Article

Patterns

Few-shot deep learning for AFM force curve characterization of single-molecule interactions

Highlights

- Data curation, annotation of AFM force measurements of single-molecule interactions
- Formulating a rupture event characterization problem as a classification problem
- Development of a few-shot deep learning algorithm for AFM force curve classification
- Software implementation and validation of proposed method on AFM experimental data

Authors

Joshua R. Waite, Sin Yong Tan, Homagni Saha, Soumik Sarkar, Anwesha Sarkar

Correspondence

anweshas@iastate.edu

In brief

Atomic force microscopy is an important tool to understand nano-scale properties of biological samples. However, analyzing large datasets produced by atomic force microscopy can be an extremely laborious and time-consuming process for experimentalists. Waite et al. propose a machine learning tool to partially automate this process to save time and effort.

Waite et al., 2023, Patterns *4*, 100672
January 13, 2023
https://doi.org/10.1016/j.patter.2022.100672

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Few-shot deep learning for AFM force curve characterization of single-molecule interactions

Joshua R. Waite,[1] Sin Yong Tan,[1] Homagni Saha,[1] Soumik Sarkar,[1] and Anwesha Sarkar[2,3,*]
[1]Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA
[2]Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA
[3]Lead contact
*Correspondence: anweshas@iastate.edu
https://doi.org/10.1016/j.patter.2022.100672

THE BIGGER PICTURE AFM plays a significant role in characterizing soft biological samples; e.g., proteins, DNA, and live cells. Many such characterizations demand expert scanning of measurement samples (e.g., AFM force-distance curves) for statistical analysis. However, manual analyses are often time consuming, laborious, qualitative, and affected by subjective human biases. Machine learning (ML) can alleviate many such issues, leading to fast, automated, and bias-free analyses. However, because many scientific problems, including ours, suffer from a lack of annotated data, ML approaches also need to be sample efficient. This study demonstrates the efficacy of such an approach for AFM force curve characterization. In the future, this research can be expanded to explore more complex AFM-based characterization tasks, leading to bidirectional innovation in biophysics and ML. Efficient biophysical characterization can lead to accelerated drug delivery studies that have a significant societal impact.

1 **2** 3 4 5 **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Deep learning (DL)-based analytics has the scope to transform the field of atomic force microscopy (AFM) with regard to fast and bias-free measurement characterization. For example, AFM force-distance curves can help estimate important parameters of binding kinetics, such as the most probable rupture force, binding probability, association, and dissociation constants, as well as receptor density on live cells. Other than the ideal single-rupture event in the force-distance curves, there can be no-rupture, double-rupture, or multiple-rupture events. The current practice is to go through such datasets manually, which can be extremely tedious work for the experimentalists. We address this issue by adopting a few-shot learning approach to build sample-efficient DL models that demonstrate better performance than shallow ML models while matching the performance of moderately trained humans. We also release our AFM force curve dataset and annotations publicly as a benchmark for the research community.

INTRODUCTION

Deep learning (DL)-based analytics has been extremely useful for many bioimaging and bio-image analytics problems. However, DL or machine learning (ML) in general has rarely been explored for atomic force microscopy (AFM) characterization and standardization of biological samples.[1] One example is characterization of single-molecule interactions of ligand-receptor pairs on live cells.

Single-molecule techniques, such as AFM,[2–5] optical tweezers,[6,7] and magnetic tweezers,[8,9] have established themselves as powerful methods to examine the properties of individual or single molecules at a time, yielding important information about topographical features, binding parameters, and dissociation kinetics.[10,11] AFM[3,5,12–18] has several advantages over other single-molecule techniques, such as the ability to perform high-resolution imaging of conducting as well as non-conducting samples (biological samples like DNA, protein, and live cells) in ambient air, liquids, or a vacuum without using any dye and performing force measurements in the piconewton range. Many ligand-receptor systems[11,19–21] have been investigated with AFM to shed light on the interaction forces,[22–24] binding

probability, dissociation and association rates, affinity, etc. These ligands and receptors are attached to the AFM cantilever tip and substrate, respectively, using flexible polymeric linkers. Low ligand density is usually used on the cantilever tip to lower multiple attachments among molecules. Typically, in AFM force measurement studies, a ligand-functionalized cantilever is initially brought close to live cells with complementary receptors or a functionalized substrate so that the ligand can bind to the receptor. Upon cantilever retraction at constant speed, the main observation is the minimum amount of force that is needed to break the bond: unbinding or rupture force. The key measurement in this regard is the force-distance curve, which tracks the interaction forces corresponding to the distances between the cantilever tip and the substrate.

Upon collection of a large number (e.g., 1,000) of AFM force-distance curves (or simply force curves) from selected cell samples, the data need to be analyzed to first identify different types of rupture events, such single-, double-, multiple-, or even no-rupture events. The current practice is to manually go through the force curves and classify them into different rupture types based on visual inspection. This process is not only time consuming but also requires a significant level of experience and expertise because data can be corrupted by spurious noise, disturbances, and other artifacts. Here we propose a DL approach to classify the force curves automatically to save significant expert time and effort. However, generating a large, annotated dataset appropriate for training typical supervised DL models can be extremely difficult and time consuming. First of all, obtaining individual force curves is time consuming. The problem is exacerbated by the tedious annotation task, as discussed above. Use of simulated or synthetic data to increase the amount of data may be a possible solution that is explored in many applications.[25] However, in our case, the force curves, and specifically the noise in the force curves, have complex characteristics depending on the AFM instrumentation, cantilever properties, and soft biological samples. This implies that a simulated dataset would have limited usefulness unless the signal and noise characteristics are accurately captured. Therefore, we would require an extremely high-fidelity, physics-based simulator to accurately represent the data, which, unfortunately, may be intractable. Therefore, we explore sample-efficient DL approaches for this task.

One such sample-efficient DL approach is semi-supervised learning. This approach is best suited for when unlabeled data are readily available and labeled data are difficult to gather.[26–28] Semi-supervised learning aims to improve performance achievable using the limited labeled data by also utilizing information learnable from the unlabeled data. However, in our case, unlabeled data are time consuming to gather and not available in sufficient quantities. Another approach is weakly supervised learning. Although semi-supervised learning is a type of weakly supervised learning with incomplete supervision, there are also two other types: inexact supervision and inaccurate supervision.[29] Inexact supervision considers the level of detail in labels, especially for object classification. Although it is ideal to have every object in an image labeled, it is sometimes more practical to label an entire image as containing such objects. In our case, the samples are relatively simple, but noise complicates focusing on individual rupture events rather than the entire force curve. Inaccurate supervision simply refers to working with data that may contain errors in the labels, which is an issue we face in this study as well. We attempt to minimize inaccurate labels in our data by downselecting samples that have been annotated more than once with consistent labels. There has been a lot of recent interest in self-supervised learning approaches that generate pseudolabels or pretext tasks to use in a supervised learning setting.[30–32] Generating accurate pseudolabels for very similar classes, such as in our case, can be very challenging. Self-supervised pre-training only starts to become effective when the dataset of unlabeled samples becomes significantly large. In contrast, few-shot learning is able to be trained with only a few labeled samples without the need for additional unlabeled samples.[31,33,34] By using a few-shot approach, we only need to determine whether a sample is similar or dissimilar to other samples rather than directly determining the class of a sample. This helps to create a generalizable model while only training with a few samples. Therefore, we adopt a few-shot learning approach to build sample-efficient DL models for this application. To the best of our knowledge, this is the first time a few-shot DL framework is being proposed for AFM force curve analysis.

