they indeed break (red, “true break”) or stay intact (green, “true intact”). The ratio of green to red drops represents the accuracy of the prediction for the top array in Figure 3B, and the ratio of red to green drops represents the accuracy of the prediction for the bottom array in Figure 3B. The scalar shape descriptors essentially describe how elongated the drops are. These descriptors predict that the more elongated the drops, the more likely they are going to break. However, as can be seen, elongation alone was not effective in predicting whether the drop would break or not. Figure 3C further shows examples of droplet shapes that have different outcomes even though they have the same elongation.

On the other hand, the accuracy of the 8D descriptor was high at 91.7%. This high accuracy indicates that the shape of the drops incorporated sufficient information about the local stress fields arising from the interaction with other drops and the channel wall to predict its subsequent outcome, even though the shape of the drop continued to evolve past the point where its shape was extracted (at $x = 0$). This fact implies that whatever interaction that led to droplet break-up was already determined at the entrance of the constriction ($x = 0$), and was therefore captured by our machine learning model. In our previous studies, we have found that for the drops that underwent break-up, the actual droplet pinch-off often occurred when the leading edge of the drop was at $x = 0.5 - 1$ droplet diameter into the constriction.¹²⁻¹⁵ It would be the subject of a separate study, however, to examine the dependence of prediction accuracy on the time and corresponding location where the drop shape was extracted to train the model. We note that although it may not be surprising that our 8-dimensional shape descriptor outperformed the scalar descriptors, they were evaluated on the same predictive task and thus we believe the comparison is fair.

Figure 3D visualizes the drop shapes that break and those that stay intact. The color convention is the same as that in Figure 3B. In general, while the drops that break tend to appear more deformed from a circular shape than those that stay intact, there are subtlety that are not easy to identify by eye. To attempt to better identify features that make droplets more prone to break-up, we rely on the fact that the droplet classifier already achieved a high prediction accuracy of 91.7%. It is thus possible to feed the classifier with artificially generated drop shapes with controlled features such as droplet elongation, throat size, and surface curvature using an operation similar to that in Figure 2B, and ask the classifier to predict which of these generated shapes would break.

Figure 4A shows the prediction when droplet elongation increased horizontally (from left to right), and when the curvature of the drop increased vertically to contain more sharp corners (from top to bottom). Figure 4B shows the prediction when droplet elongation increased horizontally (from left to right), and when the throat size decreased (from top to bottom). As can be seen, elongated drops with sharp curvature and small throat sizes are predicted to be more prone to break-up than those that are more rounded.

These trends are expected since drops under no external stress are spherical in shape as held by interfacial tension between the drop and the outer fluid. The introduction of external stress leads to droplet deformation and shapes that deviate from a spherical one. The degree of deformation typically increases with the magnitude of the stress experienced by the drop. Beyond a critical stress level, interfacial tension can no longer sustain the drop's stable shape and the drop eventually breaks up. As such, drops with large deformation are generally more prone to break-up than those with small deformation. Depending on the type of flow and therefore the details of the stress field at the surface of the drop, the deformation can take many forms. Due to the interactions with the neighboring drops and the channel wall and the heterogeneity and time dependence in the local stress fields, the drops in our concentrated emulsion take on more forms of deformation (e.g., having a range of sharp curvatures and throats) than previous studies on single drops under simple shear or extensional flows. We note that while this paper does not focus on the mechanisms of break-up, the discovery of a new shape descriptor and the ability to determine shapes that are prone to break-up will lay the groundwork for mapping local interactions to resulting droplet shapes, as well as for potentially identifying new modes of break-up.

3.3. Prediction of droplet break-up: effect of the number of dimensions of descriptor

The results from Figure 2A indicate that 8 dimensions of the shape descriptor were needed for faithful reconstruction of droplet shape. Nevertheless, it was unclear if all 8 dimensions were equally important in predicting droplet break-up. To probe this question, we trained the droplet outcome classifier with a subset of the 8 dimensions and compared their prediction accuracy using training paradigm II (Table 1). For all cases here, training phases 1 and 2 were identical where all 8 dimensions of the descriptor were generated and used. Training phase 3 was added to train the classifier using a subset of the 8 dimensions only.

First, we trained the droplet outcome classifier in phase 3 with the four interpretable dimensions of the shape descriptor (i.e., drop elongation, drop skewness, drop throat size, and drop surface curvature) (training paradigm IIa, case A2 in Table 3). Figure 5 and Table 3 show that the prediction accuracy was 84.2%, 7.5% lower than the case when all 8 dimensions of the shape descriptor were used in the baseline case (training paradigm I, case A1 in Table 3). When we trained the model with the four “non-interpretable” dimensions (training paradigm IIb, case A3 in Table 3), the prediction accuracy was 64.3% only. These results indicate that the interpretable dimensions were more important in predicting droplet break-up than the non-interpretable dimensions, even though the latter still played a role in predicting break-up.

