Soft Matter

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5sm00943j

Machine learning approaches to quantify nanoscale variations in the mechanical properties of soft nanoparticles

Benjamin Baylis and John R. Dutcher *

We use machine learning to analyze atomic force microscopy-force spectroscopy (AFM-FS) measurements of the mechanical properties of soft nanoparticles on a hard substrate. We compare two approaches based on the manual selection of features that describe various aspects of the mechanical properties measured using AFM-FS – one that uses an extreme gradient boosting algorithm (supervised learning), and the other based on *k*-means clustering (unsupervised learning) to classify the force–distance curves according to the features. We used these approaches to generate two machine learning (ML) classifiers – one to differentiate between the soft nanoparticles and the hard substrate, and the other to identify structure within individual nanoparticles based on nanoscale variations in their mechanical properties. After training the classifiers on data from just two AFM images, we found that both approaches were successful in correctly identifying individual soft nanoparticles within the AFM-FS scans, whereas the supervised approach was more successful in correctly identifying and quantifying stiffer inner regions within the nanoparticles. The results of our study show that ML strategies can be used to accurately and efficiently characterize nanoscale variations in the mechanical properties of soft, biological materials.

Received 17th September 2025,
Accepted 23rd February 2026

DOI: 10.1039/d5sm00943j

rsc.li/soft-matter-journal

Introduction

Artificial intelligence (AI) and machine learning (ML) are being used increasingly to analyze large datasets in scientific fields such as drug discovery, medical imaging and materials science.[1–9] These approaches have led to advances ranging from the prediction of drug–protein interactions[1,2] to the detection and classification of cancerous tumours, melanoma and Alzheimer's disease[3,5] to the prediction of the physical properties of nanoparticles[6,7] to the mapping of the properties of eukaryotic cells.[8] Models based on these approaches can quickly, efficiently and accurately analyze very large datasets to identify and correlate the most important parameters that characterize these complex systems, providing new and useful insights. In the field of materials science, it is common to collect large amounts of data such as microscopy images and spectra on a wide range of samples. The use of ML can simplify and improve the analysis and interpretation of these large collections of data.[9]

An important use of ML approaches in materials science is to classify measured data to gain useful insights into the physical properties of different materials. Applying a ML classifier model to materials science data involves several steps: processing the data to identify a set of features (feature engineering) that describes the range of properties of the material, using an appropriate ML model to classify materials according to these features, training the model using the selected features, and evaluating the model on new, unseen data.[9] To be successful, it is necessary to use high-quality data, feature engineering to accurately and adequately describe the range of the properties of the material, and a sophisticated, robust ML model that can efficiently identify relationships between the input features from new, unseen data and accurately relate them to a target classification or cluster.

Unsupervised and supervised learning are two widely used ML approaches to classify large amounts of data. Unsupervised approaches such as clustering techniques like *k*-means clustering are very effective in identifying patterns and have the advantage that they do not require labelled training data. Supervised approaches require labelled training data but can provide better interpretability compared to unsupervised approaches.[10] For example, supervised models that use decision trees can provide information on the gain, or contribution to the performance of the model, for each input feature.[10,11]

Atomic force microscopy (AFM) is used extensively in materials science to characterize samples at micrometer and nanometer length scales, and its use in measuring variations in nanoscale

Department of Physics, University of Guelph, Guelph, ON, N1G 2W1, Canada.
E-mail: dutcher@uoguelph.ca

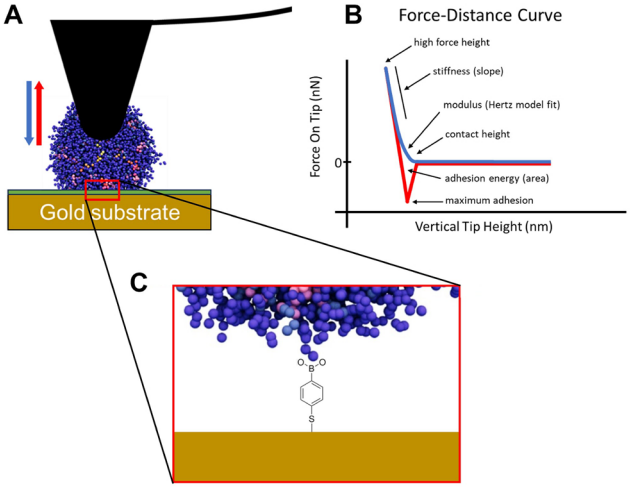

Fig. 1 (A) Schematic diagram of an AFM-FS measurement on a phytoglycogen (PG) nanoparticle. The AFM tip (black) is pressed into the particle and then retracted. The radius of curvature of the AFM tip and the PG nanoparticle are roughly to scale. (B) Schematic diagram of a force–distance curve from a single approach (blue) and retraction (red) motion of the AFM cantilever tip showing the force on the tip *versus* the vertical tip height. Several parameters extracted from the force–distance curve are indicated. (C) Schematic diagram (not to scale) of a 4-mercaptophenylboronic acid (4-MPBA) molecule forming covalent bonds with a PG nanoparticle and the gold surface to immobilize the PG nanoparticle onto the surface in water.

mechanical properties has resulted in a better understanding of the morphology of soft, biological materials.[12,13] In AFM force spectroscopy (AFM-FS), data is collected in a grid-like pattern with each pixel in the pattern corresponding to a force–distance curve collected as the AFM tip is pressed into and retracted from the sample (Fig. 1A). The sharp AFM tip and high force sensitivity allow the topography and mechanical properties of the sample to be mapped at high spatial resolution. Analysis of different regions of the approach and retraction curves (Fig. 1B) yields a variety of mechanical properties of the sample for each pixel of the images, including stiffness, modulus, adhesion, and sample height at high applied force.

The large amount of topographical and mechanical information contained in AFM-FS images make it an excellent candidate for the application of ML models. Recently, ML techniques have been developed to analyze AFM-FS measurements on a variety of soft and biological samples, allowing the distinction between live and dead bacteria,[14] the identification of different cancerous cells,[15–17] the location of features on cell surfaces responsible for aggressive cancers,[18] and the identification of different polymers in a polymer blend.[19] Petrov *et al.*[19] demonstrated that values of mechanical properties determined using Bruker's proprietary Peak Force Tapping™ mode can be used as inputs to supervised ML classifiers to predict the location of polymers on a surface. A more robust and flexible approach is to engineer features directly from AFM-FS force–distance curves. This provides the advantage that ML techniques can be applied to AFM-FS data acquired with any commercial AFM with a force spectroscopy mode, and it allows the user the possibility of fine tuning the choice of properties for use in expanded ML frameworks. In addition, user-controlled features combined with interpretable ML classifiers provide additional insights into material properties.

In the present work, we use AFM-FS to characterize the nanoscale mechanical properties of phytoglycogen (PG), a highly branched, glucose-based polysaccharide produced as soft, dense, compact, hydrated, hairy nanoparticles with a diameter of 42 nm and an underlying dendritic architecture.[19–25] These physical characteristics of the PG particles lead to unique and useful properties, *e.g.*, water within the particles that is much more ordered than bulk water,[26–29] zero-shear viscosity that increases much more gradually with concentration than that for typical polysaccharides,[24,30] and strong associations that can be achieved between the particles and small bioactive molecules.[31,32] The morphology and functionality of PG nanoparticles can also be tuned through simple mechanical or chemical modifications.[30,33–38] The properties of native and modified PG, together with their biodegradability and non-toxicity, make the particles desirable as unique additives in personal care formulations[39,40] and as nanocarriers for bioactive molecules.[41–43]

We have recently used AFM-FS to study the nanoscale mechanical properties of PG nanoparticles in water immobilized on smooth hard gold substrates (Fig. 1).[21] Measurements of high-resolution maps of force–distance curves of many individual PG nanoparticles revealed the high deformability of the hydrated nanoparticles, with the highly branched, dendritic chain architecture resulting in a stiffer inner structure surrounded by a softer outer structure within each particle. In the AFM-FS images, we observed stiffer inner regions, corresponding to the interaction of the AFM tip with the stiffer inner structure through the softer outer structure, and softer outer regions, corresponding to the interaction of the AFM tip with only the softer outer structure. AFM-FS measurements also allowed the determination of the average equivalent spherical radius ($\bar{r}$ = 23.0 nm), in close agreement with values determined using small angle neutron scattering[20,22] and dynamic light scattering,[30] and the average Young's modulus ($\bar{E}$ = 688 kPa), which was consistent with values of the bulk modulus determined using osmotic pressure techniques.[28,29] In this previous AFM-FS study, the PG nanoparticles on the underlying hard gold substrate were identified by examining the fit of each force–distance curve to a hard substrate model and manually selecting regions with a relatively low goodness-of-fit that likely corresponded to the measurement of soft PG nanoparticles on the hard substrate. The dendritic stiffer inner structure was observed in height images collected at high force values (~2 nN) when the particles are compressed onto the hard substrate. The stiffer inner and softer outer regions within each nanoparticle were identified by fitting height images collected at high force to a mixed Gaussian model to determine an appropriate height threshold between the stiffer and softer regions within the particles. These manual methods of identifying the particles and the spatial extent of their stiffer inner regions are time consuming and prone to uncertainties due to

slight variations in the underlying substrate topography and the subjective choice of threshold values.