The proposed framework will save a lot of time for future researchers in related fields of experimental microscopy and make the data screening process of large datasets of force curves highly efficient for statistical analyses. We would like to highlight the fact that our algorithm can be applied as a generalized approach for classifying rupture events and determining the binding kinetics for any specific interaction (in DNA-protein interactions, protein-protein interactions, or ligand-receptor interactions) between functionalized AFM tip and functionalized substrate. Annotated data obtained from AFM single-molecule interaction experiments made available through this endeavor would also serve as a useful benchmark for the research community.

RESULTS

In this section, we first formulate the problem and then present our solution approach, which is evaluated empirically using AFM experimental data.

Problem formulation

The objective of this study is to automatically characterize (i.e., classify) the force-distance curves based on the types of rupture events, as shown with examples in Figure 1. The force-distance curve in Figure 1A shows how the force on the AFM probe varies with the tip-sample separation distance. The region where the bond is ruptured is of primary interest, and it is manifested as an event of nonlinear nature (rupture/unbinding event). The location offset as well as the nonlinear nature of these type of events set it apart from the non-specific adhesions.[35,36] Rupture between the attached molecules on the substrate and the tip is represented by these rupture/unbinding events. In the example force-distance curve shown in Figure 1A, location (a) shows where the bond is formed, (b) shows where the bond experiences the maximum force, and (c) shows where the bond breaks. Other than the ideal specific

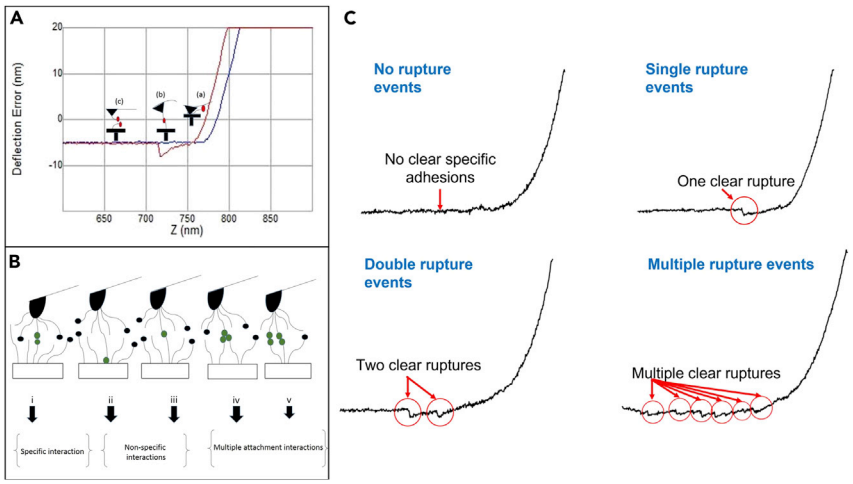

Figure 1. Problem formulation: Classifying the AFM force-distance curves from single-molecule interaction experiments based on the types of rupture events

single-rupture event depicted in Figure 1A, there can be other events in a large set of experiments, as shown in Figure 1B. The ideal single-rupture event (unbinding/rupture of two single molecules [attached to the cantilever tip and substrate]) is shown in Figure 1B, and the maximum extension of the linkers shown here is approximately twice the contour length of the linkers involved. In this scenario, the applied force on the cantilever tip is the same as that on the molecular bond formed because of protein-protein or ligand-receptor interaction. The length of the stretched linker is the distance between the point where the bond ruptures and the point where the AFM cantilever tip detaches from the substrate. In Figure 1B, ii and iii show undesired non-specific adhesions because of rupture of protein molecule-gold-coated substrate and rupture of protein molecule-linker molecule, respectively. Reduction of these type of events is necessary to build a statistically sound and useful dataset.

Therefore, any rupture event located at a distance smaller than the linker contour length must be excluded while estimating the distribution of single-rupture forces. In Figure 1B, iv and v show unbinding of multiple bonds/attachments involving multiple protein molecules and extension of multiple linkers associated with detected rupture forces much larger than rupture force of a single protein bond. The density of protein molecules on the cantilever tip as well as the substrate must be reduced to reduce the multiple attachments. Multiple rounds of experiments are performed to gather multiple sets of force curves under the same conditions using the same cantilever to build up the distribution (i.e., a histogram) of measured specific single-rupture forces. Such a distribution can be used to estimate the probability of a given rupture force interval as well as the most probable rupture force. Finally, Figure 1C shows force curves with different types of rupture events; i.e., no-rupture, single-rupture, double-rupture, and multiple-rupture events. Among these types, identifying single-rupture events from a large number of experiments (i.e., force curves) is the most critical task, as described above. Double-rupture events could also be useful; for example, to estimate multiple attachment probability.[11,24] Therefore, we define the classification problem as a 3-class classification to distinguish single- and double-rupture events, with the third class corresponding to no-rupture or multiple-rupture (i.e., more than two ruptures) events.

Few-shot learning framework for force curve characterization

We propose a sample-efficient framework that can learn to classify force curves with complex features with fewer labels than what is required by conventional DL techniques. The training and testing process of our algorithm is summarized in Figure 2. The crux of our framework is training a convolutional neural network model using triplet loss, as described below.

Triplet convolutional neural network

There have been many variations of few-shot DL proposed over the years, but the core of all of these variations is the two architectures: Siamese network and triplet network. The Siamese network and triplet network are typically trained using contrastive loss and triplet loss, respectively. Although contrastive loss compares two samples, triplet loss involves three (triplet) samples in one go. We explored using contrastive and triplet loss and found that triplet loss consistently performed better. This is expected because of the nature of the data, where there can be significant variation within a class.[37] For example, our class 0 contains no rupture and multiple ruptures, and the size and location of ruptures also vary significantly from sample to sample. We are able to use this approach to predict the class of a sample by measuring its similarity to the classes of known samples. For example, if we compare an unknown sample and determine that it is very similar to single-rupture samples, then it is also likely to be a single-rupture sample. Details about how the degree of similarity is computed are provided below.

