

RESEARCH ARTICLE | OCTOBER 13 2021

# Feature importance in multi-dimensional tissue-engineering datasets: Random forest assisted optimization of experimental variables for collagen scaffolds <sup>EP</sup>

Malavika Nair   ; Ioana Bica  ; Serena M. Best  ; Ruth E. Cameron 

 Check for updates



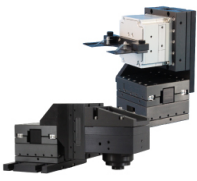
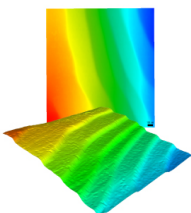
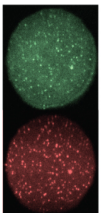
*Appl. Phys. Rev.* 8, 041403 (2021)

<https://doi.org/10.1063/5.0059724>

  
View  
Online

  
Export  
Citation

CrossMark

 <p><b>MCL</b> MAD CITY LABS INC. www.madcitylabs.com</p>	<p>Nanopositioning Systems</p> 	<p>Modular Motion Control</p> 	<p>AFM and NSOM Instruments</p> 	<p>Single Molecule Microscopes</p> 
--	--	--	---	--

# Feature importance in multi-dimensional tissue-engineering datasets: Random forest assisted optimization of experimental variables for collagen scaffolds

Cite as: Appl. Phys. Rev. 8, 041403 (2021); doi: 10.1063/5.0059724

Submitted: 11 June 2021 · Accepted: 13 September 2021 ·

Published Online: 13 October 2021



View Online



Export Citation



CrossMark

Malavika Nair,<sup>1,a)</sup>  Ioana Bica,<sup>2,3</sup>  Serena M. Best,<sup>1</sup>  and Ruth E. Cameron<sup>1</sup> 

## AFFILIATIONS

<sup>1</sup>Cambridge Centre for Medical Materials, Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, United Kingdom

<sup>2</sup>Department of Engineering Science, University of Oxford, Oxford, United Kingdom

<sup>3</sup>The Alan Turing Institute, London, United Kingdom

<sup>a)</sup> Author to whom correspondence should be addressed: [mn404@cam.ac.uk](mailto:mn404@cam.ac.uk)

## ABSTRACT

Ice-templated collagen-based tissue-engineering scaffolds are ideal for controlled tissue regeneration since they mimic the micro-environment experienced *in vivo*. The structure and properties of scaffolds are fine-tuned during fabrication by controlling a number of experimental parameters. However, this parameter space is large and complex, rendering the interpretation of results and selection of optimal parameters to be challenging in practice. This paper investigates the impact of a cross section of this parameter space (drying conditions and solute environment) on the scaffold microstructure. Qualitative assessment revealed the previously unreported impact of drying temperature and pressure on pore wall roughness, and confirmed the influence of collagen concentration, solvent type, and solute addition on pore morphology. For quantitative comparison, we demonstrate the novel application of random forest regression to analyze multi-dimensional bio-materials datasets, and predict microstructural attributes for a scaffold. Using these regression models, we assessed the relative importance of the input experimental parameters on quantitative pore measurements. Collagen concentration and pH were found to be the largest factors in determining pore size and connectivity. Furthermore, circular dichroism peak intensities were also revealed to be a good predictor for structural variations, which is a parameter that has not previously been investigated for its effect on a scaffold microstructure. Thus, this paper demonstrates the potential for predictive models such as random forest regressors to discover novel relationships in biomaterials datasets. These relationships between parameters (such as circular dichroism spectra and pore connectivity) can therefore also be used to identify and design further avenues of investigation within biomaterials.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0059724>

## I. INTRODUCTION

Collagen-based implants have emerged as viable tissue-engineering constructs, since they offer a similar model of the surface chemistry of natural tissues *in vivo*,<sup>1–3</sup> and can be modified to match the mechanical and biochemical properties required for a given application.<sup>4,5</sup> The process of freeze-drying offers a scalable method to create high porosity scaffolds, eliciting a positive cell response across a variety of cell types.<sup>6,7</sup> Different cell types have also been proposed to infiltrate a scaffold to differing extents depending on its interconnectivity,<sup>8,9</sup> which can be controlled through the experimental variables in

the freeze-drying process.<sup>9</sup> This process involves two main stages: freezing and drying.<sup>10,11</sup> During the first freezing stage, ice crystals nucleate and grow, rejecting any added solutes in the collagen suspension into the grain boundaries between ice crystals.<sup>10</sup> The next stage comprises *primary* drying which refers to the sublimation of ice, and *secondary* drying involving the desorption of unfrozen water.<sup>10,12</sup>

A large body of literature has been devoted to understanding the influence of numerous experimental variables on the resulting structural parameters of a scaffold, including solutes,<sup>13</sup> solvents,<sup>6,14,15</sup> biomolecular additives,<sup>16,17</sup> and, to a lesser extent, drying conditions.<sup>18</sup>

Recently investigations of ice crystals during sublimation have been shown to result in mesostructural features,<sup>19,20</sup> suggesting that the drying process during freeze-drying may impart some of these structural attributes. In all of these studies, a single independent experimental variable is varied systematically in order to extract the relationship between the measured structural attribute and the input parameters. In spite of the number of variables which have been modified individually during freeze-drying, there is a need for a controlled, systematic investigation on the sublimation pressures and any morphological changes that may consequently arise in the ice or collagen microstructures.

With a growing number of variables that can be modified, datasets in tissue-engineering have become both large and complex. While this may lead to some difficulty in data interpretation, they also present the opportunity to uncover underlying non-linear relationships between variables. Machine learning methods based on supervised learning can be trained to map a set of input parameter to output features and have the power to extract such relationships from large datasets. Random forests are one such type of supervised learning model which are advantageous for their ability to account for non-linear interactions in complex datasets.<sup>21</sup> Random forests are also a more suitable machine learning method for smaller datasets since they are less prone to overfitting than alternative methods such as neural networks.<sup>22,23</sup> Although random forests have been used extensively for classification within biomedical and genetic datasets,<sup>24–28</sup> and to a lesser degree for regression,<sup>29–32</sup> their potential with tissue-engineering datasets remains untapped.

In this paper, we investigate two case studies: first, the influence of primary and secondary drying on scaffold microstructure and second the influence of collagen–solvent–solute interactions on these microstructural attributes. This paper then demonstrates the use of random forest regression on these datasets to assess the relative influence of a range of experimental variables on the structural attributes of collagen scaffolds. Understanding these relationships between input and output parameters can aid in the targeted design of future experiments, and lead to the development of a scaffold toolkit. Such design toolkits have the potential to identify the parameters required to create tissue-engineering scaffolds which are optimized for a specific biological application.

## II. MATERIALS AND METHODS

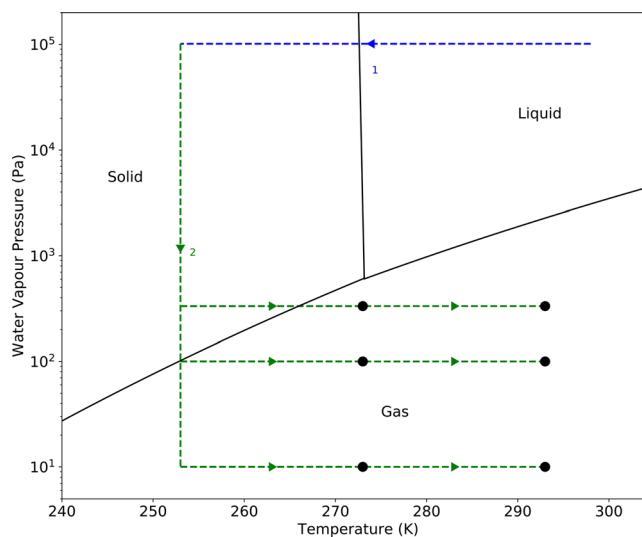
### A. Conditions of study

#### 1. Case 1: Primary and secondary drying of collagen scaffolds

Scaffolds made from 1.0 w/v%, undialyzed collagen hydrated in acetic acid were studied under the six drying conditions illustrated in Fig. 1, including two temperatures of 273 K (0 °C) and 293 K (20 °C) at three chamber pressures of 80 mTorr (10 Pa), 100 Pa (750 mTorr), and 333 Pa (2500 mTorr). The chamber pressure values are also listed in mTorr for comparison with usual freeze-drying protocols, although the results and discussions will refer to the pressures in Pa for consistency.

#### 2. Case 2: Solvent–solute interactions with collagen

33 conditions were chosen from a selection of four collagen concentrations (0.5, 1.0, 1.5, and 2.0 w/v%) and three solvents (0.05 M acetic acid, 0.001 M hydrochloric acid, water), shown in Table I.



**FIG. 1.** Schematic of the freezing (blue) and drying (green) cycles employed in the freeze-drier. Since the scaffolds were allowed to equilibrate to the chamber pressures and temperatures, the drying can be considered to occur at the conditions as set in the freeze-drier (black dots). The black dots represent the six conditions considered, with three pressures (10, 100, and 333 Pa) and two temperatures (273 and 293 K) considered in this study.

The concentrations of both the acid solvents were selected to ensure they possessed the same pH (pH 3) but were comprised of different chemical moieties. The effect of solute interaction with collagen was examined through the dialysis of the as-supplied collagen to remove the presence of any bound salts, as well as through the subsequent re-introduction of a precise quantity (0.5 w/v%; 85 mM) of sodium chloride in the dialyzed collagen. Scaffolds are identified by codes of the form “A-B-C” as indicated in Table I, where

- A is the concentration of collagen in w/v%,
- B is the first letter of the solvent used, and
- C denotes the solute conditions:
  - X for undialyzed with no additives,
  - D for dialyzed with no additives, and
  - N for dialyzed with 0.5 w/v% NaCl added.

Where general behavior at a given solute condition is considered irrespective of the concentration, the notation is reduced simply to “B-C.”

### B. Collagen dialysis

Type I insoluble microfibrillar bovine dermal collagen (Devro Medical, Collagen Solutions) was packed into a semi-permeable cellulose Visking tubing (permeability limit of molecules larger than 14 kDa, Sigma Aldrich) and hydrated with water. The Visking tubing was placed in a large water bath for 72 h, with frequent replacement of the water bath to ensure that a concentration was established to drive the diffusion of solutes out of the collagen. Residual water was removed from the dialyzed collagen by freeze-drying at 80 mTorr (VirTis Advantage).

**TABLE I.** 33 conditions of study with varying collagen concentration, solvent and solute content (through dialysis and sodium chloride addition) investigated in this study. Scaffolds are identified by codes of the form A-B-C, where X is the concentration of collagen in w/v%; Y is the first letter of the solvent used (A for acetic acid, H for hydrochloric acid, and W for water); and Z denotes the solute conditions (X for no additives, D for dialyzed, and N for dialyzed with 0.5 w/v% NaCl added).

		Dialysis	Solvent					
			Acetic acid		Hydrochloric acid		Water	
			No solute	+ NaCl	No solute	+ NaCl	No solute	+ NaCl
Collagen concentration (w/v%)	0.5	×	0.5-A-X		0.5-H-X		0.5-W-X	
		✓	0.5-A-D		0.5-H-D		0.5-W-D	
	1.0	×	1-A-X		1-H-X		1-W-X	
		✓	1-A-D	1-A-N	1-H-D	1-H-N	1-W-D	1-W-N
	1.5	×	1.5-A-X		1.5-H-X		1.5-W-X	
		✓	1.5-A-D	1.5-A-N	1.5-H-D	1.5-H-N	1.5-W-D	1.5-W-N
	2.0	×	2-A-X		2-H-X		2-W-X	
		✓	2-A-D	2-A-N	2-H-D	2-H-N	2-W-D	2-W-N

### C. Collagen slurry preparation

Suspensions of Type I microfibrillar bovine dermal collagen (Devro, Collagen Solutions) of collagen concentrations of 0.5, 1.0, 1.5, and 2.0 w/v% were hydrated in the solvent of choice as defined in Subsection II A 2 and left to swell overnight. Suspensions were blended using a commercial blender (Waring model 8011EG) at 22 000 rpm in two, two-minute bursts, with one minute of rest between each burst to cool the motor down. They were then centrifuged at 2500 rpm for 10 min to allow any trapped air bubbles to rise and coagulate. The top layer in the collagen samples (containing air bubbles) was subsequently separated and discarded.

