Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Data-driven approach for rapid prediction of strength scatter in brittle ceramics using deep learning and swarm optimization

Taiyo Maeda, Muhammad Aiman Bin Musa, Toshio Osada & Shingo Ozaki

To cite this article: Taiyo Maeda, Muhammad Aiman Bin Musa, Toshio Osada & Shingo Ozaki (2026) Data-driven approach for rapid prediction of strength scatter in brittle ceramics using deep learning and swarm optimization, Science and Technology of Advanced Materials: Methods, 6:1, 2656050, DOI: 10.1080/27660400.2026.2656050

To link to this article: https://doi.org/10.1080/27660400.2026.2656050

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2026, VOL. 6, NO. 1, 2656050
https://doi.org/10.1080/27660400.2026.2656050

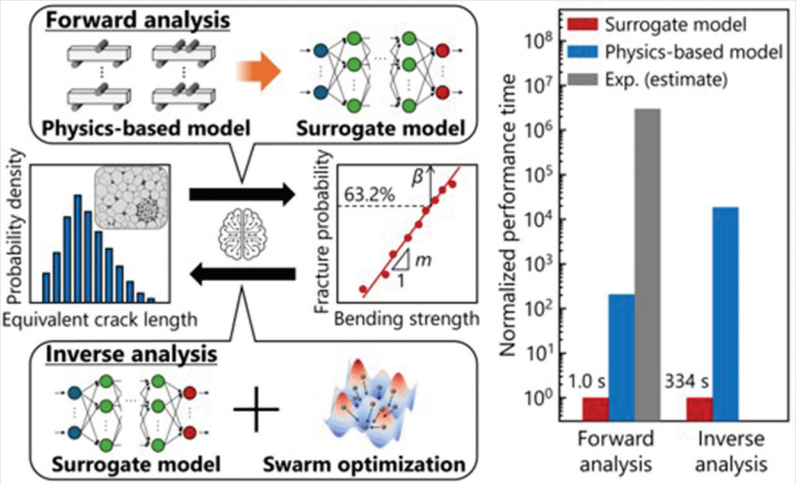
Data-driven approach for rapid prediction of strength scatter in brittle ceramics using deep learning and swarm optimization

Taiyo Maeda[a*], Muhammad Aiman Bin Musa[a*], Toshio Osada[b,c] and Shingo Ozaki[b,c]

[a]Graduate School of Engineering Science, Yokohama National University, Yokohama, Japan; [b]Research Center for Structural Materials, National Institute for Materials Science, Tsukuba, Japan; [c]Faculty of Engineering, Yokohama National University, Yokohama, Japan

ABSTRACT

The intrinsic brittleness and high defect sensitivity of ceramics result in significant strength variability, presenting substantial challenges for structural reliability assessments. Experimental characterization of strength scatter in ceramic components is both time-consuming and costly. Conventional physics-based (forward) analyses can model strength scatter based on microstructural data; these methods are computationally intensive. This study introduces a deep learning – based surrogate model that directly predicts the Weibull distribution parameters of ceramic bending strength from equivalent crack length distributions, achieving substantial reductions in computational cost without compromising predictive accuracy. Additionally, an inverse analysis framework is developed by integrating the surrogate model with intelligent swarm optimization, enabling the estimation of defect distributions from reference strength measurements. The proposed approach demonstrates high accuracy and efficiency, achieving over 200-fold speedup in forward analysis and 18,000-fold in inverse analysis. It facilitates a rapid and reliable evaluation of the relationship between defect distributions and strength scatter as characterized by Weibull parameters. This methodology provides a robust tool for accelerating the design and development of high-performance ceramic materials.



IMPACT STATEMENT

Innovative deep learning approach predicts Weibull parameters for ceramic strength, enabling rapid, accurate evaluation of defect-strength relationships and accelerating high-performance ceramic material development

ARTICLE HISTORY
Received 20 January 2026
Revised 29 March 2026
Accepted 2 April 2026

KEYWORDS
Ceramics; fracture statistics; defect distribution; multilayer perceptron; inverse analysis

1. Introduction

Ceramic materials possess a distinctive combination of desirable properties, exceptional high-temperature resistance, excellent electrical performance, superior wear resistance, and high specific strength. These attributes render ceramics essential for a wide range of advanced technological applications [1–9]. However, the mechanical behaviors differ fundamentally from those of most engineering materials. Their inherent brittleness and pronounced sensitivity to defects mean

CONTACT Shingo Ozaki s-ozaki@ynu.ac.jp Tokiwadai 79-5, Hodogaya-ku, Yokohama 240-8501, Japan
*Co-first authors: Taiyo Maeda and Muhammad Aiman bin Musa contributed equally to this work.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/27660400.2026.2656050

that the strength of ceramics is highly dependent on the presence of internal and surface flaws, such as machining cracks, inclusions, coarse grains, and pores [10–13]. Furthermore, the random nature of defect distribution and size results in significant variability in the measured strength of nominally identical specimens, resulting in probabilistic fracture behavior. The pronounced size effect further complicates the reliable use in structural applications [14,15]. Accurately assessing this strength variability typically necessitates extensive experimental testing, which can be both time-consuming and costly.

To overcome these challenges, Ozaki et al. introduced a 'forward analysis' methodology for predicting the strength distribution of ceramic components based on microstructural characteristics, such as the distribution of pore size, pore aspect ratio, and grain size [16–18]. In this approach, random values reflecting the probability distributions of the observed microstructural features are assigned to each finite element within the component model. These values were then translated into local fracture stresses using a fracture mechanics framework, enabling the construction of a component model that simulates the actual microstructure. By generating multiple models using consistent methodologies and conducting strength analysis on each model, the strength variability of the target components can be evaluated. Although this computational approach is robust, it incurs high costs when numerous samples are required to accurately characterize the true strength scatter characteristics.