Figure 2 provides an overall illustration of the training and testing process of the triplet convolutional neural network. In the training process, a triplet convolutional neural network[38] accepts three distinct inputs: an anchor, a positive sample, and a negative sample. The anchor is the sample being examined, the positive sample is a sample of the same class as the anchor, and the negative sample is a sample of a different class from the anchor. All three inputs are passed through convolutional neural networks with shared parameters to get embeddings respective to each input. The distance between the anchor and positive embeddings $(D_{(x^+,anchor)})$, and the distance between the anchor and negative embeddings $(D_{(x^-,anchor)})$ are calculated using squared $L2$ norm. The triplet loss, L is simply the difference between the two distances plus a margin (m):

$$L = \max\left(0, D_{(x^+,anchor)} - D_{(x^-,anchor)} + m\right).$$

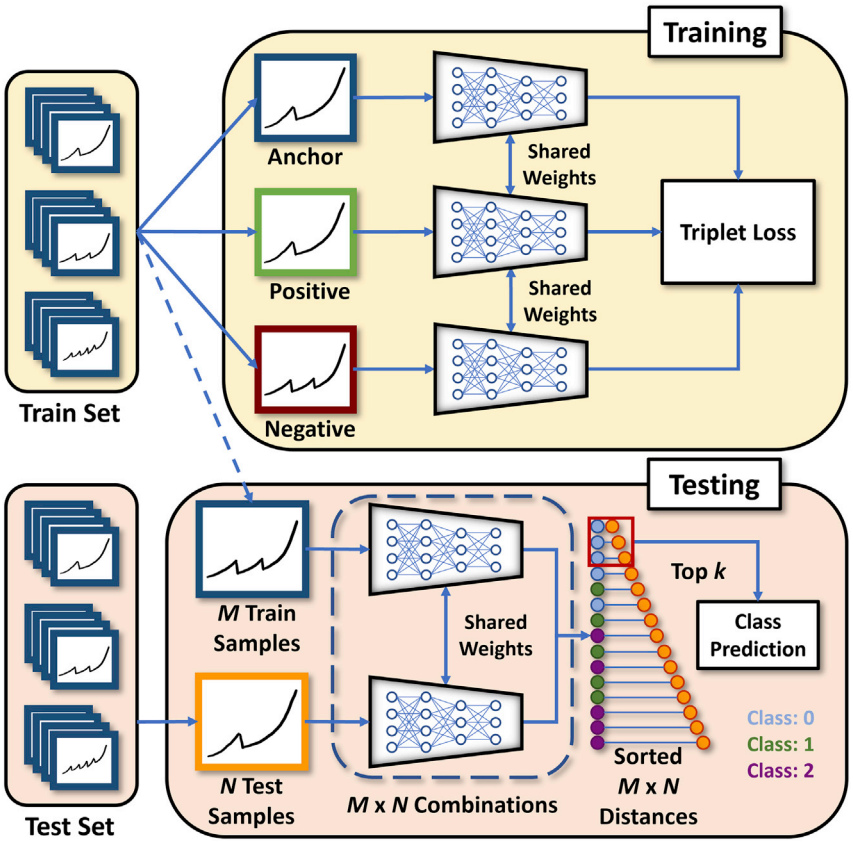

Figure 2. Schematic of our few-shot learning algorithm using triplet loss for training

Training, shown in the top block, is done by entering three samples from the train set into a triplet loss architecture. Testing, shown in the bottom block, is done by entering train-test sample pairs into the model to compare embedding distances.

The margin (m) is a value that ensures the minimum distance between similar and dissimilar pairs so that the similar sample is closer to the anchor than the dissimilar sample. If the margin is too small, then it becomes difficult to distinguish between the classes.

When the model parameters are learned, we can move on to the testing or inferencing phase, where class predictions can be made by comparing an unknown sample with known samples and classifying it to the class with the shortest distances. From the testing block in Figure 2, N test samples are paired with M train samples. The train samples are the anchors with known labels from the training process, and the test samples are the unknown samples whose classes are being predicted. Similar to the training process, these samples are passed through the trained convolutional neural networks with shared parameters to get the embeddings. The distances between the embeddings of the unknown samples and every known sample is computed using the squared $L2$ norm, which gives $(M \times N)$ distances corresponding to each train-test pair. The distances between the embeddings represent the degree of similarity; thus, a shorter distance between two embeddings also indicates that the samples are more likely to belong to the same class. Therefore, after sorting the pairs in ascending order of distance, the first k class of the train (known class) samples are returned as the class prediction of the test (unknown class) samples. For example, assuming $k = 1$, the pair with the shortest distance is used to predict the class of the unknown sample as the same class of the known sample. For $k > 1$, a majority voting was performed for the first k class to determine the final predicted class.

In terms of model architecture, we performed an extensive hyper-parameter search involving 2D and 1D convolution layers, varying the number of layers and number of units in a layer, to name a few. Additional details can be found in Table S1. One of the best-performing models (Figure 3) consists of a vector input, 1D convolutional layer, 1D max-pooling layer, another 1D convolutional layer, another max-pooling layer, and four fully connected layers, leading to a 10-dimension vector output. More details about each of these layers is provided in the few-shot parameter exploration section (Table 2).

Empirical evaluation

We evaluate our proposed few-shot learning framework using AFM experimental data along with performance comparison with some shallow ML methods. We begin the discussion with a description of the dataset.

AFM experimental data

As described earlier, we focus on classifying force curves into three classes: single rupture, double rupture, and a third class referring to no-rupture and multiple-rupture events. The experimental data generation process is detailed in the experimental procedures section. Although an entire force curve has scientific relevance, the rupture event(s) show up in the flat portion of the force curve at the bottom, as shown in Figure 1. Therefore, we only take the lower third of a force curve and use it as input to the ML model as a vector, which is done by removing 100 data points from the start of the vector.

For empirical validation, we required a labeled and balanced dataset of force curves. In general, with our experimental specification, we observe a reasonable number of single-rupture, no-rupture and multiple-rupture events. However, the number of double-rupture events is usually low. Many force curves suffer from measurement noise or spurious disturbances. Therefore, annotating the force curves with high confidence becomes difficult, and the dataset usually suffers from significant labeling noise that is often difficult to avoid. To construct a reliable (without labeling noise) dataset, we first remove samples with a high level of noise and then go through multiple rounds of expert annotation to identify force curves with consistent labels. For the present study, we begin with a set of about 1,600 force curves and create a completely balanced 3-class dataset of 216 force curves with highly reliable labels. Such significant downselection demonstrates the difficulty of generating a large

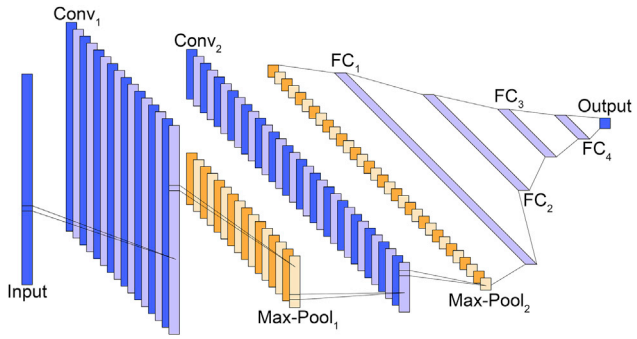

Figure 3. 1D convolutional neural network (CNN) model architecture for few-shot learning

Layer details can be found in Table 2.

annotated dataset, justifying the need for sample-efficient learning strategies. The downselection process is explained in more detail in Section S3. The dataset is then divided into a training set and test set. The training set contains 42 samples, each with class labels of 0 (no rupture or multiple rupture), 1 (single rupture), and 2 (double rupture). The test set contains 30 samples for each of the three classes.

