This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Karakoç, Alp; Keleş, Özgür

A predictive failure framework for brittle porous materials via machine learning and geometric matching methods

Published in:
Journal of Materials Science

DOI:
[10.1007/s10853-019-04339-1](https://doi.org/10.1007/s10853-019-04339-1)

Published: 01/04/2020

Document Version
Peer reviewed version

Published under the following license:
CC BY

Please cite the original version:
Karakoç, A., & Kele, Ö. (2020). A predictive failure framework for brittle porous materials via machine learning and geometric matching methods. *Journal of Materials Science*, 55(11), 4734-4747.
<https://doi.org/10.1007/s10853-019-04339-1>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

A PREDICTIVE FAILURE FRAMEWORK FOR BRITTLE POROUS MATERIALS VIA MACHINE LEARNING AND GEOMETRIC MATCHING METHODS

Alp Karakoç¹, Özgür Keleş^{2,*}

¹Aalto University, Department of Bioproducts and Biosystems, Vuorimiehentie 1, Espoo, Finland

²San Jose State University, Chemical and Materials Engineering Department, One Washington Square, San Jose, CA 95192, USA

ABSTRACT

Brittle porous materials are used in many applications, such as molten metal filter, battery, fuel cell, catalyst, membrane, and insulator. The porous structure of these materials causes variations in their fracture strength that is known as the mechanical reliability problem. Despite the importance of brittle porous materials, the origin of the strength variations is still unclear. The current study presents a machine learning approach to characterize the stochastic fracture of porous ceramics and glasses. A combined finite element modeling and fracture mechanics approach was used to generate a unique empirical data set consisting of normalized stress intensity factors (nSIFs, $K_I/\sigma_\infty = Y\sqrt{\pi a}$) that define fracture strength of brittle systems under uniaxial and biaxial tensile loading. These empirical data sets were used to generate prediction functions and validate their accuracy. Monte-Carlo simulations with two machine learning algorithms, random forests (RF) and artificial neural networks (ANN), were used to simultaneously determine the optimum percentages for the training and test data set split and the prediction function validation. The constraint was taken to be the mean absolute percentage error (MAPE) during the process. In the

implementation step, new porous media with uniformly distributed pores were created and the prediction functions were used to obtain nSIFs and characterize the media. As a novelty of this approach, which ensures the predictive characterization of the generated media, a geometric matching method by means of the Euclidean bipartite matching between the empirical and the generated media was presented and the nSIFs were compared by means of MAPE. As a result of the study, MAPE ranges are 3.4-17.93% (uniaxial load) and 2.83-19.42% (biaxial load) for RF, 3.79-17.43% and 3.39-21.43 for ANN at the validation step; 3.54-18.20% (uniaxial load) and 3.06-21.60% (biaxial load) for RF, 3.57-18.26% and 3.43-21.76% for ANN at the implementation step. Proposed approach can be thus used as a predictive characterization tool, especially for the analysis and Weibull statistics of porous media subjected to brittle failure.

Keywords: Porous materials, ceramics, glasses, stress intensity factor, machine learning, Weibull statistics.

*Corresponding author(s): Özgür Keleş, ozgur.keles@sjsu.edu.

1. INTRODUCTION

Porous materials are ubiquitous in nature. Porosity functionalizes biological materials by enabling mass transportation and storage. For example, porous structure of wood, crab exoskeleton, and diatoms allows nutrition transportation; pores in bone allow storage of cells and reduce weight [1, 2]. Industrial porous materials are also used in many applications, including battery, solid oxide fuel cell, filter, insulation, and bio-scaffolds [1, 3–6]. However, porosity reduces mechanical

performance and reliability especially in brittle porous materials (BPMs), e.g., porous ceramics and glasses. Prediction of brittle failure in porous systems is challenging due to complex stress interactions between cracks, pores, and grains [7–13]. Various analytical and numerical approaches have been used to quantify the fracture behaviour of BPMs, but the complex stress interactions between pores and crack size distributions limit these investigations to single/few pore(s) and small simulation domains with limited number of pores [7, 9, 12, 14, 15]. Spatial distribution of pores can be obtained via X-ray computed tomography [16], but accurate description of initial defect size and position distributions is unobtainable. Moreover, reliability analyses—stochastic investigation of strength values—require sample size (N) >30 [17]. This large N restricts (a) experimental investigations due to sample preparation and testing costs and (b) deterministic numerical approaches due to computational cost. In this context, machine learning (ML) approaches arise as an extension of numerical investigations [18–21]. However, predictive capabilities of ML for stochastic brittle fracture problems are unknown.

Brittle fracture is the main problem of ceramics and glasses, which is amplified for porous systems. Therefore, in our previous work, we combined fracture mechanics (FM) and finite element method (FEM) to better understand stochastic fracture in porous alumina and glasses [8]. Isotropic systems such as glass eliminate elastic mismatch stresses between grains, which simplifies the FEM implementation by avoiding elastically anisotropic grains. This approach was developed to observe the effects of the most significant stress concentrators, which were the pores and cracks. The thermal expansion anisotropy and elastic stiffness anisotropy effects are not as severe as the pore-crack combination, but can be implemented in the FEM scheme in the future studies. The combined FM-FEM approach allowed us to predict the changes in fracture strength (σ_f) with

increasing porosity in porous glasses and porous alumina. The results showed that the σ_f slightly decreased with increasing number of pores for 1-2 vol. % porosity. On the other hand, σ_f decreased significantly up to 10 vol. % porosity due to strong stress interactions between the pores. Beyond 10 vol. % porosity and less than ~ 30 vol. %, the σ_f did not change as strongly as the initial decrease for up to 10 vol. %, because the stress interactions were saturated, i.e., pore-to-pore stress interactions did not increase significantly because the structure already includes strongly interacting pore clusters for porosity > 10 vol. %. We proposed a new equation to capture σ_f dependence on porosity based on our numerical observation

$$\sigma_f = \sigma_k - c \exp(-d/\psi) \quad (1)$$

Here, σ_k is related to the strength of pore-free material; c and d are empirical constants; and ψ represents the fraction of porosity.