To address this challenge, the present study focuses on alumina – the most versatile ceramic material – and aims to develop a deep learning-based surrogate model capable of directly predicting the Weibull distribution parameters for bending strength from defect size distributions. This approach can significantly reduce the computational cost while maintaining accuracy. In recent years, machine learning has been increasingly integrated into materials informatics to predict various mechanical properties from microstructural features [19]. Similar advancements have been observed in the ceramic field [20–25]. For example, Furushima et al. utilized convolutional neural networks to predict the fracture toughness and bending strength of silicon nitride ceramics from microstructural images [21,22]. Zong et al. employed XGBoost to predict thermal conductivity and bending strength based on sintering aids, raw material properties, and processing conditions for three functional ceramics [23]. However, direct prediction of strength scatter characteristics – an essential consideration in strength design – remains rare, highlighting the novelty and potential impact of the proposed methodology.

Beyond forward analysis, inverse analysis is also critical in material design, as high-accuracy inverse approaches can facilitate the identification of optimal microstructures to achieve target strength properties. In this context, our previous research introduced a reliability evaluation scheme for ceramics that inversely estimates defect distribution from standardized reference test results and applies these estimations in forward analysis of different components produced within the same lot [26]. However, such inverse analysis, which combines forward analysis with an optimization algorithm, can be computationally intensive. By substituting the forward analysis with the proposed surrogate model and integrating it with particle swarm optimization (PSO) [27], a type of intelligent swarm optimization, the inverse estimation of defect distributions can be performed with significantly greater efficiency.

2. Construction of supervised datasets

2.1. Target material

In this study, high-purity alumina AS999 (Ferrotec Material Technologies Corporation, Japan) was selected as the target material for both experiments and numerical simulations. Microstructural observations and bending tests under various conditions were performed on sintered specimens with widths of 4 mm and thicknesses of 3 mm by Ito et al. [18]. In this study, both forward and inverse models were validated using the equivalent crack length distribution derived from microstructural observations, alongside the results from three distinct bending test configurations (3pb-16: three-point bending with an external span length of 16 mm, 3pb-30: three-point bending with an external span length of 30 mm, and 4pb-3010: four-point bending with an external span length of 30 mm and an internal span length of 10 mm), with 30 specimens tested under each condition. Specimen surface preparation and testing procedures adhered to the Japanese Industrial Standard (JIS R 1601). The results of the bending tests are presented in Section 4.2. To ensure consistency in material properties and minimize variability owing to manufacturing differences, all specimens were sourced from the same production lot. The fracture origins of all test specimens were associated with the same defect type, specifically internal defects containing unsintered grains [18]. The Mode I fracture toughness, K_{IC} of AS999, was reported as 4.0 MPa $m^{0.5}$ [18].

2.2. Probability distribution of equivalent crack length

The distribution of defects in ceramic materials is crucial in determining their mechanical properties and reliability. To accurately and efficiently characterize this distribution, the concept of equivalent crack length a_e is employed, which incorporates the effects of defect

geometry. The use of equivalent crack length facilitates direct comparison among various defect types and streamlines the calculation of fracture strength.

Because only relatively large equivalent cracks among those present in a specimen are likely to serve as the fracture origin, the distribution of equivalent crack length in this study was characterized using extreme value statistics. Specifically, the generalized extreme value (GEV) distribution, which models block maxima, was employed. To construct the GEV distribution, the dataset was partitioned into blocks of equal size, and the maximum value of each block was extracted. In the context of ceramic defect data, each block corresponds to a defined volume. The cumulative distribution function G and probability density function g of the GEV distribution are expressed as follows [28,29]:

$$G(a_e) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{a_e-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{for} \quad \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{a_e-\mu}{\sigma}\right)\right]\right\} & \text{for} \quad \xi = 0 \end{cases}, \tag{1}$$

$$g(a_e) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{a_e-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \frac{1}{\sigma}\left[1+\xi\left(\frac{a_e-\mu}{\sigma}\right)\right]^{-1/\xi-1} & \text{for} \quad \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{a_e-\mu}{\sigma}\right)\right]\right\} \frac{1}{\sigma}\exp\left(-\frac{a_e-\mu}{\sigma}\right) & \text{for} \quad \xi = 0 \end{cases}, \tag{2}$$

where μ, σ, and ξ represent the location, scale, and shape parameters, respectively. μ, σ, and ξ are related to the mode, width, and tail properties of the distribution, respectively. The shape parameter ξ is critical, as its sign determines whether the distribution has a finite upper bound.

2.3. Dataset generation

In this study, input-output datasets were numerically generated to match the neural network structure described in Section 3.1.

Supervised datasets were obtained using the finite element method-inspired bending strength simulation method of Ito et al. [18] (Section S1) to determine the two-parameter Weibull distribution parameters for various test conditions corresponding to the equivalent crack-length distribution. Here, two bending test configurations (3pb-40: three-point bending with an external span length of 40 mm, and 4pb-4020: four-point bending with an external span length of 40 mm and an internal span length of 20 mm) were employed in accordance with the American Society for Testing and Materials (ASTM C1161) in addition to the 3pb-16, 3pb-30, and 4pb-3010 described in Section 2.1.

We used Latin hypercube sampling to generate a diverse and representative set of input samples [30]. This technique enables efficient coverage of the parameter space while minimizing computational time. Based on previous experience and knowledge in ceramics research [26], appropriate ranges for each GEV distribution parameter were defined as μ = 10–30 μm, σ = 0–10 μm, and ξ = −0.5–0.5. Additionally, the variation in the strength distribution of the ceramics (Weibull distribution parameters) is significantly impacted by the number of specimens [14,31]. Our previous study demonstrated that the Weibull distribution parameters demonstrated minimal variation when the number of specimens is N = 10,000 or greater [26]. Accordingly, to ensure the reliability of the supervised datasets, the number of specimens was set to N = 10,000. Furthermore, the element size utilized in the bending strength analysis also influences the reliability of the analysis results, as it affects the reproducibility of the stress distribution. Ito and Maeda et al. validated that, for bending analysis of smooth specimens, the strength analysis results are largely independent of the element size when the element size is less than h_e = 0.25 mm [18,26]. Therefore, the element size of h_e = 0.25 mm was adopted in this study.