Data pre-processing

Before passing the data to the model, we first apply some processing to improve the clarity of the ruptures after trimming the data to focus on the lower third of the curve. This section describes the processing step by step, and the results of each processing step are illustrated in Figure 4. First, the difference (between consecutive data points on the curve) is computed to remove the overall trend and emphasize the spikes in the data. The plot in the middle of Figure 4 shows a large spike around 200 nm from the single rupture, but the plot is visibly noisy and could make learning slightly difficult for the few-shot model. To reduce the noise and make the actual ruptures more prominent, we apply a moving average filter with a window size of 13. Last, the data are scaled to between 0 and 1, as shown in the last plot in Figure 4.

Few-shot parameter exploration

Although we initially explored using 2D convolution layers on images of the force curves, the model seemed to be unable to isolate the rupture events in the images. This is likely due to the rupture events occupying an extremely small portion of the image, even after the pre-processing steps to focus on the lower third section. Hence, we switched to using 1D convolution layers on the force curves represented as a vector of y values. From here, we performed an extensive, although not exhaustive, parameter search, as shown in Table 1.

Table 2 shows the details for the best set of hyper-parameters of the few-shot model architecture shown in Figure 3. In addition to the three-class classification problem, we also explored two other problem formulations that are relevant for asking different scientific questions, both with binary classification problems. We found slight variations in the best architecture we obtained for these different cases. The values without brackets represent the details for the three class cases: no/multiple rupture, single rupture, and double rupture. Values in parentheses, including the dropout layer, represent the binary class case with no rupture versus any (at least one) rupture. Values in brackets, which do not include the dropout layer, represent the binary class case with single rupture versus rest. For the other parameters, we found that a learning rate of 1e−2, learning rate gamma of 1, and the stochastic gradient descent (SGD) optimizer were best for all cases. A learning rate gamma less than 1 affects training by lowering the learning rate every step, which, in our case, is every epoch. Reducing the learning rate as training progresses can help fine-tune weights. A batch size of 32 was best for the 3 class and single versus rest case, and a batch size of 16 was best for the no versus any rupture case.

Few-shot DL performance and comparison with shallow models

We present the results of our few-shot DL framework based on the training and testing sets as discussed above.

We obtain an overall accuracy of 66.7% with 73.3%, 63.3%, and 63.3% per class accuracy for classes 0, 1, and 2 respectively. In addition to accuracy values, other metrics, such as precision, recall, and F-1 score for our few-shot approach are presented in Table 3. We compare the performance of few-shot deep model with a few popular shallow learning approaches, such as k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). Similar to our few-shot approach, we performed an extensive hyperparameter search for the shallow models to obtain the best performance as well. Details of the shallow model hyper-parameters are provided in the supplementary section (Table S1). From the performance metrics listed in Table 3, it is evident that the few-shot

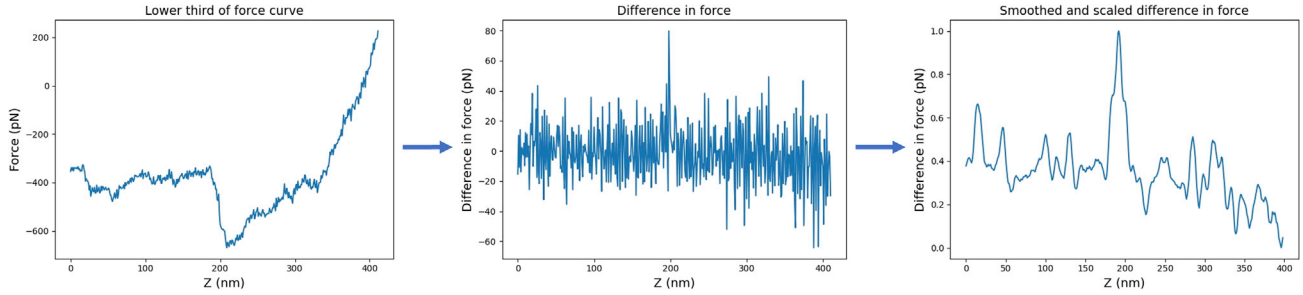

Figure 4. Visual overview of data pre-processing

The left plot shows the force curve after trimming 100 points to focus on the lower portion of the curve. The center plot shows the force curve after taking a difference between the points. This removes the overall trend but amplifies the noise. The right plot shows the curve after the three rounds of rolling means and scaling.

Table 1. Parameters explored for the few-shot method

Parameter	Number of FC layers	Number of Conv layers	Conv channel sizes	Conv kernel sizes	Loss function	Batch size	Optimizer	Learning rate	Learning rate gamma
Values	2–4	0–3	8–64	3, 5, 8	contrastive, triplet	16–64	SGD, ADAM	1e−2, 3e−4	1, 0.977

FC, fully connected; Conv, convolution.

model performs significantly better (~15% gain in accuracy) compared with the shallow models with respect to most of the performance metrics. The class accuracies for few-shot learning models also show that the performance on each of the classes were fairly balanced, whereas the shallow models have biased predictions, resulting in high accuracy in one of the classes but at the cost of having other class accuracies collapse almost entirely.

To understand the dependency of the few-shot deep model performance on training data size, we reduced the sample size from 40 samples per class to 30, 20, and 10 samples per class by randomly selecting subsets of the original training data. For each case, we repeated the training process 3 times to estimate the uncertainty in model performance. Figure 5A shows how accuracy changes over different sizes of labeled training sets. As expected, performance drops quite significantly for the few-shot approach (becomes similar to that of the shallow models with the smallest data size) with reduction in training data size. The uncertainty in model performance increases at the same time. On the other hand, the shallow approaches are relatively stable in terms of performance with a change in the number of labeled training samples. This suggests that there is a minimum labeled data size necessary to start benefitting from the few-shot DL approach compared with the shallow models. We performed extensive parameter searches for the shallow models at each amount of training samples, whereas the best-performing few-shot model, when using all 42 training samples per class, was retrained with the lower data amounts.

