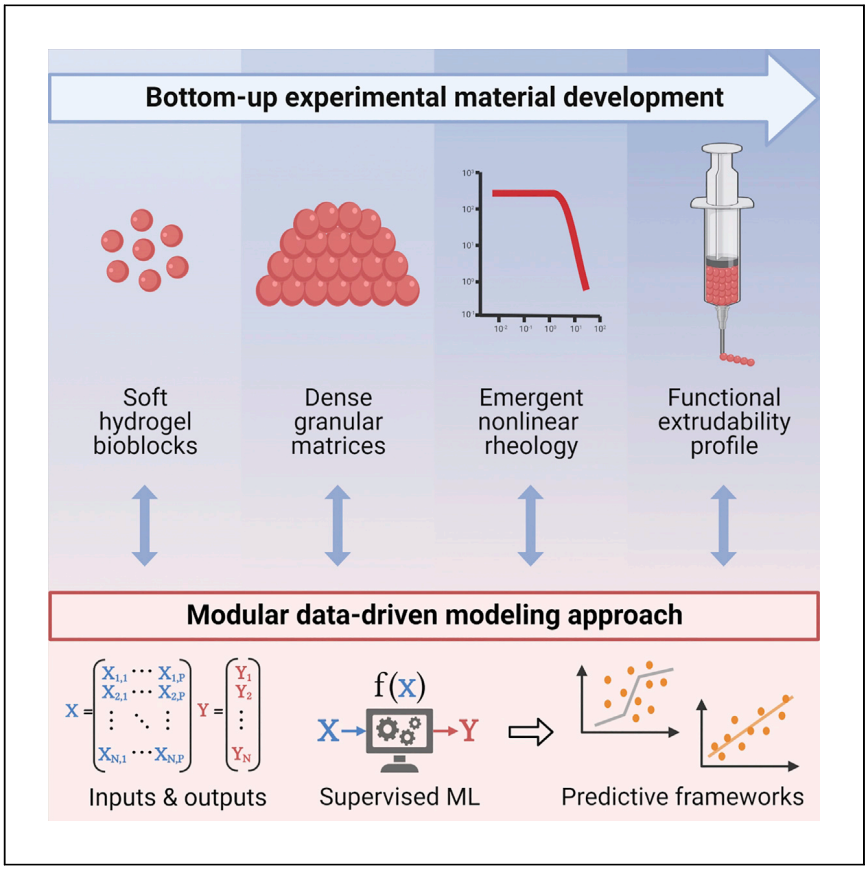
Matter

CellPress
OPEN ACCESS

Article

Integrated data-driven modeling and experimental optimization of granular hydrogel matrices



We tackled the challenge of designing soft granular matrices with predictable structures, properties, and performance for use in biomedical applications. Experimentally, we report the assembly of hydrogel bioblocks and their consolidation into granular matrices with emergent non-linear rheological behavior and functional extrudability. At each stage, our modular ML approach is leveraged to derive data-driven predictive models and design rules. Our generalizable approach should be applicable for data-driven advancement of any complex material system.

Demonstrate
Proof-of-concept of performance with intended application/response

Connor A. Verheyen, Sebastien G.M. Uzel, Armand Kurum, Ellen T. Roche, Jennifer A. Lewis

connorv@mit.edu

Highlights

Built flexible ML pipeline for robust model selection, validation, and explanation

Applied modular ML approach to stepwise empirical material development workflow

Optimized hydrogel bioblocks, granular matrices, complex rheology, and extrudability

Produced data-driven models and extracted human-readable predictive design insights

Verheyen et al., Matter 6, 1–22
March 1, 2023
https://doi.org/10.1016/j.matt.2023.01.011

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

Matter

CellPress
OPEN ACCESS

Article

Integrated data-driven modeling and experimental optimization of granular hydrogel matrices

Connor A. Verheyen,[1,2,3,6,*] Sebastien G.M. Uzel,[3,4] Armand Kurum,[3,4] Ellen T. Roche,[1,2,5] and Jennifer A. Lewis[3,4]

SUMMARY

Granular hydrogel matrices have emerged as promising candidates for cell encapsulation, bioprinting, and tissue engineering. However, it remains challenging to design and optimize these materials given their broad compositional and processing parameter space. Here, we combine experimentation and computation to create granular matrices composed of alginate-based bioblocks with controlled structure, rheological properties, and injectability profiles. A custom machine learning pipeline is applied after each phase of experimentation to automatically map the multidimensional input-output patterns into condensed data-driven models. These models are used to assess generalizable predictability and define high-level design rules to guide subsequent phases of development and characterization. Our integrated, modular approach opens new avenues to understanding and controlling the behavior of complex soft materials.

INTRODUCTION

Granular hydrogel matrices are an emerging class of soft matter that offer several advantages over traditional biomaterials. Composed of discrete, yet densely packed building blocks, these materials are promising for a wide range of biomedical applications.[1,2] For example, granular matrices composed of hydrogel bioblocks can encapsulate drugs, biologics, and cells,[3–5] serve as an ink or support matrix for *in vitro* bioprinting,[6–9] or be injected into cavities, open wounds, or damaged cardiac tissue for *in vivo* tissue engineering.[10–13] Despite such promise, they remain challenging to design, assemble, and optimize. Individual hydrogel bioblocks must first be generated (e.g., via microfluidics, fragmentation, bulk emulsion) and then consolidated into densely packed granular matrices that exhibit the reversible yielding and shear-thinning behavior required for bioprinting and injectability.[1–3,14–18] Open challenges in this workflow include the scalable and tunable formation of user-defined hydrogel bioblocks, the dynamic evolution of bioblocks and their compaction into densely packed granular matrices, the emergent non-linear rheology of soft granular matrices, and the controlled flow of soft granular matrices in confined geometries.[19–25]

In data-driven modeling (Figure 1A), supervised machine learning (ML) is applied to material databases to automatically build predictive frameworks directly from the data itself.[26–29] Unlike humans, computers can learn arbitrarily complex patterns from heterogeneous and high-dimensional data without pre-defined theoretical frameworks and without a bias toward positive or recent trials. To date, researchers have successfully harnessed this approach at the atomic and molecular levels, often

PROGRESS AND POTENTIAL

Granular hydrogel matrices are promising for biomedical applications ranging from extrusion-based bioprinting to injectable tissue engineering. However, they remain challenging to design, assemble, and optimize. Each development stage involves multidimensional input-output spaces affected by poorly understood multi-scale, multi-physics phenomena. Here, we demonstrate the utility of a flexible and modular machine learning (ML) approach to advance complex materials in a stepwise fashion. We apply our ML approach to automatically construct, validate, and explain predictive design frameworks for each set of empirical results. These data-driven models allow one to assess each experimental design space and provide condensed design insights extracted from high-dimensional input-output maps. The resulting bioblock materials have broad biomedical applications, yet our approach should be applicable for data-driven advancement of any complex material system.

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

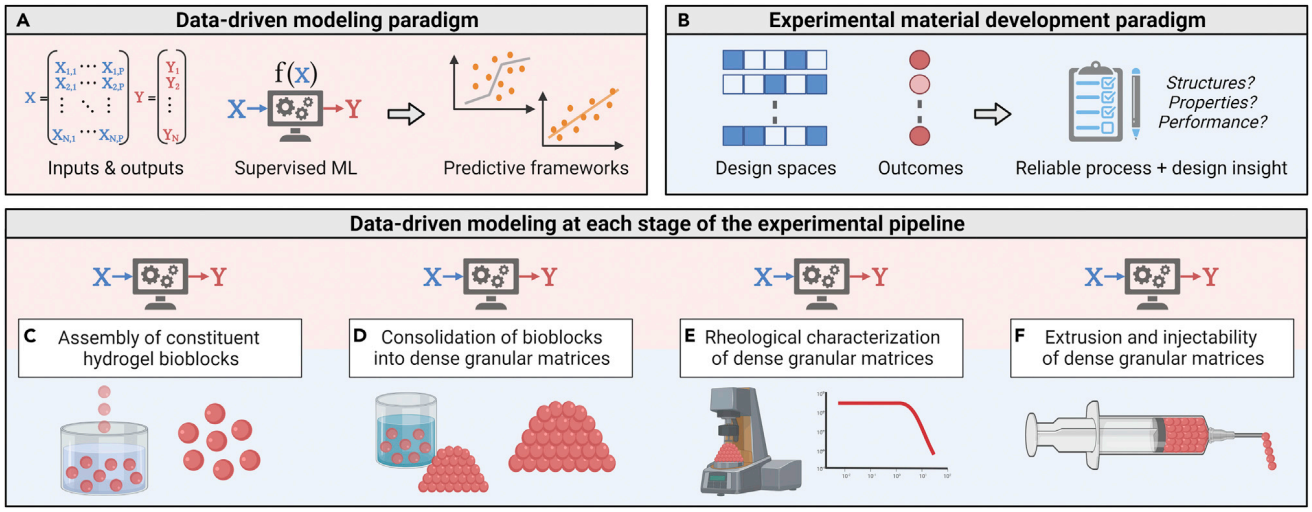


Figure 1. An integrated data-driven experimental pipeline for soft granular matrices

(A and B) The data-driven paradigm leverages machine learning to build predictive frameworks directly from data itself (A). The experimental paradigm relies on empirical exploration to optimize structures, properties, and performance of complex materials (B).

(C–F) Our modular approach integrates data-driven modeling with experimentation to map the input-output relationships at each stage of soft granular matrix development.

relying on large simulation-derived databases.[28–30] By contrast, less attention has been given to using ML for experimental granular-scale soft matter.[30–32] Further, ML tools are often used to map a single input-output space, whereas full material lifecycles involve many potential input-output spaces at many different stages of development.[33]

In experimental optimization (Figure 1B), researchers explore potential design spaces to uncover reliable processing routes and structure, property, and performance insights.[34,35] Links between parameters and outcomes are informally encoded as "expertise" or formally encoded in design plots and mathematical models. Many materials were advanced this way, but there are limitations. First, humans are not adept at handling high-dimensional data, so processes with multiple inputs pose challenges for evaluation.[31,33] Second, humans are subject to positive-results bias, recency bias, or confirmation bias that can inadvertently distort analysis.[26,36] The omission of negative results, reporting of subsets, or failure to include confounders can skew conclusions. Finally, complex materials like granular bioblock matrices exhibit multi-scale, multi-physics phenomena that are difficult to describe or model, hindering the translation of experimental results into predictive design frameworks.[31,33,37] Indeed, manual derivation of governing equations at each step of development would be intractable.[38]

We posit that data-driven modeling could be coupled with experimental optimization to assess the predictability of soft granular material design spaces and delineate high-level input-output relationships. Hence, we propose to combine a flexible ML workflow with structured empirical results to derive predictive design frameworks at each stage of material development. Specifically, we integrate data-driven modeling with experiments to create extrudable and injectable granular matrices composed of alginate-based bioblocks (Figures 1C–1F). Computationally, we focused on automated tuning and selection of algorithms, rigorous evaluation via multi-metric grouped and nested cross-validation, and simplified predictive maps for human-readable insight into n-dimensional design spaces. Experimentally, we

[1]Harvard–MIT Program in Health Sciences and Technology, Cambridge, MA 02139, USA

[2]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA

[4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[5]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Lead contact

*Correspondence: connorv@mit.edu

https://doi.org/10.1016/j.matt.2023.01.011

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

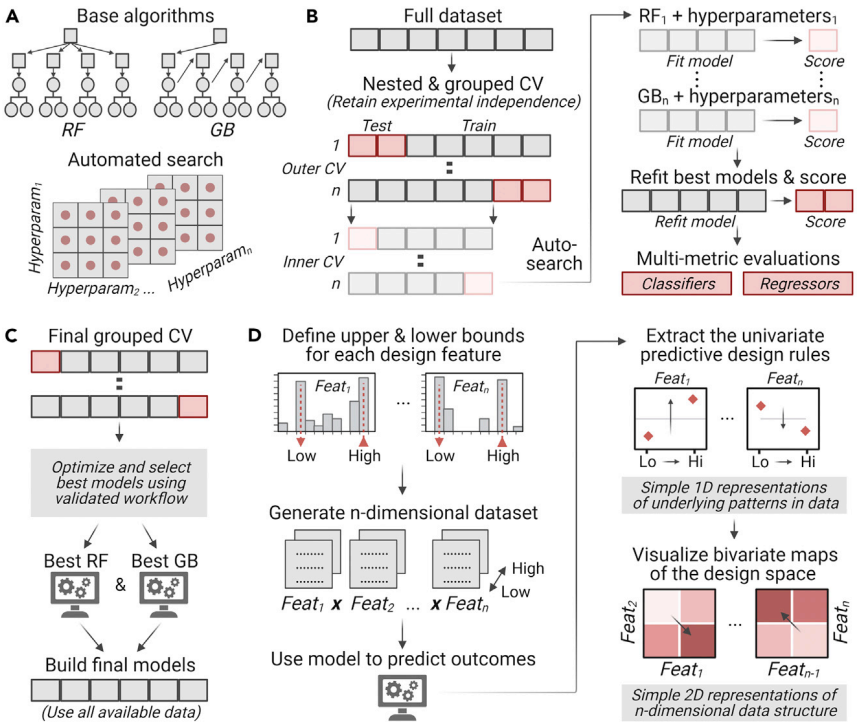


Figure 2. Flexible data-driven modeling pipeline for complex materials

(A) Base algorithms (random forest, gradient boosting) were tuned and selected in an automated grid search.