### D. Scaffold fabrication

The choice of mold, filling volume, and thermal profiles was based on prior work<sup>33,34</sup> to fabricate an isotropic pore architecture. 8 ml of collagen slurry was pipetted into each well of a polystyrene 6-well plate (CytoOne<sup>®</sup>, STARLAB, filling height approximately 8 mm assuming a 17.4 mm diameter well plate with cylindrical geometry). The well plates were equilibrated to room temperature in the freeze-drier (VirTis Advantage), then cooled at 0.5 °C min<sup>-1</sup> to -20 °C. The scaffolds for the drying study were allowed to equilibrate, then dried at the chamber pressures and temperatures of choice as illustrated in Fig. 1, as set through the freeze-drier. The scaffolds created for the solvent study referred to in Sec. II A 2 were sublimed at 0 °C for 20 h under a vacuum less than 100 mTorr (10 Pa).

### E. Scanning electron microscopy

Scanning electron microscopy (Phenom X Pro) was performed on scaffolds produced through all sublimation conditions and a subset of solvent-solute conditions. Scaffold production conditions were chosen to ensure that two distinct relationships were explored: the effect of various solvents on the standard collagen concentration (1.0 w/v%) and the effect of various concentrations with a standard solvent (0.05 M AcOH). Longitudinal cross sections from each sample were sectioned and attached to a pin-style stub using carbon tape. Samples were not sputter-coated with a conductive material to prevent

obscuration of structural features as previously reported by Ashworth.<sup>35</sup> Images were captured using a variable pressure stage with the Phenom Pro X with a full BSE detector at a high voltage of 10 kV at ~ × 1100 magnification. Micrographs from two distinct replicates were imaged to verify batch-to-batch reproducibility.

### F. Atomic force microscopy

Collagen films from a slurry of 0.5 w/v% collagen in acetic acid were cast on microscope glass cover slips and imaged using the Dimension V Scanning Probe Microscope (Veeco) under tapping mode after air-drying overnight, then subsequently imaged after vacuum drying at 50 mTorr overnight. 5 × 5 μm areas were imaged for an overview of the sample topography, and 1 × 1 μm areas were imaged to verify the presence of D-banded collagen fibrils. Complete AFM micrographs are shown in the supplementary material, Fig. 1.

### G. Circular dichroism

Dialyzed and non-dialyzed collagen suspensions in 0.05 M acetic acid, 0.001 M hydrochloric acid and water as prepared in Sec. II C were diluted to 0.0125 w/v%, and suspensions with 0.5 w/v% sodium chloride were diluted to 0.025 w/v%. The concentrations were selected after an optimization of the CD signal collected. Suspensions were then transferred to a quartz cuvette with a path length of 10 mm (Hellma Analytics) and placed in the CD spectrometer (ChiraScan-Plus CD Spectrometer, Applied Photophysics).

A sample of 0.0125 w/v% of soluble collagen solution in 0.05 M acetic acid was used as a positive control and the acetic acid, hydrochloric acid, and water solvents were used as background controls. The elliptical polarization of all samples was measured using a sampling rate of 2 s per point, between 200 and 280 nm at a bandwidth of 1 nm. Traces of the background solvents were removed from the traces of samples, following a Savitsky-Golay polynomial smoothing of window size 3 to correct for the baseline of the signal. Data were plotted as the mean ± standard deviation of three independent measurements.

## H MicroCT

### 1. Imaging

To eliminate edge effects, sections of each scaffold were biopsy punched from the same central location to ensure consistency when comparing locations. A 3D projection of the scaffold architecture was imaged without a filter at 25 kV, 138  $\mu$ A, with a 0.2° step size between projections obtained across a 180° rotation using the Skyscan 1172 (Bruker) MicroCT. The binning size was set to 4000  $\times$  2000 and the pixel size was set to 3  $\mu$ m following the guidelines for MicroCT data acquisition as explored by Nair *et al.*<sup>36</sup>

### 2. Reconstruction

Cross sections were reconstructed using a Feldkamp algorithm with 30% beam hardening correction with ring artifact reduction set to level 4 out of 10 in the Skyscan NRecon software. Beam hardening correction reduces the effect of the edges appearing brighter than the center due to differential x-ray absorption, whereas the ring artifact reduction was set to correct for the presence of ring-like artifacts arising from the rotation of dead camera pixels during the imaging process, without overcompensating and resulting in a blurry image.

### 3. Analysis

Using the *Bruker CT-Analyser* software, 1 mm<sup>3</sup> volumes of interest (VOIs) were selected from each scaffold from the top, middle, and bottom of the scanned datasets. These VOIs were thresholded and segmented using the automatic Otsu algorithm. The average pore sizes and pore size distributions were determined using the 3D Analyser tool in *CT-Analyser*, producing a histogram of pore diameters binned at approximately every 6  $\mu$ m (twice the pixel size). Pore sizes were plotted as the mean  $\pm$  the standard deviation measured across nine VOIs.

Two measures of interconnectivity, the median interconnection diameter, and the percolation diameter were used to quantify microstructural connectivity in this work. The median interconnection diameter is the voxel cluster diameter at which the volume interconnectivity is 50%, and the percolation diameter is a scale-invariant parameter quantifying the maximum diameter of a sphere that can traverse through the entire structure.

To obtain the connectivity values, A 3D region of interest (ROI) shrink-wrap analysis was then performed using *CT-Analyser* at 2–100 voxels, and the total VOI was calculated at each voxel size. The percentage interconnectivity ( $I$ ) at a given connection size  $s$  can be expressed as<sup>37</sup>

$$I = \frac{V - V_c}{V - V_m} \times 100\%,$$

where  $V$  is the total volume of interest,  $V_m$  is the volume of the solid material, and  $V_c$  is the volume measured after a shrink-wrap operation at voxel size  $c$ . Then, using the custom python script and subdivision method developed by Nair *et al.*,<sup>36</sup> The median interconnection diameter (probe size at which only 50% of the scaffold is accessible) was calculated using a linear interpolation for each interconnectivity measurement. The median interconnection diameter was represented

as a violin plot depicting the distribution of connectivities obtained in the scaffolds.

### I. Random Forest Regression

This study considers a supervised learning problem which involves a dataset with 11 input features and three output parameters (pore size, percolation diameter, and median interconnection diameter) and 37 data points representing 37 distinct conditions (averages from three independent sections from three scaffolds each). Of these 11 input parameters, seven are derived from experimental conditions (drying pressure, temperatures, choice of solvent, presence of additives, collagen concentration, and whether dialysis was performed) which are varied during fabrication, two are values extracted from a circular dichroism dataset (minimum and maximum spectral intensity), and two are highly correlated “supervariables” taking into account existing experimental input parameters (pH, which is correlated with solvent choice, and solute content, which is correlated with additives and solvent choice).

A random forest model involves an ensemble of individual decision trees, each of which is trained on a random section of the training data. The final prediction of the random forest is then decided by majority voting by the trees in a classification problem, or the mean prediction for a regression problem. This allows random forests to reduce problems of overfitting while maintaining a good performance. Random forests also offer high interpretability of the results, since the features influencing the predictions can be identified and ranked by their importance. Random forests are particularly suited to this machine learning problem of predicting scaffold structural parameters due to the small size of the dataset, and the ability to estimate the relative importance of the input features.

A random forest regressor was implemented for each pore size, percolation diameter, and median interconnection diameter, using the *scikit-learn* package in python. 80% of the dataset was assigned for training the final regressor and the remaining 20% as testing. The hyperparameters of the regressor were tuned through threefold cross-validation on the training data using a random search and localized grid search of the parameter space. The data were pre-processed by applying min-max normalization, and one-hot encoding of categorical variables (solvents) with the first column excluded to prevent multicollinearity. The results were obtained from 10 independent training runs of the regressors with different training-validation-test splits. The training and testing datasets are independent from each other, with the training set only being used as part of the cross-validation procedure for hyperparameter selection and for fitting the model. During each run, the test dataset is only used for evaluation. By repeating this process 10 times to sample different independent training and testing datasets, we can assess the performance of the model when evaluated on different subsets of the full dataset.

The performance of the random forest model is provided as the mean  $\pm$  standard deviation of 10 independent runs of the root mean squared error (RMSE) in prediction for each regressor. The feature importances were plotted as the mean  $\pm$  standard deviation of both the impurity-based and permutation-based importances of each input feature, obtained over the 10 repeats of the training and fitting process. Permutation-based importances were calculated on the training data to assess the importance assigned to each feature in developing the regressors.



### III. RESULTS

#### A. Case 1: Primary and secondary drying of collagen scaffolds

In this section, the effects of the two drying processes were uncoupled through a layered characterization of morphology development. First, the effect of secondary drying on collagen ultrastructure was investigated using AFM. Second, collagen scaffolds were dried under six distinct chamber conditions (including two chamber temperatures and three chamber pressures) in order to elucidate the influence of the two drying processes on scaffold microstructures and connectivity.

The phase channel from the AFM micrographs of cast collagen films is shown in Fig. 2 after air-drying at atmospheric pressure and after vacuum drying. Collagen samples displayed little or no fibrillar features in the AFM after air-drying overnight as seen in Fig. 2(a) across all samples. Figure 2(a) reveals the presence of several thin fragments of collagen which comprised the film (marked by black arrows in the zoomed inset).

Upon imaging the same films after vacuum drying, more, albeit a relatively low amount, of collagen fibers and fibrils were readily observable as seen in Fig. 2(b), exhibiting the characteristic D-banding. However, in spite of the increased incidence of such D-banded fibrils in the vacuum dried films, similar to the air-dried films, these samples were also primarily comprised of smaller collagen fragments as seen in the phase channel.

High magnification SEM images revealed that scaffolds produced at the chamber pressure of 333 Pa at 273 K possessed smooth pore walls, whereas a decrease in chamber pressure and, most importantly, drying temperature resulted in significant roughening of the pore wall topography.

The features arising from the roughening of the pore wall are indicated with arrows in Fig. 3(a). These include raised circular areas (blue arrows), a vein-like distribution of collagen struts (yellow arrows) or the faint aligned grooves of collagen (purple arrows).

The increase in roughness was most prominent when comparing the two drying temperatures (293 and 273 K). The differences in topography when comparing the highest chamber pressure (333 Pa) to the lowest (10 Pa) at a given temperature were predominantly through

the introduction of the small-scaled features from the circular areas of roughness and faint grooves of collagen alignment on a pore wall.

The mean pore sizes are plotted in Fig. 4(a) as a function of the drying temperatures and pressures. Most samples possessed similar pore sizes regardless of the drying pressure or temperature used, which is consistent with the low magnification SEM micrographs (supplementary material, Fig. 2). Notable exceptions to this include samples dried at 10 Pa at 273 and 293 K, samples dried at 333 Pa at 293 K, where some pore size distributions were skewed to smaller pore diameters (from approximately 120 to 50  $\mu\text{m}$ ), which is further supported by the full pore size distribution shown in the supplementary material, Fig. 3(a).