In the present study, we developed a ML approach to analyze large numbers of force–distance curves collected in AFM-FS measurements of PG nanoparticles on hard gold substrates, which allows the identification of the particles and the quantification of nanoscale variations in their mechanical properties. We began by manually selecting different measures of the mechanical properties (features) in different regimes within the force–distance curves using a feature engineering process. We then trained two supervised ML classifiers on a limited amount of data (just two 512 pixel × 512 pixel AFM images) using the selected features: the first ML classifier was trained to distinguish between the soft PG nanoparticles and the surrounding hard substrate, and the second ML classifier was trained to locate the spatial extent of the stiffer inner regions within the PG nanoparticles. We then tested the two supervised ML classifiers on new, unseen data to evaluate their accuracy by qualitatively comparing height images from the AFM-FS measurements to the classifier outputs and by quantitatively comparing the values of the particle radius and the Young's modulus to average values determined previously using the manual approach described in ref. 21 We also compared the performance of the two supervised ML classifiers with that of an unsupervised learning approach based on *k*-means clustering using the same features and training data. The unsupervised approach removes the need for labelled training data that is required in the supervised approach. This comparison showed that both approaches were successful in identifying the location of the soft PG nanoparticles on the hard substrate, whereas the supervised learning approach was more successful in determining small nanoscale variations in the mechanical properties within the PG nanoparticles.

Materials and methods

Sample preparation

Monodisperse phytoglycogen (PG) nanoparticles were extracted and purified from sweet corn in the form of a lyophilized powder using the procedure described in ref. 21. Dilute dispersions were prepared by adding the powder to Milli-Q water (resistivity of 18.2 MΩ-cm) and mixing on a Stovall shaker.

The substrates for the AFM-FS experiments were prepared using the following procedure. Glass slides were cleaned by first submerging them in a sulfuric acid (H_2SO_4) bath for 15 min, rinsing thoroughly with Milli-Q water, then submerging them in a heated 1 : 1 solution of chloroform ($CHCl_3$) and methanol (CH_3OH) for an additional 15 min. Slides were removed and thoroughly rinsed with Milli-Q water and methanol before being dried with a stream of nitrogen gas. A titanium adhesion layer (~5 nm thick) was sputter deposited onto the clean glass slides, followed by the sputter deposition of a layer of gold (~200 nm thick) to create stable gold-coated substrates. Small sections (~3 mm × 3 mm) were cut using a diamond tip pen and smooth terraces of gold (lateral extent ~350 nm) were created by annealing the small sections of gold-coated glass slides at 675 °C for 65 s.[44] The annealed substrates were then submerged in a 0.4% w/w solution of 4-mercaptophenylboronic acid (4-MPBA) in methanol for 24 h to allow the self-assembly of a monolayer of 4-MPBA on the gold surface *via* the formation of gold-thiol bonds between the gold and 4-MPBA. To measure isolated PG nanoparticles, a drop of a dilute (0.005% w/w) dispersion of PG in Milli-Q water was placed on the surface of the 4-MPBA/gold substrate for 12 min and then rinsed with water. This allowed the covalent bonding of an optimal number density of isolated PG nanoparticles to the 4-MPBA molecules (Fig. 1C). Additional details on the preparation of the samples can be found in ref. 21

Atomic force microscopy force spectroscopy (AFM-FS)

AFM-FS measurements were performed on bare 4-MPBA/gold substrates (which we refer to as hard substrates) as well as hard substrates with covalently bonded PG nanoparticles. Samples were measured in water using the quantitative imaging (QI) mode (an optimized force spectroscopy mode) on a JPK NanoWizard ULTRA Speed atomic force microscope, using a NanoWorld Pyrex-Nitride silicon cantilever with a chromium/gold coating on the detector side (PNP-TR) and a nominal spring constant of 0.32 N m^{-1}. The approach of the cantilever toward the sample surface and the retraction of the cantilever away from the sample surface was performed using a 150 nm piezo range at an approach/retraction rate of 62.5 μm s^{-1} with a 2 nN maximum force setpoint. A scan area of 1 μm × 1 μm (512 pixels × 512 pixels) was used for all measurements.

Data analysis

Selection of features for characterization of mechanical properties. In the AFM-FS images, we analyzed each force–distance curve using a feature engineering approach to select a comprehensive set of parameters/features that described the mechanical properties of the sample at that position (pixel) in the AFM scan such as deformations at low and high forces, stiffness, and modulus values (Fig. 1B and 2, and Table S1). Force–distance curves were processed and analyzed using the JPK Data Processing Software (version 6.1.198). Curves were converted from piezo height (z_{piezo}) to vertical tip height (z_{tip}) using the spring constant of the cantilever that was calculated from force–distance curves collected on the hard substrate surrounding the particles and using the thermal noise method.[45] The baseline force of the force–distance curves was determined by fitting a portion of the approach curve for large separations between the cantilever and sample surface to a straight line and subtracting this line from the data. Any hysteresis/offset between the baseline force in the approach and retract curves was determined and accounted for by fitting a portion of the retraction curve for large separations between the cantilever and sample surface to a horizontal line and subtracting this value from the retraction curve data. The contact point was defined as the smallest value of the vertical tip height for which the force was less than the baseline force

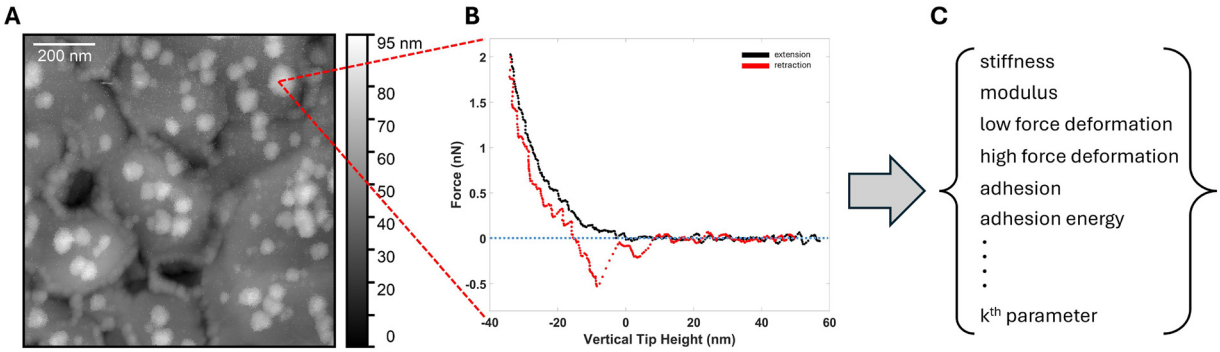

Fig. 2 (A) AFM contact height image of PG nanoparticles (white objects) on flat terraces of a 4-MPBA/gold (hard) substrate. (B) Processed force–distance curve for a single pixel from the AFM scan in A. The approach data is shown in black and the retraction data is shown in red. The vertical tip height values of the retraction curve have been shifted to match the approach curve values at 2 nN. Values above 2 nN are not shown. (C) List of a subset of the parameters calculated from each force–distance curve that are used to construct a feature vector describing the sample at that location.

measured in the approach curve. An example of a processed force–distance curve is shown in Fig. 2B.

To create a robust ML classifier, it is necessary to construct a feature vector with values that represent the different distinctive mechanical properties of a material. For AFM-FS measurements, materials may have similar properties in one or more regimes of the force–distance curves but different properties in others, *e.g.*, similar surface characteristics resulting in similar adhesion values, but different stiffness or modulus values due to different structure. The feature vector should account for these differences, and this is especially important when classifying different regions within a single object such as identifying the location of the stiffer inner regions within the PG nanoparticles investigated in this work.