The second problem of ceramics and glasses is the variation in their σ_f . This second problem is known as the mechanical reliability problem that limits the use of brittle materials in critical applications. The variation in σ_f has been quantified using extreme value statistics, specifically Weibull statistics. The two-parameter Weibull distribution provides a link between the probability of failure F and applied stress and volume of the specimen, which follows

$$F(\sigma, V) = 1 - \exp\left(-V/V_0 (\sigma/\sigma_0)^m\right) \quad (2)$$

where the V_0 and σ_0 are the normalizing constants. The “ m ” is Weibull modulus, which describes the degree of variation in σ_f . The V can be assumed to equal to V_0 for samples of the same volume, making the σ_0 and m as the only parameters to be fitted to any data. Common values of m for ceramics ranges from 5 to 20; whereas, metals can have $m=90-100$. The lower the m , the higher the variation in σ_f .

Our FM-FEM approach was performed on unique structures containing pores, which was the first approach to include both the effects of random porous structure and crack sizes on the variation in σ_f . We predicted σ_f of at least 40 structures for fixed vol. % of porosity, but including random spatial distribution of pores. The use of 40 structures allows estimation of Weibull distribution parameters with a low error. The predicted Weibull modulus depended on the vol. % of porosity and initial crack size distribution. We observed a decrease in m value with increasing porosity even for a change in porosity from 1 to 2 vol. %. This result shows that the decrease in reliability depends on pore-to-pore stress interactions. The experimental m value for BPMs are below 10 for structures with porosity >10 vol. %. We did not observe statistically significant change in m for porosity >10 vol. %, which is consistent with the experimental observations. We also run simulations to investigate the effect of crack size distributions in porous glasses, as BPMs contain a distribution of crack sizes and pore sizes. These simulations contained pores of fixed size, but randomly distributed in the structure. Four types of crack size distributions were used: Gaussian, Pareto, and two different experimental distributions. We revealed that the effect of crack size distribution on the stochastic fracture decreases for larger pores, because the crack size to pore diameter ratio decreases and the crack tip stresses become similar. As a result, pore-to-pore stress interactions dominate fracture behavior for crack size to pore diameter ratio smaller than 0.1.

In addition to our FM-FEM approach, porous media was also simulated using a lattice spring model and Brazilian disk tests were virtually performed on porous media by Zhiwei et al. [12]. This lattice model of BPM fracture also showed decreasing Weibull modulus with increasing porosity up to 10 vol. % and showed insensitivity of Weibull modulus to porosity between 20 and

50 vol. % [12]. However, the reported Weibull moduli were higher than ~60 for 12 sets of porous media with different porosity and pore size, except one set with the largest pore size (pore sizes were relative in this study without any absolute value). These Weibull moduli are higher than reported values of less than 10 for porosity > 10 vol. %, indicating the complexity of the problem[8]. Moreover, stochastic FEM models were used to predict fracture behavior of porous ceramic matrix composites, where the focus was on the largest pores with circumferential cracks [22]. In this study, pores were not explicitly simulated, but stochastic material properties were distributed in a rectangular 3-point bend specimen. The distribution of fracture strength throughout the specimen generated a Weibull modulus of 12 to 16 that was representative of the experimental system, but the effect of pores size was not discussed. Other Monte Carlo-fracture mechanics simulation approaches describing fracture behavior of brittle porous systems were also reported without pore-to-pore stress interactions [23].

Although the FM-FEM combined approach was successful in predicting the relationships between fracture behavior and porosity in sets of 40 explicit microstructures, computational cost of FEM was prohibitive for an investigation of higher number of larger microstructures with larger number of pores. Reliable and safe use of BPMs require better understanding of their stochastic fracture; hence, large datasets. In this context, informatics can provide computationally cheaper approaches for the analysis of stochastic fracture, using limited number of FEM simulations as the training dataset. For example, Gu et al. used FEM to observe crack tip stress fields and used ML approaches to predict local microstructure near crack tip to reduce near-tip stresses and toughen the composite material [18]. Moore et al. applied decision trees (DT), random forests (RF) and artificial neural network (ANN) algorithms for the prediction of dynamic fracture in terms of the fracture path,

where they predicted the fracture path with ~85% accuracy [24]. Mangal and Holm also performed a study, where they used the RF model to predict stress hot spots in grains of metallic materials [25]. These stress hot spots are rare events in materials under stress, as they occur in regions representing less than 10 vol. % of the metal. On the other hand, from a fracture mechanics point of view, fracture in brittle materials depend on the maximum normalized stress intensity factor nSIF ($Y\sqrt{\pi a} = K_I/\sigma_\infty$), assuming fracture toughness is independent of crack size. We used the maximum nSIFs in our previous study to build ML models [21]. The RF model resulted in prediction accuracy of ~85% for simulation results both under uniaxial and biaxial tensile loads. However, this work was based on the maximum nSIF in a microstructure, not in a single pore.