(B) The modeling workflow was evaluated via nested and grouped cross-validation with multi-metric evaluation.

(C) Optimal configurations were identified in a final cross-validation. Optimized algorithms were trained on all data and individual predictions were combined.

(D) Simplified synthetic datasets were constructed and fed into trained models to make new predictions. Human-readable 1D and 2D predictive maps were extracted from the n-dimensional design space to intuitively visualize the learned patterns.

focused on the scalable generation of alginate-based bioblocks (Figure 1C), which are subsequently compacted into dense granular matrices (Figure 1D) with tunable rheological properties (Figure 1E) that facilitate controlled delivery during extrusion or injection (Figure 1F). At each step, we leverage our modular ML approach to (1) assess whether the empirical data structure is learnable and generalizable, and (2) identify the underlying relationships among design, structure, property, and performance outcomes. We find that these ML models facilitate transparent data-driven progression through the material processing pipeline, from initial bioblock assembly to final functional characterization. We expect this integrated approach will be applicable for a broad range of soft and living materials.

RESULTS

Data-driven modeling pipeline for complex materials

We converted unstructured experimental results from physical and digital records into structured machine-readable datasets with input design matrices and corresponding output vectors.[39] We based our computational workflow on tree-based ensemble algorithms, namely random forest (RF) and gradient boosting (GB) (Figure 2A). These non-parametric algorithms are chosen because they can flexibly handle classification or regression tasks, non-linear relationships, high-dimensional data, and mixed datatypes.[40] RF constructs a simple-averaged ensemble in parallel,

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

using random bootstrapped data with random feature subsets.[41] GB constructs a weighted-average ensemble in series, using weighted bootstrapped data with all features.[42] In general, ensembles provide better predictive capacity and stability than single models.[43–45] Given the variability between datasets, base algorithms are tuned for optimal performance on each problem.[46,47] We applied an automated grid search to appraise hyperparameter configurations and select models with sufficient flexibility to fit signal without overfitting to noise[48] (Figure 2A). Because our datasets are empirically derived, we must consider issues like repeated measures, batch effects, and uneven sampling.[49–51] To rigorously evaluate the model-building process, we applied a nested and grouped cross-validation (CV) procedure[49,52,53] (Figure 2B). The full dataset is subjected to an "outer" k-fold CV and each outer training fold is subjected to an "inner" k-fold CV. Configured algorithms are trained, scored, and selected via the inner protocol, then top performers are re-fit and scored on unseen outer test data. Because no scoring metric is perfect, we used three different metrics for each problem to obtain more comprehensive estimates of model performance (accuracy, area under the receiver operating characteristic curve [ROC-AUC], F1 for classification; r^2, mean absolute error [MAE], median absolute error [AE] for regression).[54–56] To avoid data leakage, we assigned unique IDs for each experiment and used a grouping procedure to ensure the same experiment could not appear in both train and test simultaneously.[57] With a single holdout set, performance estimates may display high variance or optimistic bias (if experimental errors are minimized over time or if the samples are in a densely sampled region of the training distribution).[58,59] However, our approach synthetically used all available experiments as holdouts to create a composite score independent of (1) when the trials are conducted, and (2) what the specific trial conditions are. Accordingly, if some trials display significant noise or batch effects, or if a sparsely sampled region is unpredictable, the composite score would be properly penalized (unlike a single holdout that may not capture such variability or failure).[58,60] Thus, we report scores that should reflect the generalizability and reproducibility of our data-driven models for each particular experimental phase.

Next, we applied a standard k-fold CV to automatically configure and select top-performing algorithms using the validated modeling process (Figure 2C). These final configurations are trained on all available data and averaged into a final ensemble to further improve stability.[43–45] At this stage, our data-driven models can be deployed directly for predictive analytics (i.e., predicting material outcomes for candidate design inputs) or combinatorial optimization (i.e., searching for optimal design inputs for target material outcomes). However, black-box models and high-dimensional patterns are incomprehensible to human users.[61–63] As experimentalists, we desired a human-in-the-loop approach to develop data-driven intuition about the input-output associations in our problem domain, identify promising avenues for further exploration, and explain the models via condensed design summaries.[61,63] Accordingly, we extracted human-readable predictive maps of the design space. First, simplified 2xn synthetic datasets are created by selecting low and high values for each input (within the training distributions) and then generating a matrix containing pairwise combinations (to obtain coverage of the design space) (Figure 2D). These simplified datasets are fed into our trained models to generate predictions for each set of inputs. Visual maps are then created by plotting predicted outputs against the bilevel inputs. These maps represent 1D or 2D slices of the n-dimensional patterns learned by the data-driven model (with contributions from other variables averaged out by the combinatorial strategy). Such visualizations provide human users with easily understood, high-level takeaways about the direction and magnitude of data-driven input-output relationships[62,64] (Figure 2D).

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

Finally, we apply this customized ML workflow to granular matrix development, covering the entire continuum from initial bioblock assembly to final performance. Below we provide overview schematics, representative results, aggregated evaluations, and high-level design takeaways for each integrated experiment and model. These condensed highlights are paired with in-depth supplemental information covering both experimental and computational results (e.g., empirical dataset breakdowns, full model evaluations, feature importance values, all univariate and bivariate design maps).

Controlled assembly of hydrogel bioblocks

We assembled hydrogel bioblocks composed of ionically crosslinked alginate via a droplet-based method.[3,21,65,66] Alginate precursor solutions are extruded from a syringe into a perpendicular airstream, where droplets form, detach, and are ultimately gelled within an aqueous reservoir that contains calcium ions (Ca^{2+}) (Figure 3A). In our first set of experiments, we tested combinations of intrinsic (e.g., alginate and ion concentration) and extrinsic (e.g., extrusion rate and air pressure) parameters and recorded formation success in a binary manner (Figures 3A, 3B, and S1–S6). We trained a classifier (Tables S1 and S2) to predict bioblock formation as a function of six different assembly inputs and observed high accuracy (0.912), ROC-AUC (0.902), and F1 scores (0.898), suggesting the underlying patterns could be effectively learned. (Figure 3C). Univariate maps revealed the importance of intrinsic parameters, with large increases in the predicted probability of formation if alginate concentration or precursor solution viscosity are increased (Figures 3D and S5). Extrinsic assembly parameters had a smaller effect, with the exception of air pressure, which displayed a sizable inverse relationship. The bivariate map showed that low-viscosity precursor and high pressure would likely end in failure (P(formation) = 0.23), while the opposite conditions would typically lead to stable bioblock formation (P(formation) = 0.94) (Figures 3E and S6).

We categorized these alginate bioblocks by their shape (tail, oblong, or sphere) and built a multi-class classifier (Tables S1 and S2) to model assembly shape relationships (Figures 3A, 3F, and S7–S12). The scores are good (Acc = 0.797, ROC-AUC = 0.829, F1 = 0.742), but we noted a decrease in predictive capacity compared with the formation classifier (Figures 3C and 3G). This decline is attributed to the non-deterministic nature of the shape data, where identical assembly conditions yielded differently shaped bioblocks (this stochastic phenomenon is also present in the formation data but is much less prevalent) (Figures S2 and S8). Precursor solution viscosity again emerged as a dominant parameter, whereas alginate concentration and air pressure have a smaller effect on shape than on formation (Figure 3H). Instead, other inputs, such as the surface tension of the gelling reservoir, are key determinants of bioblock shape (Figure S11). Hence, the relative impact of a given design parameter depends on the specific input-output relationship being studied. In the 2D design maps, we observed that increased precursor viscosity and decreased surface tension favored spherical bioblocks (P(spherical) = 0.87), while the opposite conditions favored those that are oblong or contain tails (P(spherical) = 0.40) (Figures 3I and S12).

Next, we excluded low-performance regions of the design space (associated with failed or aspherical bioblocks) in these experiments, while freely exploring the effects of assembly parameters on bioblock size (Figures 3J and S13–S18). We built a regressor (Tables S1 and S2) to predict their size as a function of nine different inputs, finding a high r^2 value (0.897) and mean (45.4 μm) and median absolute errors (32.8 μm) below 50 μm (Figure 3K). These errors were small relative to their size

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

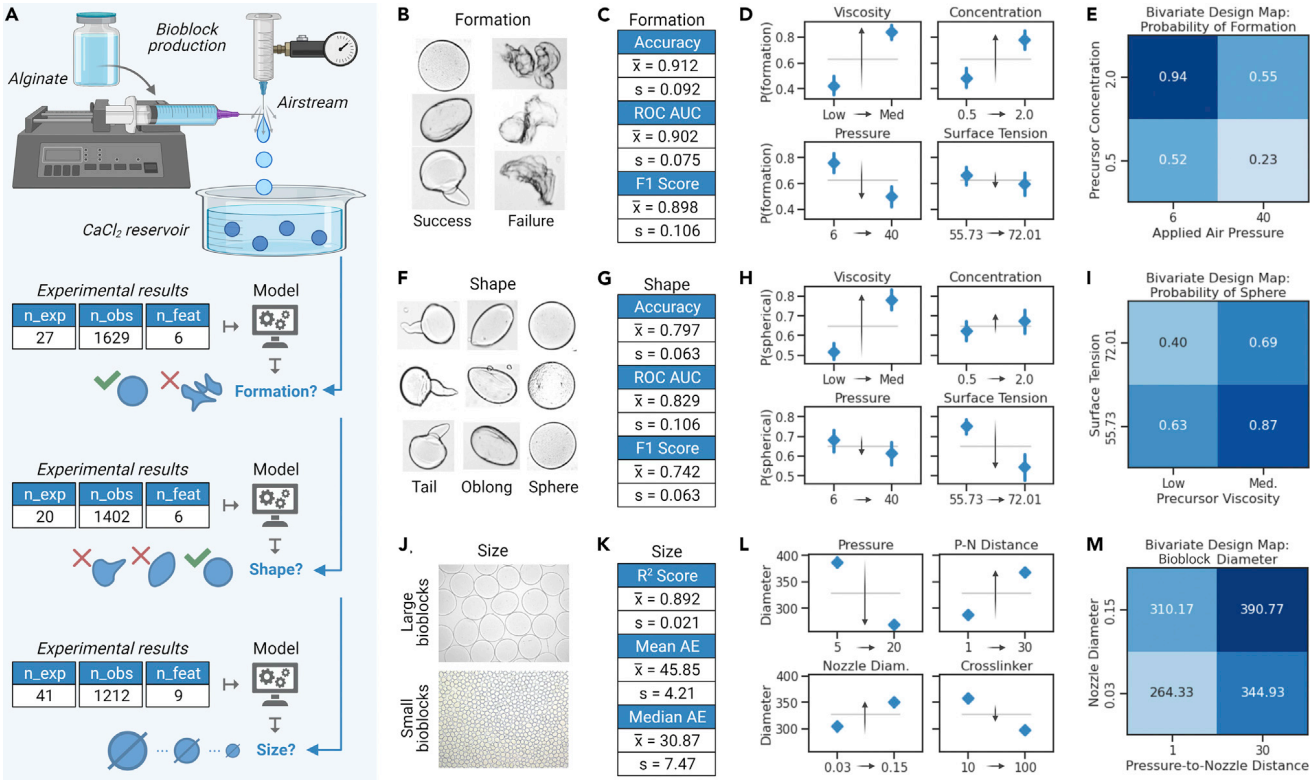


Figure 3. Controlled assembly of hydrogel bioblocks

(A) Overview of the droplet-based bioblock fabrication and corresponding data-driven models. Summary boxes display the number of unique experiments, number of recorded observations, and number of input design parameters.