A supervised dataset comprising 2,000 samples was generated for each test condition, resulting in a total dataset of 10,000 samples. The dataset used in this study was generated using a program for the finite element method-inspired bending strength simulation written in Fortran 90, and the total computational time required to generate them was approximately 1 day. During the generation of supervised data, certain cases were observed in which the Weibull modulus diverged to positive infinity, particularly when both the scale parameter σ and shape parameter ξ of the GEV distribution were simultaneously small. In these instances, finite Weibull distribution parameters could not be obtained. Specifically, such cases occurred when $0 \leq \sigma \leq 1$ and $-0.5 \leq \xi \leq 0$ concurrently. When σ and ξ fall within these specified ranges, the resulting equivalent crack length distribution becomes narrow and demonstrates a defined upper bound. Therefore, the variation in strength was minimal, with the Weibull modulus approaching positive infinity.

2.4. Data preprocessing and feature engineering

Data preprocessing and feature engineering are crucial in preparing supervised data for model development. Initially, the generated supervised datasets undergo a cleaning process to eliminate inconsistencies and errors that could negatively impact model performance. In particular, datasets with a Weibull modulus diverging to positive infinity, as described in Section 2.3, were excluded from the supervised data. This filtering reduced the total number of valid supervised datasets from 10,000 to 7,755. To mitigate potential bias in the supervised datasets, the data were supplemented such that the number of valid datasets for each test condition was 2,000. This approach was

essential for maintaining data integrity and ensuring the reliability of subsequent modeling efforts.

Subsequently, feature engineering techniques were applied to address disparities in the scales and units of the various parameters. Specifically, the MinMaxScaler was employed to normalize the features to a range of 0–1. This normalization enhanced model prediction accuracy on observed data and promotes stability during the development. By scaling all features to a common range, each parameter contributed proportionally to the learning process of the model, preventing features with larger magnitudes from disproportionately influencing the model development. The generated dataset can also be reused for training or validating other machine learning models under the same modeling assumptions.

After preprocessing, the dataset was randomly split into two subsets: a development dataset comprising 80% of the entire dataset and a testing dataset containing the remaining 20%. This 80–20 split is a widely adopted practice in machine learning, providing a balance between sufficient data for model development and reserving an adequate portion for unbiased performance evaluation [32–34]. By maintaining a clear separation between development and testing data, the ability of the model to generalize unseen data can be effectively assessed.

3. Data-driven models

3.1. Surrogate model for forward analysis

3.1.1. Multilayer perceptron

In this study, a multilayer perceptron (MLP), a fundamental deep learning architecture, is employed to predict the Weibull distribution parameters of the bending strength of alumina based on defect distribution data. An MLP was adopted as a simple yet effective surrogate model because the relationship between the GEV distribution parameters of the equivalent crack length distribution and the Weibull distribution parameters of bending strengths exhibits a relatively smooth nonlinear correspondence. The basic building block of neural networks is the perceptron, which is a simple artificial neuron that underpins more complex architectures [35]. A perceptron functions as a computational unit that receives multiple inputs, applies corresponding weights, sums them, and processes the result through an activation function to generate an output. A single perceptron can be expressed as follows:

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{3}$$

where y represents the output, ϕ represents the activation function, w_i represents the weight, x_i represents the input, b represents the bias, and n represents the number of inputs. The structure of a perceptron, in which multiple input nodes (x_1, x_2, …, x_n) are connected to a single output node y, is shown in Figure 1. Each connection is associated with a weight (w_1, w_2, …, w_n), and a bias term with a fixed value of 1 is also included. The weighted sum of the inputs and bias is then passed through the activation function ϕ to compute the output.

Expanding upon this concept, the MLP represents a more advanced architecture composed of multiple layers of interconnected perceptrons [36], as shown in Figure 2. A standard MLP is structured with an input layer, one or more hidden layers, and an output layer. These layers receive the input data, process the intermediate representations, and generate the final predictions. By leveraging learning algorithms, such as backpropagation [37], the MLP can learn complex, nonlinear mappings between inputs and outputs. During development, the network updates its internal parameters – namely, weights and biases – by minimizing the discrepancy between the predicted and actual values, typically through gradient-based optimization methods. The general formulation for the MLP is expressed as follows:

$$h_j^{(l)} = \phi^{(l)}\left(\sum_{i=1}^{n_{l-1}} w_{ji}^{(l)} h_i^{(l-1)} + b_j^{(l)}\right), \tag{4}$$

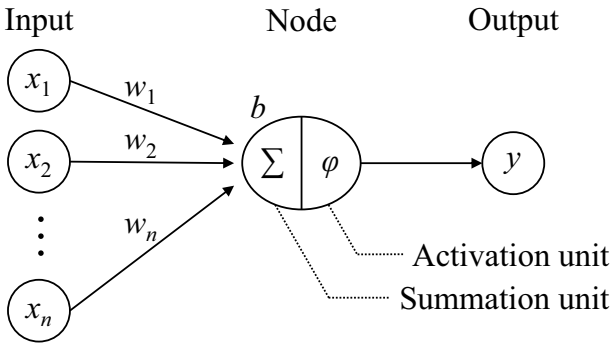


Figure 1. Schematic of a perceptron with multiple inputs and a single output.