Therefore, in this study, we used the nSIF data, which included the maximum nSIF values for each pore in a microstructure and pore position, a total of 14880 pores from 680 unique microstructures. This FM-FEM generated data was used to build ML predictors using RF and ANN. We used the predicted nSIFs to obtain distribution of σ_f and related Weibull statistics for porous structures. As a novelty of the present study, we conducted the predictive failure characterization and error quantification for the generated microstructures in comparison with the previously simulated microstructures, which was depicted in Figure 1. The error quantification was based on the mean absolute percentage error (MAPE), for which the empirical results were taken from the previously conducted FM-FEM studies and the predicted results were obtained from the generated prediction functions implemented on the generated microstructures. Here, the relationship between the empirical and predicted results were established on the geometric likelihood, i.e. the Euclidean bipartite matching of the pore centroid proximities of the previously investigated and generated microstructures. This enabled us to compare the microstructures with a known metric. This metric

can be extended to other types of microstructural features, such as pore shape/orientation, crack size/orientation, and other geometrically representable features.

Figure 1: The flowchart showing the validation and implementation steps followed in the present study.

2. METHODOLOGY

FEM simulations on porous two-dimensional microstructures were previously performed using OOF2 program [26]. Details of FEM simulations were reported in [7, 8]. In these simulations, porous material was assumed to be a brittle isotropic material with elastic modulus $E=100$ GPa and Poisson's ratio $\nu=0.3$. Uniaxial and biaxial tensile loads were investigated to understand the effects of uniformly distributed pores on the stochastic fracture of the BPMs. Uniaxial tensile stress was applied by keeping the bottom boundary fixed and displacing the top boundary by $0.2 \mu\text{m}$ (0.01 % displacement); the left and right boundaries were mechanically free. Equibiaxial stress was applied by keeping the bottom and left boundaries fixed and displacing the top and right boundaries by $0.2 \mu\text{m}$. The linear systems of equations were solved using a conjugate gradient method with an incomplete lower-upper (LU)-factorization as the preconditioner. The pore size kept constant in each microstructure of size $2000 \mu\text{m} \times 2000 \mu\text{m}$. The pore centroids were randomly generated within this microstructure. The minimum distance between non-overlapping pores was $15 \mu\text{m}$, where the pores were generated $200 \mu\text{m}$ away from the edges to reduce the stress interactions with free surfaces. Each simulation had $\sim 50,000$ quadratic triangular elements. Adaptive meshing scheme in OOF2 was used to increase the mesh density near pores, which corresponded to a segment size of $\sim 6 \mu\text{m}$ along the perimeter of pores. The simulated stress distributions were used to calculate geometric factor Y for each pore centroid positions (x_i, y_i) [8]. The number and radii of pores were taken as $n_{\text{pores}} = \{5, 10, 20, 30, 40, 50\}$ and $R = \{40, 60, 80\} \mu\text{m}$, respectively. Sets of 40 unique microstructures were simulated with total number of 680 unique microstructures.

Cracks were assumed to be present perpendicular to the pores. A normal distribution (ND) of crack sizes was assumed with a mean = 8.8 μm and standard deviation = 2.8 μm . This ND was bounded at zero and 15 μm that avoided negative crack sizes and pore-connecting cracks. The grain size distribution is taken to be equal to the crack size distribution. The use of grain size as initial crack size was also suggested by Evans and Langdon [27]. These cracks were randomly generated along the perimeter of the pore. The crack length was used as the distance to the next crack.

Once the crack size and orientation is known, the stress distributions perpendicular to the crack face was used to calculate the geometric factor according to the method given by Shah [28]. Linear elastic fracture mechanics establishes the basis for the relationship, for which σ_f distribution is equal to the $K_{Ic}/Y\sqrt{\pi a}$ distribution. As a result, stochastic fracture behavior depends on the distribution of $Y\sqrt{\pi a}$, which we called as the normalized stress intensity factor (nSIF). In the present study, maximum nSIF for each pore were used as a part of the training/test data, a total of 14880 pores for 680 microstructures as it can be deduced from Table 1. We performed 1360 FEM simulations under uniaxial loading and biaxial loading using 680 unique microstructures.

Table 1. Design of experiments with 40 repetitions in each set.

Set	Number of holes	Radius (μm)	Set	Number of holes	Radius (μm)
1	5	40	10	20	80
2	5	60	11	30	40
3	5	80	12	30	60
4	7	40	13	30	80
5	10	40	14	40	40
6	10	60	15	40	60
7	10	80	16	40	80
8	20	40	17	50	40
9	20	60			

2.1. Prediction and validation

We used two machine learning algorithms—RF and ANN—to generate prediction functions. RF, or more specifically Breiman-Cutler ensembles of decision trees, was initially introduced by Breiman based on decision trees for classification and regression analysis [29]. A decision combination model P is formed from a sequence of models f , or known as decision trees, dependent on input data x_i where $i = \{1, \dots, n\}$ and can be broadly represented as

$$P = \sum_j f_j(x_1, \dots, x_n) . \quad (3)$$

Initially, RF produces bootstrapped data and assigns a decision tree f for each data subset. These decision trees (i.e. single classifiers) lay the ground for the final prediction by means of forming the forest (i.e. group of classifiers) through bootstrap aggregation (bagging) samples, which means fitting trees to the subsampled data and averaging all the trees (here, used for the regression purpose). During the process, each tree is independently formed through individual data sampling, which makes RF an efficient approach for numerical data handling and capturing linear and non-linear interaction between the input and output data [30]. In the present study, 100 trees were used based on the default value of RF method in Mathematica “Predict” function [31].