Finally, as discussed earlier, avoiding label noise in training data is quite challenging for this application. Therefore, we explore the change in few-shot model performance while increasingly corrupting the training data with possibly erroneous labels. We implemented this by using labels from a moderately trained human annotator. For example, in the case of 10% noisy labels, we replace 10% of the original training samples (with highly reliable labels provided by an expert) with samples labeled by the moderately trained human annotator. Similar to the previous experiments, we repeated the training process 3 times to estimate the uncertainty in model performance. Figure 5B shows how increasing the percentage of noisy labels affects the performance of our few-shot approach as well as the shallow methods. Performance of the few-shot model decreases as the percentage of noisy labels increases. The shallow methods are not affected as much, but the uncertainty in performance grows in some cases. However, even with 50% noisy labels, the performance of the few-shot method remains similar to if not above that of the shallow methods. We performed extensive parameter searches for the shallow models at each percentage of noisy data, whereas the best-performing few-shot model, when using all 42 training samples per class, was retrained with the different percentages of noisy data.

Table 2. Few-shot model layer details

Layer	Dimension	Input features	Output features	Kernel size	Stride
Input	399	–	–	–	–
$Conv_1$	395	1	16 (32) [32]	5	1
$Max\text{-}Pool_1$	98	16 (32) [32]	16 (32) [32]	4	4
$Conv_2$	94	16 (32) [32]	32 (64) [64]	5	1
$Max\text{-}Pool_2$	23	32 (64) [64]	32 (64) [64]	4	4
FC_1	–	736 (1,472) [1,472]	512	–	–
FC_2	–	512	256	–	–
(Dropout p = 0.25)	–	–	–	–	–
FC_3	–	256	128	–	–
FC_4	–	128	10	–	–
Output	10	–	–	–	–

1D convolution layers are notated as $Conv_n$, and fully connected layers are notated as FC_n. Differences for the no rupture versus any rupture case are shown in parentheses, including a dropout layer with p = 0.25 between FC_2 and FC_3. Differences for the single rupture versus rest case are shown in brackets.

Anecdotal example analysis

We further investigated the wrongly classified samples anecdotally to understand the less-than-ideal performance. Figure 6 shows one correctly classified example for each class of force curves (in the diagonal green boxes). With a quick visual inspection, it can be verified that these examples are accurately classified. These curves are less noisy compared with the few representative examples of wrongly classified force curve samples shown here (in the off-diagonal red boxes). The "incorrectly" classified examples are not quite obvious upon careful visual inspection. For example, consider the force curve that is predicted to be showing norupture or multiple-rupture (N/MR) events, whereas the ground truth is that it is a double-rupture event force curve. Visual inspection suggests that the force curve is either relatively noisy or showing many rupture events. Similarly, the ground truth N/MR but predicted single rupture (SR) force curve shows one clear rupture event and possibly many other rupture events or just observation noise. The other force curve that is incorrectly classified as an SR curve (ground truth DR) also shows only one clear rupture event. Therefore, these are also non-obvious examples of incorrectly classified force curves. The visual confusion suggests that many such force curves could also have been annotated wrongly by the human expert(s). The anecdotal analysis indicates that the quality

Table 3. Overall accuracy, class accuracy, precision, recall, and F-1 score for our few-shot approach, SVM, RF, and KNN

Algorithm	Overall accuracy (%)	Class accuracy (%)			Precision (%)	Recall (%)	F-1 score (%)
		N/MR	SR	DR			
Few-shot (ours)	*66.7*	*73.3*	63.3	63.3	*66.8*	*66.7*	*66.7*
SVM	53.3	60.0	10.0	90.0	39.8	53.3	42.0
RF	46.7	43.3	0.0	*96.7*	32.7	46.7	36.2
KNN	53.3	56.7	*80.0*	23.3	57.4	53.3	51.3

of the few-shot learner could be better than what the confusion matrix in Figure 6 reflects (the number of test samples for each category in the confusion matrix is noted in the corresponding cell).

We performed t-distributed stochastic neighbor embedding (t-SNE) and plotted (in 2D) the resulting dimension reduction, as shown in Figure 7A. As expected, samples from different classes are quite mixed up, suggesting the difficulty of classifying or distinguishing the force curves to begin with. Performing t-SNE analysis on the embedding obtained from our few-shot approach shows improved separation in Figure 7B, but it is still challenging to fully separate the classes. This t-SNE visualization is consistent with the results in Figure 6, where class 0 has the highest class accuracy and is more clearly separated on the right in Figure S1B. Classes 1 and 2 have lower class accuracies and are more often misclassified as each other rather than class 0, which is reflected in the overlap of the classes on the left in Figure S1B.

DISCUSSION

In this section, we discuss how the proposed ML framework can support statistical analyses of force curves to characterize single-molecule interactions. This also helps with contextualizing the quantitative results presented above.

Rupture force is a stochastic variable because dissociation of the complex bond is a thermally activated process. So, to obtain a statistically sound analysis, experimentalists typically generate multiple sets of a statistically significant number (e.g., 1,000) of force curves on different locations of the cell and on multiple cells. As mentioned earlier, a dataset of a large number of force curves performed with, for example, a collagen-functionalized cantilever on live cells (as done in this study) will contain single-rupture events, double-rupture events, multiple-rupture events, and no-rupture events. Among these events, identifying single- and double-rupture events particularly help with estimating important parameters of binding kinetics, such as the most probable rupture force, association and dissociation constants, as well as receptor density on live cells. Therefore, the proposed ML framework can be used as an initial screening tool to isolate single- and double-rupture events from the dataset. Upon ML-based screening, although missed single- or double-rupture events can be difficult to recover, a human expert still has the opportunity to get rid of the falsely detected events during further analysis. Therefore, the few-shot ML model presented here can still be useful as an initial screening tool despite the 66.7% classification accuracy. In fact, this performance is comparable with a moderately trained human annotator (an engineering graduate student), who could also achieve only 65.6% classification accuracy on the same test set.