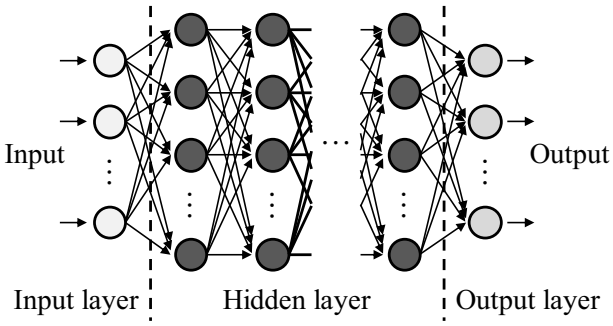


Figure 2. Schematic of an MLP comprising an input layer, a hidden layer, and an output layer.

where $h^{(l)}{}_j$ represents the output of the j-th neuron in the l-th layer, $\phi^{(l)}$ represents the activation function for the l-th layer, $w_{ji}^{(l)}$ represents the weight connecting the i-th neuron in the $(l-1)$-th layer to the j-th neuron in the l-th layer, $h_i^{(l-1)}$ represents the output of the i-th neuron in the $(l-1)$-th layer, $b_j^{(l)}$ represents the bias for the j-th neuron in the l-th layer, and n_{l-1} represents the number of neurons in the $(l-1)$-th layer.

3.1.2. Neural network architecture

The architecture of the MLP surrogate model is detailed as follows. In this study, the model was constructed using TensorFlow [38], a widely used open-source machine learning library for Python. The rectified linear unit activation function was applied to the hidden layers, whereas a linear activation function was applied to the output layer. The mean squared error (MSE) was used as a loss function, guiding the learning process by quantifying the difference between the predicted and actual values during development. To comprehensively evaluate the predictive performance of the models, additional metrics, such as the mean absolute error (MAE), correlation coefficient (R), and coefficient of determination (R^2) were employed. These metrics provide a multidimensional evaluation of the predictive capabilities of the models, ensuring a robust and reliable evaluation of their predictive performance.

The surrogate model developed for forward analysis incorporates five input variables: GEV distribution parameters (location parameter μ, scale parameter σ, and shape parameter ξ), which characterize the equivalent crack length distribution, as well as two geometric parameters from the bending test, namely the internal span length L_i and external span length L_o. The model outputs two variables: the Weibull distribution parameters (Weibull modulus m and scale parameter β), which describe the statistical characteristics of the bending strength scatter as shown in Figure 3(a). Details regarding the hidden layer architecture and convergence criteria, which are consistent across all forward model configurations, are listed in Table 1. This configuration was determined through iterative refinement to achieve an appropriate balance between predictive accuracy and computational efficiency. Because the dimensionality of the input parameters is relatively small and the relationship between the defect distribution parameters and the resulting Weibull distribution parameters is smooth, a relatively shallow network was found to be sufficient for accurately approximating the forward model.

3.1.3. Training and validation procedure

To rigorously assess the generalization capability of the model and minimize the risk of overfitting, K-fold cross-validation was employed. This procedure, implemented using the K-fold class from the scikit-learn library [39], subdivides the dataset into K folds. In each iteration, the model was trained on $(K-1)$ folds and validated on the remaining fold, ensuring that each subset served as the validation set exactly once. This systematic approach provides a comprehensive evaluation of the model performance across the entire development dataset.

A value of $K = 10$ was selected to balance computational demands with model robustness. Additionally, early stopping was utilized during development, with a patience parameter set to 50 epochs. Development was terminated if the validation loss did not improve for 50 consecutive epochs, thereby reducing the likelihood of overfitting and yielding more reliable performance estimates by mitigating the effects of data variability.

3.2. Inverse analysis model

The direct application of a machine learning model for inverse analysis, that is, predicting input parameters from output values by simply reversing the forward model, is generally challenging. This difficulty stems from the non-uniqueness inherent in inverse problems, in which multiple combinations of input variables can yield similar output values [40–42]. Furthermore, this issue was highlighted in our previous study [26], in which we demonstrated that multiple distinct patterns of equivalent crack length distributions can yield identical strength scatter characteristics. To address this challenge, we developed an inverse analysis framework that integrates a forward surrogate model and PSO, following the methodology established in our previous research [26]. This approach enables robust estimation of input variables – specifically, the equivalent crack length distribution – that can accurately reproduce the experimentally observed strength distribution.

The flowchart of the inverse model is shown in Figure 3(b). In this framework, particles representing candidate input parameter values (that is, the GEV distribution parameters for equivalent crack length distributions) were initially generated within the search space using the Latin hypercube sampling method [30]. Subsequently, for each particle, the corresponding Weibull distribution parameters were computed by the forward surrogate model in Figure 3(a) based on the position of the particle in the parameter space. The predicted strength distribution characteristics were then compared with those obtained from reference experimental tests, and the fitness of the particles was evaluated using an objective function.