ANN is a systematic approach consisting of interconnected neuron-like structures in a multi layered system with an input layer, an output layer and hidden layers facilitating the association of input and output [19]. Its architecture with connected supervised feed-forward neural network containing minimum one hidden layer is called multilayer perceptron (MLP). Additionally, the gradient of the error function can be calculated by back-propagation algorithm (BPA) by propagating towards the hidden layer from the outermost layer [30]. The mathematical expression

of how MLP maps the input value of i^{th} independent value X_i and prediction value output P can be presented as [32, 33]

$$P = g \left(\theta + \sum_{j=1}^m v_j \left[\sum_{i=1}^n h(w_{ij} X_i + \beta_j) \right] \right) . \quad (4)$$

Here, n refers to the number of input parameters, which is equal to the number of samples and m is the number of hidden neurons whereas g and h are the activation functions of output and hidden neurons, respectively. In the present study, activation functions are chosen to be hyperbolic tangent function ($\tanh$) while number of hidden neurons are 10 with two hidden layers following the default values implemented by Mathworks [34]. The weights of the connection values, i.e. w_{ij} between the i^{th} input neuron and j^{th} hidden neuron and v_j between j^{th} hidden neuron and output neuron, are inserted to increase the precision of P and β_j , which is the bias of the j^{th} hidden neuron. In addition, an offsetting factor θ for a better fit is also available in Equation (4) [20].

In order to train these two models, the empirical data were chosen to be the previously conducted nSIFs from the FM-FEM simulations on the porous ceramics subjected to uniaxial and biaxial loading cases and used as the input data (x_i) [8, 35]. In order to simultaneously determine the optimum percentages for the training and test data split and the prediction function validation, Monte-Carlo simulations were used with RF and ANN that were readily available in Mathematica [31]. The mean absolute percentage error MAPE for each individual specimen was taken to be the decisive parameter in these simulations, for which is expressed as

$$\text{MAPE} = \frac{100}{n_{\text{pores}}} \sum_{i=1}^{n_{\text{pores}}} \left| \frac{E_i - P_i}{E_i} \right|. \quad (5)$$

Here, n_{pores} refers to the number of pores in the porous medium, E is the empirical results and P is the predicted output for each specimen. Following the Monte-Carlo simulations of several

repetitions—e.g., 10 repetitions in the present study—, the minimum MAPE value determined and the corresponding training set percentage was used for the split. An illustration of this process can be found in Figure 2.

Figure 2. Determination of training and test data split based on MAPE. The dots refer to the mean and filled background represents the standard deviation. The highlighted rectangle exemplifies the selected training set percentage of 82 % for the split based on the minimum MAPE value.

2.2. Implementation

In the implementation step, the validated prediction functions were used to characterize the generated media with randomly distributed uniform size pores. The pore centroids were randomly generated for both validation and implementation steps. ML models were not trained to predict the pore positions. In order to ensure the characterization, pore centroid proximities indicating the

geometric likelihood and the corresponding nSIFs between the empirical data and the predictive results of the generated media were analysed by means of the Euclidean bipartite matching as illustrated in Figure 3. Thus, $n_{\text{pores}} \times n_{\text{pores}}$ distance matrix was generated based on the Euclidean distances $\mathbf{d}$ between the pore centroid coordinates of the empirical data, denoted with p , and the generated pore coordinates, denoted with q , as

$$\begin{bmatrix} \mathbf{d}(p_1, q_1) & \mathbf{d}(p_1, q_2) & \dots & \mathbf{d}(p_1, q_{n_{\text{pores}}}) \\ \mathbf{d}(p_2, q_1) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \mathbf{d}(p_{n_{\text{pores}}}, q_1) & \dots & \dots & \mathbf{d}(p_{n_{\text{pores}}}, q_{n_{\text{pores}}}) \end{bmatrix}. \quad (6)$$

Thereafter, the minimal cost flow problem was set for determination of the optimal permutation based on the total Euclidean distance T of the centroids as [36]

$$T = \min_{\Pi} \sum \mathbf{d}(p, q), \quad (7)$$

where Π refers to the permutations for one-to-one correspondence. In order to solve the problem given in Eq. (7), Kuhn-Munkres algorithm was conducted to determine the non-overlapping path resulting in the minimum total Euclidean distance as seen in Figure 3 [37].

Figure 3. Euclidean bipartite matching: a) empirical domain, b) generated domain, and c) pore matching based on the minimization of total centroid distances.

3. RESULTS AND DISCUSSIONS

3.1. Error quantification

We generated the prediction functions based on empirical nSIF data sets, where the specimens were subjected to uniaxial and biaxial loading. Monte-Carlo simulations with the decisive parameter as MAPE and 10 repetitions resulted in 70%-85% of the data sets to be used as the training data listed in detail in the Appendix Tables A.1 and A.2. The prediction functions, thereafter, were used to predict nSIFs of the generated specimens and compared with the empirical data by using the Euclidean bipartite matching method as depicted for different cases in Figure 4. As depicted in Figure 5 and listed in Tables A.1 and A.2, MAPE ranges are 3.4-17.93% (uniaxial load) and 2.83-19.42% (biaxial load) for RF, 3.79-17.43% and 3.39-21.43 for ANN at the validation step; 3.54-18.20% (uniaxial load) and 3.06-21.60% (biaxial load) for RF, 3.57-18.26% and 3.43-21.76% for ANN at the implementation step. The results show that MAPE increases faster under biaxial loading as the stress interactions increase faster than uniaxial loading, where the amplifying stress interactions are limited to plane perpendicular to tensile stress. Biaxial loading allows amplifying stress interactions in all directions; hence, the increased variation in strength under biaxial tension.

Figure 4. Implementation step with Euclidean bipartite matching and comparison: a) specimen with 5 pores with radius of $40 \mu\text{m}$ under uniaxial loading, b) specimen with 5 pores with radius of $40 \mu\text{m}$ under biaxial loading, c) specimen with 50 pores with radius of $40 \mu\text{m}$ under uniaxial

loading, and d) specimen with 50 pores with radius of 40 μm under biaxial loading. Here, nSIF refers to the normalized stress intensity factor. Each microstructure represents a 2000x2000 μm domain—same as Figure 3.