(B) Optical images of successfully formed or failed alginate bioblocks.

(C) Accuracy, ROC-AUC, and F1 scores for the data-driven classifier predicting bioblock formation.

(D) Predicted probability of bioblock formation given hypothetical changes in precursor viscosity, precursor concentration, applied air pressure, or gelation bath surface tension.

(E) Predicted probability of bioblock formation given hypothetical changes in both precursor concentration and applied air pressure.

(F) Optical images depicting examples of alginate bioblocks with tails, oblong bioblocks, and spherical bioblocks.

(G) Accuracy, ROC-AUC, and F1 scores for the data-driven classifier predicting bioblock shapes.

(H) Predicted probability of a spherical bioblock given hypothetical changes in precursor viscosity, precursor concentration, applied air pressure, or gelation bath surface tension.

(I) Predicted probability of a spherical bioblock given hypothetical changes in both bath surface tension and precursor viscosity.

(J) Optical images depicting small (D = 89.9 ± 15.4 μm) and larger (D = 834.2 ± 31.0 μm) alginate bioblocks.

(K) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting bioblock size.

(L) Predicted bioblock diameter given hypothetical changes in applied air pressure, air pressure to extrusion nozzle distance, extrusion nozzle diameter, or bath crosslinker concentration.

(M) Predicted bioblock diameter given hypothetical changes in both nozzle diameter and air pressure-to-nozzle distance.

For all univariate design plots (D, H, L), gray lines represent the mean prediction for all conditions in the synthetic 2 × N dataset, and blue diamonds with vertical error bars represent the mean predictions and bootstrapped 95% confidence intervals for the specific conditions being compared. For in-depth experimental and computational results for bioblock assembly, see also Figures S1–S18.

range (953 μm) and SD (218 μm), indicating that the model had extracted useful predictive patterns (Figures 3L and S14). Not surprisingly, there is a strongly negative relationship between diameter and air pressure. Since high pressures are linked to unstable or aspherical bioblocks (Figures 3E and 3H), the model identified other associations from the multidimensional design space that could also be leveraged for size control (Figures 3L and S17). For example, increasing both nozzle diameter (30 μm–150 μm) and pressure-to-nozzle distance (1 mm–30 mm) would yield significantly larger (>60%) alginate bioblocks compared with the opposite conditions (Figures 3M and S18).

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

CellPress
OPEN ACCESS

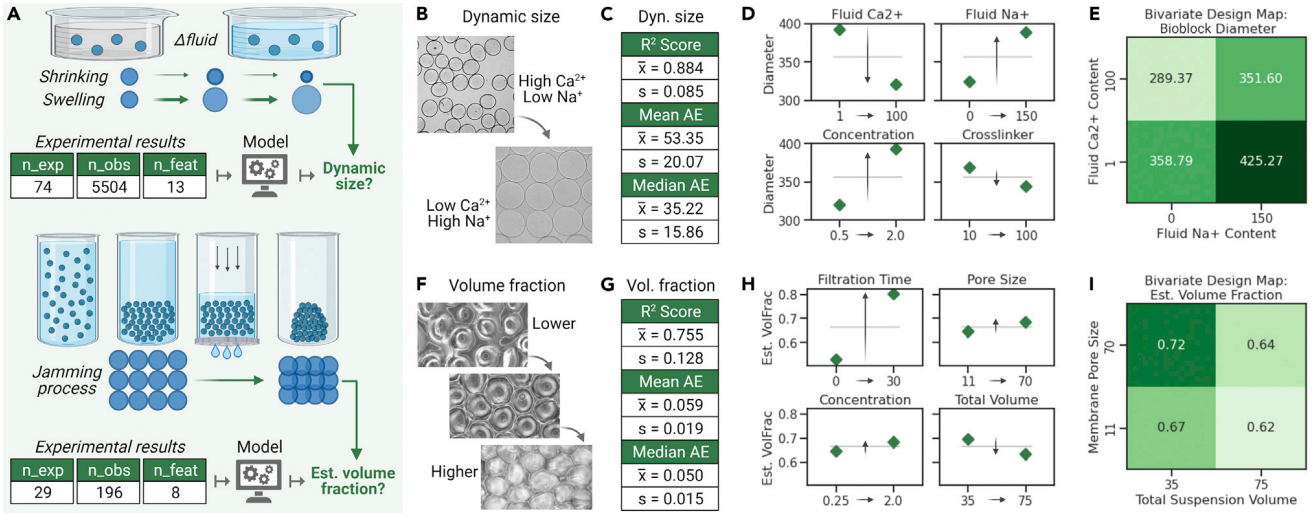


Figure 4. Controlled consolidation of hydrogel bioblocks into dense granular matrices

(A) Overview of the processing of alginate bioblocks into dense granular matrices and corresponding data-driven models. Summary boxes display the number of unique experiments, number of recorded observations, and number of input design parameters.

(B) Optical images depicting highly contracted bioblocks in high-Ca^{2+}/low-Na^{+} suspending medium and highly swollen bioblocks in a low-Ca^{2+}/high-Na^{+} suspending medium.

(C) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting bioblock size over time in different media.

(D) Predicted diameter given hypothetical changes in fluid-phase Ca^{2+} and Na^{+} concentration, solid-phase alginate concentration, or crosslinker concentration (Ca^{2+}) in the initial gelling reservoir.

(E) Predicted diameter given hypothetical changes in Ca^{2+} and Na^{+} concentrations of the fluid phase.

(F) Optical images depicting jammed bioblocks at increasing volume fractions, with gradual removal of the fluid phase and notable compaction and deformation of solid-phase bioblocks.

(G) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting the granular matrix's approximate volume fraction.

(H) Predicted volume fraction given hypothetical changes in filtration time, membrane pore size, solid-phase alginate concentration, or total suspension volume.

(I) Predicted volume fraction given hypothetical changes in membrane pore size and total suspension volume.

For all univariate design plots (D and H), gray lines represent the mean prediction for all conditions in the synthetic 2×N dataset, and green diamonds with vertical error bars represent the mean predictions and bootstrapped 95% confidence intervals for the specific conditions being compared. For in-depth experimental and computational results for granular matrix preparation, see also Figures S19–S35.

Controlled consolidation of hydrogel bioblocks into dense granular matrices

In the second phase of the study, we compacted individual alginate bioblocks into dense granular matrices that exhibit viscoplasticity.[7,8,11,67–69] First, because hydrogels can swell or de-swell based on environmental conditions,[21,22] we measured bioblock size as a function of swelling time under different media conditions (Figures 4A, 4B, and S19–S24). We built a regressor (Tables S1 and S2) to predict bioblock size as a function of 13 different inputs from their initial droplet-based assembly and subsequent incubation. The r^2 was high (0.884) and the mean (53.4 μm) and median (35.2 μm) absolute error are low given the considerable range (972 μm) and SD (225 μm) of the dataset (Figures 4C and S20). The alginate bioblocks did change over time, and their size depended on the Ca^{2+} and sodium ion (Na^{+}) concentrations in the fluid phase (Figures 4D and S23). While the degree of swelling depended strongly on the alginate concentration, there was a much smaller dependence on initial crosslinker concentration. The bivariate map highlighted the divergent effect of differing ionic profiles—bioblocks suspended in a low-Ca^{2+}, high-Na^{+} environment are predicted to have over a 5-fold greater volume compared with those in a high-Ca^{2+}, low-Na^{+} environment (Figures 4E and S24). Since highly contracted bioblocks will exhibit greater stiffness and friction than highly swollen particles, intermediate structural changes should be carefully monitored throughout the bioblock

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

design process to enable accurate forecasting of the downstream matrix properties.[15,23]

To obtain granular matrices, we needed to convert dilute suspensions of free-floating bioblocks into highly compacted (i.e., "jammed") matrices, where neighboring particles are in contact with one another[1,2] (Figures 4F and S25–S28; Tables S3 and S4). Because these soft bioblocks can deform, it is difficult to precisely measure the bioblock volume fraction within the dense granular matrices[17,70] (Figure 4F). We therefore used a practical estimate of their volume fraction to enable discrimination between different levels of compaction (Figure S29; Tables S5 and S6). Ultimately, we used gravity- or pressure-driven filtration to remove interstitial fluid phase and jam the remaining solid-phase bioblocks. Before jamming, bioblock volumes are recorded and the solid-phase fraction is approximated.[71,72] We then recorded the fluid-phase volume lost during filtration, which enabled on-the-fly estimation of their solid volume fraction. This heuristic allowed us to compact free-floating bioblocks into dense matrices and compute approximations of the matrix packing state simultaneously (Figure 4A). The time-dependent volume fraction curves are non-linear, with rapid fluid loss during the initial transition from dilute to jammed suspensions (Figure S30). We built a regressor (Tables S1 and S2) to predict the volume fraction of granular bioblock matrices as a function of eight different assembly and processing inputs. The r^2 was 0.755 and mean (0.059) and median (0.05) absolute error are around 0.05, which demonstrated that the time-dependent jamming process could be predicted reasonably well (Figures 4G and S31–S35). Filtration time had the greatest impact on volume fraction, while the other inputs served to manipulate the trajectory of the jamming process (Figures 4H, S34, and S35). Increasing pore size or increasing pressure accelerated fluid passage through the membrane and increased the predicted volume fraction. However, extrinsic parameters are not the only relevant features (Figures S34 and S35). For example, large volumes of soft bioblocks would be at lower volume fractions than smaller volumes of stiff bioblocks, all other things being equal (Figure 4I).

Rheological behavior of granular matrices composed of hydrogel bioblocks

Next, we explored the rheological properties of granular matrices composed of densely packed, hydrogel bioblocks. Specifically, we measured their shear moduli, shear yielding, and shear-thinning behavior (Figure 5A). The ability to tailor these critical properties is important for many biomedical applications.[1,2] We first measured storage (G′), loss (G″), and complex ($G^* = G' + iG''$) moduli via oscillatory amplitude sweeps (Figure 5A; full data in Figures S36–S39). Each granular matrix exhibited a linear elastic region in which G′ exceeded G″, indicative of a solid-like response.[15,23,24] However, by varying material inputs, we could generate markedly different matrices with moduli spanning several orders of magnitude (Figures 5B and S37). Beyond the linear elastic regime, each matrix exhibited a yielding transition and subsequent viscous flow[15,23,24] (Figure S38).