To select features that describe the distinctive mechanical properties of the PG particles, we considered different portions of the approach and retraction curves. For the approach curves, we considered features that characterized three regimes: near the contact point, near the maximum applied force, and within the range from low to high forces. Features calculated near the contact point describe the material properties near the sample surface such as the height deformation at low values of the applied force and the modulus corresponding to small indentations. Features calculated near the maximum applied force describe the material properties under compressive stress such as the height deformation at high values of the applied force and the stiffness (slope of the force–distance curve near the maximum applied force). Features calculated within the range from low to high forces are influenced by the material properties near its surface in combination with properties that describe the material under compressive stress, such as height deformation and the modulus calculated within this force range. For the retraction curves, we considered features that characterized three regimes: near the maximum applied force, near the surface of the sample, and within the range from low to high forces. Features calculated near the maximum applied force include the recovery of the sample height after reducing the applied force at high forces. Features calculated near the surface of the material include the recovery of the sample height at low forces. Features calculated within the range from low to high forces include the recovery of the sample height after reducing the applied force within this force range. In addition, we characterized the adhesion between the AFM tip and the sample by calculating the adhesion force, which is the magnitude of the maximum negative force in the retraction curve, and the adhesion energy, which is the area of the retraction curve below zero force. These adhesion-related quantities are significantly larger for measurements on the PG particles than for those on the hard substrate. We note that the adhesion between the AFM tip and the sample is influenced by the surface chemistry of the tip. Because of this, comparisons between measurements should involve the same type of cantilever tip, as we have done in the present study. Also, all measurements in the present study were performed in water to eliminate the influence of environmental conditions on the tip-sample adhesion that are present when performing measurements in air. The different height deformation and recovery values described above were obtained by examining the difference in tip height at six different applied forces for both the approach and retraction curves: 2% (0.04 nN), 5% (0.1 nN), 25% (0.5 nN), 50% (1 nN), 75% (1.5 nN), and 100% (2 nN) of the maximum applied force.

For force–distance curves collected on PG nanoparticles, we note that features calculated at higher forces can be influenced by the stiffer inner structure of the particles as well as the underlying hard substrate, whereas features calculated at lower forces are only influenced by the softer outer structure of the particles. For force–distance curves collected on the periphery of the PG nanoparticles, corresponding to their softer outer region, the influence of the underlying hard substrate dominates the force on the AFM tip at lower forces than for force–distance curves collected within the PG nanoparticles due to their stiffer inner structure. This is because it requires more force to compress the particle when pressing on the stiffer inner region. For force–distance curves collected on the hard substrate, the height deformation values are much smaller than those corresponding to force–distance curves collected on the PG particles, especially at low forces.

We constructed a feature vector for each pixel in the AFM scan that described the mechanical properties of the sample at each pixel location using a set of 37 features (Table S1). We chose the engineered features manually, instead of using AI generated features (by either generative or non-generative means), since the manually selected features had well-defined physical interpretations that allow more direct comparisons with results obtained using other techniques. In addition, the manual selection of features reduced the amount of data used as input to the machine learning classifiers, which correspondingly reduced the computational time and cost.

Machine learning classifiers. In the present work, we developed two supervised ML classifiers: one to distinguish the PG particles from the surrounding hard substrate (supervised particle/substrate ML classifier), and the other to distinguish the stiffer inner region from the softer outer region within each particle (supervised inner particle structure ML classifier). In both cases, we classified individual force–distance curves using an extreme gradient boosting algorithm available in the library XGBoost (version 2.0.3) and implemented in Python[46] (Fig. S1). The XGBoost classifier is widely used because of its computational speed and accuracy.[46–48] It provides an output of the probability of a classification for each force–distance curve, and we used a 50% probability threshold for binary classifications. Further details on the XGBoost algorithm used in this study are provided in the SI.

We also used two unsupervised ML classifiers to distinguish the PG particles from the surrounding hard substrate (unsupervised particle/substrate ML classifier) as well as to distinguish the stiffer inner region from the softer outer region within each particle (unsupervised inner particle structure ML classifier). In these cases, we classified individual force–distance curves using *k*-means clustering[49] in the library Scikit-learn (version 1.0.2) and implemented in Python.

To train and test the particle/substrate ML classifiers, we used two AFM-FS images (the training/testing dataset): one that contained ~60 PG nanoparticles and one that contained only the 4-MPBA/gold substrate. For the particle/substrate ML classifiers, we used 80% of the data contained in the two images to train the ML classifier (training dataset) and the remaining 20% of the data to test the ML classifier (testing dataset). To train and test the inner particle structure ML classifiers, we used one AFM-FS image that contained the same PG nanoparticles used in the particle/substrate ML classifier from the training/testing dataset, using only pixels corresponding to PG nanoparticles in this image. For the inner particle structure ML classifiers, we used 90% of the data contained in the determined PG nanoparticle pixels to train the ML classifier (training dataset) and the remaining 10% of the data to test the ML classifier (testing dataset). The ML classifiers were then evaluated on new, unseen data from three different AFM-FS images of PG nanoparticles on the 4-MPBA/gold substrate.

Training the particle/substrate machine learning classifiers. We trained the supervised particle/substrate ML classifier using the selected input features (Table S1) by manually labelling each force–distance curve (pixel) in the AFM-FS images as either belonging to a PG nanoparticle or the hard substrate, creating a binary target for each pixel. This was done by manually masking the PG nanoparticle regions in a measurement of the nanoparticles on the hard substrate (Fig. 3B and E). Force–distance curves corresponding to the hard substrate consisted of the regions of the substrate surrounding the masked particles (Fig. 3D) as well as a measurement of the bare hard substrate (Fig. 3C). The training and testing datasets for the supervised particle/substrate ML classifier were randomly selected from 262 144 input vectors from measurements of ~60 PG particles on a hard substrate, as well as 262 144 input vectors from measurements of a bare hard substrate. We used 80% (419 430 force–distance curves) of the labelled curves to train the classifier, which resulted in 37 155 force–distance curves labelled as corresponding to PG nanoparticles and 382 275 force–distance curves labelled as corresponding to the hard substrate.

We trained the unsupervised particle/substrate ML classifier using the same selected input features (Table S1) and the same training dataset (Fig. 3) used to train the supervised ML classifier with a set number of 2 clusters.

Training the inner particle structure machine learning classifiers. We trained the supervised inner particle structure ML classifier using only the force–distance curves that corresponded to the PG nanoparticles as determined using the same manual particle mask used in the supervised particle/substrate ML classifier. Training was performed using the selected input features (Table S1) together with a new manually labelled binary target (Fig. 4C): each force–distance curve collected on the PG nanoparticles in the AFM-FS image was labelled as either belonging to the stiffer inner region or the softer outer region of the particle. We identified the stiffer inner regions as those corresponding to the PG regions visible in the high applied force setpoint height images, which show the height of the particle when it is compressed against the underlying hard substrate (Fig. 4B). The training/testing dataset for the supervised inner particle structure ML classifier consisted of 46 462 input vectors from measurements of ~60 PG nanoparticles on a hard substrate. We used 90% (41 815 force–distance curves) of the labelled curves to train the classifier, which resulted in 16 009 force–distance curves corresponding to the stiffer inner region, and 25 806 force–distance curves corresponding to the softer outer region of the particles.

We trained the unsupervised inner particle structure ML classifier using only the force–distance curves that corresponded to the PG nanoparticles as determined by the unsupervised particle/substrate ML classifier. Training was performed using the same selected input features (Table S1) used for the supervised ML classifier with a set number of 2 clusters. We used 90% (39 236 force–distance curves) of the curves to train the classifier.