Figure 5. The mean absolute percentage error (MAPE) for each set in the course of the validation steps: a) uniaxial loading and b) biaxial loading (data from RF implementation). Here, $\triangle$ refers to $r = 40 \mu\text{m}$, $\circ$ refers to $r = 60 \mu\text{m}$, and $+$ refers to $r = 80 \mu\text{m}$.

3.2. Failure probability predictions

The Weibull plots showing the distribution of σ_f calculated from the FM-FEM simulations and predicted from the RF method are given in Figure 6 (c-d) for uniaxial loading and in Figure 7 (c-d) for biaxial loading. The results from ANN predictions are given in Figure 6 and 7 (e-f). The RF and ANN models captured the correct decrease in σ_f with increasing porosity for all the 42 sets of data under uniaxial and biaxial loading. Both of the RF and ANN models had limited prediction of the lower and upper tail deviations in σ_f distributions. These tail deviations can be seen as rare

fracture instances that are even more challenging to be predicted by the ML models, as the rare events require larger sets of data for accurate prediction.

The Weibull moduli for the validation data were the closest to the empirical Weibull moduli compared to the moduli estimated from the implementation data. The m values for the RF validation model were as low as 20, which are rarely observed in experiments. Chao and Shetty reported a Weibull modulus of 22 under biaxial loading and 24 under 4-point bending for alumina containing ~ 6 vol. % porosity [38]. All the RF and ANN predictions of σ_f generated Weibull moduli were larger than the FM-FEM simulated σ_f data. More accurate predictions were obtained for the structures containing the lowest number of pores=5. The changes in fracture strength variations with increasing porosity are visualized in Weibull plots (Figures 6 and 7). In these figures, the slopes of the fitted Weibull distribution are the Weibull moduli. Therefore, similar slopes mean similar reliability. The similar slopes of the high strength data set—corresponding to low porosity—for empirical and RF predicted values can be seen in Figures 6 and 7. On the other hand, majority of the ANN σ_f predictions resulted in higher Weibull modulus, 33 sets out of 42, compared to RF predictions (Figures 6 and 7). These high Weibull moduli are not representative of the real systems with m lower than 20. Yet, the RF model qualitatively predicted the decreasing m values with increasing pore size under uniaxial and biaxial loading (Figures 6 and 7). ANN predicted σ_f data also showed similar trend in decreasing Weibull modulus with increasing pore size, but not for all the sets. Similar decrease in Weibull modulus with increasing pore size was also numerically predicted by lattice spring models [12].

Figure 6. Weibull plots showing uniaxial σ_f data for pore size 40 μm (a, c, e) and 80 μm radius (b, d, f). The FM-FEM generated σ_f data are shown in a) and b); RF predicted σ_f data in c) and d); ANN predicted σ_f data in e) and f). Number of pores increases from right to left for all the figures a-f; blue $\bullet$, orange $\blacksquare$, green $\blacklozenge$, red $\blacktriangle$, and purple $\blacktriangledown$ represent 5, 10, 20, 30, and 40 pores, respectively. The Weibull moduli are reported as m_5 to m_{40} for corresponding number of pores. Dotted lines represent the fitted 2-parameter Weibull distribution. Sample size is 40 for each set.

Figure 7. Weibull plots showing biaxial σ_f data for pore size 40 μm (a, c, e) and 80 μm radius (b, d, f). The FM-FEM generated σ_f data are shown in a) and b); RF predicted σ_f data in c) and d); ANN predicted σ_f data in e) and f). Number of pores increases from right to left for all the figures a-f; blue $\bullet$, orange $\blacksquare$, green $\blacklozenge$, red $\blacktriangle$, and purple $\blacktriangledown$ represent 5, 10, 20, 30, and 40 pores, respectively. The Weibull moduli are reported as m5 to m40 for corresponding number of pores. Dotted lines represent the fitted 2-parameter Weibull distribution. Sample size is 40 for each set.

The prediction accuracy of the RF and ANN model was higher for smaller pores at 40 μm radius. The origin of this observation was attributed to the lower probability of finding strongly interacting pores in structures with the same amount of smaller size pores, that is, lower porosity. The maximum porosity of the 40 μm radius pores were ~ 10 vol. %; whereas, the largest pore system with 80 μm radius pores had a porosity of ~ 31 vol. % (Figure 5). This high porosity introduces stress amplification and stress shielding based on the spatial distribution of pores [8, 35]. As a result, the distribution of fracture strength is wider in systems with high porosity and high number of pores. Hence, the higher porosity systems demand even larger datasets to deliver more accurate predictions.

4. CONCLUSIONS

An investigation of the stochastic fracture behavior of porous ceramics and glasses is challenging due to complex interactions between pores, pore and crack, and far field stresses. This complexity and requirement of large number of virtual tests prohibit the use of high-fidelity simulations to predict the reliability of brittle porous materials. In this context, we used the results of FM-FEM simulations on 680 unique porous structures to develop RF and ANN models. The RF model correctly captured the general decreasing trend of σ_f values and decreasing Weibull modulus with increasing porosity. The ANN model was also predicted the decreasing σ_f with increasing porosity, but was not accurate enough to predict the variations in σ_f quantified by the Weibull modulus. The RF model had relatively better prediction of the Weibull modulus and its change with porosity. The prediction accuracies of both RF and ANN were higher for the lowest porosity systems.