Second, we hypothesize that among the four interpretable dimensions of the descriptor, droplet skewness should not play a role in predicting droplet break-up as skewness or orientation was primarily determined by whether the drop entered the constriction from above or below the channel centerline. Indeed, when we trained the model with three interpretable dimensions of the shape descriptor (drop elongation, drop throat size, and drop surface curvature) only (training paradigm IIc, case A4 in Table 3), the prediction accuracy (84.9%) was comparable to the

accuracy (84.2%) when the model was trained with the four interpretable dimensions of the shape descriptor (training paradigm IIa, case A2 in Table 3). This result is expected, and shows that even though this dimension was needed for faithful droplet reconstruction, it was not important in determining drop break-up. Similar trends were observed when training was performed at a higher flow rate of 1000 $\mu\text{L/hr}$ (training paradigms IIa-IId, cases B2-B4 in Table 3).

Third, to check if droplet break-up can be predicted by an even smaller number of dimensions, we trained the drop outcome classifier with two of the three interpretable dimensions (drop elongation, drop throat size, drop curvature sharpness) only (training paradigms IIe-IIg, cases A5-A7 and B5-B7 in Table 3). The classification accuracy decreased significantly to $< 73\%$ in these cases. For the 400 $\mu\text{L/hr}$ dataset, drop elongation and throat size achieved slightly higher classification accuracy (72.9%) than other cases (67.7% and 67.8%). For the 1000 $\mu\text{L/hr}$ dataset, all combinations achieved approximately the same classification accuracies (64.6% - 67.3%). These results indicate that all three dimensions were important in predicting break-up, with drop elongation and throat size being slightly more important at the low flow rate case.

Fourth, we verified that using a single dimension of the shape descriptor was insufficient to predict break-up. The prediction accuracy of the drop outcome classifier decreased significantly to 58.8% - 61.9% (training paradigms IIh-IIj, cases A8-A10 in Table 3), and were not much better than the performance of conventional scalar descriptors $SS_1 - SS_3$. Similar trends were observed at a higher flow rate of 1000 $\mu\text{L/hr}$ (training paradigms IIk-IIm, cases B8- B10 in Table 3).

3.4. Comparison of model accuracy at different flow rates

Thus far we have focused on cases where both the training dataset and the testing dataset were derived from the same flow rate. Here, we are interested in seeing if the learned shape descriptor is also able to predict break-up at a flow rate different from the one that it is trained on.

Practically, this ability would be useful for reducing the burden of data collection, allowing data at different flow rates to be leveraged, and for saving time and reducing the need to retrain models for every new flow rate.

Figure 6A and Table 4 show the accuracies of the prediction for six cases, where the autoencoder and the droplet outcome classifier were trained and then tested using datasets at different flow rates. The details of the training paradigms are listed in Table 1. We can make three observations.

First, the baseline cases (training paradigm I) have the highest accuracy as expected, since both the autoencoder and the droplet outcome classifier were trained on data from the same flow rate as the test set. The reason for why the baseline case for 1000 $\mu\text{L/hr}$ had a lower accuracy than that at 400 $\mu\text{L/hr}$ is still under investigation. We suspect that it is due to, in part, the fact that break-up events tended to occur further downstream of the constriction than that at 400 $\mu\text{L/hr}$,¹² and the droplet shape at $x = 0$ was less predictive of its outcome. Second, the accuracy was the lowest when both the autoencoder and the droplet outcome classifier were trained at a flow rate different from that tested (training paradigm III). This result is expected, since the classification threshold to predict droplet break-up was different for each flow rate. Third, the accuracy improved (though still below the baseline level) when a third training phase was added to fine-tune the droplet outcome classifier on the same flow rate as the test set, even though the autoencoder was trained at a different flow rate (training paradigm IV).

The last observation is unexpected, but could be explained when we consider the similarity of drop shapes at the two flow rates at 400 $\mu\text{L/hr}$ and 1000 $\mu\text{L/hr}$ respectively. First, we verified that we were able to achieve a small MSE loss of ~ 0.0027 in image reconstruction using an 8-dimensional shape descriptor for the drop shapes extracted from the experiment at 1000 $\mu\text{L/hr}$. Second, using the same procedure for generating Figure 2B, we found that 4 dimensions of the descriptor also described drop elongation, skewness, throat size, and the surface curvature respectively. Third, while there is no simple way to compare quantitatively the 8D shape descriptors from the two flow rates, a t-SNE plot allows us to visualize how the two sets of drop shapes overlap (see details in section 2.3.5).

Figure 6B shows that the set of drop shapes at 400 $\mu\text{L/hr}$ significantly overlap, and appear to be a subset of, the set of drop shapes at 1000 $\mu\text{L/hr}$. This result indicates that drops at a high flow rate displayed a wider range of shapes than at a low flow rate, while including most of the drop shapes at a low flow rate. Indeed, this trend has been observed in our prior work,¹²⁻¹⁵ although we had only examined drop shapes using a scalar descriptor SS_3 , the normalized droplet perimeter. As flow rate increased, an increasing proportion of drops became highly deformed that underwent break-up, but there remained a finite proportion of drops with little deformation that stayed intact. This observation contrasts studies on single drops or dilute emulsions, where an increase in flow rate tends to increase the deformation for all drops. One would not expect the drop shapes at different flow rates to overlap as much as they did here. In our concentrated emulsion, the overlap in drop shapes implies that even when the applied flow rate was high, the configurations of the drops—and the resulting interactions among the drops—allowed for cases where the local strain remained small so that some drops did not undergo much deformation.