Evaluation of the machine learning classifiers. We examined the ML classifiers using outputs from the training datasets, testing datasets, and a dataset containing new, unseen data outside of the training/testing dataset. Evaluating the classifiers using the testing datasets allowed us to compare the output classifications to the true classifications (as determined by

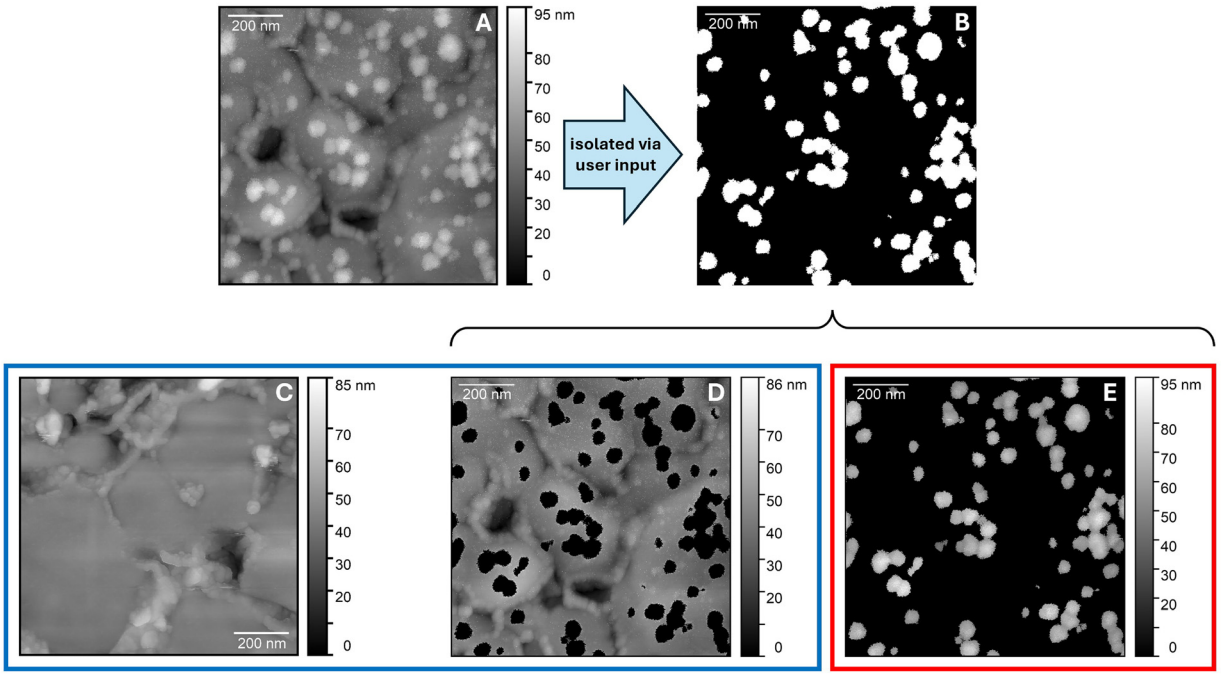

Fig. 3 AFM-FS measurements representing the training and testing datasets. (A) AFM contact height image of PG nanoparticles on a hard substrate (4-MPBA/gold). (B) Manually created mask of the particles (white) in A with the surrounding hard substrate (black). (C) AFM height image of the bare hard substrate. (D) AFM height image of the hard substrate surrounding the PG nanoparticles (black regions) after manually removing the regions corresponding to the PG nanoparticles using the mask in B. (E) AFM height image of the PG nanoparticles after manually isolating them using the mask in B. The non-black pixels (force–distance curves) in D and all pixels in C were classified as corresponding to the hard substrate. The non-black pixels (force–distance curves) in E were classified as corresponding to PG nanoparticles.

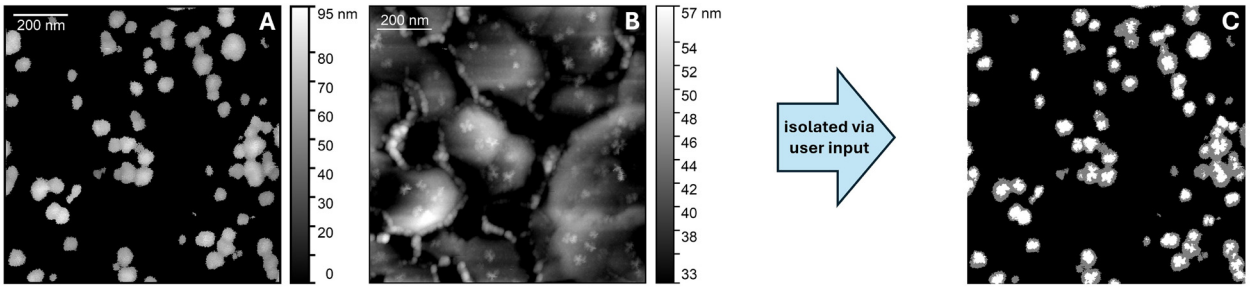

Fig. 4 (A) AFM contact height image of isolated PG nanoparticles. The isolated nanoparticles correspond to the same regions used to train the supervised particle/substrate ML classifier. (B) AFM peak force height image of the same scan in A revealing the stiffer inner region of the PG nanoparticles when they are compressed onto the surface, as well as the surrounding hard substrate (4-MPBA/gold). (C) The particle regions in A and the stiffer inner regions in B were used to manually produce this masked image. The inner (white) and outer (grey) regions were used to train and test the supervised inner particle structure ML classifier.

manual labelling) and to determine the accuracy through metrics such as the total percentage of correctly identified force–distance curves (pixels), percentage of false positives, and percentage of false negatives. For the supervised ML classifiers, the XGBoost algorithm also provides a measure of the relative importance, or gain, of each feature expressed as a percentage of the total contribution to the improvement in the accuracy of the classifier.[11] To interpret the supervised ML classifiers, we examined the features most important to each classifier after training. Because we manually engineered these features, we can then relate the most important features to the different regimes of the force–distance curves to gain insight into the most important mechanical properties. We evaluated all ML classifiers using new, unseen data from three AFM-FS images measured in three separate experiments performed on three different days using three different AFM cantilevers. This dataset is separate from the training and testing datasets, and it allows us to gain insights into the generalizability and robustness of the ML classifiers. We used both qualitative comparisons between height images and the output from the ML classifiers, as well as quantitative comparisons of Young's modulus and radius values to results from previous work.[21,22,30]

Results and discussion

Accuracy of the supervised particle/substrate machine learning classifier

For the 104 858 force–distance curves used to test the supervised particle/substrate ML classifier (testing dataset), we obtained an overall pixel classification accuracy of 99.5% with a false positive percentage of 0.3% and a false negative percentage of 3.1% when classifying the force–distance curves labelled as corresponding to PG (Fig. S2A and B). The features that contributed the largest improvement in the accuracy of the supervised particle/substrate ML classifier were obtained from the low force regime of the approach portion of the force–distance curves and describe the deformations at low applied forces (deformations measured for forces between 0.04 nN and 1 nN). The top three most important features describe this low force regime and accounted for 83.7% of the improvement in the accuracy of the supervised particle/substrate ML classifier (Fig. S2C). The large improvement in the accuracy of classifying the pixels resulting from features describing the low force regime is reasonable given the large difference between the mechanical properties of the soft PG nanoparticles and the hard substrates.

Application of the supervised particle/substrate machine learning classifier to new, unseen data

The results of applying the supervised particle/substrate ML classifier to new, unseen data can be represented as a probability map in which the value of each pixel represents the probability of it being classified as corresponding to the PG nanoparticles; an example of this is shown in Fig. 5B. Using a probability threshold of 50%, we generated a classification map of the locations of the PG nanoparticles (white regions in Fig. 5C) and the surrounding hard substrate (black regions in Fig. 5C). To include only entire PG nanoparticles, we removed small clusters of pixels (<35 pixels in area) that the classifier identified as corresponding to PG (Fig. 5D). These small areas, which could be due to small pieces of PG or noisy/irregular force–distance curves arising from measurement artifacts, corresponded to only 1.6% of the pixels attributed to PG in the example in Fig. 5C and 1.2% of the pixels attributed to PG for all three AFM scans of new, unseen data (Fig. 5C, Fig. S4C and Fig. S5C).