Although the 3-class model can be used as a general purpose screening tool for single- and double-rupture events, as mentioned in Section 2.3.3, we also explored other models with specific scientific objectives to check whether we could

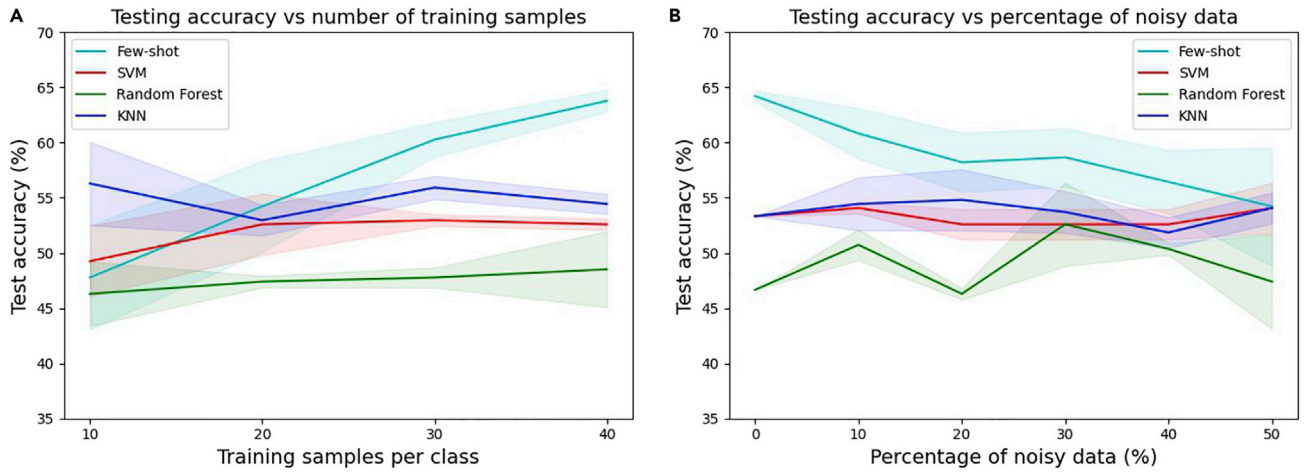

Figure 5. Few-shot comparison with shallow models.

(A) Number of training samples per class, explored from 10–40 in steps of 10.

(B) Effect of increasing the percentage of noisy data (i.e., samples with possibly erroneous labels) on the few-shot and shallow methods. Experiments were repeated 3 times with different random seeds. Plotted uncertainty is ±1 SD.

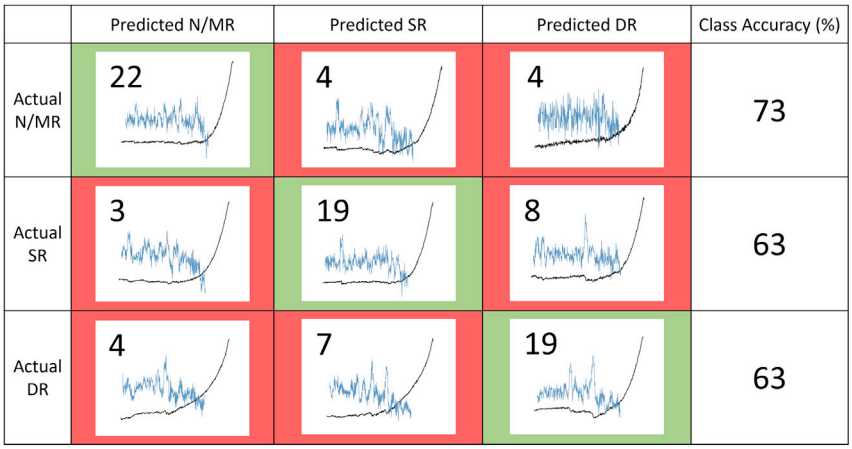

Figure 6. Anecdotal examples from the test set of correctly and incorrectly classified force curve examples

Many of the incorrectly classified force curves are demonstrated to be confusing even to the human experts. We also show the confusion matrix for the test set and the class specific accuracies. The original data are shown in black, and the data after pre-processing are shown in blue.

obtain even better performance. For example, we attempted to build a few-shot model only to separate single-rupture events from the rest of the event types. With this problem formulation, we were able to achieve an overall accuracy of 78.3% with a single rupture detection accuracy of 73.3% and an accuracy of 83.3% for the rest. We also looked at another problem formulation, where no-rupture events need to be separated from events with at least one rupture. Such a tool could be useful to calculate parameters such as binding probability.

In this case, a few-shot deep model was able to achieve an overall classification accuracy of 91.7% with per-class accuracy of 93.3% and 90.0% for no-rupture class and the class with at least one rupture, respectively. These results are summarized in Table 4. Hence, it is evident that targeted few-shot modeling may be performing better. However, multiple such models may have to be trained for different purposes. We also performed principal-component analysis (PCA) on the single rupture versus rest and no rupture versus rest cases, which can be seen in Figures S2 and S3, respectively. An additional exploration of performance on a larger but low-confidence test set is included in Section S4.

It is clear that our proposed ML framework can be used as a screening tool for a large dataset of AFM force curves generated to characterize single-molecule interactions. Such a tool may save a lot of time and effort of experts and make the analysis free of human subjectivity. Future work will focus on developing other ML-based tools for automated extraction of biophysical properties from individual force curves and from distribution of selected force curves.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, A.S. (anweshas@iastate.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data and original code have been deposited at https://doi.org/10.5281/zenodo.7301801 and are publicly available as of the date of publication.

Cell lines used for experiments

Overexpression of Discoidin domain receptors[39–42] (DDRs) is associated with multiple human diseases, such as fibrosis and cancer. To develop novel therapeutic strategies that target DDRs, we studied interaction forces and binding kinetics between DDRs and their ligand, collagen, at the single-molecule level on live cells using AFM-based single-molecule force spectroscopy experiments. The cell lines used in this study are human benign prostatic hyperplasia (BPH1) cells and human pancreatic cancer (MiaPaCa-2) cells. The sources of these cell lines are mentioned in our previous publication.[24] These cell lines were maintained using RPMI medium and DMEM supplemented with 10% fetal bovine serum and 1% streptomycin/penicillin

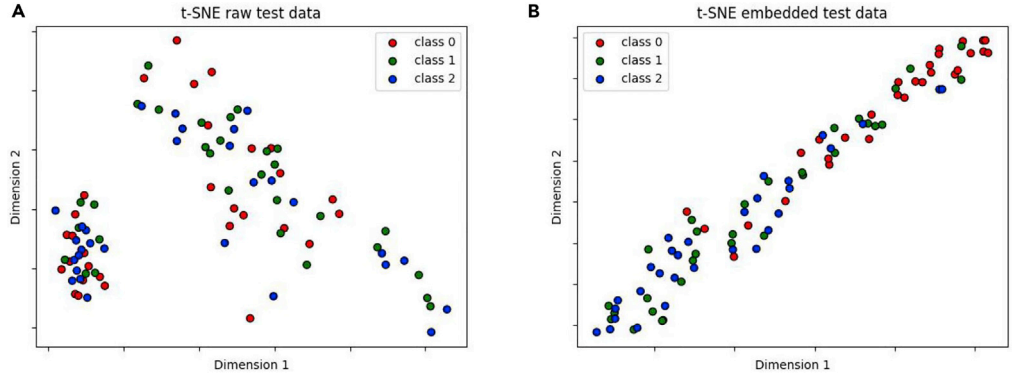

Figure 7. Few-shot t-SNE visualization

(A) t-SNE separation on raw test data for the 3-class case.