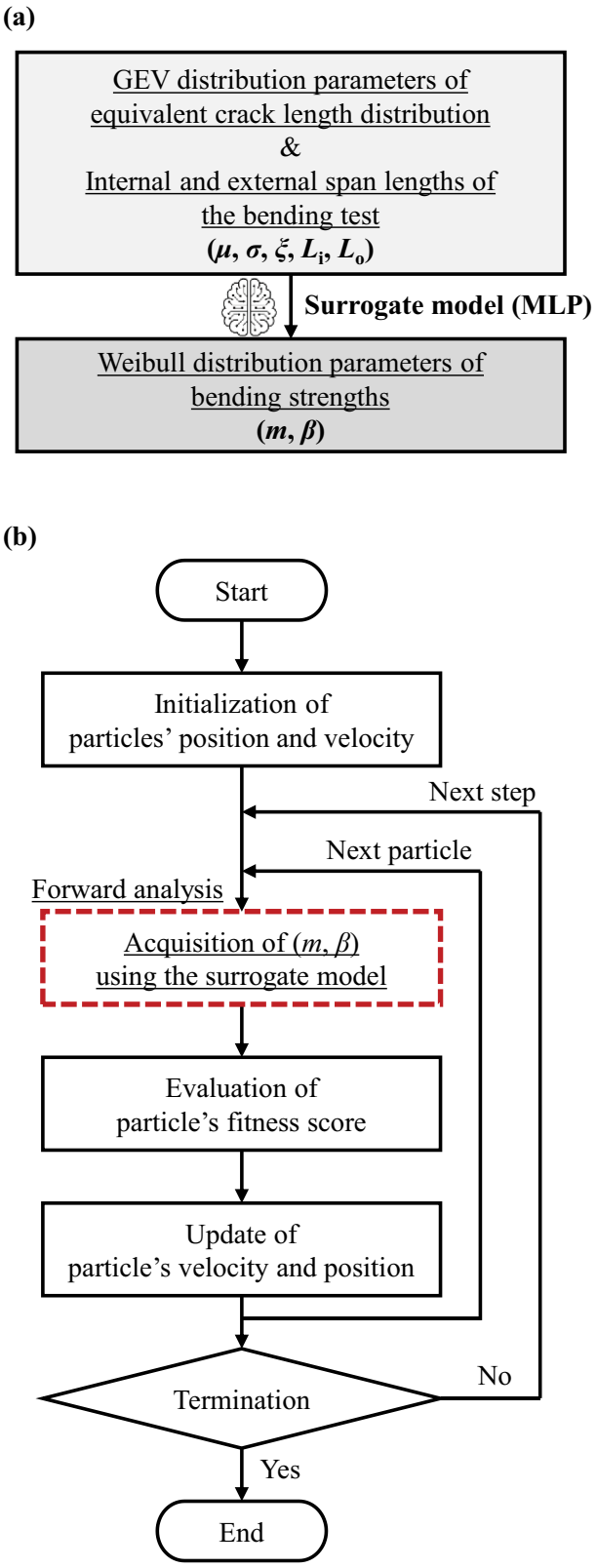


Figure 3. Overall framework of the proposed data-driven approach based on a surrogate model: (a) input – output relationship of the surrogate model for forward analysis; (b) flowchart of the inverse analysis integrating a surrogate model with PSO.

The objective function (*Error*) adopted in this study, as referenced in our previous research [26], is defined as follows:

$$Error = \sum_{i=1}^{n_{ref}} \left[\left(\frac{\ln \sigma_{Exp}^{0.007\%} - \ln \sigma_{Sim}^{0.007\%}}{\ln \sigma_{Exp}^{0.007\%}} \right)^2 \right] \left[+ \left(\frac{\ln \sigma_{Exp}^{99.99\%} - \ln \sigma_{Sim}^{99.99\%}}{\ln \sigma_{Exp}^{99.99\%}} \right)^2 \right], \quad (5)$$

where $\sigma^{0.007\%}$ and $\sigma^{99.99\%}$ represent the strengths corresponding to 0.007% and 99.99% cumulative fracture probabilities, which reflect the weakest and strongest strengths in $N = 10{,}000$ specimens, respectively. The subscripts 'Exp' and 'Sim' refer to values obtained from experimental measurements and simulation, respectively. The parameters $\sigma^{0.007\%}$ and $\sigma^{99.99\%}$ are calculated using the Weibull distribution parameters m and β. The variable n_{ref} denotes the number of reference experimental conditions. The velocity (movement amount) and position of each particle were updated based on the fitness score to minimize the *Error*. The update rules for particle velocity and position adhered to the standard PSO formulation, as outlined in Section S2. This iterative process was repeated until a predefined number of iterations was completed. The search ranges for the GEV distribution parameters were set to $\mu = 10$–30 μm, $\sigma = 0$–10 μm, and $\xi = -0.5$–0.5, consistent with those used in the development of the forward surrogate model. The PSO parameters adopted from our previous study are listed in Table 2.

4. Results and discussion

4.1. Surrogate model performance

4.1.1. Model validation

The generalization capability of the surrogate model was evaluated using K-fold cross-validation ($K = 10$), as described in Section 3.1.3 [39]. The training and validation loss curves for each fold are shown in Figure 4(a,b), respectively, with losses calculated using normalized output variables. The curves demonstrate consistent convergence across all folds: training losses decrease steadily, whereas validation losses display only minor fluctuations before approaching zero. This pattern is a strong indicator of robust generalization and reflects the capacity of the model to learn effectively from diverse data subsets. The mean and standard deviation of the evaluation metrics computed on the validation folds, based on the normalized outputs, are listed in Table 3. Generally, low standard deviations indicate consistent model performance and high reliability. Although the standard deviation of MSE appears relatively large compared with its mean, this is attributed to the extremely small magnitude of the MSE itself and does not raise concerns. The low MSE and MAE values indicate minimal prediction errors, whereas R and R^2 values close to one suggest

Table 1. Configuration of the neural network architecture and development settings.

Parameters	Values
Number of hidden layers	5
Arrangement of nodes across hidden layers (Input layer to output layer)	256–128–64–32–8
Batch size	128
Maximum epoch	500
Optimizer	Adam

Table 2. Parameters for PSO.

Parameters	Values
Particle population N_p	200
Maximum iteration	50
Inertia weight w	0.5
Cognitive coefficient c_1	0.5
Social coefficient c_2	0.5