To qualitatively assess the performance of the supervised particle/substrate ML classifier, we compared the contact height image of the entire scan (Fig. 5A) to the PG nanoparticle

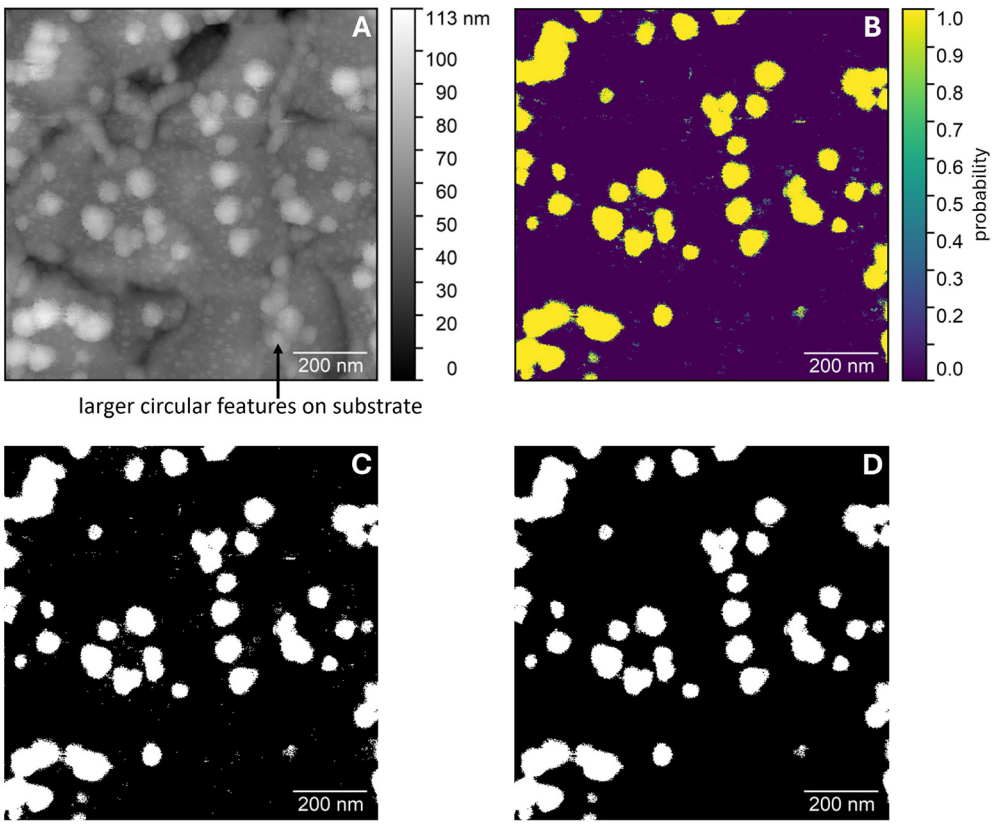

Fig. 5 (A) AFM contact height image of PG nanoparticles on the hard substrate (4-MPBA/gold). The region indicated by the black arrow highlights circular features on the substrate that are similar in shape and size to PG nanoparticles in the contact height image but are identified as part of the hard substrate by the supervised ML classifier. (B) The probability map output of the supervised particle/substrate ML classifier using the force–distance curves from the measurement in A, showing the probability that each pixel corresponds to an interaction of the AFM tip with PG. (C) The classification map output using a threshold probability of 50% to classify each pixel as corresponding to either PG (white pixels) or the hard substrate (black pixels). (D) The same classification map shown in C with the removal of small clusters (<35 pixels) of white pixels that accounted for only ~1.6% of PG pixels in C.

regions, as determined using the supervised particle/substrate ML classifier (Fig. 5D). From this comparison, we can see that the supervised particle/substrate ML classifier is able to accurately determine the locations of the PG nanoparticles on the substrate: PG nanoparticle regions in Fig. 5D correspond to an average probability of 98% in Fig. 5B. Fig. 5 also highlights a region in the contact height image that contains circular objects that closely resemble the PG nanoparticles (indicated by the black arrow in Fig. 5A). By comparing the contact height image of the input data in Fig. 5A to the particle regions determined by the supervised particle/substrate ML classifier output in Fig. 5D, we see that these circular objects, which could be mistaken for PG nanoparticles, were correctly classified by the supervised particle/substrate ML classifier as corresponding to the hard substrate. For completeness, the contact height image and the output from the supervised particle/substrate ML classifier for the additional new, unseen data are shown in Fig. S4A–D and S5A–D.

To quantitatively assess the performance of the supervised particle/substrate ML classifier, we compared the median of the Young's modulus values E of the pixels corresponding to the PG nanoparticles in the new, unseen data to the mean value $\bar{E}$ of the median Young's modulus values of isolated PG nanoparticles determined manually in previous work.[21] The values of E were determined using the fitting procedure outlined in ref. 21 in which a modified Hertz model was used to account for the close proximity of the underlying hard substrate.[50] A histogram of the values of E for all pixels corresponding to the PG nanoparticles (N = 112 897) as determined by the supervised particle/substrate ML classifier is shown in Fig. S7B. The median value of this distribution is 693 kPa with a standard error of the median of 2 kPa, which is in excellent agreement with the mean value $\bar{E}$ of the Young's modulus values of isolated PG particles reported previously ($\bar{E}$ = 688 kPa with a standard error of 23 kPa).[21] We also compared the average equivalent spherical radius, $\bar{r}$, of individual isolated PG nanoparticles in the new, unseen data to the average value $\bar{r}$ of hydrated PG nanoparticles reported previously.[21] The output of the supervised particle/substrate ML classifier consisted of 79 isolated PG nanoparticles (not on the boundary of the scan and not touching other particles) (Fig. S6). We calculated the equivalent spherical radius r of each of these particles using the method outlined in ref. 21 in which the volume V, corresponding to the sum of the volume of each pixel within the particle region, was calculated using the contact height image and the height of the underlying substrate. We then calculated the value of r using the relationship $r = (3V/4\pi)^{1/3}$. A histogram of the resulting values of r is shown in Fig. S7A for which the mean radius value $\bar{r}$ is 20.2 nm with a standard error of 0.7 nm. This value of $\bar{r}$ is in good agreement with the $\bar{r}$ value of 23.0 nm (with a standard error of 0.5 nm) determined previously using manual methods[21] as well as the nanoparticle radius determined using small angle neutron scattering[20,22] and dynamic light scattering.[30] The excellent agreement between the values of the average Young's modulus E and the equivalent spherical radius r of isolated PG nanoparticles found in the present study and those determined previously, including those using manual methods, provides confidence that the supervised particle/substrate ML classifier correctly identifies the location and lateral extent of the PG nanoparticles on the substrate.

Accuracy of the unsupervised particle/substrate machine learning classifier

To make a direct comparison between the supervised and unsupervised ML classifiers, we used the same testing dataset (104 858 force–distance curves) for testing both classifiers and compared the output from the unsupervised ML classifier to the labels used to test the supervised ML classifier. For the unsupervised particle/substrate ML classifier, we obtained an overall pixel classification accuracy of 99.0% with a false positive percentage of 0.4% and a false negative percentage of 7.2% when classifying the force–distance curves labelled as corresponding to PG (Fig. S8A and B). A visual comparison between the output for the supervised and unsupervised particle/substrate ML classifiers on the training/testing dataset is provided in Fig. S9. The large difference in the mechanical properties between the soft PG nanoparticles and the hard substrate results in a high classification accuracy of the unsupervised ML classifier with results that were very similar to those obtained for the supervised particle/substrate ML classifier.

Application of the unsupervised particle/substrate machine learning classifier to new, unseen data

The results of applying the trained unsupervised particle/substrate ML classifier to new, unseen data can be represented as a classification map of the locations of the PG nanoparticles (white regions in Fig. S10D–F) and the surrounding hard substrate (black regions in Fig. S10D–F). To include only entire PG nanoparticles, we removed small clusters of pixels ($<$35 pixels in area) that the classifier identified as corresponding to PG (Fig. S10G–I); this was the same procedure used for the supervised ML classifier. These small areas corresponded to only 1.3% of the pixels attributed to PG for all three AFM scans of new, unseen data (Fig. S10D–F).

To qualitatively assess the performance of the unsupervised particle/substrate ML classifier, we compared the contact height images of the three AFM scans of new, unseen data to the PG nanoparticle regions, as determined using the trained unsupervised particle/substrate ML classifier (Fig. S10). From this comparison, we can see that the trained unsupervised particle/substrate ML classifier is able to accurately determine the locations of the PG nanoparticles on the substrate.

To quantitatively assess the performance of the unsupervised particle/substrate ML classifier and to compare its performance to the supervised particle/substrate ML classifier, we examined the number of matching pixel classifications between the outputs of the trained unsupervised (Fig. S10D, E and F) and supervised (Fig. 5C and Fig. S4C, S5C) particle/substrate ML classifiers. We found that 98.5% (774 359 force–distance curves) of all pixels in the three AFM scans of new, unseen data (786 432 force–distance curves) had the same

particle/substrate classification in the outputs of the two ML classifiers. We note that we achieved similar results when applying the unsupervised classifier to the new, unseen data without training on the training/testing dataset.