The large overlap in drop shapes at different flow rates is advantageous in facilitating the prediction of break-up at flow rates different from the ones that the autoencoder was trained in. The improvement in prediction accuracy using training paradigm IV suggests that the learned shape descriptors can generalize across different flow rates as long as we have some data with which to fine-tune our classification model. Practically, this result is useful as it allows one to perform the bulk of the training on a dataset which may have more data points than the test dataset. It also suggests that there is potential for applying the model to datasets collected at flow rates that have never been seen during training.

4. CONCLUSIONS

Unlike single drops or dilute emulsions, the many-drop interactions in a concentrated emulsion give rise to a wide range of drop shapes and complex phenomena that cannot be predicted easily. Previous shape descriptors and methods developed to examine drops in isolated or dilute systems are often inadequate to fully capture the diversity and complexity of the features. This work represents our first step in applying machine learning methods to identify a new shape descriptor to represent the broad range of drop shapes in a concentrated emulsion confined in a microfluidic system. By using an 8-dimensional shape descriptor, we show that droplet shapes can be reconstructed with high fidelity. The new shape descriptor is also capable of achieving a high classification accuracy in the prediction of droplet break-up. We further found that three of the 8 dimensions (droplet elongation, curvature, and throat size) were particularly important in predicting break-up. The classification accuracy is found to be lower for experiments performed at a high flow rate than that at a low flow rate, likely due to the shifting of the break-up location downstream of the constriction, thereby making the drop shape measured at the constriction

entrance less predictive of the outcome. Furthermore, we show that using a learned shape descriptor to predict break-up at a flow rate different from the one that it is trained on reduces the classification accuracy from the baseline case, where the training and test datasets are performed at the same flow rate. Nevertheless, including an additional training phase to minimize the classifier loss using the test flow rate restores the classification accuracy modestly. This result suggests that one can perform the bulk of the training on a dataset which may have more data points than the test dataset, and can reduce the burden to retrain the entire model for every new flow rate.

While this work did not aim to discover new physics, the methodology developed is expected to facilitate follow-on studies to understand how droplet-droplet interactions lead to the range of droplet shapes observed, as well as how they act together to cause the break-up of droplets. Ongoing work includes the evaluation of advanced machine learning models to quantify the evolution of drop shapes and break-up conditions across different flow rates. Finally, we note that the method developed here is not limited to concentrated emulsions, but also applies to other soft materials such as foams, gels, and cells and tissues.

ACKNOWLEDGEMENT

The work is supported by the National Science Foundation through the NSF CAREER Award No. 1454542. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152.

References

1. H. M. Princen and A. D. Kiss, *J. Colloid Interface Sci.*, 1986, **112**, 2, 427-437.
2. H. M. Princen and A. D. Kiss, *J. Colloid Interface Sci.*, 1989, **128**, 1, 176-187.
3. H. M. Princen, *J. Colloid Interface Sci.*, 1983, **91**, 1, 160-175.
4. H. M. Princen, *J. Colloid Interface Sci.*, 1985, **105**, 1, 150-171.
5. T. G. Mason, *Curr. Opin. Colloid Interface Sci.*, 1999, **4**, 231-238.
6. P. Friedl and K. Wolf, *Nat. Rev. Cancer*, 2003, **3**, 362-374.
7. N. Reymond, B. B. d'Agua and A. J. Ridley, *Nat. Rev. Cancer*, 2013, **13**, 858-870.
8. A. Mogilner and K. Keren, *Curr. Biol.*, 2009, **19**, R762-771.
9. L. Rosenfeld, T. Lin, R. Derda and S. K. Y. Tang, *Microfluid Nanofluid*, 2014, **16**, 921-939.
10. J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, D. A. Weitz, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 9, 4004-4009.
11. M. T. Guo, A. Rotem, J. A. Heyman and D. A. Weitz, *Lab Chip*, 2012, **12**, 2146-2155.
12. L. Rosenfeld, L. Fan, Y. Chen, R. Swoboda, S. K. Y. Tang, *Soft Matter*, 2014, **10**, 421-430.
13. Y. Gai, J. W. Khor and S. K. Tang, *Lab Chip*, 2016, **16**, 3058-3064.
14. Y. Gai, M. Kim, M. Pan and S. K. Y. Tang, *Biomicrofluidics*, 2017, **11**, 034117-1-17.
15. J. W. Khor, M. Kim, S. S. Schütz, T. M. Schneider and S. K. Y. Tang, *Appl. Phys. Lett.*, 2017, **111**, 124102-1-4.
16. G. I. Taylor, *Proc. Royal Soc. Lond.*, 1932, **138**, 41-48.
17. G. I. Taylor, *Proc. Royal Soc. Lond.*, 1934, **146**, 501-523.