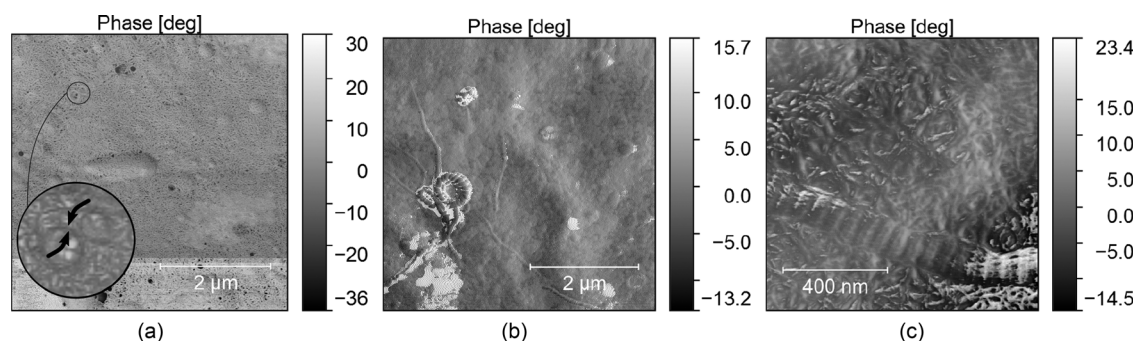
Figures 4(b) and 4(c) show violin plots of the percolation and median interconnection diameters, respectively. The violin plots include a box plot of the quantitative data, with the “violins” representing the probability distribution of the structural parameter at each condition. The percolation and median interconnection diameters are well matched to the pore sizes, although the plots reveal the presence of significant sample-to-sample variability and spread in measuring the percolation diameter.

Higher chamber pressures (333 Pa) resulted in scaffolds possessing a higher median interconnection diameter than scaffolds produced at lower chamber pressures (10 Pa), particularly at 273 K. Scaffolds produced at higher drying temperatures (293 K) also demonstrated a wider spread in the median interconnection diameters observed within the scaffolds.

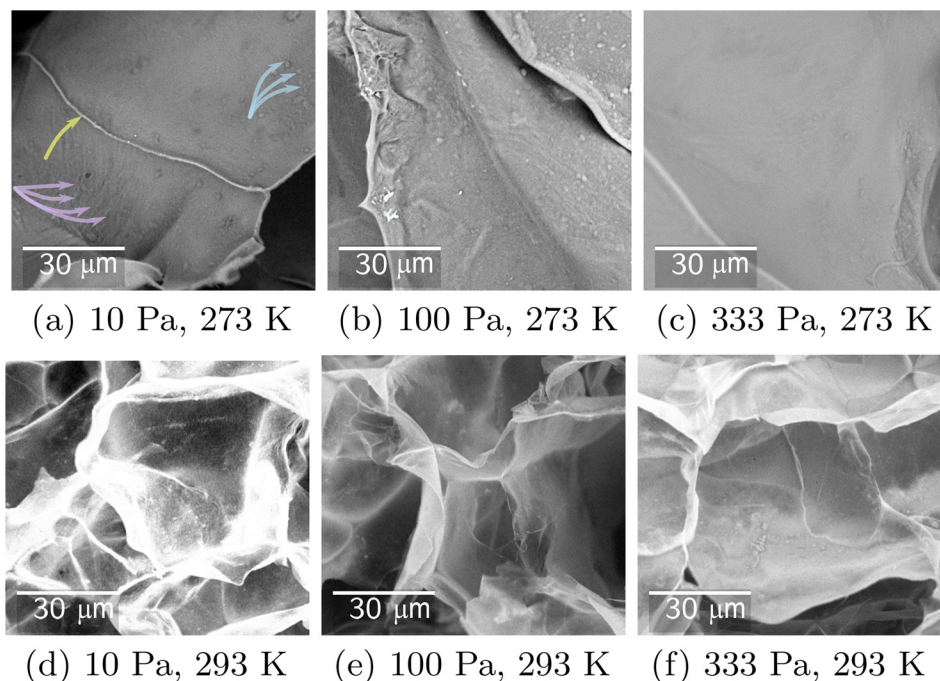
Some spread within each condition was observed, particularly at 333 and 10 Pa at both temperatures. This may be attributed to the observation of two distinct sets of interconnectivity plots comparing the accessible volume of a scaffold (interconnected volume) as a function of a probe size (interconnection diameter) in the supplementary material, Fig. 3(b).

#### B. Case 2: Solvent-solute interactions with collagen

This study considers a systematic variation in solute content and solvent type in the fabrication of collagen scaffolds. In addition, collagen concentration was also modified in each of these conditions with the hypothesis that features developing through the interaction of the solutes with collagen will be primarily influenced by collagen



**FIG. 2.** Comparison of  $5 \times 5 \mu\text{m}$  AFM images of collagen films (a) after air-drying and (b) after additional vacuum drying, revealing the inherent banding. (c) Higher magnification  $1 \times 1 \mu\text{m}$  AFM image of a typical fibril found in the vacuum dried collagen films in (b), with fibrils displaying the characteristic D-banding. Black arrows in the zoomed inset: thin fragments of collagen. White arrows in the height channel indicate a period in the banded collagen fibril.



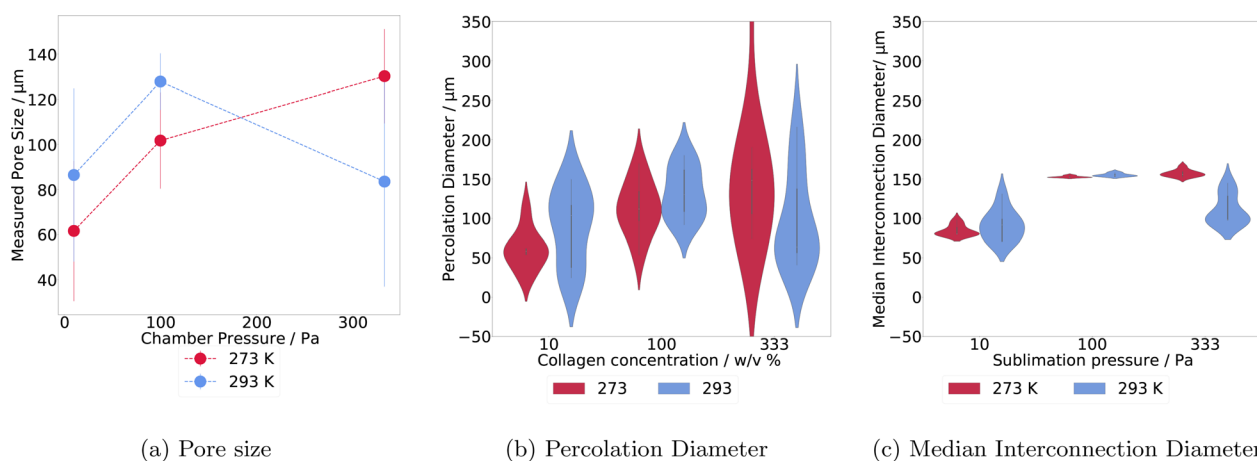
**FIG. 3.** High magnification SEM transverse micrographs of collagen scaffolds dried at various chamber pressures and temperatures. Mesoscale features were observed on the pore walls, with periodic features and roughness increasing at lower chamber pressures, and higher sublimation temperatures. Blue arrows: raised circular areas of roughness. Yellow arrows: vein-like collagen struts. Purple arrows: faint aligned grooves of collagen.

concentration, whereas those driven by changes to the ice crystal growth environment will be largely independent of collagen concentration. The microstructure and connectivity of the scaffolds were investigated through SEM and MicroCT, and any effects of solute interactions on collagen conformation were confirmed through CD.

Representative high magnification images of collagen scaffolds produced using acetic acid at various collagen concentrations, and with various solvents at 1.0 w/v% are shown in Fig. 5. The pore morphologies did not display significant changes with concentration at a given solute or solvent condition, except for the small pores seen with 1.5-A-X [Fig. 5(e)] and 2-A-D [Fig. 5(f)].

Although no significant changes are seen in the pore morphologies with increasing collagen concentration, the pore walls were observed to increase in thickness. For instance, the pore walls seen in the micrographs revealed that at 0.5 and 1.0 w/v% the walls were thin and fibrous [Figs. 5(a) and 5(b)], unlike the scaffolds fabricated at the higher concentrations where pore walls were thick and smooth [Figs. 5(e) and 5(f)].

Undialyzed scaffolds were found to possess slightly thinner fibers of collagen [Fig. 5(b)] when compared with thicker bundles of collagen that make up the pore walls in the dialyzed collagen scaffolds [Fig. 5(h)]. The incorporation of sodium chloride also resulted in several



**FIG. 4.** Pore size, median interconnection diameter, and percolation diameter of scaffolds. All three structural parameters indicate that scaffolds produced under various drying conditions reveal that higher chamber pressures resulted in scaffolds possessing a higher median interconnectivity. Higher drying temperatures (293 K) produce scaffold with a wider spread in the median interconnection diameters.

thin strands of collagen, coated in salt crystals [Fig. 5(m)]. Pore walls for these scaffolds were also highly faceted and polygonal in morphology when compared with the smoother walls in the dialyzed and undialyzed scaffolds.

A comparison of the pore morphology of scaffolds produced in various solvents revealed smooth collagen struts for A-X and A-D scaffolds [Figs. 5(b) and 5(h)] with several openings in the pore wall, whereas thin fibrillar struts were observed in H-X and H-D scaffolds [Figs. 5(c) and 5(i)]. In comparison, W-X and W-D had a closed, collapsed structure with thread-like pore walls [Figs. 5(d) and 5(j)].

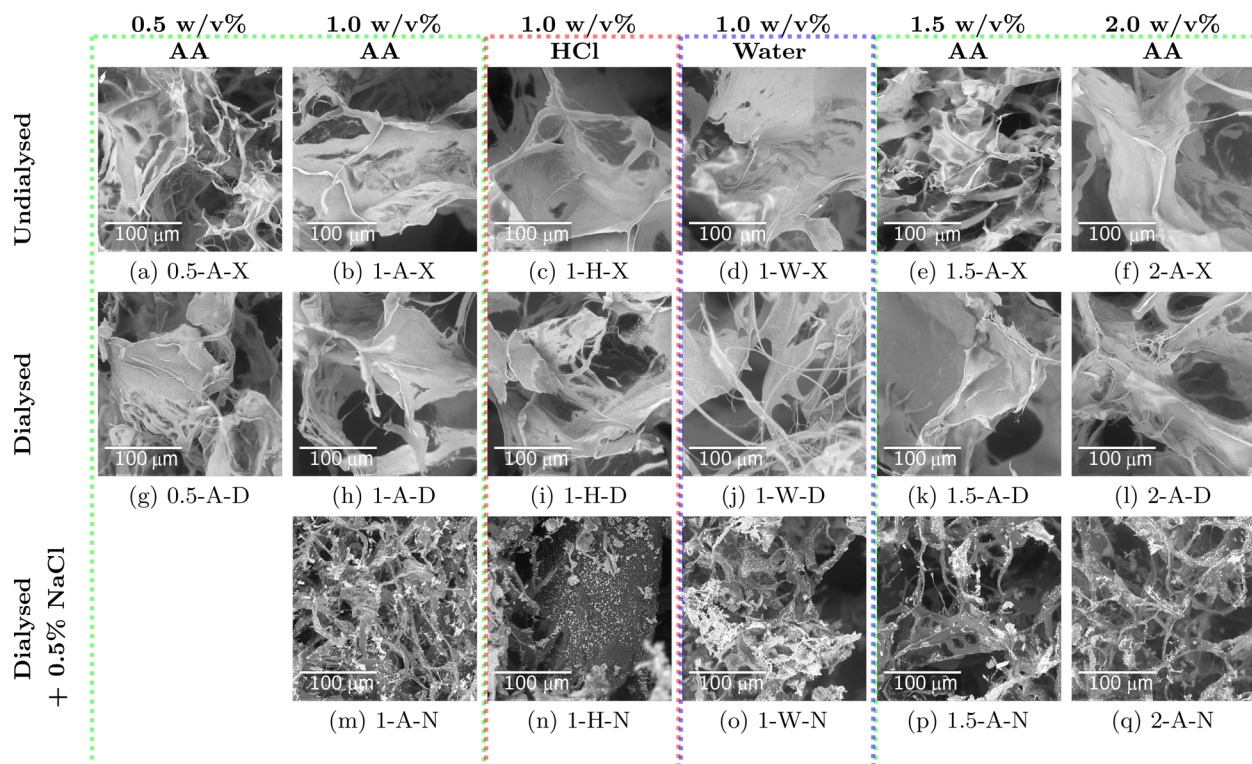
The samples with incorporated salt, A-N, H-N, and W-N [Figs. 5(m)–5(o)], possessed wire-like fibrous pore walls with a closed pore morphology unlike the dialyzed and non-dialyzed samples. The most significant impact arises from the inclusion of sodium chloride as seen from the fibrous collagen strands observed in the W-N samples [Fig. 5(o)] and to a lesser extent in the A-N or H-N samples [Figs. 5(m) and 5(n)].