We built a multi-output regressor (Tables S1 and S2) to predict the complete viscoelastic profile (G′, G″, G*), and when using oscillatory stress as a predictor, we obtained r^2 of 0.944, MAE of 136.98, and median AE of 18.49 (Figures 5C and S40–S44). Given the complexity and the range of possible output values (from ~0.1 Pa to ~11 kPa), we concluded the model effectively learned the non-linear viscoplastic responses (Figure S40). We find that alginate concentration, ion concentration, and volume fraction are key modulators of matrix stiffness (Figures 5D and S43). Increasing the alginate and Ca^{2+} concentrations (or decreasing Na^+ concentration)

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

CellPress
OPEN ACCESS

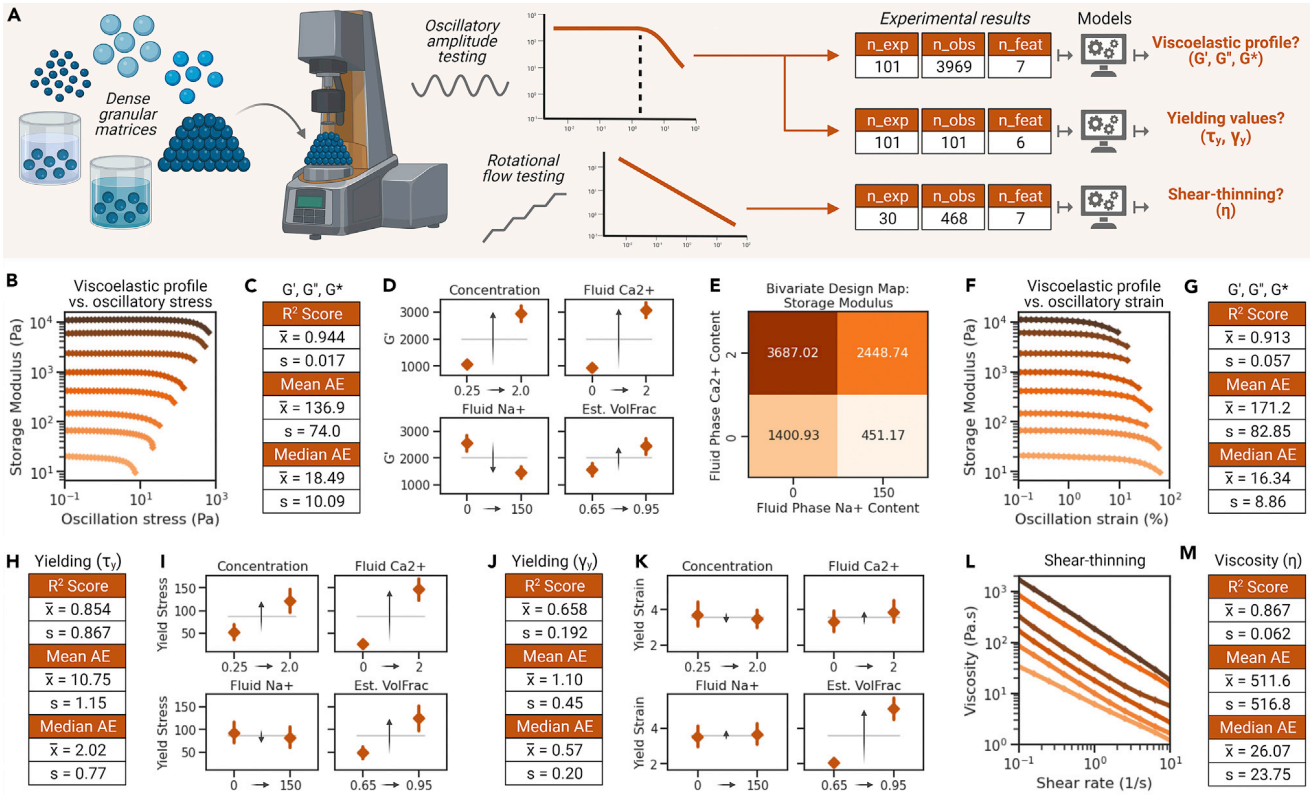


Figure 5. Rheological behavior of dense granular matrices composed of hydrogel bioblocks

(A) Overview of the granular matrix characterization approach and corresponding data-driven models. Summary boxes display the number of unique experiments, number of recorded observations, and number of input design parameters.

(B) Representative curves of storage modulus versus oscillation stress.

(C) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting storage, loss, and complex moduli versus oscillation stress.

(D) Predicted storage modulus given hypothetical changes in bioblock concentration, fluid-phase Ca^{2+} and Na^{+} concentrations, and estimated solid-phase volume fraction.

(E) Predicted storage modulus given hypothetical changes in the Ca^{2+} and Na^{+} concentrations of the fluid phase.

(F) Representative curves of storage modulus versus oscillation strain.

(G) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting storage, loss, and complex moduli versus oscillation strain.

(H) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting yield stress.

(I) Predicted yield stress given hypothetical changes in alginate concentration, fluid-phase Ca^{2+} and Na^{+} concentrations, and estimated solid-phase volume fraction.

(J) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting yield strain.

(K) Predicted yield strain given hypothetical changes in alginate concentration, fluid-phase Ca^{2+} and Na^{+} concentrations, and estimated solid-phase volume fraction.

(L) Representative curves of apparent viscosity versus rotational shear rate.

(M) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting apparent viscosity versus shear rate.

For all univariate design plots (D, I, K), gray lines represent the mean prediction for all conditions in the synthetic 2×N dataset, and orange diamonds with vertical error bars represent the mean predictions and bootstrapped 95% confidence intervals for the specific conditions being compared. For in-depth experimental and computational results for rheological characterization, see also Figures S36–S66.

produced bioblocks with higher stiffness, resulting in much stiffer granular matrices. Not surprisingly, stiff granular matrices also arise when the bioblock volume fraction is under further compaction. The combined effect of ion concentration is clear in bivariate maps, where granular matrices composed of highly swollen bioblocks (in high Na^{+}, low Ca^{2+}) are much softer than those formed in high Ca^{2+} and low Na^{+} (Figures 5E and S44). We repeated the model-building process (Tables S1 and S2) using oscillatory strain as a predictor (instead of stress) and again found high

evaluation scores with r^2 = 0.913, MAE = 171.22, and median AE = 16.34 (Figures 5F, 5G, and S45–S47). Hence, their dynamic viscoplastic profile can be predicted quite accurately given the upstream processing and structural history.

After modeling the full oscillatory response, we extracted shear yielding values to better understand the yielding behavior of dense granular matrices (Figure S36). We define the yielding point as the value at which G′ is 0.9 × G'_o (where G'_o is the plateau modulus).[73,74] The yield stress spans well over 3 orders of magnitude for different matrices, while the yield strain varies by less than 2 orders of magnitude (Figure S49). We built a regressor (Tables S1 and S2) to predict the shear yield stress (Figure 5A) and observed an r^2 of 0.854, MAE of 10.75, and median AE of 2.02, indicating accurate mapping of upstream material inputs to downstream yielding (Figures 5H and S50). We observed similar trends as the storage modulus, with greater alginate concentration, Ca^{2+} concentration, or volume fraction leading to greater predicted yield stresses (Figures 5I and S51–S53). After building a regressor for yield strain (Tables S1 and S2), we observed r^2 of 0.658, MAE of 1.10, and median AE of 0.57 (Figures 5J and S54), indicating that yield strain is more difficult to predict than yield stress, potentially due to a less deterministic input-output map and a relatively greater contribution of noise and batch effects to the recorded output values. Several univariate yield strain patterns are also different from those observed for yield stress or storage modulus (Figures 5K, 5I, and 5D). In this case, the volume fraction is the most critical parameter, while alginate concentration and fluid-phase ions had a negligible effect (Figures 5K, S55, and S56). Interestingly, bioblock size had a minimal impact on yield stress or stiffness, but there is a sizable negative relationship between bioblock size and predicted yield strain (Figures S56 and S57). These results reinforce how the relative effects of a parameter depend on the specific input-output relationship being studied. We also encoded a binary variable tracking the bioblock resuspension protocol (where 1 = single wash, 2 = multiple washes), and though it is not critical for predicting stiffness or yield stress, it is useful for predicting yield strain (Figures S56 and S57). This highlighted the importance of monitoring intermediate steps when developing complex, hierarchical materials like granular bioblock matrices. Even low-precision observations can help to explain outcome variance in a data-driven way, which in turn can suggest future avenues for troubleshooting and optimization.

Next, we characterized the shear-thinning profiles of these dense granular matrices (Figures S58–S62). Each matrix displayed a pronounced decay in apparent viscosity with increasing shear rate (Figures 5L and S37) and all extracted flow behavior indices were <0.4 (Figure S36). We developed a regressor (Tables S1 and S2) to predict the shear-thinning response (indicated by the log of the viscosity) and observed an r^2 of 0.867, MAE of 511.59, and median AE of 26.07 (Figure 5N). While the median error is quite low, the mean error is roughly 20-fold higher, indicating that the error distribution is skewed by the presence of outlier values (Figures S58 and S59). The bioblock concentration and washing protocol are critical for determining viscosity, with softer and more swollen matrices exhibiting a lower resistance to flow (Figures S60–S62). Lower apparent viscosity is also predicted for matrices composed of either larger bioblocks or lower bioblock volume fraction, suggesting that the number density of bioblock "contacts" within these matrices is a key parameter (similar to yield strain above). Last, we extrapolated the limiting stress from the flow curves to obtain estimated shear yield stress values[73,74] (Figures S63–S66). We developed a data-driven model and found we could accurately predict yield stresses from flow curves, and in general the bioblock concentration, size, and resuspension protocol emerged as key parameters.

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

Injectability and extrudability of dense granular matrices composed of hydrogel bioblocks

Finally, we explored the extrudability and injectability of dense granular matrices composed of alginate bioblocks. The ability to optimize their performance is essential for both bioprinting and injectable therapies.[7,12,75] Since performance depends on both material attributes and functional context, the precise combination of intrinsic material parameters and extrinsic (extrusion/injection) parameters may yield a complex design space with low- and high-performing regions. To elucidate this further, we attempted to inject a range of granular matrices under different extrusion conditions and categorized the outcomes as either "success" (e.g., smooth and controlled delivery of material) or "failure" (e.g., material clogging, uncontrollable delivery) (Figures 6A, 6B, and S67–S74). More than one-third of these trials resulted in failure, confirming our intuition that certain regions of the performance space are incompatible with injectability (Figure 6B). We developed a classifier (Tables S1 and S2) to predict extrusion outcomes as a function of intrinsic material parameters and extrinsic extrusion parameters (Figures S74–S79). After applying the modeling pipeline, we observed high accuracy (0.885), ROC-AUC (0.955), and F1 scores (0.881), indicating that the patterns from the multidimensional performance space were effectively learned (Figure 6C). The relative mechanics of the granular matrix phases (solid-phase bioblock stiffness, fluid-phase viscosity) emerged as key modulators of extrudability[75] (Figures 6D, S77, and S78). While densely jammed matrices are less injectable, the packing state had a smaller effect than the solid and fluid-phase mechanics. As highlighted in the 2D map, stiffer bioblocks suspended in a low-viscosity fluid tend to clog immediately (fail), while softer bioblocks suspended in higher viscosity fluids exhibit smooth and reliable injection (success) (Figures 6E and S79). As expected, extrusion parameters are also important determinants of injectability (Figure 6D). Small bioblock-to-nozzle ratios favored success, while matrix injectability diminished when the bioblock and nozzle diameters approached one another. The use of longer nozzles also markedly decreases success. Interestingly, the model did not predict much difference between manual and non-manual injection, though we perceived that the latter was more reliable during our trials (Figure 6D). By learning patterns across all dimensions simultaneously, ML can help to challenge and/or refine human analysis of multifaceted processes.

To complete this effort, we quantified the extrusion profiles of different granular matrices (Figures 6F and S80). The prior qualitative model showed that the relative solid- and fluid-phase mechanics are critical (Figure 6E). However, high-viscosity fluids pose handling challenges, so here we focused on regions of the design space that did not require viscous fluid for injection. Specifically, we assessed granular matrices with different bioblock composition, volume fraction, and fluid-phase ions, along with an array of syringe volumes, nozzle diameters, and extrusion speeds (Figures S83–S85). Successful injections exhibited smooth force-versus-displacement curves, while failed (unstable) injections had erratic jumps or gradual increases in applied force due to matrix clogging within the nozzle (Figures 6F, 6G, and S81). Such instabilities are undesirable, as they can result in defects during bioprinting or potential patient harm in injectable therapeutics.[7,12,75] Of the 108 trials performed, 20 resulted in unstable injection curves (including all formulations predicted to be "low-performance"). We devised a classifier (Tables S1 and S2) to predict extrusion stability as a function of 12 different material and extrusion parameters (Figures 6F, 6G, and S86–S90). After executing the computational pipeline, we find that our data-driven model achieved very high accuracy (0.982), ROC-AUC (0.967), and F1 score (0.981) (Figure 6H) This represents a sizable improvement compared with the previous extrusion model (Figure 6), highlighting that the incorporation of quantitative