18. D. R. Link, S. L. Anna, D. A. Weitz and H. A. Stone, *Phys. Rev. Lett.*, 2004, **92**, 5, 054503-1-4.
19. H. A. Stone, *Annu. Rev. Fluid Mech.*, 1994, **26**, 65-102.
20. H. A. Stone, B. J. Bentley and L. G. Leal, *J. Fluid Mech.*, 1986, **173**, 131-158.
21. H. A. Stone and L. G. Leal, *J. Fluid Mech.*, 1990, **220**, 161-186.
22. H. A. Stone and L. G. Leal, *J. Fluid Mech.*, 1989, **198**, 399-427.
23. A. Acrivos, *Ann. N. Y. Acad. Sci.*, 1983, **404**, 1-11.
24. A. Acrivos and T. S. Lo, *J. Fluid Mech.*, 1978, **86**, 641-672.
25. B. J. Bentley and L. G. Leal, *J. Fluid Mech.*, 1986, **167**, 241-283.
26. H. P. Grace, *Chem. Eng. Commun.*, 1982, **14**, 225-277.
27. S. Belongie, J. Malik, J. Puzicha, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, 24, 509- 522.
28. H. Su, S. Maji, E. Kalogerakis and E. Learned-Miller, *arXiv:1505.00880v3 [cs.CV]*, 2015, 1-12.
29. C. Doersch, *arXiv:1606.05908v2 [stat.ML]*, 2016, 1-23.
30. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, *J. Mach. Learn. Res.*, 2010, **11**, 3371-3408.
31. D. P. Kingma and M. Welling, *arXiv:1312.6114v10 [stat.ML]*, 2014, 1-14.
32. I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, 2016.
33. V. Turchenko, E. Chalmers, A. Luczak, *arXiv:1701.04949v1 [cs.NE]*, 2017, 1-21.
34. Y. Xia and G. M. Whitesides, *Annu. Rev. Mater. Sci.*, 1998, **28**, 153-184.
35. A. R. Abate, D. Lee, T. Do, C. Holtze and D. A. Weitz, *Lab Chip*, 2008, **8**, 516-518.

36. S. L. Anna, N. Bontoux and H. A. Stone, *Appl. Phys. Lett.*, 2003, **82**, 3, 364-366.
37. C. Holtze, A. C. Rowat, J. J. Agresti, J. B. Hutchison, F. E. Angile, C. H. Schmitz, S. Koster, H. Duan, K. J. Humphry, R. A. Scanga, J. S. Johnson, D. Pisignano and D. A. Weitz, *Lab Chip*, 2008, **8**, 1632-1639.
38. D. P. Kingma and J. L. Ba, *arXiv:1412.6980v9 [cs.LG]*, 2015, 1-15.
39. L. v. d. Maaten and G. Hinton, *J. Mach. Learn. Res.*, 2008, **9**, 2579-2605.

Figure 1. **A.** Scheme of the tapered microfluidic channel. **B.** A series of snapshots showing the break-up process of the drop highlighted in red within the concentrated emulsion. **C.** Process flow of our machine learning model.

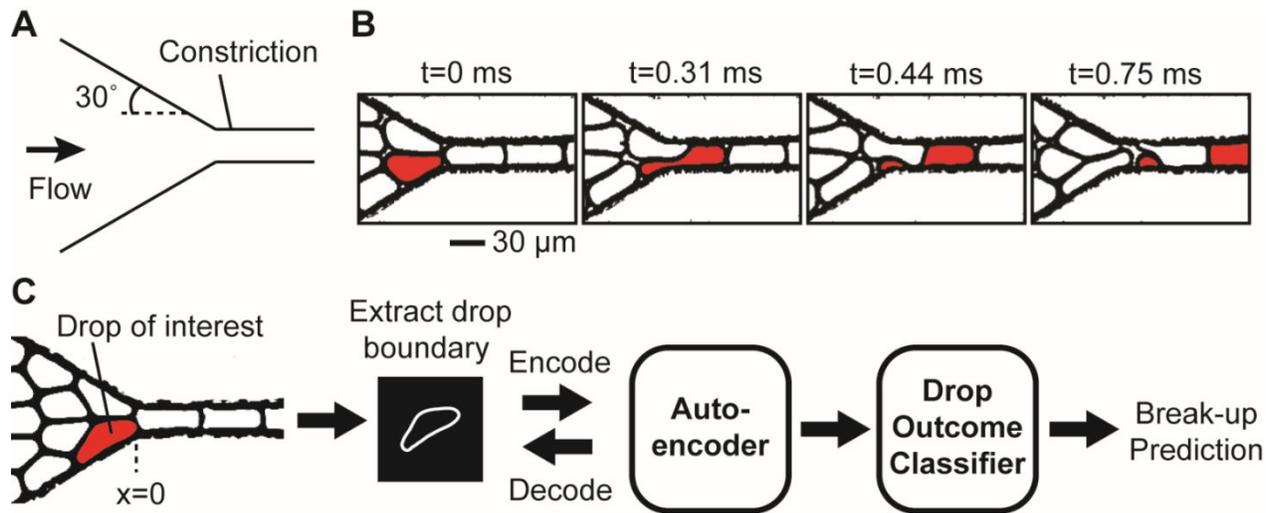


Figure 2. A. MSE loss as a function of the number of code dimensions. The images in the inset show examples of images before and after reconstruction for a (i) 2-dimensional, (ii) 4-dimensional, and (iii) 8-dimensional code respectively. **B.** Visualization of drop reconstructions to show how drop shapes change as the value of individual dimensions of the shape descriptor is perturbed (4 dimensions shown). Each dimension of the shape descriptor has an approximately normal distribution; on the horizontal axis, 0 represents the mean value and Δ represents one standard deviation from the mean for the dimension being perturbed. The borders of the drops were thickened after image reconstruction to facilitate visualization.