The similarity in the outputs of the unsupervised and supervised particle/substrate ML classifiers indicates that materials with large differences in mechanical properties can easily be distinguished using the methods outlined in this work and may not require the need for supervised training using labelled data.

Accuracy of the supervised inner particle structure machine learning classifier

For the 4647 force–distance curves used to test the supervised inner particle structure ML classifier (testing dataset), we obtained an overall pixel classification accuracy of 86.0% with a false positive percentage of 11.1% and a false negative percentage of 18.4% when classifying force–distance curves labelled as corresponding to the stiffer inner region (Fig. S3A and B). We note that the difficulty in the manual labelling of the stiffer inner regions results in some human error that impacts the percentages of false positives and negatives in this analysis. The features that contributed the largest improvement in the accuracy of the supervised inner particle structure ML classifier span both the low and high force regimes in the retract and approach portion of the force–distance curves, describing (1) the recovery of the particle height as the applied force was reduced, and (2) deformations in the particle height as the applied force was increased. The top three most important features described the recovery of the particle height for forces between 2 nN and 0.5 nN and accounted for 55.4% of the improvement in the accuracy of the supervised inner particle structure ML classifier (Fig. S3C). The next three most important features described the deformation of the particles for forces between 0.5 nN and 2 nN and accounted for 13.1% of the improvement in the accuracy of the supervised inner particle structure ML classifier (Fig. S3C). All six of these features described changes in the particle height for forces greater than 0.5 nN, indicating that the stiffer inner structure of the PG nanoparticles has a significant influence on the force exerted on the AFM cantilever tip when the particle is compressed onto the substrate.

Application of the supervised internal particle structure machine learning classifier to new, unseen data

The results of applying the supervised internal particle structure classifier to new, unseen data consisting of force–distance curves from the PG nanoparticle regions determined from the supervised particle/substrate ML classifier can be represented as a probability map in which the value of each pixel represents the probability of being classified as corresponding to the stiffer inner region of the PG nanoparticles; an example of this is shown in Fig. 6C. Using a probability threshold of 50%, we generated a classification map of the locations of the stiffer inner regions of the PG nanoparticles, corresponding to the stiffer inner structure (white regions in Fig. 6D), and the surrounding softer outer regions (grey regions in Fig. 6D).

To qualitatively assess the performance of the supervised inner particle structure ML classifier, we compared the peak force height images (for which the PG nanoparticles are compressed onto the substrate revealing the stiffer inner structure), such as that shown in Fig. 6B, to the corresponding images of the stiffer inner regions determined by the supervised inner particle structure ML classifier in the classification map (Fig. 6D). From this comparison, we observed that the supervised inner particle structure ML classifier can accurately determine the location of the stiffer inner regions in PG (the white inner regions in Fig. 6D). For completeness, the peak force height image and the output from the supervised inner particle structure ML classifier for the additional new, unseen data are shown in Fig. S4E–G and S5E–G.

To quantitatively assess the performance of the supervised inner particle structure ML classifier, we compared the median of the Young's modulus values E of the pixels corresponding to the stiffer inner and softer outer regions in the new, unseen data to the mean value $\tilde{E}$ of the median Young's modulus values of the stiffer inner and softer outer regions determined previously using a manual approach in previous work.[21] A histogram of the Young's modulus E values corresponding to the stiffer inner (orange) and softer outer (blue) regions of the PG nanoparticles as determined by the supervised inner particle structure ML classifier is shown in Fig. S7C. The median E value for the stiffer inner region (N = 47 797) was 1006 kPa with a standard error of the median of 3 kPa and the median E value for the softer outer region (N = 65 100) was half that of the stiffer inner region, with a value of 503 kPa with a standard error of the median of 2 kPa. These absolute and relative values of the moduli are in excellent agreement with the average median E values of isolated PG nanoparticles determined previously for the stiffer inner region ($\tilde{E}$ = 1053 kPa with a standard error of 33 kPa) and softer outer region ($\tilde{E}$ = 518 kPa with a standard error of 20 kPa). This excellent agreement between the average Young's modulus E values for the stiffer inner and softer outer regions determined using the supervised inner particle structure ML classifier and those determined previously using a manual approach provides confidence that the supervised inner particle structure ML classifier correctly identifies the location and lateral extent of the stiffer inner regions within the PG nanoparticles.

In addition to the high level of accuracy and the elimination of user bias, the supervised ML based analysis also results in a significant reduction in the analysis time. After training the two supervised classifiers on a relatively small, labelled database (just two 512 pixel × 512 pixel AFM images), the time required to identify the particle locations and the stiffer inner and softer outer regions within the particles in new, unseen data was reduced from hours to tens of seconds.

Accuracy of the unsupervised inner particle structure machine learning classifier

For the 4360 force–distance curves used to test the unsupervised inner particle structure ML classifier (testing dataset), we obtained an overall pixel classification accuracy of 68.6% with a

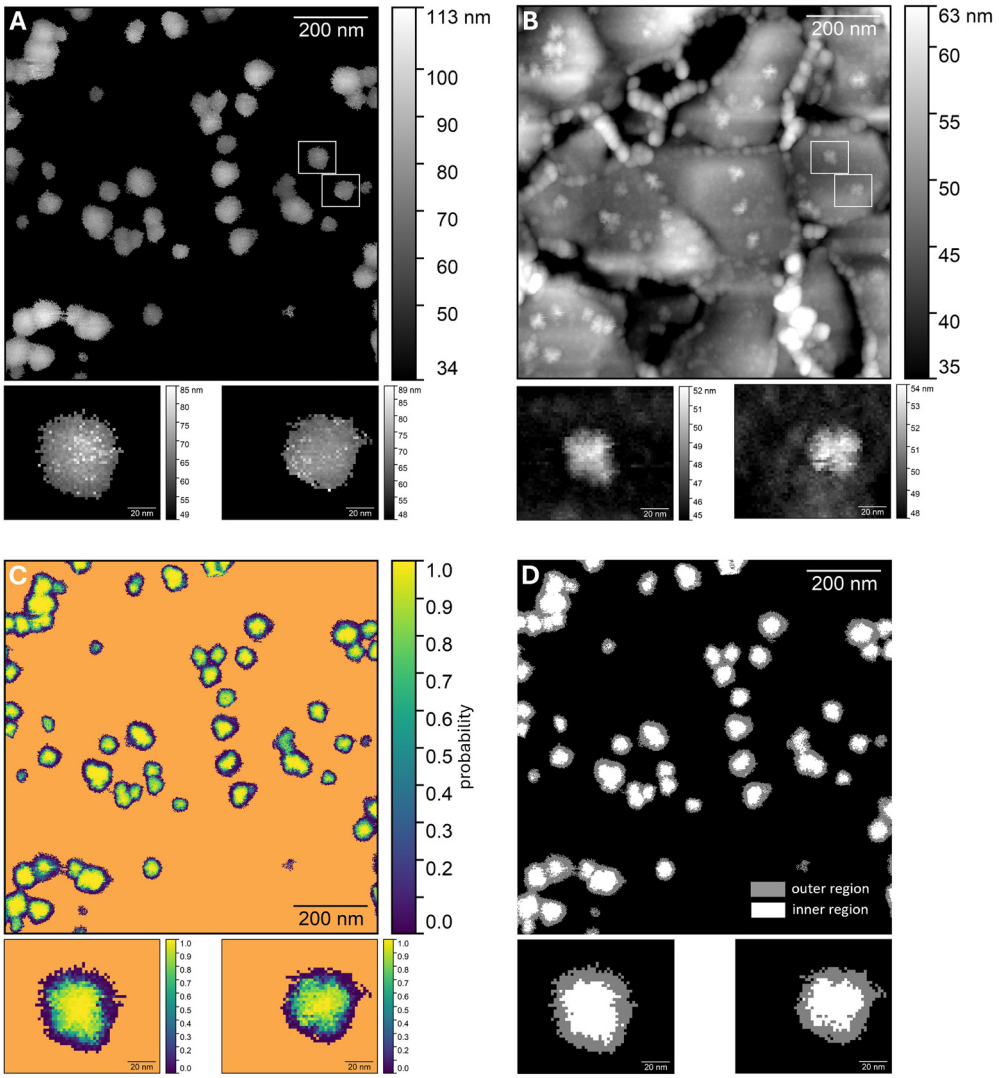

Fig. 6 (A, top) AFM contact height image of the regions corresponding to the PG nanoparticles, as determined by the supervised particle/substrate ML classifier. The area of the scan corresponds to 1 μm × 1 μm. (A, bottom) AFM contact height images at higher spatial resolution of two PG nanoparticles from the scanned region in A, highlighted by the white rectangles. (B, top) AFM peak force height image of the same scan in A revealing the stiffer inner structure within the PG nanoparticles when compressed onto the surface with an applied force of 2 nN, as well as the surrounding hard substrate (4-MPBA/gold). (B, bottom) AFM peak force height images at higher spatial resolution of the same two PG nanoparticles as in A, highlighted by the white rectangles. (C, top) The output of the supervised inner particle structure ML classifier using force–distance curves from the measurements shown in A and B, indicating the probability that each pixel corresponds to the stiffer inner region of the PG nanoparticles. (C, bottom) Output at higher spatial resolution for the same two PG nanoparticles as in A and B. (D, top) The output of the supervised inner particle structure ML classifier for the same region as shown in A, B and C, indicating the stiffer inner (white) and softer outer (grey) regions of the particles using a probability threshold of 50% to classify the two regions. (D, bottom) Output at higher spatial resolution for the same two PG nanoparticles as in A, B and C.