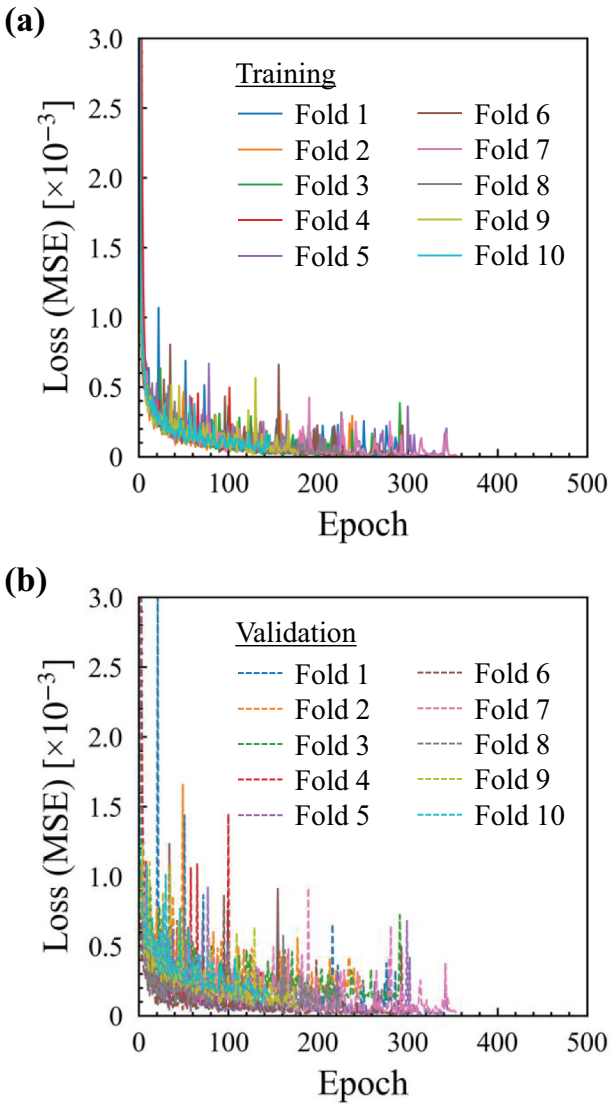


Figure 4. Loss curves for *K*-fold cross-validation: (a) training; (b) validation.

a strong linear relationship between the predicted and actual values.

The model architecture and hyperparameters were determined based on the results of the *K*-fold cross-validation. The final model was then trained using the hold-out method with early stopping, utilizing 80% of the development dataset, with 20% of this subset reserved for internal validation. The training and validation loss curves computed using the normalized output variables are shown in Figure 5. Both curves demonstrated smooth and stable convergence toward zero, with minimal noise in the validation loss, indicating that the model effectively captured the underlying data patterns without overfitting.

4.1.2. Test results

The performance of the model was then evaluated using the remaining 20% of the prepared data, which was reserved as a testing dataset. The correlation plots between the actual and predicted values for the Weibull modulus and scale parameter in the development and testing datasets are shown in Figure 6(a,b), respectively. Both plots demonstrate excellent alignment with the ideal 1:1 line, with nearly all predictions falling within ±10% relative error bounds. Notably, the prediction accuracy was particularly high within the range of m = 5–20 and β = 200–600 MPa, which represents the typical region for pressureless sintered alumina. Furthermore, accurate predictions were obtained across different testing conditions with varying span lengths, indicating that the surrogate model can reliably account for the influence of testing geometry.

The evaluation metrics for the development and testing datasets are listed in Table 4. All primary values were computed using normalized output variables, whereas values in parentheses represent results computed in the original (unnormalized) scale, where m is dimensionless, and β is expressed in MPa. For both datasets, the MSE and MAE values remained small, whereas the R and R^2 values were close to 1. These results validated the outstanding predictive performance of the model and its potential for practical applications in materials science. The strong generalization capability of the model from development to unseen testing data further validates its robustness as a surrogate model for understanding and optimizing ceramic material properties.

The present framework represents defect sizes using the concept of equivalent crack length, which enables both internal and surface defects to be described within a unified fracture mechanics formulation. However, the applicability of the model depends on the assumption that the defect population

Table 3. Mean and standard deviation of evaluation metrics across all folds in *K*-fold cross-validation, based on the validation data.

Metrics	Mean	Standard deviation
MSE	7.57×10^{-5}	6.58×10^{-5}
MAE	2.49×10^{-3}	5.31×10^{-4}
R	0.999	1.15×10^{-3}
R^2	0.997	2.30×10^{-3}

These results correspond to the validation loss curves shown in Figure 4(b). All metrics are computed using normalized outputs.

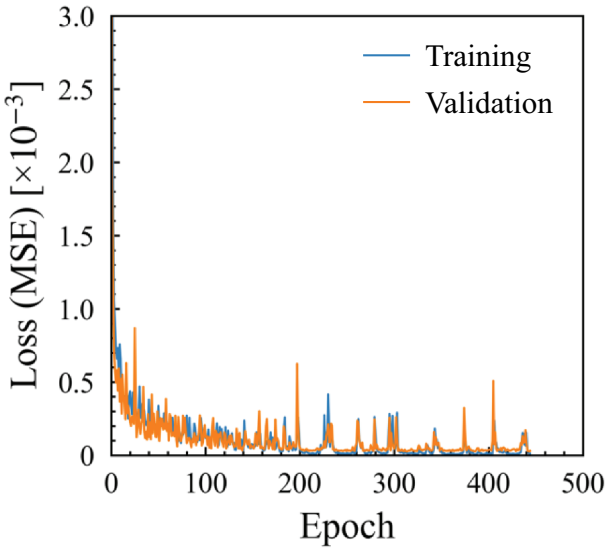


Figure 5. Training and validation loss curves.

is statistically consistent. When different types of defects dominate the fracture origins, the strength scatter characteristics may not be accurately described. This limitation is inherent not only to the surrogate model but also to the underlying physics-based forward analysis [26]. Furthermore, regions such as specimen edges, where the stress field cannot be approximated by the tensile stress state of an infinite plate, are not explicitly modeled in the present framework. In the experimental data analyzed in this study, however, all fracture origins were identified as internal defects associated with agglomerated pores around the unsintered grains, and therefore the assumption of a consistent defect population is satisfied.