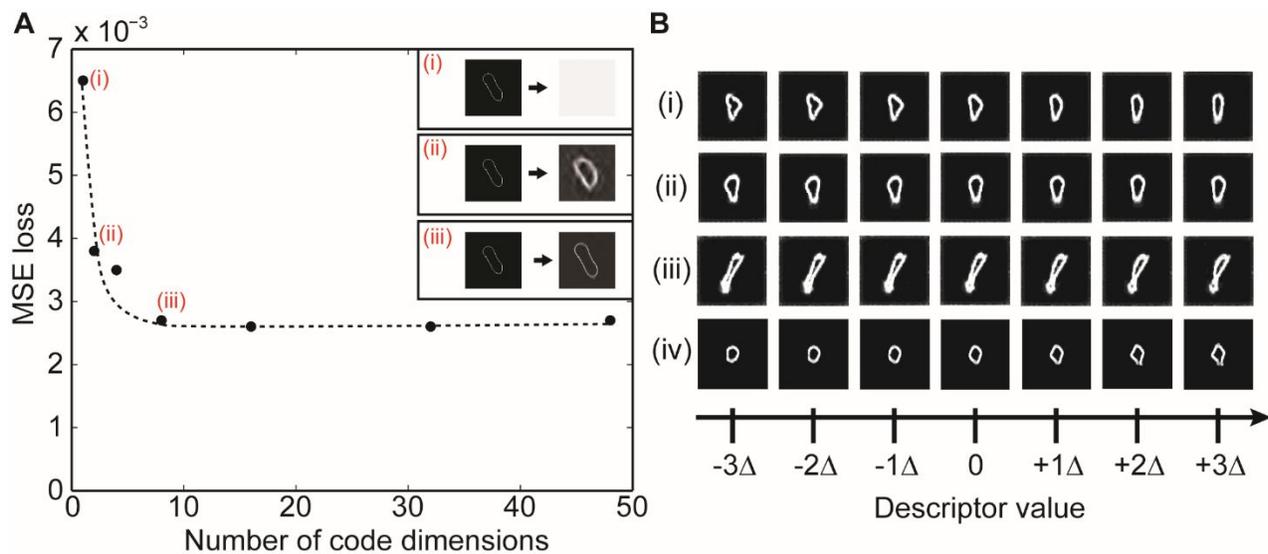


Figure 3. **A.** Prediction accuracies of scalar descriptors SS_1 , SS_2 , SS_3 , and our machine learning model respectively. **B.** Montage of 20 x 20 drop shapes predicted by SS_2 to stay intact (top) and break (bottom). **C.** Drop shapes that were predicted to break or stay intact by SS_2 . **D.** Montage of 20 x 20 drop shapes predicted by our model to stay intact (top) and break (bottom). For **B**, **C**, and **D**, drops colored in green represent true intact drops, and drops colored in red represent true broken drops.