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

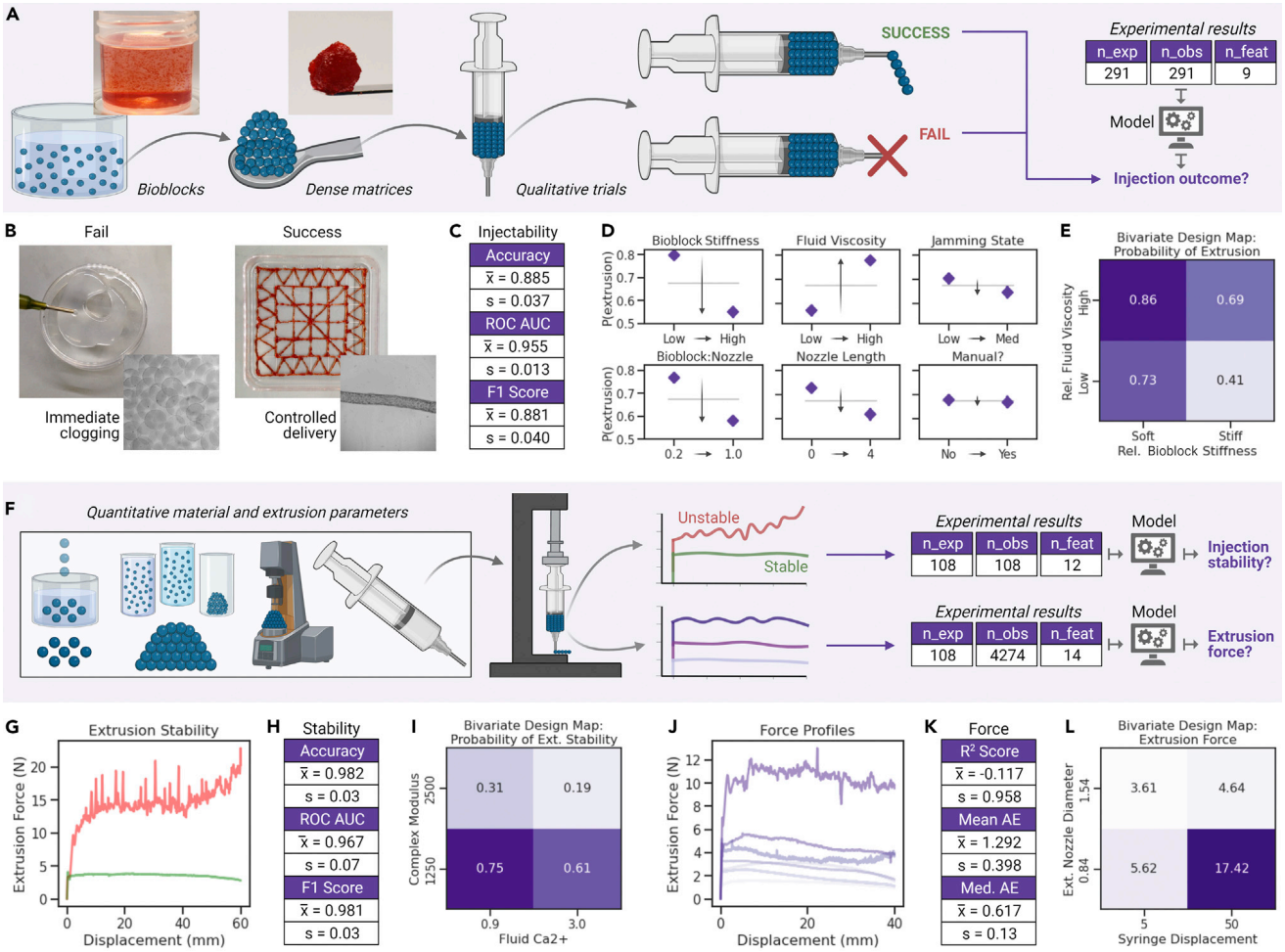


Figure 6. Injectability and extrudability of dense granular matrices composed of hydrogel bioblocks

(A) Overview of the qualitative approach for assessing granular matrix injectability and corresponding data-driven model. Summary boxes display the number of unique experiments, number of recorded observations, and number of input design parameters.

(B) Macroscopic and microscopic optical images of failed injection (immediate clogging and delivery of watery, un-jammed bioblocks) and successful injection (controlled delivery of dense granular matrices).

(C) Accuracy, ROC-AUC, and F1 scores for the data-driven classifier qualitatively predicting the injection performance of these granular matrices.

(D) Predicted probability of successful granular matrix extrusion given hypothetical changes in the relative solid-phase bioblock stiffness, relative fluid-phase viscosity, relative granular matrix jamming status, ratio of bioblock diameter-to-nozzle diameter, nozzle length, and injection method (pneumatic/mechanical or manual). Gray lines represent the mean prediction for all conditions in the synthetic 2×N dataset, and purple diamonds with vertical error bars represent the mean predictions and bootstrapped 95% confidence intervals for the specific conditions being compared.

(E) Predicted probability of successful granular matrix extrusion given hypothetical changes in the relative solid-phase bioblock stiffness and relative fluid-phase viscosity.

(F) Overview of the quantitative approach for granular matrix injectability assessment and corresponding data-driven models. Summary boxes display the number of unique experiments, the number of recorded observations, and the number of design features used to predict the outcome.

(G) Representative plots of extrusion stability (in green) and instability (in red).

(H) Accuracy, ROC-AUC, and F1 scores for the data-driven classifier predicting granular bioblock matrix injection stability.

(I) Predicted probability of stable extrusion given hypothetical changes in the granular bioblock matrix complex modulus and fluid-phase Ca^{2+} concentration.

(J) Representative plots of differing extrusion force curves.

(K) r^2, mean absolute error, and median absolute error for the data-driven regressor predicting the complete granular bioblock matrix extrusion force profile.

(L) Predicted extrusion force given hypothetical changes in the extrusion nozzle diameter and the displacement of the syringe plunger.

For in-depth experimental and computational results for extrudability assessment, see also Figures S67–S95.

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

process, structure, and property data enables better estimates of performance in disparate functional contexts. We observed that granular matrix injection stability depends on multiple inputs simultaneously (Figures S88 and S89). From a material standpoint, a greater complex modulus magnitude led to a sizable decrease in the probability of stable extrusion. A similar trend is observed with yield stress, highlighting the strong link between rheology and performance. However, rheology alone does not always determine injectability. For example, stiff bioblocks at low volume fraction displayed a storage modulus very similar to soft bioblocks at high volume fraction; however, the former matrix is non-injectable while the latter can be injected reliably (Figure S82). This finding is reflected in the design map, where matrices composed of contracted bioblocks (in higher Ca^{2+}) had a lower probability of stability than those composed of swollen bioblocks (in lower Ca^{2+}) even if the storage moduli are nearly identical (Figures 6I and S90). Hence, the entire processing and structural histories are required to accurately predict granular matrix injectability (Figure S89). Predictably, smaller extrusion diameters are strongly associated with lower extrusion stability. We also found that higher rates of injection favored stability, while larger syringe barrel diameters favored instability (Figures S89 and S90). Finally, we surprisingly observed that nozzle length did not have much of a predicted effect on stability.

After classifying injection stability, we built a regressor to predict the complete time-varying force profile for each extrusion (Figures 6F, 6J, and S91–S95). The mean injection force was 4.2 ± 4.5 N but the maximum force was 76.1 N (due to material clogging events) (Figure S91). Although the unstable trials contributed significant variability, most granular matrices could be extruded with limited application of force. This ease of delivery is conducive to safe manual injection in the clinic and controlled pneumatic or mechanical extrusion during bioprinting.[76] We built a data-driven model (Tables S1 and S2) and interestingly found that the r^2 was -0.117, but the mean and median absolute error were 1.292 and 0.617 (Figure 6K). This highlights the importance of a multi-metric evaluation approach (Figure 2A). The r^2 and MAE were skewed by poor performance on outlier test samples, but the low median AE indicates that many force profiles could actually be predicted quite accurately. Unsurprisingly, increasing the magnitude of the material's complex modulus substantially increased the predicted force, whereas the yield stress had a negligible effect (Figures S93 and S94). The upstream bioblock processing (e.g., manipulating fluid-phase ions) also played a role, with granular matrices composed of contracted bioblocks (in higher Ca^{2+}) having greater predicted injection forces than granular matrices composed of swollen bioblocks (in lower Ca^{2+}) (Figures S94 and S95). Increased nozzle diameter decreased the force while increased plunger displacement increased the force, and there was a strong interaction between the variables (Figure 6L). For large-diameter nozzles, similar forces were required at both the start (5-mm displacement) and end (50-mm displacement) of injections. However, for smaller diameters, the predicted force was much greater at the end of an injection, suggesting gradual nozzle clogging or phase separation of the matrix, which drastically increased the force required to maintain extrusion. Other parameters were not as critical, but they still interacted to modulate extrusion force (Figures S93–S95).

DISCUSSION

Granular matrices composed of hydrogel building blocks are promising, yet complex materials due to the multidimensional parameter spaces and multi-scale phenomena associated with each step in their development pipeline. Through a

combination of data-driven modeling and experiment, we explored the scalable formation of hydrogel bioblocks, their compaction into dense granular matrices, their emergent non-linear rheology, and their delivery through fine nozzles. Our bottom-up empirical progression was supported by a flexible pipeline that enables automated generation of predictive input-output frameworks for a wide range of datatypes and ML tasks. This data-driven approach allows one to assess the predictability of each design stage, identify critical parameters, perform predictive analytics or combinatorial optimization, and extract high-level data-driven insights. We used these data-driven material maps to optimize the design and assembly of both bioblocks and granular matrices to obtain predictable rheological behavior and extrudability. With refinement, such insights could support the rapid design and fabrication of granular matrices with controlled structure, properties, and performance for applications ranging from drug delivery to tissue engineering.[77–80]

We showed that data-driven approaches could be effectively applied to a suite of complex problems across the soft granular matter life cycle that remain difficult to simulate or analyze.[19,25,31,33] Key limitations of data-driven modeling are the reliance on data/problem representation and the inability to differentiate between correlation and causation.[27,28,81] We find that the data-driven models consistently uncovered reliable design insights, many of which could be linked to patterns from other empirical reports focused on a single phase of the material pipeline.[82–89] Future efforts could use hybrid frameworks with physical equations to more explicitly integrate mechanistic understanding into the modeling process.[37,90–92] Eventually, it may be feasible to directly build inverted models or build fully connected forward models that can be combined with powerful algorithms for optimization.[27,29] Further, closed-loop active learning could be used to interactively improve predictive capacity and efficiently sample within each experimental space, while transfer learning could be applied to transfer knowledge from related domains to small empirical datasets.[32,93] Finally, efforts to map the supervised ML space, develop problem-specific recommendations, and automate modeling could facilitate more rigorous adoption of ML by the materials science community.[33,94]

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Connor Verheyen (connorv@mit.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All original data (baseline data, processed data, model prediction matrices) and original code (custom machine learning pipelines and complete analytical workbooks for each data-driven model) have been deposited in a public GitHub repository: https://github.com/connor-verheyen/DataDrivenGranularHydrogels and permanently archived in a public Zenodo repository: https://doi.org/10.5281/zenodo.7506819.

Hydrogel bioblocks

Alginate precursor stock solutions were prepared by mixing dry sodium alginate (Sigma-Aldrich A2033-MV or A1112-LV) with deionized water in a FlackTek SpeedMixer to obtain 0.25%, 0.50%, 1.00%, or 2.00% solutions by weight. After complete dissolution of the alginate, all precursor solutions were passed through syringe filters

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

and stored at 4°C. Calcium chloride stock solutions were prepared by dissolving calcium chloride dihydrate (Sigma-Aldrich) in deionized water to obtain 500-mM solutions. To produce a batch of bioblocks, alginate precursor solutions were first loaded into 25-mL glass syringes (Hamilton 1000 series). The alginate syringe was loaded into a syringe pump (Harvard Apparatus PHD ULTRA), which was placed on a surface with adjustable height to control the drop distance (30–280 mm). An extrusion nozzle was then attached to the alginate syringe (EFD Nordson, 30–150 μm diameter, 0–45-degree angle). A gelation bath was prepared by diluting the $CaCl_2$ stock solution to 100 or 10 mM, with or without the addition of 5% ethanol to modify bath surface tension. The gelation batch was placed underneath the alginate syringe nozzle to collect the falling microdroplets. A ring stand and clamp was used to position an air pressure cartridge and nozzle above the alginate syringe nozzle (0.8–30 mm above, at a 70–90° angle), and the air pressure cartridge was attached to a pressure box (Nordson EFD Ultimus V) to maintain a desired air pressure (1–40 PSI). Once the setup was completed, an assembly run was started by first initiating the air pressure and then initiating an infusion on the syringe pump at a programmed volumetric flow rate (25–200 μL/min). Nanoliter to microliter-scale droplets were extruded from the alginate syringe into the airstream, which induced the droplets to detach and fall into the downstream gelation bath. Alginate bioblocks formed as the precursor droplets were crosslinked by calcium ions in the bath. Bioblocks were incubated in the gelation bath for at least 1 h to allow for sufficient crosslinking before any further characterization or manipulation. After crosslinking, bioblock batches were removed from the gelation bath, transferred into storage containers, and resuspended with 2 to 3 mM $CaCl_2$. Upon completion of the empirical exploratory phase, the various qualitative and quantitative attributes from each production trial were retrospectively aggregated into a machine-readable spreadsheet for data-driven modeling.