The CD spectra for dialyzed and undialyzed insoluble collagen slurries hydrated in hydrochloric acid, acetic acid, and water are plotted in Fig. 6, along with the control spectrum for soluble collagen in acetic acid. The soluble collagen spectrum possessed a strong negative ellipticity at 200 nm and a positive band at 222 nm, with a crossover point at approximately 215 nm. Similar negative and positive peaks at 210 and 225 nm, respectively, were seen with insoluble collagen slurries.

Strong peaks were observed with A-X and H-X, whereas negative and positive peaks of diminished intensities were observed with the A-D, H-D, W-X, and W-D. All spectra intersected at the isoelliptical crossover point at approximately 215 nm. All decreases in the intensity of the negative band were accompanied by concurrent decreases in the intensity of the positive peaks, indicating that no significant denaturation has occurred during the processing of the slurries.

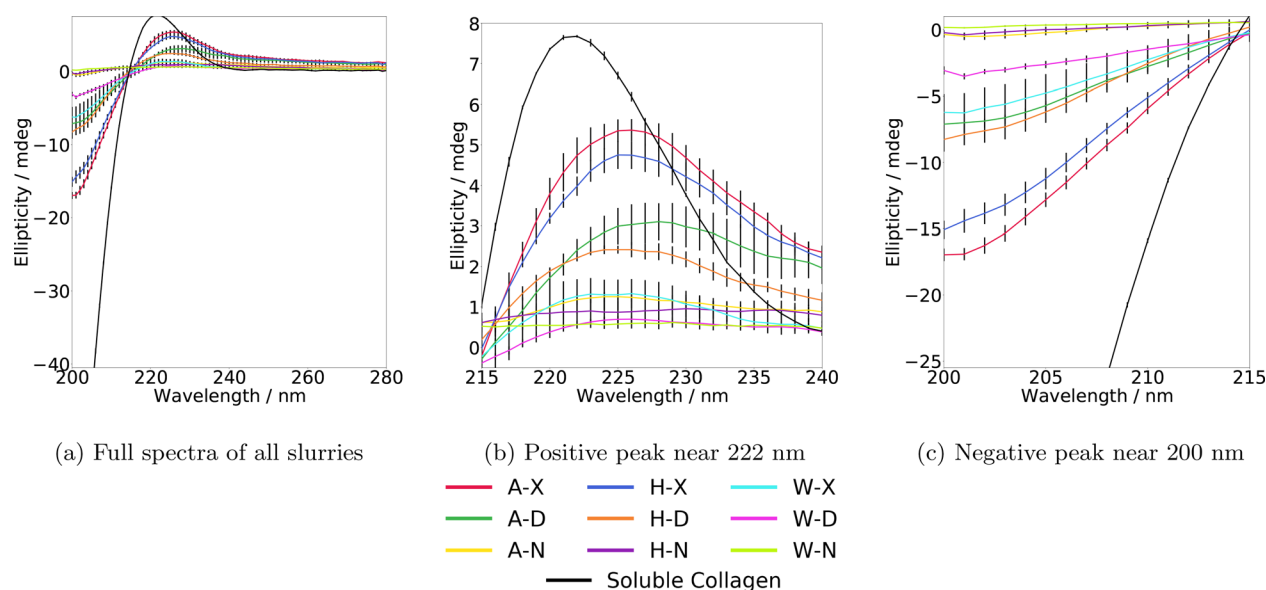
Dialysis as well as the inclusion of sodium chloride resulted in a significant reduction of ellipticity when compared with their undialyzed counterparts. In particular, although the solid concentration used in A-N, H-N, and W-N was double of that used with the other solute conditions, these samples produced signals with very low intensity. These spectra revealed the triple-helical peak as seen in Fig. 6(b), but exhibit a redshift in the isoelliptical point. Notably, slurries hydrated in W-N did not exhibit any negative ellipticity across the measurement range.

The mean pore sizes, percolation diameters, and median interconnection diameters at each condition are plotted in Fig. 7, with full distributions provided in the supplementary material, Figs. 4 and 5. No significant variation was observed in the scaffold microstructure across the different ROIs selected. Although most distributions are unimodal, they exhibit significant spread. In general, trends were only observed for the interconnection and percolation diameters with solvent choice and sodium chloride incorporation. The median



**FIG. 5.** Scanning electron micrographs of collagen scaffolds produced at varying concentrations, solvent conditions, and solute contents. Representative images are shown from left to right, increasing in concentration of collagen hydrated in acetic acid (green boxes) from 0.5 to 2.0 w/v%. Images of samples produced from hydrochloric acid (red box) and water (blue box) at 1.0 w/v% are also shown for comparison of solvents at a constant collagen concentration.





**FIG. 6.** Circular dichroism spectra of dialyzed and undialyzed insoluble collagen slurries hydrated in hydrochloric acid, acetic acid, and water as well as soluble collagen in acetic acid.

interconnection and percolation diameter of samples produced in water were lower than those produced in acetic and hydrochloric acid. A-N, H-N, and W-N all demonstrated low values of interconnection diameters and percolation diameters. A comparison of the pore sizes with the percolation diameter in Fig. 7 indicates a good match in the trends observed, particularly for acetic acid and water.

### C. Feature importance

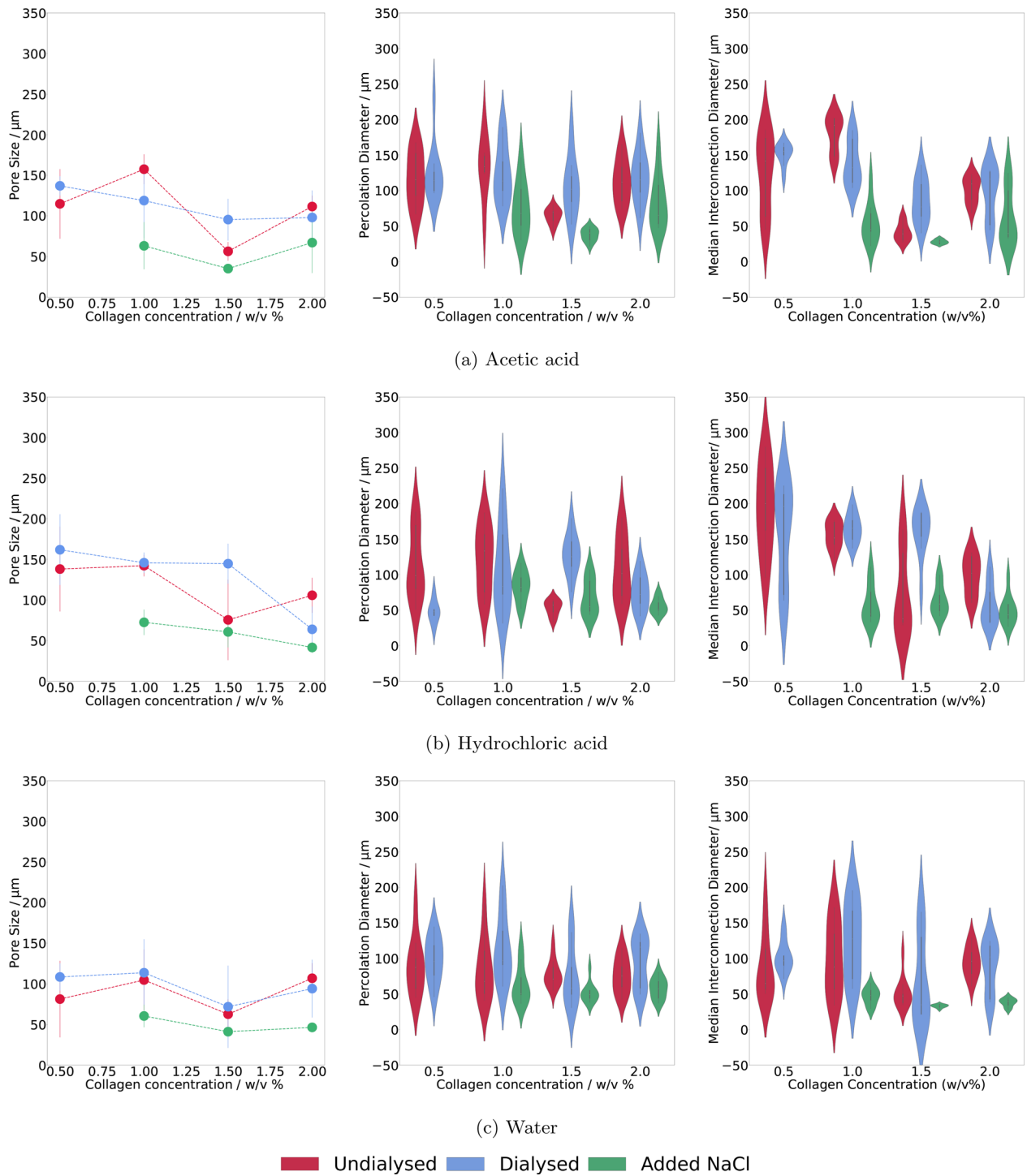
Three random forest regressors were trained on the combined sublimation and solvent datasets obtained for each structural parameter. The root mean squared error of the pore size, median interconnection, and percolation diameter models are shown in Table II, with the normalized errors also represented in brackets as the standardized percentage error of the total range. The RMSE of the test set is close to that of the training data within error for all three regressors, indicating that the models were able to fit the data well. Furthermore, the typical spread of pore sizes in the scaffolds investigated (see supplementary material Fig. 4 for pore size distributions) is comparable to these RMSEs, indicating a suitable model fit. Table II also shows that the inclusion of the supervariables in the model does not influence the predictor performance, within error.

The feature importance of the models trained on each parameter is illustrated in Fig. 8. Two feature importances are included here: first the mean impurity decrease-based importance and second, the permutation importance on the training data. The impurity-based feature importance measures how each feature decreases the variance of the splits and highlights how influential each feature is in determining the splits of the random forest model. On the other hand, permutation-based importances measure the decrease in the model performance when randomly shuffling each feature's values, but have the advantage that they are not biased by high cardinality values. Thus, the two

feature importances can be used together to ascertain the features with the greatest impact on the microstructural attributes.