(B) t-SNE separation on embedded test data for the 3-class case using our few-shot approach.

Table 4. accuracy and class accuracy for our few-shot approach on binary class scenarios

Classes	Overall accuracy (%)	Class accuracy (%)			
		NR	Any	SR	Rest
NR versus any	91.7	93.3	90.0	–	–
SR versus rest	78.3	–	–	73.3	83.3

antibiotics, respectively. BPH1 cells expressing endogenous DDR1 were employed to be transfected with scrambled short hairpin RNA (BPH1scr cell line), or these cells were used to downregulate DDR1 expression via short hairpin RNA (BPH1shDDR1 cells). Similarly, MiaPaCa-2 cells were used to express wild-type DDR1b (MiaPaCa-2 DDR1b) or mutated DDR1b isoforms (MiaPaCa-2 DDR1b R105A). These MiaPaCa-2 cells were also transfected to express MiaPaCa-2-EV cells. Three different antibody conditions (DDR-blocking antibody, integrin-blocking antibody, and no antibody) were also used for the MiaPaCa-2 and BPH1 cell lines to perform force curve measurements.

Functionalization of the cantilever tip with rat tail collagen I

In the first step, rat tail collagen type I (Advanced Biomatrix) was diluted in pH 5 solution at a 1:2 ratio. In the second step, succinimidyl 6-(3(2-pyridyldithio) propionamido)hexanoate) (LC-SPDP; Thermo Fisher Scientific) diluted in dimethyl sulfoxide (DMSO) solvent was mixed with collagen[43–47] solution at a 1:25 ratio. pH 5.2 solution of Cleland's reagent or DTT (Thermo Fisher Scientific) was added to the combined solution prepared in step 1 and 2. OBL-10 (Bruker AFM Probes) was incubated in the final solution for 2 h at room temperature on a shaker. Then the cantilever was washed carefully and preserved in 1× PBS buffer at 4°C until the beginning of the experiment on the same day.

Force measurements on live cells

Live cells (the cell lines mentioned previously) were plated on 60-mm culture plates, and the plates were placed on a Bruker BioScope catalyst AFM base plate (containing an XY motorized stage and XY piezo scanner). After alignment of the laser on the cantilever tip and obtaining the maximum sum signal, the AFM head was placed on top of the base plate. The cantilever was first approached manually and then using NanoScope software control toward the live cell surface so that collagen on the cantilever tip and cell surface receptors on the live cell surface can bind. Force measurements were achieved using constant speed ramping of the cantilever with lower scanning forces of 700 pN, employing an approach and retraction speed of 2 μ m/s and a lower scan rate of 0.3 Hz on various locations of the same cell and different cells. Multiple control experiments were performed, using a non-functionalized cantilever or denatured collagen-functionalized cantilever. Multiple DDR- and integrin-blocking antibodies were also used in the force measurements.

Scope of experimental validation

As described above, we used multiple cell lines under multiple experimental conditions (determined by different antibodies) to demonstrate the generalizability of our approach across different types of samples. We used a general experimental setup that is typical for measurement of piconewton-scale forces and binding kinetics for all protein-protein or ligand-receptor interaction at the single-molecule level using AFM. We used the same cantilever for all experiments in this study to avoid additional inconsistencies. Experimental system variability (different instruments, devices, or setups) may result in different data distributions, and we do not claim any generalizability in this context. However, the proposed learning framework should still remain relevant. It would be interesting to explore whether we would need different DL models for different instruments or experimental setups. Even if this is the case, the proposed few-shot learning paradigm will be helpful to learn models specific to experimental devices and setups with minimal effort.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100672.

ACKNOWLEDGMENTS

We thank Prof. Peter Hoffmann (Department of Physics, Wayne State University) and Prof. Rafael Fridman (Department of Pathology, Wayne State University) for help with culturing cells with overexpressed receptors and performing AFM force-distance measurements. This work was supported by the Iowa State College of Engineering Exploratory Research Program (to A.S. and S.S.).

AUTHOR CONTRIBUTIONS

Conceptualization, H.S., S.S., and A.S.; methodology, J.R.W., S.Y.T., and H.S.; software, J.R.W. and S.Y.T.; validation, J.R.W. and S.Y.T.; formal analysis, J.R.W. and S.Y.T.; investigation, J.R.W., S.Y.T., and A.S.; resources, S.S. and A.S.; data curation, J.R.W. and A.S.; writing – original draft, J.R.W., H.S., S.S., and A.S.; writing – review & editing, J.R.W., S.Y.T., and S.S.; visualization, J.R.W. and A.S.; supervision, S.S. and A.S.; project administration, S.S. and A.S.; funding acquisition, S.S. and A.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 23, 2022
Revised: September 29, 2022
Accepted: December 7, 2022
Published: January 6, 2023

REFERENCES

1. Rade, J., Zhang, J., Sarkar, S., Krishnamurthy, A., Ren, J., and Sarkar, A. (2022). Deep learning for live cell shape detection and automated afm navigation. Bioengineering *9*, 522.
2. Gaczynska, M., and Osmulski, P.A. (2008). Afm of biological complexes: what can we learn? Curr. Opin. Colloid Interface Sci. *13*, 351–367.
3. Muller, D.J. (2008). Afm: a nanotool in membrane biology. Biochemistry *47*, 7986–7998.
4. Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Miyagi, A., Taniguchi, M., and Yamashita, H. (2008). High-speed afm and nano-visualization of biomolecular processes. Pflugers Arch. *456*, 211–225.
5. Kreplak, L., Scheuring, S., and Kreplak, L. (2016). Introduction to atomic force microscopy (afm) in biology. Curr. Protoc. Protein Sci. *85*, 17.7.1–17.7.21.
6. Jones, P., Maragó, O., and Volpe, G. (2015). Optical Tweezers (Cambridge, UK: Cambridge University Press).
7. Moffitt, J.R., Chemla, Y.R., Smith, S.B., and Bustamante, C. (2008). Recent advances in optical tweezers. Annu. Rev. Biochem. *77*, 205–228.
8. De Vlaminck, I., and Dekker, C. (2012). Recent advances in magnetic tweezers. Annu. Rev. Biophys. *41*, 453–472.
9. Tanase, M., Biais, N., and Sheetz, M. (2007). Magnetic tweezers in cell biology. Methods Cell Biol. *83*, 473–493.
10. Jones, S.K., Sarkar, A., Feldmann, D.P., Hoffmann, P., and Merkel, O.M. (2017). Revisiting the value of competition assays in folate receptor-mediated drug delivery. Biomaterials *138*, 35–45.
11. Mayyas, E., Bernardo, M., Runyan, L., Sohail, A., Subba-Rao, V., Pantea, M., Fridman, R., and Hoffmann, P.M. (2010). Dissociation kinetics of an enzyme- inhibitor system using single-molecule force measurements. Biomacromolecules *11*, 3352–3358.
12. Burnham, N.A., and Colton, R.J. (1989). Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope. J. Vac. Sci. Technol.: Vacuum, Surfaces, and Films *7*, 2906–2913.
13. Kurland, N.E., Drira, Z., and Yadavalli, V.K. (2012). Measurement of nanomechanical properties of biomolecules using atomic force microscopy. Micron *43*, 116–128.
14. Dufrêne, Y.F. (2002). Atomic force microscopy, a powerful tool in microbiology. J. Bacteriol. *184*, 5205–5213.