As an advancement in the predictive materials characterization, a geometric matching method by means of the Euclidean bipartite matching between the empirical and the generated media was also presented to be able to quantify the error (here, MAPE was selected to be the error measure). The nSIFs were compared by means of MAPE which are in ranges of 3.4-17.93% (uniaxial load) and 2.83-19.42% (biaxial load) for RF, 3.79-17.43% and 3.39-21.43 for ANN at the validation step; 3.54-18.20% (uniaxial load) and 3.06-21.60% (biaxial load) for RF, 3.57-18.26% and 3.43-21.76% for ANN at the implementation step. The present study thus indicates that further empirical investigations with large number of microstructures ($>10,000$) are needed to validate the predictive capabilities of ML for the fracture mechanics problems.

REFERENCES

1. Roohani-Esfahani SI, Newman P, Zreiqat H (2016) Design and Fabrication of 3D printed Scaffolds with a Mechanical Strength Comparable to Cortical Bone to Repair Large Bone Defects. *Sci Rep*. <https://doi.org/10.1038/srep19468>
2. Meyers MA, Chen PY (2014) *Biological materials science: Biological materials, bioinspired materials, and biomaterials*. Cambridge University Press
3. Ebner M, Chung DW, García RE, Wood V (2014) Tortuosity anisotropy in lithium-ion battery electrodes. *Adv Energy Mater*. <https://doi.org/10.1002/aenm.201301278>
4. Xu H, Chen B, Tan P, et al (2018) Modeling of all porous solid oxide fuel cells. *Appl Energy*. <https://doi.org/10.1016/j.apenergy.2018.03.037>
5. Brodnik NR, Faber KT (2019) Out-of-plane mechanical characterization of acicular mullite and aluminum titanate diesel particulate filters. *Int J Appl Ceram Technol*. <https://doi.org/10.1111/ijac.13161>
6. Han L, Li F, Deng X, et al (2017) Foam-gelcasting preparation, microstructure and thermal insulation performance of porous diatomite ceramics with hierarchical pore structures. *J Eur Ceram Soc*. <https://doi.org/10.1016/j.jeurceramsoc.2017.02.032>
7. Keleş Ö, Edwin García R, Bowman KJ (2014) Failure variability in porous glasses: Stress interactions, crack orientation, and crack size distributions. *J Am Ceram Soc*. <https://doi.org/10.1111/jace.13214>
8. Keleş Ö, García RE, Bowman KJ (2013) Stochastic failure of isotropic, brittle materials with uniform porosity. *Acta Mater*. <https://doi.org/10.1016/j.actamat.2013.01.024>

9. Keleş Ö, García RE, Bowman KJ (2014) Pore-crack orientation effects on fracture behavior of brittle porous materials. *Int J Fract.* <https://doi.org/10.1007/s10704-014-9934-8>
10. Rice RW (2017) *Porosity of Ceramics*. CRC Press, Boca Raton
11. Omer N, Yosibash Z (2019) Extracting stochastic stress intensity factors using generalized polynomial chaos. *Eng Fract Mech.* <https://doi.org/10.1016/j.engfracmech.2018.12.001>
12. Cui Z, Huang Y, Liu H (2017) Predicting the mechanical properties of brittle porous materials with various porosity and pore sizes. *J Mech Behav Biomed Mater.* <https://doi.org/10.1016/j.jmbbm.2017.02.014>
13. Biswas D (1981) Crack-void interaction in polycrystalline alumina. *J Mater Sci* 16:2434–2438
14. Evans A, Biswas D, Fulrath R (1979) Some Effects of Cavities on the Fracture of Ceramics: I, Cylindrical Cavities. *J Am Ceram Soc* 62:95–100. <https://doi.org/10.1111/j.1151-2916.1979.tb18814.x>
15. Evans A, Biswas D, Fulrath R (1979) Some Effects of Cavities on the Fracture of Ceramics: II, Spherical Cavities. *J Am Ceram Soc* 62:101–106. <https://doi.org/10.1111/j.1151-2916.1979.tb18815.x>
16. Keleş Ö, Anderson EH, Huynh J, et al (2018) Stochastic fracture of additively manufactured porous composites. *Sci Rep.* <https://doi.org/10.1038/s41598-018-33863-4>
17. Xu Y, Cheng L, Zhang L, et al (2001) Optimization of sample number for Weibull function of brittle materials strength. *Ceram Int.* [https://doi.org/10.1016/S0272-8842\(00\)00056-0](https://doi.org/10.1016/S0272-8842(00)00056-0)
18. Gu GX, Chen CT, Richmond DJ, Buehler MJ (2018) Bioinspired hierarchical composite

- design using machine learning: Simulation, additive manufacturing, and experiment.
Mater Horizons. <https://doi.org/10.1039/c8mh00653a>
19. Özkan M, Borghei M, Karakoç A, et al (2018) Films based on crosslinked TEMPO-oxidized cellulose and predictive analysis via machine learning. Sci Rep 8:.
<https://doi.org/10.1038/s41598-018-23114-x>
 20. Özkan M, Karakoç A, Borghei M, et al (2019) Machine Learning assisted design of tailor-made nanocellulose films: A combination of experimental and computational studies.
Polym Compos. <https://doi.org/10.1002/pc.25262>
 21. Swaminathan S, Shah T, Sirkeci-Mergen B, Keles O (2018) Machine Learning Models for Predicting Fracture Strength of Porous Ceramics and Glasses. In: Proceedings of the 2018 International Conference on Data Science
 22. Ozaki S, Aoki Y, Osada T, et al (2018) Finite element analysis of fracture statistics of ceramics: Effects of grain size and pore size distributions. J Am Ceram Soc.
<https://doi.org/10.1111/jace.15468>
 23. Zimmermann A, Rödel J (2004) Fracture Statistics Based on Pore/Grain-Size Interaction. J Am Ceram Soc. <https://doi.org/10.1111/j.1151-2916.1999.tb02080.x>
 24. Moore BA, Rougier E, O'Malley D, et al (2018) Predictive modeling of dynamic fracture growth in brittle materials with machine learning. Comput Mater Sci.
<https://doi.org/10.1016/j.commatsci.2018.01.056>
 25. Mangal A, Holm EA (2018) Applied machine learning to predict stress hotspots I: Face centered cubic materials. Int J Plast. <https://doi.org/10.1016/j.ijplas.2018.07.013>
 26. Reid ACE, Langer SA, Lua RC, et al (2008) Image-based finite element mesh construction for material microstructures. Comput Mater Sci.