For comparison between the variables, the mean feature importance for each regressor is plotted as a red vertical line. As a first estimate, the most influential features can be identified as those with importances above the mean. The results from training the model are represented by plotting the feature importance of 11 variables. These include experimental variables that were directly modified such as the drying temperature, the drying pressure, the presence of additives, the use of water/hydrochloric acid as a solvent, and whether the collagen was dialyzed. Other calculated variables were also included in the model, such as the intensity of the maximum peak and minimum peaks of the CD spectra, the total solute content, and solvent pH. The total solute content was calculated as the sum total of the concentrations of all solutes in the solvent. The pH variable normalized to "0" for the acidic solvents and "1" for water.

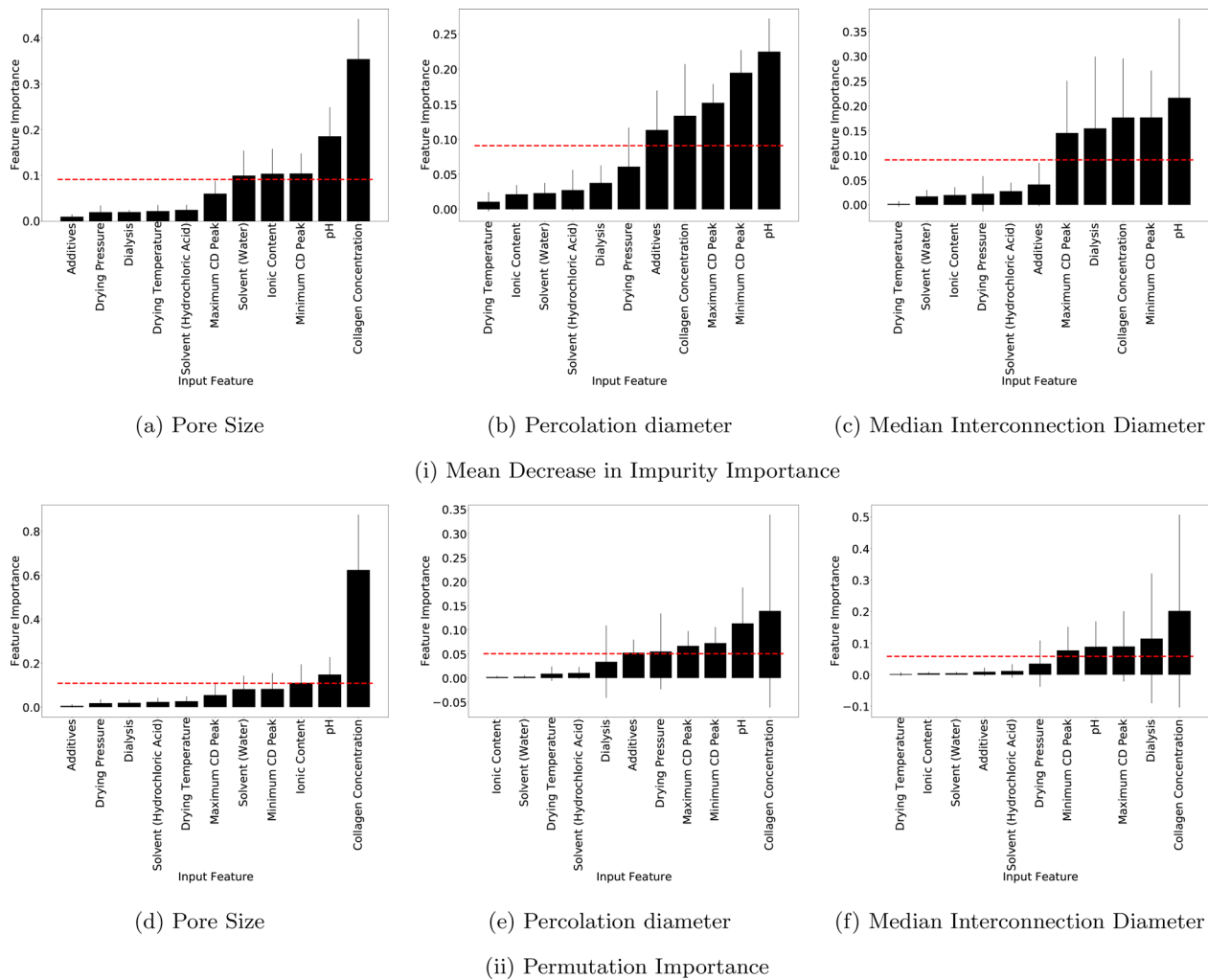
The feature importance graphs highlight that the collagen concentration and pH are the only experimental variables to influence all three structural parameters to the same extent. The feature importance graphs also show that the conformation of collagen as measured through circular dichroism spectral intensity may be capable of accounting for the changes caused by other experimental variables, such as solvent, solute incorporation, or the dialysis process. The solutes in the system are also shown to have an impact on the three structural attributes. The feature importances highlight the subtle differences caused by different solute environments. For instance, the presence of sodium chloride is an important feature for percolation diameter, dialysis is important for median interconnection diameter, whereas for pore size, the supervariable "total solute content" (i.e., both the molarity of sodium chloride and the molarity of the solvent) is seen to have a large impact.



**FIG. 7.** Pore size, median interconnection diameter, and percolation diameters of scaffolds produced with varying solvents and solute content. Pore size did not display any dependence on concentration or on dialysis. Samples produced with water and with added sodium chloride were the only conditions that demonstrated distinctly reduced pore sizes in comparison with other conditions. Percolation diameter did not exhibit a strong dependence on collagen concentration, whereas the median interconnectivity was observed to decrease with collagen concentration. Samples produced with water and with added sodium chloride demonstrated distinctly reduced percolation and interconnection diameters.

**TABLE II.** Summary table of the root mean squared errors of random forest regressors. Three regressors are considered here for the pore size, median interconnection diameter, and the percolation diameter. The regressors were trained both with and without the inclusion of supervariables (pH and "ionic content"), showing a slightly higher average RMSE for the models without the supervariables, but within the variability observed. RMSEs are given as the mean  $\pm$  standard deviation of 10 training repeats on the dataset. To standardize the comparison across the different regressors, the RMSEs normalized by the min-max range of each structural attribute are also reported in brackets, to 1 s.f.

Predictor	RMSE $\pm$ standard deviation ( $\mu\text{m}$ )			
	With supervariables		Without supervariables	
	Training	Test	Training	Test
Pore size	15 $\pm$ 4 (7%)	30 $\pm$ 6 (14%)	18 $\pm$ 5 (8%)	29 $\pm$ 4 (13%)
Percolation diameter	18 $\pm$ 3 (4%)	28 $\pm$ 4 (7%)	18 $\pm$ 3 (4%)	28 $\pm$ 4 (7%)
Median interconnection diameter	34 $\pm$ 10 (13%)	48 $\pm$ 10 (18%)	33 $\pm$ 10 (13%)	50 $\pm$ 14 (20%)



**FIG. 8.** Feature importance graphs: experimental parameters are ranked in order of least influential to most influential in determining (a) the pore size, (b) median interconnection diameter, and (c) percolation diameter. The red horizontal dotted line represents the mean feature importance for each structural measurement.

The inclusion of pH as an independent parameter (which is constant for acetic acid and hydrochloric acid) also provides further insight into the role played by the solvents. In the feature importance graph for pore size, the parameters “water as a solvent” and “pH” have nearly an identical importance, whereas “hydrochloric acid as a solvent” consistently ranks as a feature of low importance. This suggests that the change in pH has a greater impact on the measured pore values, than a change in the chemical moiety of the solvent.

Finally, the relative effect of drying conditions (temperature and pressure) is seen to be insignificant when compared with the effect of solvent and solute changes for the pore size and median interconnection diameter, yet drying pressure is an influential parameter for percolation diameter.

#### IV. DISCUSSION

In this paper, the impact of new experimental variables on scaffold microstructure as well as the ability to extract relationships between variables and structural parameters have been investigated through two case studies. The discussion will therefore first attempt to provide a mechanistic understanding of the role played by primary and secondary drying, as well as solvent–solute interactions within each case study and second, consider the implications of the relationships identified by the random forest model.

#### A. Mechanistic interpretation of results

##### 1. Case 1: Primary and secondary drying of collagen scaffolds

The AFM images seen in Figs. 2(a) and 2(b) indicate that the differences in air-drying and freeze-drying alone result in large changes to the collagen surface and nanostructure. In particular, there was a lack of D-banding in air-dried films, with long, thin fibers embedded within the surface as seen with AFM phase data in Fig. 2(a). In addition to the effect of drying, differences in pH<sup>38</sup> and hydration have been known to cause restructuring within a fibril, particularly with respect to the obscuration of D-banding with pH and water-driven linking of the polypeptide chains within and across the triple-helical molecules<sup>39</sup> to create the larger packing order of collagen. The increased incidence of fibrillar features on the surface of the collagen films following vacuum drying may therefore be explained by hydration-driven restructuring of collagen fibrils.<sup>39,40</sup>

Higher magnification SEM images in Fig. 3 revealed increased facetting and roughening at higher drying temperatures. Some increase in roughness and collagen alignment within the pore walls were also observed with an increase in pressure from 273 to 293 K. The aligned collagen grooves and the thicker collagen struts on the pore walls may have arisen from either the ice templated facets during primary drying or the collagen restructuring observed using the AFM during secondary drying. However, the presence and increased incidence of raised circular features on the wall cannot be explained by the observation of ice crystal sublimation seen here, and are therefore more likely to have arisen from secondary drying. Although little prior literature exists on the role of water in the formation on fiber bundles at the micro- and sub-micro scale, one interpretation of the formation of the vein-like collagen features can arise from packing of water around collagen fibers. A theoretical and experimental study on the formation of nanofibrils with peptide amphiphiles from micelles by

Deshmukh *et al.* suggests that the self-assembly and aggregation to larger structures may be governed by the dynamics and ordering in the vicinal water near hydrophilic side groups.<sup>41</sup> Regardless, the environmental scanning electron micrographs depicting thick collagen struts emerging from the grain boundaries of subliming ice crystals in the supplementary material, Fig. 6 suggest that the facetting of the pore walls as delineated by thick collagen struts were most likely to have arisen as an ice-templated feature. Thus, one of two phenomena controlling the underlying ice crystal morphology is responsible for the facetting at higher drying temperatures: freezing or sublimation.

The process responsible for these features can then be reconciled by the structural measurements of the scaffold, illustrated in Fig. 9. The increase in facetting and roughness at higher temperatures (and lower pressures) was reflected in the 3D morphology and connectivity with little to no change in pore diameters yet a small decrease in scaffold connectivities as shown in Figs. 4(b) and 4(c). At higher temperatures, some samples demonstrate the same pore sizes whereas others display lower pore sizes which directly contrasts with the increase in pore size expected from grain growth during annealing. On the other hand, at higher temperatures and lower pressures, there was a greater drive away from equilibrium which can result in the formation of fine ridges in the ice crystals as discussed by Nair *et al.*<sup>19</sup> Thus, the ice templating of sublimation-induced mesostructure may be responsible for the pore wall roughening and small associated decrease in connectivity observed at high drying temperatures and low drying pressures.