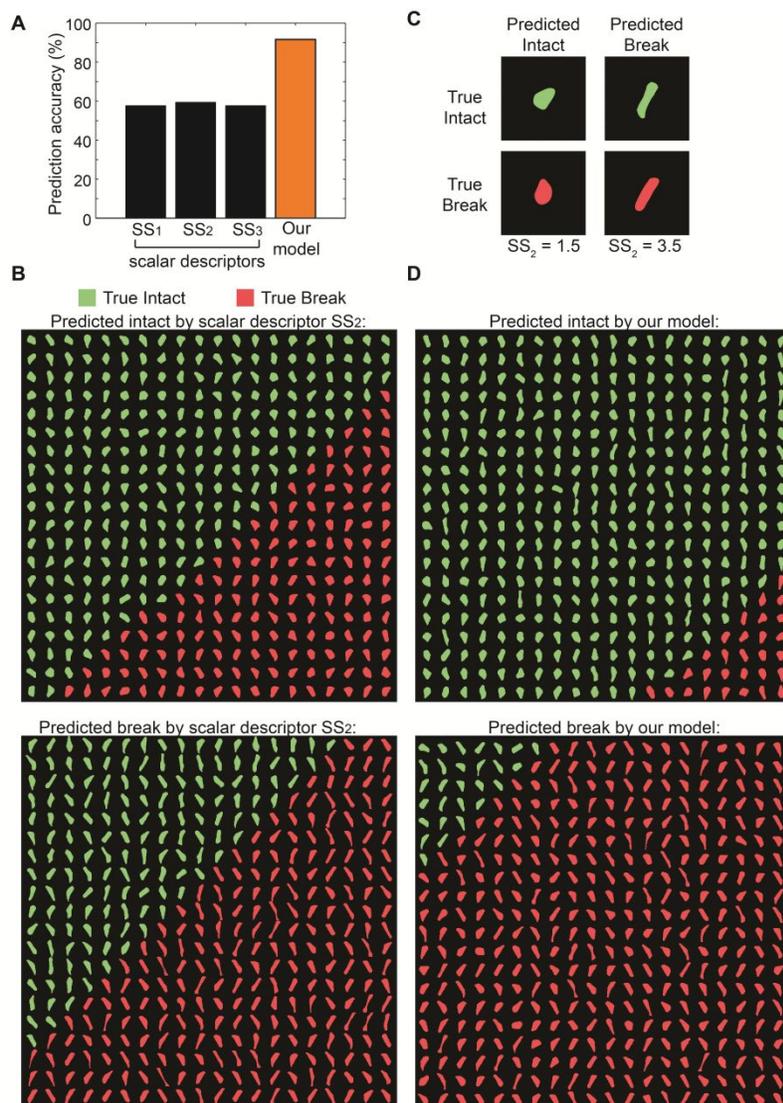


Figure 4. A. Perturbation of shape descriptor values with increasing drop elongation (left to right) and curvature (top to bottom). **B.** Perturbation of shape descriptor values with increasing drop elongation (left to right) and decreasing throat size (top to bottom). Drops that are colored green were predicted to stay intact and drops that are red were predicted to break by our model.

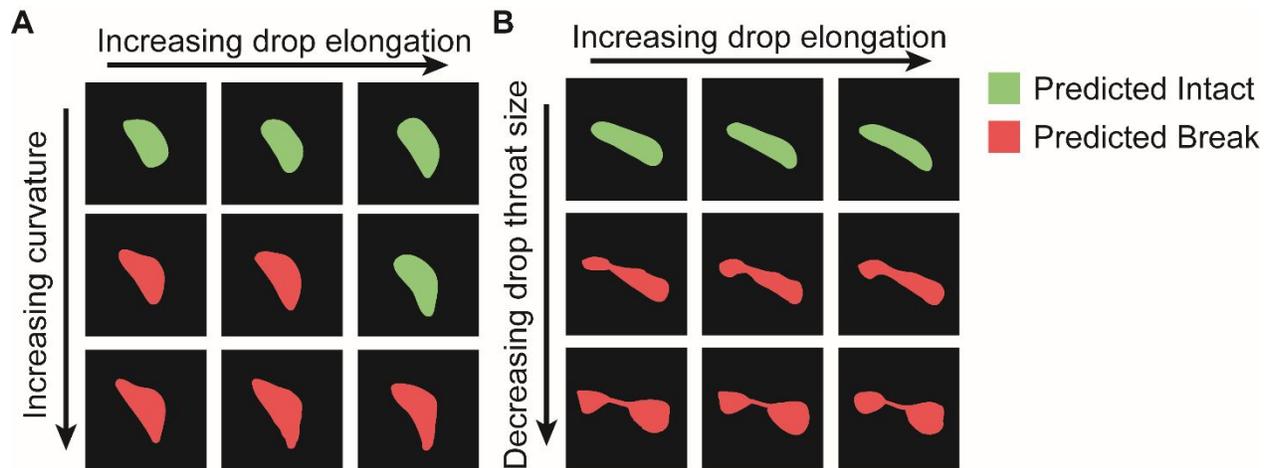


Figure 5. Prediction accuracy of our model when the droplet outcome classifier was trained with different number of dimensions of the shape descriptor (training paradigms I and II in Table 1 and Table 3) with 400 $\mu\text{L/hr}$ (orange) and 1000 $\mu\text{L/hr}$ (blue) datasets respectively. Also see text for details. We define drop elongation as DE, drop throat size as DT, and drop surface curvature as DS.