false positive percentage of 24.1% and a false negative percentage of 41.8% when classifying force–distance curves labelled as corresponding to the stiffer inner regions (Fig. S8C and D). These results are significantly worse than those obtained using the supervised inner particle structure ML classifier. The relatively small difference (~a factor of 2) in the Young's modulus of the stiffer inner region and softer outer region within the PG nanoparticles results in similar distributions of their feature values. This produces rather indistinct boundaries between the two regions within the feature space and results in a relatively poor match between the extent of the stiff inner regions identified by the unsupervised inner particle structure ML classifier and the manually labelled pixels indicating the locations of the stiff inner structure in the PG nanoparticles.

Application of the unsupervised internal particle structure machine learning classifier to new, unseen data

The results of applying the unsupervised internal particle structure classifier to new, unseen data consisting of force–distance curves from the PG nanoparticle regions determined from the unsupervised particle/substrate ML classifier can be represented as a classification map of the locations of the stiffer

inner regions of the PG nanoparticles, corresponding to the stiffer inner structure (white regions in Fig. S10J, K and L), and the surrounding softer outer regions (grey regions in Fig. S10J, K and L).

To qualitatively assess the performance of the unsupervised inner particle structure ML classifier, we compared the peak force height images (for which the PG nanoparticles are compressed onto the substrate revealing the stiffer inner structure) for the three new, unseen AFM scans (Fig. 6B and Fig. S4E and S5E) to the corresponding images of the stiffer inner regions determined by the trained unsupervised inner particle structure ML classifier in the classification maps (Fig. S10J, K and L). From this comparison, we observed that the unsupervised ML classifier inconsistently locates the stiffer inner regions within the PG nanoparticles and is less accurate than the corresponding supervised ML classifier. This is most likely due to the relatively small difference between the modulus values of the stiffer inner and softer outer regions, which results in rather indistinct boundaries between the two regions.

Summary and conclusions

We developed a supervised machine learning (ML) approach to analyze atomic force microscopy force spectroscopy (AFM-FS) data collected on soft phytoglycogen (PG) nanoparticles. This approach involved the use of two supervised ML classifiers that used features within the approach and retraction portions of the force–distance curves measured in the AFM-FS experiments: one classifier to identify the location and spatial extent of soft nanoparticles on the underlying hard substrate, and the other classifier to identify and quantify the stiffer inner and softer outer regions within individual nanoparticles. We compared the accuracy of the supervised ML classifiers with two corresponding unsupervised ML classifiers that used *k*-means clustering. After training all the classifiers using data collected in just two AFM images and applying them to new, unseen AFM-FS images, we found comparable accuracy in identifying the location and spatial extent of the soft nanoparticles for the supervised and unsupervised approaches, whereas the supervised ML classifier was significantly more successful in identifying and quantifying the stiffer inner and softer outer regions within the particles. These ML approaches significantly decrease the time to identify PG nanoparticles on the surface and the nanoscale variations in the mechanical properties within the nanoparticles compared to manual methods and eliminate user bias which could lead to incorrect identification. Our results provide an improved analysis technique for understanding the morphology of PG nanoparticles and show that ML classifiers can be used to accurately and efficiently identify and quantify nanoscale variations in the mechanical properties of soft, biological materials. More generally, our results provide a framework for characterizing sample morphology and identifying different materials within a sample using feature engineering of AFM-FS data and ML classifiers. We believe that this framework can be implemented and further developed for use in interpreting AFM-FS measurements of many different soft and biological materials including nanoparticles, cells, polymers, and proteins to reveal or classify underlying structures, and differentiate different materials in multi-component systems.

Author contributions

Benjamin Baylis: conceptualization, investigation, methodology, formal analysis, writing – original draft. John Dutcher: conceptualization, supervision, project administration, funding acquisition, formal analysis, writing – review & editing.

Conflicts of interest

We declare that J. R. D. is a Founder of Mirexus Biotechnologies Inc., who supplied the phytoglycogen nanoparticles for the present study.

Data availability

AFM images are included in the manuscript and supplementary information (SI), and the analysis code is available at: https://github.com/Dutcher-Lab/AFM-Material-Classification. Supplementary information is available. See DOI: https://doi.org/10.1039/d5sm00943j.

Acknowledgements

Mirexus Biotechnologies Inc. generously supplied the phytoglycogen nanoparticles. We thank Michael Grossutti for valuable discussions. This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada (J. R. D.). J. R. D. is the recipient of a CEPS Research Chair in Novel Sustainable Nanomaterials.

References

1 Z. Xia, L.-Y. Wu, X. Zhou and S. T. C. Wong, Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces, *BMC Syst. Biol.*, 2010, **4**, 1–16.
2 H. Zhou, M. Gao and J. Skolnick, Comprehensive prediction of drug-protein interactions and side effects for the human proteome, *Sci. Rep.*, 2015, **5**, 11090.
3 Y. Su, Y. Wang, J. Jiao and Y. Guo, Automatic detection and classification of breast tumors in ultrasonic images using texture and morphological features, *Open Med. Inform. J.*, 2011, **5**, 26–37.
4 N. Codella, Q.-B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern and J. R. Smith, Deep learning ensembles for melanoma recognition in dermoscopy images, *IBM J. Res. Dev.*, 2017, **61**(5), 1–5.
5 S. Sarraf, D. D. DeSouza, J. Anderson and G. Tofighi, DeepAD: Alzheimer's disease classification via deep