The six chamber conditions considered in this study will also correspond to six distinct drying rates. Several theoretical models have been developed to characterize the primary and secondary drying stages in the freeze-drying process, summarized in a review by Tchessalov *et al.*<sup>42</sup> In these models, for a constant inner surface area and mass transfer resistance, the drying rate  $\frac{\partial m}{\partial t}$  can be expressed to be directly proportional to the difference between the chamber pressure  $P_{ch}$  and the pressure over the subliming surface  $P_{subl}$  as follows:

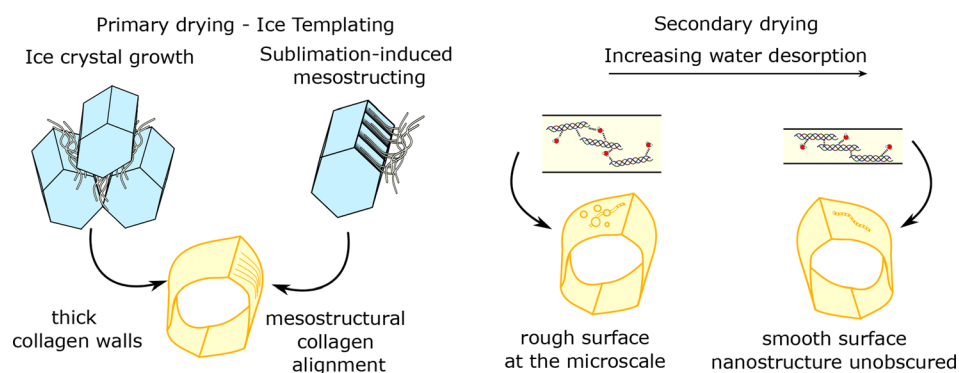
$$\frac{\partial m}{\partial t} \propto P_{subl} - P_{ch}.$$

For an estimate of the relative drying rates, the equilibrium vapor pressure of sublimation can be taken from a data table compiled by Wexler<sup>43</sup> as 611.4 Pa at 0 °C, at 2338.5 Pa at 20 °C. Compared to the slowest drying rate at 0 °C, 333 Pa as the baseline, a decrease in pressure results in an increase in the drying rate by 1.8-fold at 100 Pa and 2.1-fold at 10 Pa. Although an increase in temperature results increases the drying rate by approximately seven times at 20 °C, the changes in pressure only result in a maximum relative increase by 1.1-fold. With this analysis, the changes to pore shape and facetting at larger temperatures may be attributed to the large increase in drying rates from changing drying temperature. Similarly, the onset of finer mesoscaled features on the pore walls due to drying pressure was most obvious at 0 °C. The ability to observe these structural differences at this lower temperature may arise from the more pronounced differences in drying rates between the different pressures.

##### 2. Case 2: Solvent–solute interactions with collagen

The pore morphologies observed in this work (Fig. 5) indicate a smoother and more rounded pore morphology with acetic acid and to





**FIG. 9.** Hypothesized role of primary and secondary drying on collagen scaffold microstructure. The outer pore walls are templated by the ice crystal wall, whereas finer scaled mesostructural collagen alignment arises from sublimation-induced mesostructuring. A high water content may also result in the obscuration of nanoscaled structures, which is revealed with increasing water desorption.

a lesser extent hydrochloric acid. On the other hand, water and sodium chloride produce a more faceted morphology. This change in morphology does not appear to have a dependence on the solute content, but depends directly on the chemical moiety of the solute. Such fibrillar walls were also observed at high magnifications by Pawelec *et al.* with the incorporation of sodium chloride in scaffolds,<sup>13</sup> where the pore sizes were described to increase in size with the addition of salt. Although arguments based on the viscosity of the collagen slurry have been made to justify the larger pore sizes observed in literature,<sup>13</sup> the same argument falls short of supporting the evidence presented in this work. The results of Pawelec *et al.* are also in stark contrast to the decreased pore size and percolation diameters observed in this work. This discrepancy may be reconciled by the 1-hour washing step employed in the former to remove salts.

The addition of salt to 0.5 w/v% collagen was not considered in this experimental plan due to the fragility of the resulting samples. The incorporation of sodium chloride has been previously observed to reduce the thermal stability of collagen using CD<sup>44</sup> as compared with calcium or potassium chloride at the same ionic strength. This decrease may result in the thinning of the collagen fibrils constituting the pore wall, therefore producing the larger pore sizes that were subsequently created post-washing and cross-linking. The washing step was not adopted in this work in order to understand the direct influence of sodium chloride on pore morphology. However, the variation of pore sizes upon removal of salts serves as a further avenue for investigation.

Although the slurries in this work were produced from microfibrillar collagen, the hydration process in acetic acid and hydrochloric acid were likely to have resulted in the destruction of the fibrillar packing,<sup>45</sup> further supported by the lack of the characteristic CD peak for fibrils in these results. Additionally, the flattening and reduction in intensity in the CD spectra with the use of water as a solvent suggest that aggregates of collagen were larger in water when compared with the acidic solvents, with a more pronounced effect upon dialysis. Since hydrochloric acid and acetic acid do not exhibit significant differences in the CD spectra, the pH may primarily drive the conformational stability and aggregate formation observed for the undialyzed samples.

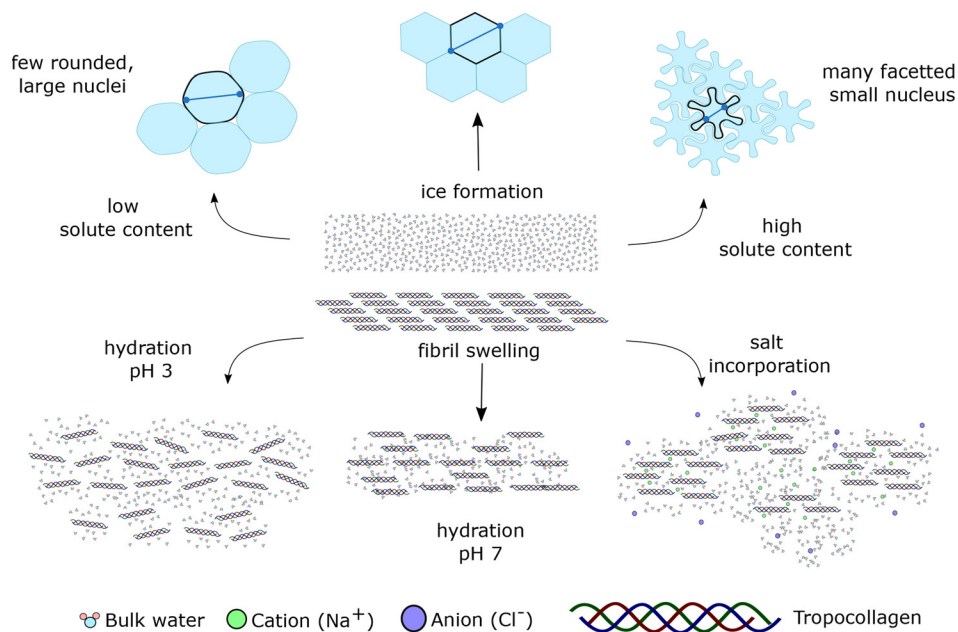
The driving mechanism for changes in the scaffold microstructure may therefore be explained through a combination ice growth which is affected by solute content, and collagen organization which is affected by collagen concentration, solute content, and pH as explored in Fig. 10.

## B. Implications of the random forest model

### 1. Interpreting feature importance graphs

Random forests are widely used as the machine learning model of choice where interpretability is desired in the analysis.<sup>46,47</sup> This interpretability is offered through the feature importance graphs that discriminate parameters based on how influential they are in determining the decision made by the trees in the random forest. While the ranking of a feature's influence in the random forest would usually directly correspond to the influence of the parameter physically, there are a few situations where the true or *physical* importance is not reflected by its *feature* importance. These include parameter spaces where several features are highly correlated.<sup>48,49</sup> For instance, the solute content is a sum of the total concentration of solutes, and is therefore an example of a variable that is derived from a linear combination of the other features (presence of water, presence of additives). The high importance of solute content for pore size measurements in Fig. 8 is accompanied by low importances for these correlated variables, since this single variable is sufficient to influence the prediction. Since the solute content is necessarily dependent on the presence of additives and specific solvents, a low feature importance does not suggest that these variables have no physical importance, but instead that the “supervariable” of solute content and pH, or other physically correlated parameters such as CD spectral intensity, may account for the combined effects. This is observed in the feature importances for models trained without these supervariables (supplementary material, Fig. 7). pH is a highly correlated extracted variable from the use of water as a solvent. In these models, the feature importances ranked above and below the mean importance are mostly preserved, except where the pH's importance is increased at the cost of the use of water as a solvent. While the interpretation of results is not impacted in this study, the use of recursive feature elimination may be helpful in datasets with highly correlated features.<sup>48,49</sup> Future investigations can also then be informed on the basis of these results, such as verifying the physical importance of supervariables such as “solute content” and pH across a larger range of solutes and solvents than those considered in this study.

Furthermore, the random forest model was trained on a single dataset combining data from the two case studies on the drying conditions and solute environments. The results from the model revealed that the relative effect of drying pressure and temperature are not as significant as varying other parameters, except in the case of the percolation diameter. It is worth noting that the case study on solvents were



**FIG. 10.** Diagrammatic representation of the hypothesized role of collagen and ice in scaffold microstructure. Ice morphology is suggested to be primarily modified through solute-driven changes at the ice–water interface: small nuclei with pronounced faceting can be observed with salt, whereas acetic acid produced a larger, rounded pore structure. The interaction of the solutes in the slurry with collagen can result in the reorganization of collagen molecules in the fibril. At low pH, this includes the lyotropic relaxation of the tropocollagen molecules within a fibril. On the other hand, at higher ionic strengths and salt content, the incorporation of salts can locally stabilize the fibrillar organization creating larger aggregates. The microstructure of the collagen struts is therefore driven by the interactions of collagen with its solute environment, whereas the morphology of the pore space is determined by ice growth.

all performed at a single drying condition, yet produced pore sizes, percolation diameters, and median interconnection diameters with significant variation. As a result, the drying parameters will therefore be assigned a low feature importance by the model, since they do not produce as significant a variation as the solvent-based parameters in the microstructure. Therefore, once again, the low *relative importance* does not imply a lack of *physical* influence, nor can be used to discount the qualitative effects observed on scaffold meso- or nanostructure with the SEM micrographs, which are not fully encompassed by measures of microstructural pore size and connectivity.

Finally, although this study considers the effect of solvent, solutes, and drying conditions, it must be noted that several other parameters have been shown to cause significant variation in the microstructural attributes considered here. These include but are not limited to nucleation temperatures,<sup>50</sup> cooling rates,<sup>51</sup> thermal conductivity, geometry, and size of the mold.<sup>33,52</sup> This study compares scaffolds which were fabricated and imaged from the same locations under conditions where all such variables are held constant, in order to ascertain the impact of the solute and drying parameters exclusively. Since varying these additional parameters will have an influence on the microstructural attributes, this will also necessarily impact the performance of the predictor. In this paper, our aim has been to leverage the ability of machine learning algorithms such as random forests to understand the relative physical influence of the specific parameters considered in the two case studies. However, with a large database of scaffold parameters, a more accurate, general-purpose predictor for microstructural attributes can be obtained. Regardless, the same principles of feature importance demonstrated here can be employed to extract parameters of interest.

## 2. Collagen concentration

Figure 8 revealed that collagen concentration was highly influential in predicting the pore size, percolation diameter, and median

interconnection diameter of a scaffold. The influence of solid concentration on the pore size previously investigated in the literature has shown that the trends have varied significantly based on the experimental setup. The range of pore sizes measured in this paper across a 1 mm<sup>3</sup> volume of interest (60–160 μm) are consistent with the slice-based estimates of pore sizes obtained at the top and base of a collagen scaffold by Pawelec *et al.*, and the values for non-dialyzed collagen scaffolds in acetic acid by Offeddu *et al.* (75–140 μm).<sup>34</sup> Offeddu *et al.* also noted an increase in pore wall and strut thickness which is consistent with the SEM images in Fig. 5. At low concentrations, as seen in Fig. 5, collagen struts possess a thin and wire-like appearance exclusively at 0.5 w/v%. Scaffolds at low concentration such as 0.25 w/v% have been found to have small pore sizes in collagen–GAG scaffolds due to their low stability and rigidity,<sup>7</sup> which is consistent with the observations of a weak pore wall that is unable to support its own weight, in this paper.