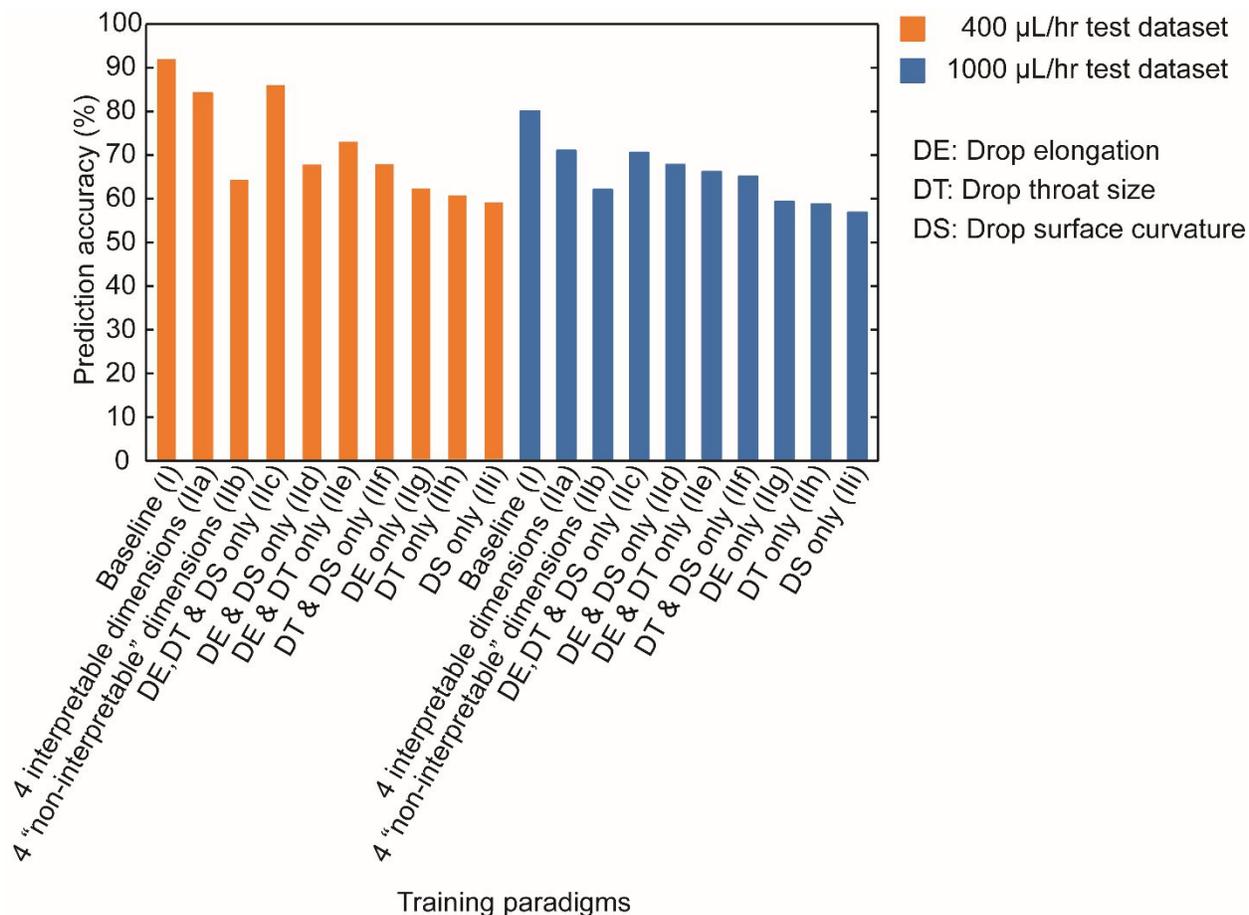


Figure 6. A. Prediction accuracy of our model using training paradigms I, III, and IV tested with 400 $\mu\text{L/hr}$ (orange) and 1000 $\mu\text{L/hr}$ (blue) datasets respectively. **B.** t-SNE plot of 400 $\mu\text{L/hr}$ dataset (orange; $N = 1500$) and 1000 $\mu\text{L/hr}$ dataset (blue; $N = 1500$) encoded by autoencoder trained with 1000 $\mu\text{L/hr}$ dataset (see section 2.3.5 for details).

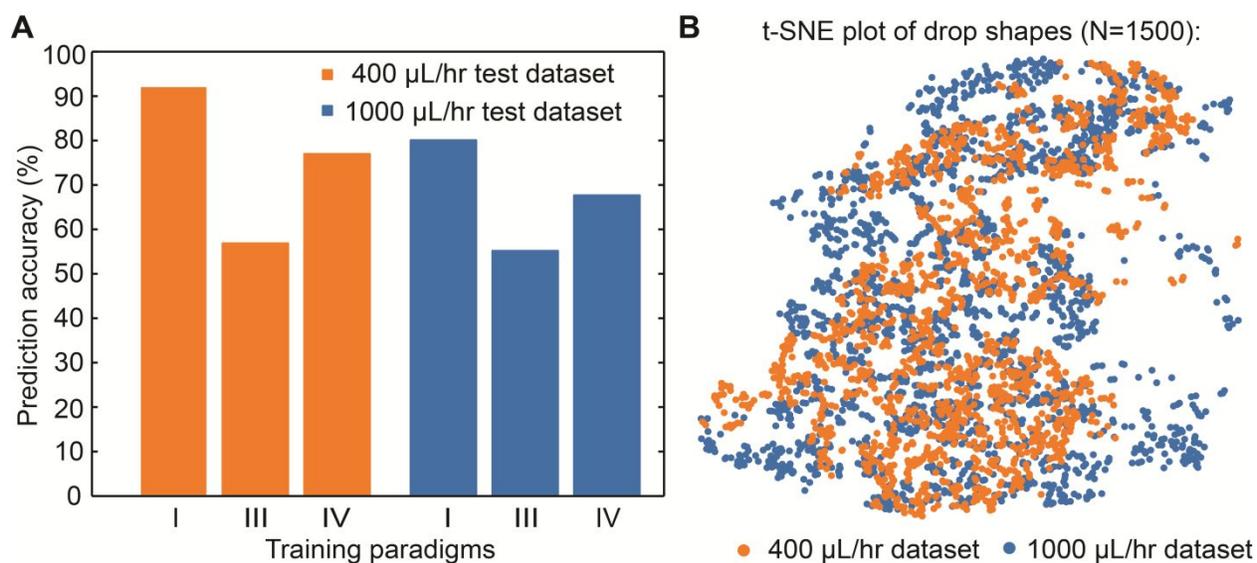


Table 1. Definitions of different training paradigms for our machine learning model.