convolutional neural networks using MRI and fMRI, *BioRxiv*, 2016, 070441.
6 J. Timoshenko, D. Lu, Y. Lin and A. I. Frenkel, Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles, *J. Phys. Chem. Lett.*, 2017, **8**, 5091–5098.
7 M. Wang, T. Wang, P. Cai and X. Chen, Nanomaterials discovery and design through machine learning, *Small Methods*, 2019, **3**, 1900025.
8 M. Checa, R. Millan-Solsona, A. G. Mares, S. Pujals and G. Gomila, Fast label-free nanoscale composition mapping of eukaryotic cells via scanning dielectric force volume microscopy and machine learning, *Small Methods*, 2021, **5**, 2100279.
9 J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei and M. Lei, Machine learning in materials science, *InfoMat*, 2019, **1**, 338–358.
10 S. S. Chong, Y. S. Ng, H.-Q. Wang and J.-C. Zheng, Advances of machine learning in materials science: Ideas and techniques, *Front. Phys.*, 2024, **19**, 13501.
11 H. Zheng, J. Yuan and L. Chen, Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation, *Energies*, 2017, **10**, 1168.
12 R. Garcia, Nanomechanical mapping of soft materials with the atomic force microscope: methods, theory and applications, *Chem. Soc. Rev.*, 2020, **49**, 5850–5884.
13 R. Garcia and J. R. Tejedor, Advances in nanomechanical property mapping by atomic force microscopy, *Nanoscale Adv.*, 2025, **7**, 6286–6307.
14 X. Xu, H. Feng, Y. Zhao, Y. Shi, W. Feng, X. J. Loh, G. J. Vancso and S. Guo, AFM-based nanomechanics and machine learning for rapid and non-destructive detection of bacterial viability, *Cell Rep. Phys. Sci.*, 2024, **5**, 101902.
15 S. Prasad, A. Rankine, T. Prasad, P. Song, M. E. Dokukin, N. Makarova, V. Backman and I. Sokolov, Atomic force microscopy detects the difference in cancer cells of different neoplastic aggressiveness via machine learning, *Adv. Nano-Biomed Res.*, 2021, **1**, 2000116.
16 M. Petrov and I. Sokolov, Machine Learning Allows for Distinguishing Precancerous and Cancerous Human Epithelial Cervical Cells Using High-Resolution AFM Imaging of Adhesion Maps, *Cells*, 2023, **12**, 2356.
17 I. Sokolov, M. E. Dokukin, V. Kalaparthi, M. Miljkovic, A. Wang, J. D. Seigne, P. Grivas and E. Demidenko, Non-invasive diagnostic imaging using machine-learning analysis of nanoresolution images of cell surfaces: Detection of bladder cancer, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 12920–12925.
18 M. Petrov and I. Sokolov, Identification of geometrical features of cell surface responsible for cancer aggressiveness: Machine learning analysis of atomic force microscopy images of human colorectal epithelial cells, *Biomedicines*, 2023, **11**, 191.
19 M. Petrov, D. Canena, N. Kulachenkov, N. Kumar, P. Nickmilder, P. Leclère and I. Sokolov, Mechanical spectroscopy of materials using atomic force microscopy (AFM-MS), *Mater. Today*, 2024, **80**, 218–225.
20 J. D. Nickels, J. Atkinson, E. Papp-Szabo, C. Stanley, S. O. Diallo, S. Perticaroli, B. Baylis, P. Mahon, G. Ehlers, J. Katsaras and J. R. Dutcher, Structure and hydration of highly-branched, monodisperse phytoglycogen nanoparticles, *Biomacromolecules*, 2016, **17**, 735–743.
21 B. Baylis, E. Shelton, M. Grossutti and J. R. Dutcher, Force Spectroscopy Mapping of the Effect of Hydration on the Stiffness and Deformability of Phytoglycogen Nanoparticles, *Biomacromolecules*, 2021, **22**, 2985–2995.
22 J. Simmons, J. D. Nickels, M. Michalski, M. Grossutti, H. Shamana, C. B. Stanley, A. L. Schwan, J. Katsaras and J. R. Dutcher, structure, hydration, and interactions of native and hydrophobically modified phytoglycogen nanoparticles, *Biomacromolecules*, 2020, **21**, 4053–4062.
23 B. Morling, S. Luyben, J. R. Dutcher and R. A. Wickham, Efficient Modeling of High-Generation Dendrimers in Solution Using Dynamical Self-Consistent Field Theory, *Macromolecules*, 2024, **57**, 4617–4628.
24 H. Shamana, M. Grossutti, E. Papp-Szabo, C. Miki and J. R. Dutcher, Unusual polysaccharide rheology of aqueous dispersions of soft phytoglycogen nanoparticles, *Soft Matter*, 2018, **14**, 6496–6505.
25 C. Miki and J. R. Dutcher, Phytoglycogen, *Handbook of Starch Science and Technology*, CRC Press, 2025, pp. 104–111.
26 M. Grossutti and J. R. Dutcher, Correlation between chain architecture and hydration water structure in polysaccharides, *Biomacromolecules*, 2016, **17**, 1198–1204.
27 M. Grossutti, E. Bergmann, B. Baylis and J. R. Dutcher, Equilibrium swelling, interstitial forces, *Langmuir*, 2017, **33**, 2810–2816.
28 M. Grossutti and J. R. Dutcher, Correlation of mechanical and hydration properties of soft phytoglycogen nanoparticles, *Carbohydr. Polym.*, 2021, **251**, 116980.
29 M. Grossutti and J. R. Dutcher, Hydration water structure, hydration forces, and mechanical properties of polysaccharide films, *Biomacromolecules*, 2020, **21**, 4871–4877.
30 H. Shamana and J. R. Dutcher, Transition in the glassy dynamics of melts of acid-hydrolyzed phytoglycogen nanoparticles, *Biomacromolecules*, 2022, **23**, 2040–2050.
31 K. Charlesworth, N. van Heijst, A. Maxwell, B. Baylis, M. Grossutti, J. J. Leitch and J. R. Dutcher, Binding affinity of concanavalin A to native and acid-hydrolyzed phytoglycogen nanoparticles, *Biomacromolecules*, 2022, **23**, 4778–4785.
32 N. van Heijst, P. Whiting and J. R. Dutcher, Solubilization of Hydrophobic Astaxanthin in Water by Physical Association with Phytoglycogen Nanoparticles, *Biomacromolecules*, 2024, **25**, 4110–4117.
33 L. Roman, B. Baylis, K. Klinger, J. D. Jong, J. R. Dutcher and M. M. Martinez, Changes to fine structure, size and mechanical modulus of phytoglycogen nanoparticles subjected to high-shear extrusion, *Carbohydr. Polym.*, 2022, **298**, 120080.
34 B. Baylis, Morphology and Mechanical Stiffness of Soft Phytoglycogen Nanoparticles Revealed by AFM Force Spectroscopy, PhD thesis, University of Guelph, Guelph, ON, 2021.

35 Y. El-Rayyes, Particle Radius and Young's Modulus of Acid-Hydrolyzed Phytoglycogen Nanoparticles, MSc thesis, University of Guelph, Guelph, ON, 2024.
36 W. Feng, Z. Wang, O. H. Campanella, T. Zhang and M. Miao, Fabrication of phytoglycogen-derived core-shell nanoparticles: Structure and characterizations, *Food Chem.*, 2023, **423**, 136317.
37 C. Miki and J. R. Dutcher, Hydration Properties of Chemically Modified Phytoglycogen Nanoparticles, *J. Phys. Chem. B*, 2025, **129**, 8573–8581.
38 C. Miki and J. R. Dutcher, Impact of Cations on the Hydration Properties of Carboxymethyl-Phytoglycogen Nanoparticles, *J. Phys. Chem. C*, 2025, **129**, 12512–12519.
39 M. Grossutti, C. Miki and J. R. Dutcher, Phytoglycogen nanoparticles: key properties relevant to its use as a natural moisturising agent, *Household Pers. Care Today*, 2017, **12**(1), 47–51.
40 J. Baumann, F. Wandrey and F. Zülli, Energizing the skin with phytoglycogen, *Personal Care*, 2024, 36–39.
41 Q. A. Besford, F. Cavalieri and F. Caruso, Glycogen as a Building Block for Advanced Biological Materials, *Adv. Mater.*, 2020, **32**, 1904625.
42 L. Chen, N. Zhao, D. J. McClements, B. R. Hamaker and M. Miao, Advanced dendritic glucan-derived biomaterials: From molecular structure to versatile applications, *Compr. Rev. Food Sci. Food Saf.*, 2023, **22**(5), 4107–4146.
43 Y. Shi, S. Chen, H. Bai, L. Chen and M. Miao, Phytoglycogen-based Systems, *Bioactive Delivery Systems for Lipophilic Nutraceuticals: Formulation, Fabrication, and Application*, Royal Society of Chemistry, 2023, pp. 322–346.
44 F. Abbasi, J. Alvarez-Malmagro, Z. Su, J. J. Leitch and J. Lipkowski, Pore forming properties of alamethicin in negatively charged floating bilayer lipid membranes supported on gold electrodes, *Langmuir*, 2018, **34**, 13754–13765.
45 J. L. Hutter and J. Bechhoefer, Calibration of Atomic-Force Microscope Tips, *Rev. Sci. Instrum.*, 1993, **64**, 1868–1873.
46 T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international Theor. Appl. conference on knowledge discovery and data mining*, 2016, pp. 785–794.
47 R. Santhanam, N. Uzir, S. Raman and S. Banerjee, Experimenting XGBoost algorithm for prediction and classification of different datasets, *Int. J. Control Theor. Appl.*, 2016, **9**(40), 651–662.
48 D. Nielsen, Tree Boosting with XGBoost: Why Does XGBoost Win Every Machine Learning Competition?, MSc thesis, NTNU, 2016.
49 A. Likas, N. Vlassis and J. J. Verbeek, The global *k*-means clustering algorithm, *Pattern Recognit.*, 2003, **36**, 451–461.
50 E. K. Dimitriadis, F. Horkay, J. Maresca, B. Kachar and R. S. Chadwick, Determination of elastic moduli of thin layers of soft material using the atomic force microscope, *Biophys. J.*, 2002, **82**, 2798–2810.