Beyond its predictive accuracy, the surrogate model offers substantial improvements in terms of its computational efficiency. The simulation time was significantly reduced by replacing the computationally intensive physics-based forward model with a trained surrogate model. For example, under the 3pb-16 condition with $N = 10{,}000$ specimens, the physics-based model required approximately 208 s, whereas the surrogate model completed the same task in just 1.0 s, corresponding to a computational speedup of more than 200-fold. Here, all computations were conducted using Python 3.12.3 (Anaconda distribution) on a machine featuring an Intel Core i7-14700KF

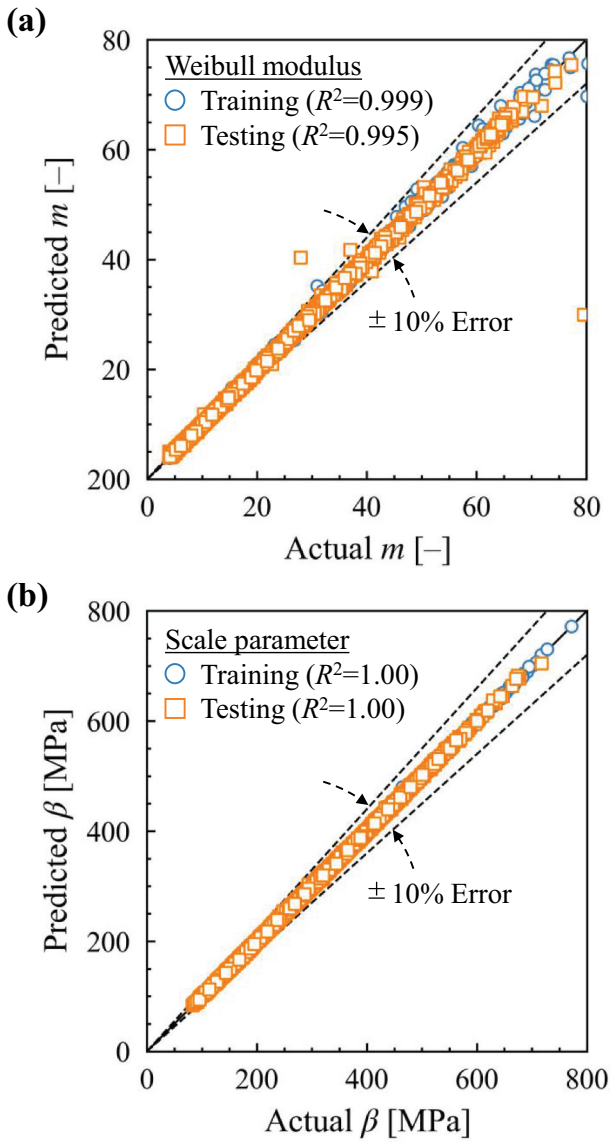


Figure 6. Correlation between the actual and predicted Weibull distribution parameters for development and testing datasets: (a) Weibull modulus; (b) scale parameter. The dashed lines represent the ±10% relative error margin.

CPU, 32 GB RAM, running Windows 11. This reduction in computational cost arises mainly from the difference between the physics-based and surrogate approaches. The physics-based simulation relies on a Monte Carlo – based procedure to evaluate strength scatter as described in Section S1, which requires a large number of simulations corresponding to many specimens to obtain statistically reliable Weibull parameters. In the case of the surrogate model, these simulations are performed in advance to generate supervised data, and the trained surrogate model can directly predict the Weibull parameters without repeating the Monte Carlo simulations.

In addition, the computational advantage becomes even more significant for testing conditions involving longer span lengths. In the physics-based model, the number of finite elements involved in the computation

Table 4. Evaluation metrics for the development and testing datasets.

	Development (80%)		Testing (20%)	
Metrics	m	β	m	β
MSE	1.81×10^{-5} (0.193)	3.38×10^{-6} (1.60)	1.36×10^{-4} (1.46)	4.59×10^{-6} (2.18)
MAE	1.97×10^{-3} (0.203)	1.37×10^{-3} (0.946)	2.45×10^{-3} (0.254)	1.53×10^{-3} (1.05)
R	1.00	1.00	0.997	1.00
R^2	0.999	1.000	0.995	1.00

All primary values are computed using normalized output variables. The values in parentheses represent metrics corresponding to the original (unnormalized) scale, where m is dimensionless and β is in MPa.

increases as the span length becomes larger because all elements subjected to the tensile component of bending stress must be considered, resulting in higher computational costs. By contrast, in the surrogate model, the span length is treated simply as input parameters, and therefore the computational time remains nearly constant regardless of the testing configuration.

4.2. Inverse analysis performance

To validate the inverse model, two types of bending tests—3pb-16 and 4pb-3010—were selected as reference experiments ($n_{ref} = 2$), in accordance with our previous study [26]. Based on the experimentally obtained strength scatter from these tests, the inverse model was employed to estimate an equivalent crack length distribution capable of reproducing the observed strength variability. The estimated distribution, alongside equivalent crack lengths derived from the results of microstructural observations [18] and those obtained in our previous inverse analysis using the physics-based model [26] is shown in Figure 7. The estimated distribution demonstrated excellent agreement with both the microstructural data and estimated results from the physics-based model. The consistency between the surrogate-based estimation and the physics-based model is attributed to the high predictive accuracy of the surrogate model, as demonstrated by the validation results shown in Figure 6, confirming the validity and robustness of the proposed approach.

To evaluate the reliability of the estimated equivalent crack length distribution further, it was input into the forward surrogate model to predict the bending strength scatter characteristics under various testing conditions. A comparison of the Weibull plots of the experimental results ($N = 30$) [18] and the predictions for 3pb-16, 3pb-30, and 4pb-3010 is shown in Figure 8. The Weibull distribution parameters obtained from the experiments and predictions shown in Figure 8 are listed in Table 5. The surrogate model accurately reproduced the reference test results (3pb-16 and 4pb-3010). For the 3pb-30 condition, which was not included in the inverse analysis, the predicted strengths were marginally higher than the experimental values. However, this discrepancy was attributed to the limited number of experimental specimens and the resulting variation in the observed strength scatter, instead of errors in the estimated equivalent crack length distribution [26]. These findings highlight the strong generalization capability of the inverse model.