15. Rugar, D., and Hansma, P. (1990). Atomic force microscopy. Phys. Today *43*, 23–30.
16. Giessibl, F.J. (2003). Advances in atomic force microscopy. Rev. Mod. Phys. *75*, 949–983.
17. Meyer, E. (1992). Atomic force microscopy. Prog. Surf. Sci. *41*, 3–49.
18. Sarkar, A. (2022). Biosensing, characterization of biosensors, and improved drug delivery approaches using atomic force microscopy: a review. Front. Nanotechnol. *3*.
19. Jobst, M.A., Schoeler, C., Malinowska, K., and Nash, M.A. (2013). Investigating receptor-ligand systems of the cellulosome with afm-based single-molecule force spectroscopy. J. Vis. Exp. *82*, e50950.
20. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., and Evans, E. (1999). Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. Nature *397*, 50–53.
21. Sundar Rajan, V., Laurent, V.M., Verdier, C., and Duperray, A. (2017). Unraveling the receptor-ligand interactions between bladder cancer cells and the endothelium using afm. Biophys. J. *112*, 1246–1257.
22. Kada, G., Kienberger, F., and Hinterdorfer, P. (2008). Atomic force microscopy in bionanotechnology. Nano Today *3*, 12–19.
23. Cappella, B., and Dietler, G. (1999). Force-distance curves by atomic force microscopy. Surf. Sci. Rep. *34*, 1–104.
24. Sarkar, A., Sohail, A., Dong, J., Prunotto, M., Shinki, K., Fridman, R., and Hoffmann, P.M. (2019). Live cell measurements of interaction forces and binding kinetics between discoidin domain receptor 1 (ddr1) and collagen i with atomic force microscopy. Biochim. Biophys. Acta Gen. Subj. *1863*, 129402.
25. Balu, A., Nallagonda, S., Xu, F., Krishnamurthy, A., Hsu, M.-C., and Sarkar, S. (2019). A deep learning framework for design and analysis of surgical bioprosthetic heart valves. Sci. Rep. *9*, 18560–18612.
26. Ouali, Y., Hudelot, C., and Tami, M. (2020). An overview of deep semi-supervised learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2006.05278.
27. Van Engelen, J.E., and Hoos, H.H. (2020). A survey on semi-supervised learning. Mach. Learn. *109*, 373–440.
28. Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. Natl. Sci. Rev. *5*, 44–53.
29. Ghosal, S., Zheng, B., Chapman, S.C., Potgieter, A.B., Jordan, D.R., Wang, X., Singh, A.K., Singh, A., Hirafuji, M., Ninomiya, S., et al. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. Plant Phenomics *2019*, 1525874.
30. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., and Makedon, F. (2020). A survey on contrastive self-supervised learning. Technologies *9*, 2.
31. Ericsson, L., Gouk, H., Loy, C.C., and Hospedales, T.M. (2022). Self-supervised representation learning: introduction, advances, and challenges. IEEE Signal Process. Mag. *39*, 42–62.
32. Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D.S., O'Neal, M.E., Singh, A.K., Sarkar, S., Ganapathysubramanian, B., and Singh, A. (2021). Self-supervised learning improves agricultural pest classification. In AI for Agriculture and Food Systems.
33. Wang, Y., Yao, Q., Kwok, J.T., and Ni, L.M. (2021). Generalizing from a few examples: a survey on few-shot learning. ACM Comput. Surv. *53*, 1–34.
34. Kadam, S., and Vaidya, V. (2020). Review and analysis of zero, one and few shot learning approaches. In Intelligent Systems Design and Applications, A. Abraham, A.K. Cherukuri, P. Melin, and N. Gandhi, eds. (Cham: Springer International Publishing), pp. 100–112.
35. Busscher, H.J., and Weerkamp, A.H. (1987). Specific and non-specific interactions in bacterial adhesion to solid substrata. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. *46*, 165–173.
36. Lee, M.H., Brass, D.A., Morris, R., Composto, R.J., and Ducheyne, P. (2005). The effect of non-specific interactions on cellular adhesion using model surfaces. Biomaterials *26*, 1721–1730.
37. Hoffer, E., and Ailon, N. (2015). Deep metric learning using triplet network. In International workshop on similarity-based pattern recognition (Springer), pp. 84–92.
38. Dong, X., and Shen, J. (2018). Triplet loss in siamese network for object tracking. In Proceedings of the European conference on computer vision (ECCV), pp. 459–474.
39. Kim, H.-G., Tan, L., Weisberg, E.L., Liu, F., Canning, P., Choi, H.G., Ezell, S.A., Wu, H., Zhao, Z., Wang, J., et al. (2013). Discovery of a potent and selective ddr1 receptor tyrosine kinase inhibitor. ACS Chem. Biol. *8*, 2145–2150.
40. Hidalgo-Carcedo, C., Hooper, S., Chaudhry, S.I., Williamson, P., Harrington, K., Leitinger, B., and Sahai, E. (2011). Collective cell migration requires suppression of actomyosin at cell–cell contacts mediated by ddr1 and the cell polarity regulators par3 and par6. Nat. Cell Biol. *13*, 49–58.
41. Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., et al. (2019). Deep learning enables rapid identification of potent ddr1 kinase inhibitors. Nat. Biotechnol. *37*, 1038–1040.
42. Leitinger, B. (2003). Molecular analysis of collagen binding by the human discoidin domain receptors, ddr1 and ddr2: identification of collagen binding sites in ddr2. J. Biol. Chem. *278*, 16761–16769.
43. Rich, A., and Crick, F. (1955). The structure of collagen. In The Excitement of Discovery: Selected Papers of Alexander Rich: A Tribute to Alexander Rich (World Scientific), pp. 103–104.
44. Van Der Rest, M., and Garrone, R. (1991). Collagen family of proteins. Faseb. J. *5*, 2814–2823.
45. Lee, C.H., Singla, A., and Lee, Y. (2001). Biomedical applications of collagen. Int. J. Pharm. *221*, 1–22.
46. Shoulders, M.D., and Raines, R.T. (2009). Collagen structure and stability. Annu. Rev. Biochem. *78*, 929–958.
47. Linsenmayer, T. (1991). Collagen. In Cell biology of extracellular matrix (Springer), pp. 7–44.