These suggest that while collagen concentration may have a significant impact on the pore size, these effects as reported in the literature are generally centered about the stability, rigidity, and loss of interchannel connectivity from thicker struts around the pore wall rather than the crystal growth during the freeze-drying process. Furthermore, the discrepancy in the effect of solid content observed in the literature may arise from the relative effect size of competing factors such as the thermal profile and collagen concentration in determining the final scaffold microstructure. Thus, the effect of collagen concentration is prominent when produced under the *same* thermal profiles as specified in this paper, whereas the same effect may be insignificant when scaffolds are produced under different thermal profiles.<sup>7,33,34</sup>

## 3. Influence of extracted parameters

Solvents and solutes have been investigated previously for their effects on collagen scaffold morphology in the literature. For instance,

prior work has established the production of thin, thread-like struts with water;<sup>6</sup> dense, smooth walled scaffolds with acetic acid;<sup>6,53</sup> fibrillar, “leaf-like,” and nanoporous walls with hydrochloric acid; and finally, small closed pores with fibrillar struts with the inclusion of sodium chloride.<sup>13</sup> The morphology of the collagen walls observed with scaffolds fabricated in this work is consistent with the literature, although the structural parameters cannot be directly compared in most cases due to minor differences in the experimental setup, such as the concentrations of solvent or thermal profiles used in fabricating the scaffolds. Since changes in the solvent result in several complex changes in the collagen environment, the phenomena driving changes in the scaffold morphologies are not clearly understood.

The random forest model in this paper was trained on a dataset with systematic modifications in solute and solvent contents, with the aim to identify the relative influence of various physical and chemical effects on the microstructure. For instance, the variables used in the model include general categorical variables such as “the presence of acetic acid,” as well as more specific, extracted variables such as the solute content. By comparing the relative influence of the general variables with the specific variables, the forces driving specific microstructural changes (such as original solute content of solvent, dialysis of collagen) may be more easily identified.

Circular dichroism is typically used to reveal the presence of the triple-helical tropocollagen molecules for collagenous materials, but also to study other conformational changes such as the denaturation of triple helices to random coils, or their packing into ordered fibrils.<sup>54–58</sup> Although CD can probe the impact of various chemical and thermal environments on collagen conformation, the scale of the technique (secondary protein structure) does not naturally lead to its use in understanding scaffold microstructure. In this work, we have demonstrated that relationships between characterization techniques which would not be typically compared in an experimental study can be included in the model, as seen with the inclusion of the maximum intensity of the CD peak for structural prediction. The CD peaks in Fig. 6 displayed the same shape with some flattening and broadening of the spectral peaks, suggesting the creation of larger aggregates in the slurry. This result is consistent with prior work where ionic strength was increased in collagen slurries resulting in a flattening and broadening of the CD spectra.<sup>59</sup> This is proposed to occur due to the unfavorable destabilizing long-range electrostatic interactions shielded by salt bridges formed with the triple-helical collagen backbone. The creation of a larger range of fiber associations due to the presence of long-range electrostatic interactions is consistent with the features seen in the scanning electron micrographs, where scaffold walls were more fibrillar with the incorporation of sodium chloride.

The loss in spectral intensity with an increase in pH and increase in ionic strength seen in Fig. 6 may be associated with the differential scattering of light through chromophore aggregation, particularly where the aggregates are larger than the wavelength of light used to acquire the CD spectra. Following this idea of aggregation-induced decreases in spectral intensity as set out by Bustamante *et al.*, the more extended or larger the aggregates are, the lower the intensity is expected to be.<sup>60</sup> Similarly, in spite of using the same concentration, the flattening and broadening of the CD spectra observed here may be related to the anisotropic distribution of the chromophores within the sample. For aggregates with a high local density of chromophores, a higher degree of spectral broadening would be expected. Therefore,

collagen aggregates formed at neutral pH and at greater ionic strengths can be concluded to be larger than in their acidic counterparts.

The high correlation observed between the CD peak intensity and values of connectivity such as the median interconnection and percolation diameter also introduces the opportunity to design new, targeted experiments between local collagen organization within the slurry and scaffold microstructure and morphology. The use of CD to characterize the organization of collagen within the slurry may also be a means to predict morphological features that were previously described qualitatively, such as the fibrillar or thin wire-like appearance of collagen struts.

#### 4. Comparing structural parameters

Although the trends observed with the median interconnection diameters were generally also observed with the percolation diameters, there were several notable exceptions. The median interconnection diameters displayed more drastic differences than the percolation diameters in scaffolds, particularly for scaffolds produced at the same solvent condition but different concentrations. One possible rationalization of these differences arises from the contributions to structural connectivity: collagen organization and ice crystal growth. There is evidence in the literature to suggest that the local solute content may influence the nucleation of ice<sup>61,62</sup> or modify thermodynamics of ice growth, which supports the role played by ionic content in the pore size, but the low impact of either additives or ionic content on the median interconnection diameter. This organization of collagen is likely to affect the pore fenestrations more than the pore space itself. This is further evidenced by the higher impact of dialysis for median interconnection diameter, and the higher impact of the CD spectral intensities on both measures of connectivity rather than pore size in the feature importance graphs (Fig. 8).

The 3D volume interconnectivity measured with shrink-wrap allows a probe to access the scaffold from all directions, unlike the percolation diameter which were measured in the direction of ice crystal growth. Although these scaffolds were polycrystalline and mostly isotropic, the pores will possess some crystallographic texture from the direction of growth from the bottom of the mold to the top. Thus, while the volume interconnectivity can access some of these larger pore fenestrations, the effect of smaller fenestrations caused by collagen organization may be limited on the percolation diameter when measured along the direction of ice growth. This rationalization is supported by the large influence of collagen concentration and drying pressure on percolation diameter where these parameters would influence the limits of ice growth and removal, unlike the median interconnection diameter, which is more heavily influenced by the pH, concentration-independent CD peak intensities and dialysis, which will impact the local organization of collagen in the slurry.

#### V. CONCLUSIONS

In this paper, we investigated the role of drying processes on the scaffold microstructure, and aimed to understand the role of solvent-solute interactions in collagen scaffolds. In this work, the drying processes were shown to introduce features on the collagen pore walls with some changes to the scaffold connectivity. In complex systems containing several solvents and solutes, globally measurable parameters such as a triple helical peak from circular dichroism, the total solute content, and the collagen concentration were found to be the most



influential parameters in determining the pore size or the pore connectivity in the scaffolds.

The results of this paper have shown first that the relative effect of distinct studies can be compared, and second that the underlying physicochemical phenomena may be elucidated by the use of random forest regressors in combination with experimental studies. This paper demonstrates the potential of using machine learning models in conjunction with experimental studies in biomaterials to obtain a deeper understanding of the underlying mechanisms during scaffold synthesis. The use of such techniques may also eventually help in the prediction of structural or biological properties of a scaffold, and aid to develop a toolkit to create application-specific tissue-engineering constructs.

## SUPPLEMENTARY MATERIAL

The [supplementary material](#) for this paper contains seven additional figures. Figure 1: height, amplitude error, and phase AFM channels of collagen films before and after air-drying, and of the characteristic D-banding in collagen. Figure 2: low magnification SEM micrographs of the longitudinal and transverse cross sections of scaffolds after drying at different chamber conditions. Figure 3: pore size distributions and volume interconnectivity of a scaffold produced under different drying conditions. Figure 4: pore size distributions of a scaffold produced under different solvent and solute conditions. Figure 5: volume interconnectivity of a scaffold produced under different solvent and solute conditions. Figure 6: environmental scanning electron micrographs of ice sublimation from solidified collagen slurries. Figure 7: feature importance graphs of models trained without supervariables.

## ACKNOWLEDGMENTS

This work was supported by the ERC Advanced Grant No. 320598 3D-E and the EPSRC Established Career Fellowship Grant No. EP/N019938/1 awarded to S.M.B. and R.E.C. M.N. additionally acknowledges the financial support provided by Gates Cambridge, Geistlich Pharma AG, and Emmanuel College to undertake this research. I.B. is funded by The Alan Turing Institute, under the EPSRC Grant No. EP/N510129/1. The authors also acknowledge Simon Griggs and Lambda Photometrics for technical support and use of the Phenom SEM.

## DATA AVAILABILITY

The data that support the findings of this study are openly available in University of Cambridge Apollo Open Data Repository at <https://doi.org/10.17863/CAM.71055>, Ref. 63. For more information on the database for collagen scaffolds being developed by the Cambridge Centre for Medical Materials, please visit <https://www.cmm.msm.cam.ac.uk/Resources/Wiki>.