Representative samples of crosslinked bioblocks from each production batch were transferred to Petri dishes and a phase contrast microscope (Leica) was used to optically image the swollen bioblocks (×4 to ×10 objective). At least four optical images were taken per batch, each in different regions of the dish to effectively survey the bioblock population (on average, images contained ~10–30 unique particles). The optical images were then analyzed using ImageJ software. Bioblock formation was recorded as a binary categorical variable, where stable and well-defined 3D particles were marked as a successful formation event and unstable and amorphous particles were marked as a failed formation event. Bioblock shape was recorded as a multi-class categorical variable, where the stable and well-defined particles selected in the previous step were further sub-classified as sphere, oblong, or tail. For both formation and shape, the ImageJ multi-point tool was used to mark the bioblocks in the field of view, and then the "measure" function was used to tally the total number of marks. For bioblock size, the pixel-to-micron conversion for the given objective was used to map the image scale to the actual scale. Then the ImageJ line and ellipse selection functions were used to demarcate the bioblocks in the field of view, and the "measure" function was used to extract the 2D cross-sectional dimensions of the alginate bioblocks. The bioblock formation, shape, and initial size outcomes were paired with the corresponding assembly parameters to enable data-driven mapping of upstream processing inputs to downstream structural outputs.

Dense granular matrices

Batches of crosslinked alginate bioblocks were filtered to remove outlier particles and then selected bioblock populations were washed and resuspended in aqueous

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

media with different ion concentration. The bioblocks were imaged over time in order to monitor the dynamic response of a bioblock to the surrounding fluid environment. As before, representative samples of the bioblocks were transferred to Petri dishes and optically imaged via phase contrast microscope (Leica). Next, the ImageJ size analysis workflow was repeated to obtain updated measurements of the bioblock dimensions (tracking the bioblock swelling or shrinking over time). The long-term bioblock size stability data and corresponding production and processing attributes were then aggregated into a machine-readable spreadsheet for data-driven modeling. Once the bioblocks had been resuspended in their desired fluid phase, dilute bioblock suspensions were allowed to settle in graduated cylinders for 1 h and the volume of settled bioblocks and the total volume in the cylinder were recorded. The suspension was then transferred to a custom jamming tube with an integrated semi-permeable membrane to enable passage of the continuous fluid phase while restricting passage of the solid-phase bioblocks. The volume of the fluid phase lost through the membrane was recorded at select timepoints throughout the jamming process. A pressure line could be attached to the cartridge to apply pressure, or gravity alone could be used to drive filtration. Over time, the gradual consolidation of the solid phase behind the membrane produced a dense suspension of jammed bioblocks—a granular bioblock matrix. The solid-phase volume fraction of the granular bioblock matrix was approximated by first assuming a settled volume fraction (~0.60) and then updating the estimate on-the-fly as the fluid phase was lost. The jamming process was halted once a granular bioblock matrix with the desired estimated volume fraction was obtained, and then the granular bioblock matrix was transferred to a container for longer-term storage in the given jammed state. For the preliminary qualitative injectability testing, capillary wicking was instead used to jam the bioblocks into granular bioblock matrices. In this case, the supernatant of settled bioblock suspensions was first aspirated, and then a wicking material (e.g., paper towel) was applied to the surface of the suspension to further withdraw the interstitial fluid via capillary action. The wicking process could be halted once a given qualitative jamming state had been reached. The volume fraction estimates were paired with the various qualitative and quantitative attributes from the jamming process and compiled into a machine-readable spreadsheet to enable data-driven mapping of upstream processing inputs to downstream structural outputs.

Rheological characterization

All rheological measurements were performed using a Haake RheoStress 6000 rheometer with 25-mm-diameter parallel plate geometry. To prevent wall slip, low-grit sand paper was glued to each parallel plate. Granular bioblock matrices were transferred onto the parallel plates via spatula, and material volume was trimmed or augmented to ensure an adequate testing fit. Hydrated paper towels were placed circumferentially around the testing site to maintain local humidity to reduce evaporation of the granular bioblock matrix. The tests performed on a given material could include oscillatory amplitude sweeps (stress, strain, frequency) as well as rotational flow sweeps (forward, reverse), and tests were performed between 22 and 25°C. Oscillatory stress and strain sweeps were conducted at a fixed frequency of 1 Hz, while oscillatory frequency sweeps were performed at a fixed strain of 1%. The plateau moduli (storage, loss, and complex modulus) were obtained by performing a linear regression on the data-points in the linear viscoelastic regime of the oscillatory curves and extracting the intercepts. The oscillatory yield points (yield stress, yield strain) were obtained by extracting the amplitudes (stress, strain) at the points on the oscillatory curves where the storage moduli had decreased by at least 30%, indicating substantial softening and a clear departure from linear viscoelasticity. The rotational yield points were obtained by linearly

Please cite this article in press as: Verheyen et al., Integrated data-driven modeling and experimental optimization of granular hydrogel matrices, Matter (2023), https://doi.org/10.1016/j.matt.2023.01.011

extrapolating the stress versus shear-rate curves and extracting the intercepts. The production, processing, and structural information for each granular bioblock matrix formulation was paired with the corresponding rheological outcomes in a machine-readable spreadsheet to enable data-driven mapping of upstream processing and structural inputs to downstream property outputs.

Qualitative assessment of granular matrix injectability

Dense granular matrices were manually loaded into standard syringes (BD Luer-Lok, 3–10 mL volume) or dispensing cartridges (Nordson EFD Optimum, 3–10 mL volume), using a spatula to transfer the material while periodically tapping the syringe to pack the material and decrease the presence of air gaps. After the material was transferred, an injection attempt was performed via manual, pneumatic, or mechanical force. For the manual approach, the syringe plungers were manually advanced in the syringe barrels in order to deliver the granular matrix. For the pneumatic approach, the cartridges were attached to pressure lines and a programmed pressure was applied to deliver the granular matrix. For the mechanical approach, the syringes were loaded into syringe pumps and a motor was used to advance the syringe plunger to deliver the granular matrix. The extrusion attempts were performed with a wide variety of potential syringe attachments (with or without a nozzle, tapered or straight, using a range of nozzle diameters, etc.) (Nordson EFD General Purpose and SmoothFlow, 0.51 mm–1.54 mm inner diameter). The extrusion outcomes were recorded in a categorical fashion: trials with stable, reliable, and controllable delivery of material through the nozzle would be marked "success," while trials with immediate or delayed clogging, uncontrollable extrusion, or other instabilities would be marked "failure." The qualitative and quantitative granular bioblock matrix attributes were paired with categorical extrusion outcomes in a machine-readable spreadsheet to enable data-driven mapping of upstream processing and structural inputs to downstream performance outputs.

Quantitative assessment of granular matrix injectability

Dense granular matrices were manually loaded into standard syringes (BD Luer-Lok, 5 or 10 mL volume). After attaching the syringe barrel, the loaded syringes were placed into a customized testing rig within the testing workspace of an Instron universal testing system. The syringes were oriented vertically (with the nozzle facing down) and the Instron test arm was lowered until it contacted the plunger face. The Instron arm was then used to advance the syringe plunger at a pre-programmed (constant) displacement rate, thereby delivering the granular bioblock matrix through the nozzle. As the extrusion progressed, the machine recorded the force required to maintain the programmed displacement rate of the plunger. Before testing the granular bioblock matrices, control extrusions were performed with air and water to define the force profile and variability of the syringe itself. After completion, the force curves were manually inspected to identify injection instability. Smooth and steady injection curves were marked "stable," while curves with erratic changes in force or increasing force over time were marked "unstable." Given the high frequency of data recording, the force curves were down-sampled prior to modeling to reduce training times. The quantitative granular bioblock matrix attributes were paired with the quantitative and qualitative extrusion outcomes in a machine-readable spreadsheet to enable data-driven mapping of upstream process, structure, and property inputs to downstream performance outputs.

Data-driven modeling workflow

The machine-readable spreadsheets compiled after each empirical data collection phase were read in and processed. For each unique dataset, basic label encoding,

data transformation, and feature engineering steps were performed to obtain a curated set of inputs and outputs for the corresponding supervised ML problem (files were stored in xlsx or csv format). RF and GB were both used for binary classification, multi-class classification, cross-sectional regression, and panel regression tasks (although only the RF supported multi-output regression). A subset of hyperparameters for each algorithm was selected and the exhaustive combination of hyperparameter configurations was assessed via automated grid search. A nested and grouped k-fold CV scheme was used to evaluate the generalizability of the automated model configuration and selection protocol in an unbiased way. The outer folds split data into train and test sets, and the inner folds further split the outer train set into inner train and test sets. Model configurations were fit on the inner train folds, the best-performing model configurations were selected based on their inner test fold scores, and then the best-performers were re-fit on the outer train fold and evaluated on unseen data in the outer test fold. Each experimental trial was assigned a unique experimental ID and the grouped data-splitting protocol ensured that data from the same trial could not be present in both train and test sets at the same time to account for batch effects, repeated measures, and other sources of systematic error that would break the independence assumption. The predictive capacity of the fitted models was scored using multiple evaluation metrics (accuracy, ROC-AUC), and F1 score for classification problems; r^2, MAE, and median absolute error for regression problems). The mean, SD, and median values across all outer (holdout) folds were computed and recorded. Then, a standard k-fold CV was applied in order to execute the automated model configuration and selection protocol that was validated in the previous step. The optimal configurations of the RF and GB algorithm were then fit to the entire dataset in order to obtain the final trained model. Then, a customized workflow was applied to extract human-readable design insights of the complex n-dimensional structure learned by the data-driven model. A simplified and subsampled synthetic dataset was generated by first identifying representative upper and lower levels for each input and then computing the Cartesian product to artificially obtain full coverage of potential input combinations. This simplified n-dimensional matrix was then fed into the trained models to generate a new set of predictions based on the patterns the model had learned from the empirical data. Then, 1D and 2D feature-target plots were generated to capture the general univariate and bivariate associations between inputs and outputs in the data. Further, a permutation feature importance protocol was applied to estimate the importance of each design input to the trained model's overall predictive capacity. All modeling work was conducted in Python and all code and results are stored in IPython notebooks in the open-access Google Colaboratory virtual environment. Libraries used for data preparation, modeling, analysis, and visualization include scikit-learn, pandas, numpy, scipy, matplotlib, and seaborn.

Note, we leveraged supervised ML to build data-driven models directly from our empirical material datasets. The use of tree-based ensembles enabled good performance across a range of supervised learning problems (binary, multi-class classification; cross-sectional, panel, multi-output regression) and datasets (diverse datatypes and distributions), but future implementations could explore the use of specific algorithms (e.g., simpler linear methods or more complex neural networks) based on problem attributes. The grouped and nested CV protocol was designed to rigorously perform model tuning during model evaluation, but certain regimes of the dataset may still be susceptible to overfitting, underfitting, and optimistic evaluation. The multi-metric performance estimates provided useful feedback about the heterogeneous predictability of any given experimental process. The simplified 1D and 2D maps allowed for facile interpretation of the design space, with the caveat

that the lower resolution may obscure more complex trends. Future implementations could also explore dimensionality reduction techniques for visualization of the learned data structure. Some limitations and assumptions of our overall approach include the following: data size and sampling, data independence, data preparation, algorithm selection, metric selection, overfitting, and correlation versus causation. First, the datasets were relatively small (hundreds to thousands of points) and tended to display uneven sampling of the entire available input-output space. Such datasets can lead to overfitting as well as unreliable predictions in the sparser regions of the input and output distributions. Second, ML algorithms assume that all data are independent, which is not the case for datapoints derived from the same experiment. We used unique experimental identifiers to guard against this issue, but similarities across trials could still produce optimistic assessments. Third, the experimental results were recorded in a range of formats and required both technical expertise and experimental familiarity to properly encode into machine-readable representations. Alternative data preparation choices by a human user could produce differing representations, which could lead to differing data-driven models and insights. Fourth, the choice of base algorithms can also impact modeling success. Here we chose to work with tree-based ensemble models, but alternative options ranging from much simpler linear methods to much more flexible deep neural networks may have been appropriate. Fifth, in a similar vein, the user's choice of an evaluation metric can lead to exacerbation or obfuscation of certain types of errors. For example, accuracy works well for balanced data while F1 score may be more appropriate for imbalanced data, and MAE will more harshly penalize outliers than median AE. Sixth, overfitting is always a challenge for any ML problem. While we used several methods to reduce the propensity to overfit, the flexibility of the models and the size and distribution of the datasets could still lead to overfitting. Finally, ML algorithms identify correlations in the data but they cannot perform causal reasoning. As such, one must be mindful when interpreting the data-driven patterns extracted from trained models. Here, technical expertise and experience are useful for critically appraising the model-derived insights.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.matt.2023.01.011.