Training paradigm	Training dataset			Testing dataset
	Phase 1: training dataset used to minimize MSE in AE (Eq.1)	Phase 2: training dataset used to minimize combined loss (Eq. 5)	Phase 3: training dataset used to minimize CE loss in classifier (Eq. 3)	
I	Flow rate A	Flow rate A	-	Flow rate A
IIa	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with four interpretable dimensions of the shape descriptor only	Flow rate A
IIb	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with four “non-interpretable” dimensions of the shape descriptor only	Flow rate A
IIc	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with three interpretable dimensions of the shape descriptor only	Flow rate A
IId	Flow rate A	Flow rate A	Flow rate A; drop outcome	Flow rate A

			classifier trained with drop elongation and drop surface curvature only	
Ile	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with drop elongation and drop throat size only	Flow rate A
IIf	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with drop throat size and drop surface curvature only	Flow rate A
IIfg	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with drop elongation only	Flow rate A
IIfh	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with drop throat size only	Flow rate A
IIfi	Flow rate A	Flow rate A	Flow rate A; drop outcome classifier trained with drop surface curvature only	Flow rate A

III	Flow rate B	Flow rate B	-	Flow rate A
IV	Flow rate B	Flow rate B	Flow rate A	Flow rate A

Table 2. Confusion matrix and prediction accuracy of scalar shape descriptors, SS_1 , SS_2 , SS_3 , and our machine learning model.

	True Intact	False Intact	True Break	False Break	Prediction Accuracy (%)
SS_1	681	524	470	325	57.6
SS_2	815	619	375	191	59.5
SS_3	581	422	572	425	57.7
Our machine learning model	931	92	902	75	91.7

Table 3. Confusion matrix and prediction accuracy of our machine learning model trained and tested with one, two, three, four, or eight dimensions of the shape descriptor using different training paradigms as defined in Table 1. The numbers in the orange and blue boxes represent datasets at 400 $\mu\text{L/hr}$ and 1000 $\mu\text{L/hr}$ respectively. The number of data points used for training is always $N = 498,002$. The number of data points used for testing is always $N = 2,000$. *indicates the case where only the one, two, three, or four dimensions of shape descriptors were used for training in Phase 3. See Table 1 and text for details.

Cases	Training paradigm	Training dataset			Testing dataset	Test Accuracy (%)	Test accuracy relative to baseline (%)	True Intact	False Intact	True Break	False Break
		Phase 1	Phase 2	Phase 3							
A1	I (Baseline)	400	400	-	400	91.7	100.0	931	92	902	75
A2	Ila	400	400	400*	400	84.2	91.8	878	145	805	172
A3	Ilb	400	400	400*	400	64.3	70.1	651	372	634	343
A4	Ilc	400	400	400*	400	84.9	93.6	922	101	776	201
A5	Ild	400	400	400*	400	67.7	73.8	639	384	714	263
A6	Ile	400	400	400*	400	72.9	79.5	778	245	680	297
A7	Ilf	400	400	400*	400	67.8	73.9	777	246	579	398
A8	Ilg	400	400	400*	400	61.9	67.5	640	383	598	379
A9	Ilh	400	400	400*	400	60.5	65.9	620	403	589	388
A10	Ili	400	400	400*	400	58.8	64.1	552	471	623	354
B1	I (Baseline)	1000	1000	-	1000	79.8	100.0	797	185	799	219
B2	Ila	1000	1000	1000*	1000	71.0	89.0	674	308	745	273
B3	Ilb	1000	1000	1000*	1000	61.8	77.4	598	384	638	380
B4	Ilc	1000	1000	1000*	1000	70.1	87.9	724	258	677	341
B5	Ild	1000	1000	1000*	1000	67.3	84.3	644	338	702	316
B6	Ile	1000	1000	1000*	1000	65.7	82.3	638	344	676	342
B7	Ilf	1000	1000	1000*	1000	64.6	80.9	652	330	639	379
B8	Ilg	1000	1000	1000*	1000	58.9	73.8	547	435	631	387
B9	Ilh	1000	1000	1000*	1000	58.4	73.1	583	399	584	434
B10	Ili	1000	1000	1000*	1000	56.4	70.7	569	413	559	459

Table 4. Confusion matrix and prediction accuracy of our machine learning model trained and tested with dataset of same and different flow rates using different training paradigms as defined in Table 1. The numbers in the orange and blue boxes represent datasets at 400 $\mu\text{L/hr}$ and 1000 $\mu\text{L/hr}$ respectively. The number of data points used for training is always $N = 498,002$. The number of data points used for testing is always $N = 2,000$.

Cases	Training paradigm	Training dataset			Testing dataset	Test Accuracy (%)	Test accuracy relative to baseline (%)	True Intact	False Intact	True Break	False Break
		Phase 1	Phase 2	Phase 3							
A1	I (Baseline)	400	400	-	400	91.7	100.0	931	92	902	75
A11	III	1000	1000	-	400	56.7	61.8	581	442	552	425
A12	IV	1000	1000	400	400	76.9	83.8	783	240	754	223
B1	I (Baseline)	1000	1000	-	1000	79.8	100.0	797	185	799	219
B11	III	400	400	-	1000	55.1	69.0	552	430	550	468
B12	IV	400	400	1000	1000	67.6	84.7	698	284	653	365

A novel shape descriptor identified by machine learning captures diverse droplet shapes and achieves high prediction accuracy of droplet instability.