- <https://doi.org/10.1016/j.commatsci.2008.02.016>
27. Evans AG, Langdon TG (1976) Structural ceramics. *Prog Mater Sci* 21:171–425.
[https://doi.org/10.1016/0079-6425\(76\)90006-2](https://doi.org/10.1016/0079-6425(76)90006-2)
 28. RC Shah Stress Intensity Factors for Through and Part-Through Cracks Originating at Fastener Holes. ASTM International, West Conshohocken, PA
 29. Breiman L (2001) Random Forests. *Machine Learning*.
 30. Matin SS, Farahzadi L, Makaremi S, et al (2018) Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2017.06.030>
 31. Wolfram Research Inc (2018) Mathematica
 32. Bose NK, Liang P (1996) Neural network fundamentals with graphs, algorithms, and applications. McGraw-Hill Ser Electr Comput Eng
 33. Tiryaki S, Özşahin Ş, Yildirim I (2014) Comparison of artificial neural network and multiple linear regression models to predict optimum bonding strength of heat treated woods. *Int J Adhes Adhes*. <https://doi.org/10.1016/j.ijadhadh.2014.07.005>
 34. Mathworks Classify Patterns with a Shallow Neural Network.
<https://www.mathworks.com/help/deeplearning/gs/classify-patterns-with-a-neural-network.html>. Accessed 23 Sep 2019
 35. Keleş Ö, García RE, Bowman KJ (2013) Deviations from Weibull statistics in brittle porous materials. *Acta Mater*. <https://doi.org/10.1016/j.actamat.2013.08.025>
 36. Karakoc A, Taciroglu E (2017) Optimal automated path planning for infinitesimal and real-sized particle assemblies. *AIMS Mater Sci*.
<https://doi.org/10.3934/matserci.2017.4.847>

37. Skala, J; Kolingerova, I; Hyka J (2009) A Monte Carlo solution to the minimal Euclidean matching. In: Handlovicova A, Mikula K, Frolkovic P, Sevcovic D (eds) ALGORITMY. Bratislava, Slovakia
38. Chao L -Y, Shetty DK (1991) Reliability Analysis of Structural Ceramics Subjected to Biaxial Flexure. J Am Ceram Soc. <https://doi.org/10.1111/j.1151-2916.1991.tb06884.x>

APPENDIX

Table A.1. Mean absolute percentage errors (MAPE) for random forests and artificial neural networks for validation and implementation steps for uniaxial loading. SD refers to the standard deviations. Sample size is 40 for each set.

Random Forests

Set	Number of holes	Radius (μm)	<i>Validation</i>		<i>Implementation</i>
			<u>Training set used out of all empirical data (%)</u>	<u>MAPE (%) Mean $\pm$ SD</u>	<u>MAPE (%) Mean $\pm$ SD</u>
1	5	40	85	3.40 $\pm$ 1.46	3.54 $\pm$ 1.39
2	5	60	80	4.80 $\pm$ 2.24	4.66 $\pm$ 2.60
3	5	80	75	8.89 $\pm$ 4.25	9.61 $\pm$ 4.69
4	7	40	85	3.77 $\pm$ 1.17	4.04 $\pm$ 0.90
5	10	40	85	5.93 $\pm$ 2.42	6.10 $\pm$ 1.97
6	10	60	85	7.88 $\pm$ 2.10	8.26 $\pm$ 2.36
7	10	80	85	10.53 $\pm$ 3.31	11.68 $\pm$ 3.37
8	20	40	85	6.54 $\pm$ 1.65	6.64 $\pm$ 1.40
9	20	60	80	10.08 $\pm$ 2.32	11.71 $\pm$ 2.81
10	20	80	70	14.49 $\pm$ 2.80	15.01 $\pm$ 2.11
11	30	40	70	8.06 $\pm$ 1.24	8.55 $\pm$ 1.08
12	30	60	85	13.02 $\pm$ 2.03	13.50 $\pm$ 1.58
13	30	80	80	16.08 $\pm$ 3.00	15.92 $\pm$ 3.48
14	40	40	85	8.84 $\pm$ 1.55	8.69 $\pm$ 1.44
15	40	60	75	14.34 $\pm$ 2.05	14.70 $\pm$ 2.10
16	40	80	75	17.93 $\pm$ 2.46	18.20 $\pm$ 2.55
17	50	40	80	10.22 $\pm$ 1.35	10.50 $\pm$ 1.35
				Max: 17.93	Max: 18.20
				Min: 3.40	Min: 3.54