ACKNOWLEDGMENTS

This work was supported by the Vannevar Bush Faculty Fellowship Program and sponsored by the Basic Research Office for the Assistant Secretary of Defense for Research and Engineering through the Office of Naval Research Grant N00014-21-1-2958 (J.A.L., S.G.M.U.) and the National Science Foundation (NSF) through the Harvard MRSEC DMR-2011754 (J.A.L., A.K.) and EFRI C3 SoRo 1935291 (E.T.R.). C.A.V. was supported by the NSF Graduate Research Fellowship, the NSF EFRI C3 SoRo 1935291, and a MathWorks Seed Fund grant. We thank Mark Skylar-Scott, Caterina Baffa, and Prof. Faez Ahmed for their experimental assistance and fruitful discussions of this work.

AUTHOR CONTRIBUTIONS

Conceptualization, C.A.V.; Methodology, C.A.V., S.G.M.U., and A.K.; Investigation, C.A.V., S.G.M.U., and A.K.; Data Curation, C.A.V.; Software, C.A.V.; Formal Analysis, C.A.V.; Writing - Original Draft, C.A.V.; Writing - Review & Editing, C.A.V., S.G.M.U., E.T.R., and J.A.L.; Supervision, E.T.R. and J.A.L.; Funding Acquisition, E.T.R. and J.A.L.

DECLARATION OF INTERESTS

The authors have a patent pending related to applications with the material. E.T.R. is scientific advisor for Pumpinheart, consults for and has equity in Holistick Medical, and is on the board of Helios Cardio and Affluent Medical. J.A.L. is a scientific advisor for Autodesk, Azul 3D, Desktop Health (a subsidiary of Desktop Metal), Mooji Meats, and Trestle Biotherapeutics and is a co-founder of AcousticaBio and Electroninks.

Received: August 5, 2022
Revised: October 21, 2022
Accepted: January 10, 2023
Published: January 31, 2023

REFERENCES

1. Riley, L., Schirmer, L., and Segura, T. (2019). Granular Hydrogels: Emergent Properties of Jammed Hydrogel Microparticles and Their Applications in Tissue Repair and Regeneration (Elsevier Ltd). https://doi.org/10.1016/j.copbio.2018.11.001.

2. Daly, A.C., Riley, L., Segura, T., and Burdick, J.A. (2020). Hydrogel microparticles for biomedical applications. Nat. Rev. Mater. *5*, 20–43. https://doi.org/10.1038/s41578-019-0148-6.

3. Newsom, J.P., Payne, K.A., and Krebs, M.D. (2019). Microgels: modular, tunable constructs for tissue regeneration. Acta Biomater. *88*, 32–41. https://doi.org/10.1016/j.actbio.2019.02.011.

4. McClements, D.J. (2017). Designing biopolymer microgels to encapsulate, protect and deliver bioactive components: physicochemical aspects. Adv. Colloid Interface Sci. *240*, 31–59. https://doi.org/10.1016/j.cis.2016.12.005.

5. Truong, N.F., Kurt, E., Tahmizyan, N., Lesher-Pérez, S.C., Chen, M., Darling, N.J., Xi, W., and Segura, T. (2019). Microporous annealed particle hydrogel stiffness, void space size, and adhesion properties impact cell proliferation, cell spreading, and gene transfer. Acta Biomater. *94*, 160–172. https://doi.org/10.1016/j.actbio.2019.02.054.

6. Shin, M., Song, K.H., Burrell, J.C., Cullen, D.K., and Burdick, J.A. (2019). Injectable and conductive granular hydrogels for 3D printing and electroactive tissue support. Adv. Sci. *6*, 1901229. https://doi.org/10.1002/advs.201901229.

7. Highley, C.B., Song, K.H., Daly, A.C., and Burdick, J.A. (2019). Jammed microgel inks for 3D printing applications. Adv. Sci. *6*, 1801076. https://doi.org/10.1002/advs.201801076.

8. Highley, C.B., Rodell, C.B., and Burdick, J.A. (2015). Direct 3D printing of shear-thinning hydrogels into self-healing hydrogels. Adv. Mater. *27*, 5075–5079. https://doi.org/10.1002/adma.201501234.

9. Xin, S., Chimene, D., Garza, J.E., Gaharwar, A.K., and Alge, D.L. (2019). Clickable PEG hydrogel microspheres as building blocks for 3D bioprinting. Biomater. Sci. *7*, 1179–1187. https://doi.org/10.1039/C8BM01286E.

10. Nih, L.R., Sideris, E., Carmichael, S.T., and Segura, T. (2017). Injection of microporous annealing particle (MAP) hydrogels in the stroke cavity reduces gliosis and inflammation and promotes NPC migration to the lesion. Adv. Mater. *29*, 1606471. https://doi.org/10.1002/adma.201606471.

11. Mealy, J.E., Chung, J.J., Jeong, H.H., Issadore, D., Lee, D., Atluri, P., and Burdick, J.A. (2018). Injectable granular hydrogels with multifunctional properties for biomedical applications. Adv. Mater. *30*, e1705912. https://doi.org/10.1002/adma.201705912.

12. Griffin, D.R., Weaver, W.M., Scumpia, P.O., di Carlo, D., and Segura, T. (2015). Accelerated wound healing by injectable microporous gel scaffolds assembled from annealed building blocks. Nat. Mater. *14*, 737–744. https://doi.org/10.1038/nmat4294.

13. Béduer, A., Bonini, F., Verheyen, C.A., Genta, M., Martins, M., Brefie-Guth, J., Tratwal, J., Filippova, A., Burch, P., Naveiras, O., and Braschler, T. (2021). An injectable meta-biomaterial: from design and simulation to in vivo shaping and tissue induction. Adv. Mater. *33*, 2102350. https://doi.org/10.1002/ADMA.202102350.

14. Cloitre, M., Borrega, R., Monti, F., and Leibler, L. (2003). Structure and flow of polyelectrolyte microgels: from suspensions to glasses. C. R. Phys. *4*, 221–230.

15. Bonnecaze, R.T., and Cloitre, M. (2010). Micromechanics of soft particle glasses. Adv. Polym. Sci. *236*, 117–161. https://doi.org/10.1007/12_2010_90.

16. Schiller, U.D., Krüger, T., and Henrich, O. (2017). Mesoscopic modelling and simulation of soft matter. Soft Matter *14*, 9–26. https://doi.org/10.1039/C7SM01711A.

17. Pellet, C., and Cloitre, M. (2016). The glass and jamming transitions of soft polyelectrolyte microgel suspensions. Soft Matter *12*, 3710–3720. https://doi.org/10.1039/c5sm03001c.

18. Muir, V.G., Qazi, T.H., Shan, J., Groll, J., and Burdick, J.A. (2021). Influence of microgel fabrication technique on granular hydrogel properties. ACS Biomater. Sci. Eng. *7*, 4269–4281. https://doi.org/10.1021/ACSBIOMATERIALS.0C01612.

19. Coussot, P. (2005). Rheometry of Pastes, Suspensions, and Granular Materials (John Wiley & Sons, Inc.). https://doi.org/10.1002/0471720577.

20. Shewan, H. (2015). Rheology of Soft Particle Suspensions. https://doi.org/10.14264/UQL.2015.533.

21. Alzanbaki, H., Moretti, M., and Hauser, C.A.E. (2021). Engineered microgels—their manufacturing and biomedical applications. Micromachines *12*, 45. https://doi.org/10.3390/MI12010045.

22. Scheffold, F. (2020). Pathways and challenges towards a complete characterization of microgels. Nat. Commun. *11*, 4315. https://doi.org/10.1038/s41467-020-17774-5.

23. Stokes, J.R., and Frith, W.J. (2008). Rheology of gelling and yielding soft matter systems. Soft Matter *4*, 1133–1140. https://doi.org/10.1039/B719677F.

24. Villone, M.M., and Maffettone, P.L. (2019). Dynamics, rheology, and applications of elastic deformable particle suspensions: a review. Rheol. Acta *58*, 109–130. https://doi.org/10.1007/S00397-019-01134-2.

25. van der Gucht, J. (2018). Grand challenges in soft matter physics. Front. Phys. *6*, 87. https://doi.org/10.3389/FPHY.2018.00087/BIBTEX.

26. de Pablo, J.J., Jackson, N.E., Webb, M.A., Chen, L.Q., Moore, J.E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D.G., Toberer, E.S., et al. (2019). New frontiers for the materials genome initiative. npj Comput. Mater. *5*. 41–23. https://doi.org/10.1038/s41524-019-0173-4.

27. Liu, Y., Zhao, T., Ju, W., Shi, S., Shi, S., and Shi, S. (2017). Materials discovery and design using machine learning. J. Materiomics *3*, 159–177. https://doi.org/10.1016/J.JMAT.2017.08.002.

28. Himanen, L., Geurts, A., Foster, A.S., and Rinke, P. (2019). Data-driven materials science: status, challenges, and perspectives. Adv. Sci. *6*, 1900808. https://doi.org/10.1002/ADVS.201900808.

29. Agrawal, A., and Choudhary, A. (2016). Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science. Apl. Mater. *4*, 053208. https://doi.org/10.1063/1.4946894.

30. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555. https://doi.org/10.1038/s41586-018-0337-2.

31. Peerless, J.S., Milliken, N.J.B., Oweida, T.J., Manning, M.D., and Yingling, Y.G. (2019). Soft matter informatics: current progress and challenges. Adv. Theory Simul. *2*, 1800129. https://doi.org/10.1002/ADTS.201800129.

32. Zhai, C., Li, T., Shi, H., and Yeo, J. (2020). Discovery and design of soft polymeric bio-inspired materials with multiscale simulations and artificial intelligence. J. Mater. Chem. B *8*, 6562–6587. https://doi.org/10.1039/D0TB00896F.

33. Li, J., Lim, K., Yang, H., Ren, Z., Raghavan, S., Chen, P.Y., Buonassisi, T., and Wang, X. (2020). AI applications through the whole life cycle of material discovery. Matter *3*, 393–432. https://doi.org/10.1016/J.MATT.2020.06.011.

34. Olson, G.B. (1997). Computational design of hierarchically structured materials. Science *277*, 1237–1242. https://doi.org/10.1126/SCIENCE.277.5330.1237.

35. Arróyave, R., and McDowell, D.L. (2019). Systems Approaches to Materials Design: Past, Present, and Future. Annu. Rev. Mater. Res. *49*, 103–126. https://doi.org/10.1146/annurev-matsci-070218-125955.

36. Raccuglia, P., Elbert, K.C., Adler, P.D.F., Falk, C., Wenny, M.B., Mollo, A., Zeller, M., Friedler, S.A., Schrier, J., and Norquist, A.J. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature *533*, 73–76. https://doi.org/10.1038/nature17439.

37. Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. Nat. Rev. Phys. *3*, 422–440. https://doi.org/10.1038/s42254-021-00314-5.

38. Radjai, F., Roux, J.-N., and Daouadji, A. (2017). Modeling granular materials: century-long Research across scales. J. Eng. Mech. *143*, 04017002. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001196.

39. Wickham, H. (2014). Tidy data. J. Stat. Softw. *59*, 1–23. https://doi.org/10.18637/JSS.V059.I10.

40. Song, Y.Y., and Lu, Y. (2015). Decision tree methods: applications for classification and prediction. Shanghai Arch. Psychiatry *27*, 130–135. https://doi.org/10.11919/J.ISSN.1002-0829.215044.

41. Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

42. Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. Ann. Statist. *29*, 1189–1232. https://doi.org/10.1214/AOS/1013203451.

43. Polikar, R., Zhang, C., and Ma, Y. (2012). Ensemble learning. Ensemble Machine Learning, 1–34. https://doi.org/10.1007/978-1-4419-9326-7_1.