## REFERENCES

- A. M. Ferreira, P. Gentile, V. Chiono, and G. Ciardelli, "Collagen for bone tissue regeneration," *Acta Biomater.* **8**, 3191–3200 (2012).
- J. Glowacki and S. Mizuno, "Collagen scaffolds for tissue engineering," *Biopolymers* **89**, 338–344 (2008).
- T. Miyata, T. Taira, and Y. Noishiki, "Collagen engineering for biomaterial use," *Clin. Mater.* **9**, 139–148 (1992).
- J.-D. Malcor, D. Bax, S. W. Hamaia, N. Davidenko, S. M. Best, R. E. Cameron, R. W. Farnedale, and D. Bihan, "The synthesis and coupling of photoreactive collagen-based peptides to restore integrin reactivity to an inert substrate, chemically-crosslinked collagen," *Biomaterials* **85**, 65–77 (2016).
- M. Nair, R. K. Johal, S. W. Hamaia, S. M. Best, and R. E. Cameron, "Tunable bioactivity and mechanics of collagen-based tissue engineering constructs: A comparison of EDC-NHS, genipin and TG2 crosslinkers," *Biomaterials* **254**, 120109 (2020).
- K. A. Faraj, T. H. Van Kuppevelt, and W. F. Daamen, "Construction of collagen scaffolds that mimic the three-dimensional architecture of specific tissues," *Tissue Eng.* **13**, 2387–2394 (2007).
- C. M. Tierney, M. G. Haugh, J. Liedl, F. Mulcahy, B. Hayes, and F. J. O'Brien, "The effects of collagen concentration and crosslink density on the biological, structural and mechanical properties of collagen-GAG scaffolds for bone tissue engineering," *J. Mech. Behav. Biomed. Mater.* **2**, 202–209 (2009).
- J. Ashworth, S. M. Best, and R. E. Cameron, "Quantitative architectural description of tissue engineering scaffolds," *Mater. Technol.* **29**, 281–295 (2014).
- J. Shepherd, S. Ghose, S. Kew, A. Moavenian, S. M. Best, and R. E. Cameron, "Effect of fiber crosslinking on collagen-fiber reinforced collagen–chondroitin-6-sulfate materials for regenerating load-bearing soft tissues," *J. Biomed. Mater. Res., Part A* **101**, 176–184 (2013).
- X. C. Tang and M. J. Pikal, "Design of freeze-drying processes for pharmaceuticals: Practical advice," *Pharm. Res.* **21**, 191–200 (2004).
- K. Pawelec, A. Husmann, S. M. Best, and R. E. Cameron, "A design protocol for tailoring ice-templated scaffold structure," *J. R. Soc. Interface* **11**, 20130958 (2014).
- J. C. Kasper and W. Friess, "The freezing step in lyophilization: Physicochemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals," *Eur. J. Pharm. Biopharm.* **78**, 248–263 (2011).
- K. Pawelec, A. Husmann, R. Wardale, S. Best, and R. Cameron, "Ionic solutes impact collagen scaffold bioactivity," *J. Mater. Sci.* **26**, 91 (2015).
- J. Ratanavaraporn, S. Kanokpanont, Y. Tabata, and S. Damrongsakul, "Effects of acid type on physical and biological properties of collagen scaffolds," *J. Biomater. Sci., Polym. Ed.* **19**, 945–952 (2008).
- Y. Li, A. Asadi, M. R. Monroe, and E. P. Douglas, "pH effects on collagen fibrillogenesis *in vitro*: Electrostatic interactions and phosphate binding," *Mater. Sci. Eng., C* **29**, 1643–1649 (2009).
- W.-H. Liang, B. L. Kienitz, K. J. Penick, J. F. Welter, T. A. Zawodzinski, and H. Baskaran, "Concentrated collagen-chondroitin sulfate scaffolds for tissue engineering applications," *J. Biomed. Mater. Res.* **94**, 1050–1060 (2010).
- C. N. Grover, R. E. Cameron, and S. M. Best, "Investigating the morphological, mechanical and degradation properties of scaffolds comprising collagen, gelatin and elastin for use in soft tissue engineering," *J. Mech. Behav. Biomed. Mater.* **10**, 62–74 (2012).
- D. Clearfield and M. Wei, "Investigation of structural collapse in unidirectionally freeze cast collagen scaffolds," *J. Mater. Sci.* **27**, 15 (2016).
- M. Nair, A. Husmann, R. E. Cameron, and S. M. Best, "*In situ* ESEM imaging of the vapor-pressure-dependent sublimation-induced morphology of ice," *Phys. Rev. Mater.* **2**, 040401 (2018).
- N. Magee, A. Miller, M. Amaral, and A. Cumiskey, "Mesoscopic surface roughness of ice crystals pervasive across a wide range of ice crystal conditions," *Atmos. Chem. Phys.* **14**, 12357–12371 (2014).
- L. Auret and C. Aldrich, "Interpretation of nonlinear relationships between process variables by use of random forests," *Miner. Eng.* **35**, 27–42 (2012).
- E. Raczko and B. Zagajewski, "Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral apex images," *Eur. J. Remote Sens.* **50**, 144–154 (2017).
- M. Liu, M. Wang, J. Wang, and D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar," *Sens. Actuators, B* **177**, 970–980 (2013).
- M. S. Briec, C. D. Waters, D. P. Drinan, and K. A. Naish, "A practical introduction to random forest for genetic association studies in ecology and evolution," *Mol. Ecol. Resour.* **18**, 755–766 (2018).



- <sup>25</sup>J. Stephan, O. Stegle, and A. Beyer, "A random forest approach to capture genetic effects in the presence of population structure," *Nat. Commun.* **6**, 7432 (2015).
- <sup>26</sup>B. A. Goldstein, E. C. Polley, and F. B. Briggs, "Random forests for genetic association studies," *Stat. Appl. Genet. Mol. Biol.* **10**, 32 (2011).
- <sup>27</sup>I. Nedjar, M. El Habib Daho, N. Settouti, S. Mahmoudi, and M. A. Chikh, "Random forest based classification of medical x-ray images using a genetic algorithm for feature selection," *J. Mech. Med. Biol.* **15**, 1540025 (2015).
- <sup>28</sup>M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Med. Inf. Decision Making* **11**, 51 (2011).
- <sup>29</sup>A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Medical Image Analysis* **35**, 475–488 (2017).
- <sup>30</sup>P. F. Smith, S. Ganesh, and P. Liu, "A comparison of random forest regression and multiple linear regression for prediction in neuroscience," *J. Neurosci. Methods* **220**, 85–91 (2013).
- <sup>31</sup>B. Singh, P. Sihag, and K. Singh, "Modelling of impact of water quality on infiltration rate of soil by random forest regression," *Modeling Earth Syst. Environ.* **3**, 999–1004 (2017).
- <sup>32</sup>O. Mutanga, E. Adam, and M. A. Cho, "High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm," *Int. J. Appl. Earth Obs. Geoinf.* **18**, 399–406 (2012).
- <sup>33</sup>K. Pawelec, A. Husmann, S. M. Best, and R. E. Cameron, "Understanding anisotropy and architecture in ice-templated biopolymer scaffolds," *Mater. Sci. Eng., C* **37**, 141–147 (2014).
- <sup>34</sup>G. Offeddu, J. Ashworth, R. Cameron, and M. Oyen, "Structural determinants of hydration, mechanics and fluid flow in freeze-dried collagen scaffolds," *Acta Biomater.* **41**, 193–203 (2016).
- <sup>35</sup>J. C. Ashworth, "Characterising the structural criteria for cell invasion in collagen scaffolds," Ph.D. thesis (University of Cambridge, 2015).
- <sup>36</sup>M. Nair, J. H. Shepherd, S. M. Best, and R. E. Cameron, "MicroCT analysis of connectivity in porous structures: Optimizing data acquisition and analytical methods in the context of tissue engineering," *J. R. Soc. Interface* **17**, 20190833 (2020).
- <sup>37</sup>G. Fostad, B. Hafell, A. Førde, R. Dittmann, R. Sabetrasekh, J. Will, J. Ellingsen, S. Lyngstadaas, and H. Haugen, "Loadable TiO<sub>2</sub> scaffolds—A correlation study between processing parameters, micro CT analysis and mechanical strength," *J. Eur. Ceram. Soc.* **29**, 2773–2781 (2009).
- <sup>38</sup>M. Venturoni, T. Gutsmann, G. E. Fantner, J. H. Kindt, and P. K. Hansma, "Investigations into the polymorphism of rat tail tendon fibrils using atomic force microscopy," *Biochem. Biophys. Res. Commun.* **303**, 508–513 (2003).
- <sup>39</sup>J. Bella, B. Brodsky, and H. M. Berman, "Hydration structure of a collagen peptide," *Structure* **3**, 893–906 (1995).
- <sup>40</sup>S. Leikin, V. Parsegian, W.-H. Yang, and G. Walrafen, "Raman spectral evidence for hydration forces between collagen triple helices," *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11312–11317 (1997).
- <sup>41</sup>S. A. Deshmukh, L. A. Solomon, G. Kamath, H. C. Fry, and S. K. Sankaranarayanan, "Water ordering controls the dynamic equilibrium of micelle-fibre formation in self-assembly of peptide amphiphiles," *Nat. Commun.* **7**, 12367 (2016).
- <sup>42</sup>S. Tchessalov, D. Latshaw II, S. Nulu, M. Bentley, T. Tharp, S. Ewan, and X. Chen, "Application of first principles primary drying model to lyophilization process design and transfer: Case studies from the industry," *J. Pharm. Sci.* **110**, 968–981 (2021).
- <sup>43</sup>A. Wexler *et al.*, "Vapor pressure formulation for ice," *J. Res. Natl. Bur. Stand., Sect. A* **81**, 5–20 (1977).
- <sup>44</sup>Y. Li and E. P. Douglas, "Effects of various salts on structural polymorphism of reconstituted type I collagen fibrils," *Colloids Surf., B* **112**, 42–50 (2013).
- <sup>45</sup>A. Ripamonti, N. Roveri, D. Braga, D. Hulmes, A. Miller, and P. Timmins, "Effects of pH and ionic strength on the structure of collagen fibrils," *Biopolymers* **19**, 965–975 (1980).
- <sup>46</sup>L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).
- <sup>47</sup>S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, and L. Hung, "Random forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships," *npj Comput. Mater.* **6**, 1–11 (2020).
- <sup>48</sup>B. F. Darst, K. C. Malecki, and C. D. Engelman, "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data," *BMC Genet.* **19**, 65 (2018).
- <sup>49</sup>B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," *Stat. Comput.* **27**, 659–678 (2017).
- <sup>50</sup>K. Pawelec, A. Husmann, S. Best, and R. Cameron, "Altering crystal growth and annealing in ice-templated scaffolds," *J. Mater. Sci.* **50**, 7537–7543 (2015).
- <sup>51</sup>F. J. O'Brien, B. A. Harley, I. V. Yannas, and L. Gibson, "Influence of freezing rate on pore structure in freeze-dried collagen-gag scaffolds," *Biomaterials* **25**, 1077–1086 (2004).
- <sup>52</sup>N. Davidenko, T. Gibb, C. Schuster, S. Best, J. Campbell, C. Watson, and R. Cameron, "Biomimetic collagen scaffolds with anisotropic pore architecture," *Acta Biomater.* **8**, 667–676 (2012).
- <sup>53</sup>J. C. Ashworth, M. Mehr, P. G. Buxton, S. M. Best, and R. E. Cameron, "Cell invasion in collagen scaffold architectures characterized by percolation theory," *Adv. Healthcare Mater.* **4**, 1317–1321 (2015).
- <sup>54</sup>K. E. Drzewiecki, D. R. Grisham, A. S. Parmar, V. Nanda, and D. I. Shreiber, "Circular dichroism spectroscopy of collagen fibrillogenesis: A new use for an old technique," *Biophys. J.* **111**, 2377–2386 (2016).
- <sup>55</sup>A. Wiecezorek, N. Rezaei, C. K. Chan, C. Xu, P. Panwar, D. Brömme, N. R. Forde *et al.*, "Development and characterization of a eukaryotic expression system for human type II procollagen," *BMC Biotechnol.* **15**, 112 (2015).
- <sup>56</sup>R. Wallis, J. M. Shaw, J. Uitdehaag, C.-B. Chen, D. Torgersen, and K. Drickamer, "Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein indirect effects of naturally occurring mutations on protease binding and activation," *J. Biol. Chem.* **279**, 14065–14073 (2004).
- <sup>57</sup>S. Perumal, K. Dubey, R. Badhwar, K. J. George, R. K. Sharma, G. Bagler, B. Madhan, and K. Kar, "Capsaicin inhibits collagen fibril formation and increases the stability of collagen fibers," *Eur. Biophys. J.* **44**, 69–76 (2015).
- <sup>58</sup>R. Usha and T. Ramasami, "Role of solvents in stability of collagen," *J. Therm. Anal. Calorim.* **93**, 541–545 (2008).
- <sup>59</sup>U. Freudenberg, S. H. Behrens, P. B. Welzel, M. Müller, M. Grimmer, K. Salchert, T. Taeger, K. Schmidt, W. Pompe, and C. Werner, "Electrostatic interactions modulate the conformation of collagen I," *Biophys. J.* **92**, 2108–2119 (2007).
- <sup>60</sup>C. Bustamante, I. Tinoco, and M. F. Maestre, "Circular differential scattering can be an important part of the circular dichroism of macromolecules," *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3568–3572 (1983).
- <sup>61</sup>P. Wilson, A. Heneghan, and A. Haymet, "Ice nucleation in nature: Supercooling point (SCP) measurements and the role of heterogeneous nucleation," *Cryobiology* **46**, 88–98 (2003).
- <sup>62</sup>G. Blond, "Velocity of linear crystallization of ice in macromolecular systems," *Cryobiology* **25**, 61–66 (1988).
- <sup>63</sup>M. Nair, I. Bica, S. Best, and R. Cameron, Research data supporting "Feature importance in multi-dimensional tissue-engineering datasets: Random forest assisted optimisation of experimental variables for collagen scaffolds," *Apollo Dataset* (2021).