Artificial Neural Networks

<u>Set</u>	<u>Number</u> <u>of holes</u>	<u>Radius</u> <u>(μm)</u>	<i>Validation</i>		<i>Implementation</i>
			<u>Training set used out of</u> <u>all empirical data (%)</u>	<u>MAPE (%)</u> <u>Mean $\pm$ SD</u>	<u>MAPE (%)</u> <u>Mean $\pm$ SD</u>
1	5	40	85	3.81 ± 1.47	3.57 ± 1.30
2	5	60	85	4.47 ± 1.85	4.91 ± 2.55
3	5	80	85	8.46 ± 4.94	9.83 ± 4.17
4	7	40	85	3.79 ± 1.35	4.01 ± 1.02
5	10	40	80	7.37 ± 2.59	6.97 ± 1.96
6	10	60	75	10.24 ± 2.21	8.24 ± 2.02
7	10	80	85	11.15 ± 3.54	11.39 ± 3.17
8	20	40	80	7.05 ± 1.64	6.82 ± 1.33
9	20	60	85	11.39 ± 2.38	11.98 ± 2.62
10	20	80	70	15.01 ± 2.94	15.13 ± 2.00
11	30	40	80	9.06 ± 1.19	8.64 ± 1.04
12	30	60	75	13.30 ± 2.13	13.57 ± 1.69
13	30	80	80	16.24 ± 3.21	15.77 ± 3.66
14	40	40	85	9.70 ± 1.53	8.82 ± 1.45
15	40	60	80	15.56 ± 2.02	14.61 ± 2.07
16	40	80	85	17.43 ± 2.39	18.26 ± 2.51
17	50	40	80	10.88 ± 1.36	10.38 ± 1.48
				Max: 17.43	Max: 18.26
				Min: 3.79	Min: 3.57

Table A.2. Mean absolute percentage errors (MAPE) for random forests and artificial neural networks for validation and implementation steps for biaxial loading. SD refers to the standard deviations. Sample size is 40 for each set.

Random Forests

<u>Set</u>	<u>Number of holes</u>	<u>Radius (μm)</u>	<i>Validation</i>		<i>Implementation</i>
			<u>Training set used out of all empirical data (%)</u>	<u>MAPE (%) Mean $\pm$ SD</u>	<u>MAPE (%) Mean $\pm$ SD</u>
1	5	40	70	2.83 $\pm$ 1.70	3.06 $\pm$ 2.04
2	5	60	85	5.08 $\pm$ 3.24	4.72 $\pm$ 2.70
3	5	80	70	8.61 $\pm$ 4.95	8.00 $\pm$ 4.69
4	7	40	80	3.73 $\pm$ 1.91	3.75 $\pm$ 1.63
5	10	40	80	7.59 $\pm$ 3.25	7.72 $\pm$ 2.95
6	10	60	70	10.75 $\pm$ 3.22	10.87 $\pm$ 2.56
7	10	80	70	12.76 $\pm$ 3.87	14.51 $\pm$ 4.05
8	20	40	75	7.46 $\pm$ 1.59	7.78 $\pm$ 1.51
9	20	60	85	13.66 $\pm$ 2.30	15.44 $\pm$ 2.84
10	20	80	85	17.60 $\pm$ 3.29	18.52 $\pm$ 3.53
11	30	40	70	9.45 $\pm$ 1.73	9.82 $\pm$ 1.89
12	30	60	75	14.69 $\pm$ 1.94	14.37 $\pm$ 3.23
13	30	80	75	19.42 $\pm$ 2.51	21.60 $\pm$ 3.18
14	40	40	80	10.04 $\pm$ 1.36	11.00 $\pm$ 1.52
15	40	60	80	16.09 $\pm$ 1.56	16.80 $\pm$ 1.46
16	40	80	75	19.36 $\pm$ 2.22	15.58 $\pm$ 2.24
17	50	40	75	11.24 $\pm$ 1.03	11.67 $\pm$ 0.98
				Max: 19.42	Max: 21.60
				Min: 2.83	Min: 3.06

Artificial Neural Networks

<u>Set</u>	<u>Number of holes</u>	<u>Radius (μm)</u>	<i>Validation</i>		<i>Implementation</i>
			<u>Training set used out of all empirical data (%)</u>	<u>MAPE (%) Mean $\pm$ SD</u>	<u>MAPE (%) Mean $\pm$ SD</u>
1	5	40	85	3.39 $\pm$ 2.34	3.43 $\pm$ 2.21
2	5	60	75	6.48 $\pm$ 3.47	5.80 $\pm$ 2.74
3	5	80	85	13.32 $\pm$ 5.04	14.22 $\pm$ 4.85
4	7	40	80	4.02 $\pm$ 2.22	4.53 $\pm$ 1.55
5	10	40	75	7.30 $\pm$ 3.26	8.71 $\pm$ 2.71
6	10	60	75	12.92 $\pm$ 3.67	11.36 $\pm$ 2.84
7	10	80	75	12.51 $\pm$ 3.64	14.44 $\pm$ 3.91
8	20	40	85	7.74 $\pm$ 1.84	7.91 $\pm$ 1.51
9	20	60	85	14.90 $\pm$ 2.54	15.37 $\pm$ 2.91
10	20	80	85	20.99 $\pm$ 3.57	18.00 $\pm$ 3.38
11	30	40	85	10.08 $\pm$ 1.81	10.09 $\pm$ 1.91
12	30	60	85	15.01 $\pm$ 2.26	14.52 $\pm$ 3.44
13	30	80	80	20.59 $\pm$ 2.77	21.76 $\pm$ 3.16
14	40	40	80	10.83 $\pm$ 1.47	10.73 $\pm$ 1.57
15	40	60	85	16.67 $\pm$ 1.66	16.60 $\pm$ 1.67
16	40	80	85	21.43 $\pm$ 2.33	19.44 $\pm$ 2.18
17	50	40	75	11.16 $\pm$ 1.15	11.66 $\pm$ 1.01
				Max: 21.43 Min: 3.39	Max: 21.76 Min: 3.43