44. Sagi, O., and Rokach, L. (2018). Ensemble learning: a survey. Wiley Interdiscip Rev Data Min Knowl Discov *8*, e1249. https://doi.org/10.1002/WIDM.1249.

45. Dietterich, T.G. (2000). Ensemble methods in machine learning. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/3-540-45014-9_1.

46. Probst, P., Wright, M.N., and Boulesteix, A.L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov *9*, e1301. https://doi.org/10.1002/WIDM.1301.

47. Natekin, A., and Knoll, A. (2013). Gradient boosting machines, a tutorial. Front. Neurorobot. *7*, 21. https://doi.org/10.3389/FNBOT.2013.00021/BIBTEX.

48. Claesen, M., and de Moor, B. (2015). Hyperparameter search in machine learning. Preprint at ArXiv. https://doi.org/10.48550/arxiv.1502.02127.

49. Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography *40*, 913–929. https://doi.org/10.1111/ECOG.02881.

50. Soneson, C., Gerster, S., and Delorenzi, M. (2014). Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. PLoS One *9*, e100335. https://doi.org/10.1371/JOURNAL.PONE.0100335.

51. Hsieh, K., Phanishayee, A., Mutlu, O., and Gibbons, P.B. (2020). The non-IID data quagmire of decentralized machine learning. In Proceedings of the 37th International Conference on Machine Learning. https://doi.org/10.5281/zenodo.3676081.

52. Vabalas, A., Gowen, E., Poliakoff, E., and Casson, A.J. (2019). Machine learning algorithm validation with a limited sample size. PLoS One *14*, e0224365. https://doi.org/10.1371/JOURNAL.PONE.0224365.

53. Cawley, G.C., and Talbot, N.L.C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. *11*, 2079–2107.

54. Caruana, R., and Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. In Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04. https://doi.org/10.1145/1014052.

55. Spuler, M., Sarasola-Sanz, A., Birbaumer, N., Rosenstiel, W., and Ramos-Murguialday, A. (2015). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. *2015*, 1083–1086. https://doi.org/10.1109/EMBC.2015.7318553.

56. Vishwakarma, G., Sonpal, A., and Hachmann, J. (2021). Metrics for benchmarking and uncertainty quantification: quality, applicability, and best practices for machine learning in chemistry. Trends Chem. *3*, 146–156. https://doi.org/10.1016/J.TRECHM.2020.12.004.

57. Jones, D.T. (2019). Setting the standards for machine learning in biology. Nat. Rev. Mol. Cell Biol. *20*, 659–660. https://doi.org/10.1038/s41580-019-0176-5.

58. Webb, G.I., and Conilione, P. (2005). Estimating Bias and Variance from Data.

59. Yadav, S., and Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In Proceedings - 6th International Advanced Computing Conference IACC 2016, pp. 78–83. https://doi.org/10.1109/IACC.2016.25.

60. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., and Kording, K.P. (2017). The need to approximate the use-case in clinical machine learning. GigaScience *6*, 1–9. https://doi.org/10.1093/GIGASCIENCE/GIX019.

61. Roscher, R., Bohn, B., Duarte, M.F., and Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. IEEE Access *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199.

62. Kovalerchuk, B., Ahmad, M.A., and Teredesai, A. (2021). Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. Stud. Comput. Intell. *937*, 217–267. https://doi.org/10.1007/978-3-030-64949-4_8.

63. Lipton, Z.C. (2018). The mythos of model interpretability. Commun. ACM *61*, 36–43. https://doi.org/10.48550/arxiv.1606.03490.

64. Sokol, K., and Flach, P. (2019). Explainability fact sheets: a framework for systematic assessment of explainable approaches. In FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 56–67. https://doi.org/10.1145/3351095.3372870.

65. Lee, K.Y., and Mooney, D.J. (2012). Alginate: properties and biomedical applications. Prog. Polym. Sci. *37*, 106–126. https://doi.org/10.1016/j.progpolymsci.2011.06.003.

66. Lee, B.B., Ravindra, P., and Chan, E.S. (2013). Size and shape of calcium alginate beads produced by extrusion dripping. Chem. Eng. Technol. *36*, 1627–1642. https://doi.org/10.1002/ceat.201300230.

67. Grosskopf, A.K., Truby, R.L., Kim, H., Perazzo, A., Lewis, J.A., and Stone, H.A. (2018). Viscoplastic matrix materials for embedded 3D printing. ACS Appl. Mater. Interfaces *10*, 23353–23361. https://doi.org/10.1021/ACSAMI.7B19818/SUPPL_FILE/AM7B19818_SI_001.PDF.

68. Bhattacharjee, T., Zehnder, S.M., Rowe, K.G., Jain, S., Nixon, R.M., Sawyer, W.G., and Angelini, T.E. (2015). Writing in the granular gel medium. Sci. Adv. *1*. https://doi.org/10.1126/SCIADV.1500655/SUPPL_FILE/MOVIE_S9.MOV.

69. O'Bryan, C.S., Bhattacharjee, T., Marshall, S.L., Gregory Sawyer, W., and Angelini, T.E. (2018). Commercially available microgels for 3D bioprinting. Bioprinting *11*, e00037. https://doi.org/10.1016/J.BPRINT.2018.E00037.

70. Conley, G.M., Aebischer, P., Nöjd, S., Schurtenberger, P., and Scheffold, F. (2017). Jamming and overpacking fuzzy microgels: deformation, interpenetration, and compression. Sci. Adv. *3*, e1700969. https://doi.org/10.1126/sciadv.1700969.

71. van Hecke, M. (2010). Jamming of soft particles: geometry, mechanics, scaling and isostaticity. J. Phys. Condens. Matter *22*, 033101. https://doi.org/10.1088/0953-8984/22/3/033101.

72. Song, C., Wang, P., and Makse, H.A. (2008). A phase diagram for jammed matter. Nature *453*, 629–632. https://doi.org/10.1038/nature06981.

73. Dinkgreve, M., Paredes, J., Denn, M.M., and Bonn, D. (2016). On different ways of measuring "the" yield stress. J. Nonnewton. Fluid Mech. *238*, 233–241. https://doi.org/10.1016/J.JNNFM.2016.11.001.

74. Larsson, M., and Duffy, J. (2013). An overview of measurement techniques for determination of yield stress. Annu. Trans. Nord. Rheol. Soc. *21*, 125–138.

75. Sarmadi, M., Behrens, A.M., McHugh, K.J., Contreras, H.T.M., Tochka, Z.L., Lu, X., Langer, R., and Jaklenec, A. (2020). Modeling, design, and machine learning-based framework for optimal injectability of microparticle-based drug formulations. Sci. Adv. *6*, eabb6594. https://doi.org/10.1126/SCIADV.ABB6594.

76. Cilurzo, F., Selmin, F., Minghetti, P., Adami, M., Bertoni, E., Lauria, S., and Montanari, L. (2011). Injectability evaluation: an open issue. AAPS PharmSciTech *12*, 604–609. https://doi.org/10.1208/S12249-011-9625-Y/FIGURES/2.

77. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2021). A comprehensive survey on transfer learning. Proc. IEEE *109*, 43–76. https://doi.org/10.1109/JPROC.2020.3004555.

78. Hinton, T.J., Jallerat, Q., Palchesko, R.N., Park, J.H., Grodzicki, M.S., Shue, H.J., Ramadan, M.H., Hudson, A.R., and Feinberg, A.W. (2015). Three-dimensional printing of complex biological structures by freeform reversible embedding of suspended hydrogels. Sci. Adv. *1*. https://doi.org/10.1126/SCIADV.1500758/SUPPL_FILE/1500758_SM.PDF.

79. Wolf, K.J., Weiss, J.D., Uzel, S.G.M., Skylar-Scott, M.A., and Lewis, J.A. (2022). Biomanufacturing human tissues via organ building blocks. Cell Stem Cell *29*, 667–677. https://doi.org/10.1016/J.STEM.2022.04.012.

80. Skylar-Scott, M.A., Uzel, S.G.M., Nam, L.L., Ahrens, J.H., Truby, R.L., Damaraju, S., and Lewis, J.A. (2019). Biomanufacturing of organ-specific tissues with high cellular density and embedded vascular channels. Sci. Adv. *5*, eaaw2459. https://doi.org/10.1126/SCIADV.AAW2459/SUPPL_FILE/AAW2459_SM.PDF.

81. Schmidt, J., Marques, M.R.G., Botti, S., and Marques, M.A.L. (2019). Recent advances and applications of machine learning in solid-state materials science. npj Comput. Mater. *5*. 83–36. https://doi.org/10.1038/s41524-019-0221-0.

82. Kuo, C.K., and Ma, P.X. (2008). Maintaining dimensions and mechanical properties of ionically crosslinked alginate hydrogel scaffoldsin vitro. J. Biomed. Mater. Res. *84*, 899–907. https://doi.org/10.1002/jbm.a.31375.

83. Davarcı, F., Turan, D., Ozcelik, B., and Poncelet, D. (2017). The influence of solution viscosities and surface tension on calcium-alginate microbead formation using dripping technique. Food Hydrocoll. *62*, 119–127. https://doi.org/10.1016/J.FOODHYD.2016.06.029.

84. Adams, S., Frith, W.J., and Stokes, J.R. (2004). Influence of particle modulus on the rheological properties of agar microgel suspensions. J. Rheol. (N. Y. N. Y). *48*, 1195–1213. https://doi.org/10.1122/1.1795193.

85. Menut, P., Seiffert, S., Sprakel, J., and Weitz, D.A. (2012). Does size matter? Elasticity of compressed suspensions of colloidal- and granular-scale microgels. Soft Matter *8*, 156–164.

86. Fernandez-Nieves, A., Wyss, H.M., Mattsson, J., and Weitz, D.A. (2011). Microgel Suspensions: Fundamentals and Applications. Microgel Suspensions: Fundamentals and Applications. https://doi.org/10.1002/9783527632992.

87. Pelaez-Fernandez, M., Souslov, A., Lyon, L.A., Goldbart, P.M., and Fernandez-Nieves, A. (2015). Impact of single-particle compressibility on the fluid-solid phase transition for ionic microgel suspensions. Phys. Rev. Lett. *114*, 098303. https://doi.org/10.1103/PhysRevLett.114.098303.

88. Harth, K., Wang, J., Börzsönyi, T., and Stannarius, R. (2020). Intermittent flow and transient congestions of soft spheres passing narrow orifices. Soft Matter *16*, 8013–8023. https://doi.org/10.1039/D0SM00938E.

89. Conley, G.M., Zhang, C., Aebischer, P., Harden, J.L., and Scheffold, F. (2019). Relationship between rheology and structure of interpenetrating, deforming and compressing microgels. Nat. Commun. *10*, 2436–2438. https://doi.org/10.1038/s41467-019-10181-5.

90. Childs, C.M., and Washburn, N.R. (2019). Embedding domain knowledge for machine learning of complex material systems. MRS Commun. *9*, 806–820. https://doi.org/10.1557/MRC.2019.90.

91. Menon, A., Gupta, C., Perkins, K.M., Decost, B.L., Budwal, N., Rios, R.T., Zhang, K., Póczos, B., and Washburn, N.R. (2017). Elucidating multi-physics interactions in suspensions for the design of polymeric dispersants: a hierarchical machine learning approach. Mol. Syst. Des. Eng. *2*, 263–273. https://doi.org/10.1039/C7ME00027H.

92. Mahmoudabadbozchelou, M., Caggioni, M., Shahsavari, S., Hartt, W.H., Em Karniadakis, G., and Jamali, S. (2021). Data-driven physics-informed constitutive metamodeling of complex fluids: a multifidelity neural network (MFNN) framework. J. Rheol. (N. Y. N. Y). *65*, 179–198. https://doi.org/10.1122/8.0000138.

93. Lookman, T., Balachandran, P.v., Xue, D., and Yuan, R. (2019). Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Comput. Mater. *5*. 21–17. https://doi.org/10.1038/s41524-019-0153-8.

94. Wang, A.Y.-T., Murdock, R.J., Kauwe, S.K., Oliynyk, A.O., Gurlo, A., Brgoch, J., Persson, K.A., and Sparks, T.D. (2020). Machine learning for materials scientists: an introductory guide toward best practices. Chem. Mater. *32*, 4954–4965. https://doi.org/10.1021/ACS.CHEMMATER.0C01907.