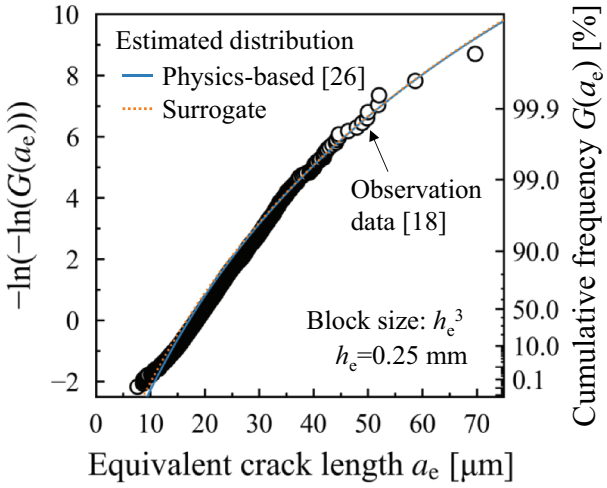


Figure 7. Estimated equivalent crack length distributions based on experimental strength data under 3pb-16 and 4pb-3010 with $h_e = 0.25$ mm using the physics-based [26] and surrogate models for the forward analysis. Data points of equivalent crack length converted from the results of microstructural observation [18] are also included.

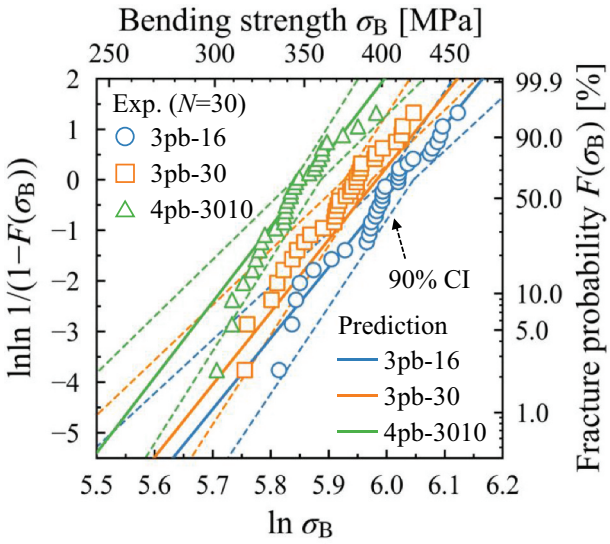


Figure 8. Comparison of Weibull plots between experimental [18] and predicted results under 3pb-16, 3pb-30, and 4pb-3010 using the equivalent crack length distribution shown in Figure 7. The dashed lines represent the 90% confidence interval of experimental results based on the maximum likelihood estimation of the Weibull distribution.

Table 5. Experimental [18] and predicted Weibull distribution parameters shown in Figure 8..

Test condition	m [–]	β [MPa]
Experiment [18]		
3pb-16	14.1	412.6
3pb-30	14.5	383.9
4pb-3010	14.9	352.1
Prediction		
3pb-16	14.0	413.0
3pb-30	14.3	396.6
4pb-3010	14.9	351.7

The significantly enhanced computational efficiency achieved through the development of the surrogate model for forward analysis is particularly advantageous for inverse analyses that require numerous forward simulations. For example, the inverse analysis to estimate the equivalent crack length distribution shown in Figure 7 was completed in approximately 334 s when the surrogate model is employed. In contrast, it is estimated that the same analysis would require approximately 72 days when the physics-based model is used, corresponding to a computational speedup of more than 18,000-fold. This estimate (72 days) is based on the assumption that the forward analysis for the 4pb-3010 configuration requires approximately twice the computational time of the 3pb-16 configuration due to the larger number of elements. Considering that the PSO algorithm uses 200 particles and 50 iterations, the total computational time for the inverse analysis using the physics-based model was estimated as $208 \times 3 \times 200 \times 50$ s $\approx$72 days.

In practical applications, when the fracture toughness of the target material is known and standardized strength test data (e.g. 3pb-16 and 4pb-3010) are already available, the total time required for the proposed approach – from generating the supervised dataset to completing the inverse analysis for estimating the equivalent crack length distribution – can be completed within 2 days. Furthermore, when the fracture toughness is identical, the trained surrogate model can be applied to other brittle materials without reconstructing the surrogate model. Consequently, the total computational cost of applying the proposed framework can be significantly reduced. In contrast, such a reduction in computational time cannot be achieved using the conventional physics-based approach, where the inverse analysis would require a large number of forward simulations each time it is performed, resulting in a large computational time (e.g. 72 days). This substantial reduction in the simulation time facilitates large-scale parametric studies and supports the practical implementation of the proposed framework in material design and structural reliability assessment.

5. Conclusion

This study developed a deep learning-based surrogate model capable of directly predicting the Weibull distribution parameters for the bending strength of alumina from equivalent crack length distributions, thereby significantly reducing computational costs compared with conventional physics – based forward analyses. Furthermore, an inverse analysis framework was established by integrating the surrogate model with PSO to estimate defect distributions from reference strength test data. The proposed methodology demonstrated high predictive accuracy for both forward and inverse analyses, with notable improvements in computational efficiency – over 200-fold acceleration in forward analysis and more than 18,000-fold in inverse analysis. This approach provides a promising avenue for accelerating material development by enabling an efficient evaluation of the relationship between defect distributions and strength scatter characteristics without compromising computational fidelity.

Nevertheless, the current surrogate model is limited to a single fracture toughness value, as fracture toughness was not included among the input parameters. Consequently, any change in fracture toughness necessitates the construction of a new surrogate model. However, because the present framework assumes a fixed fracture toughness, the trained surrogate model is expected to be applicable to brittle materials with comparable fracture toughness and similar fracture behavior, such as many pressureless sintered ceramics without significant R-curve behavior. To further enhance the efficiency and general applicability of the approach in material development, future studies should aim to extend the current model by incorporating fracture toughness as an additional input variable. Here, because fracture toughness strongly influences the scale parameter β, the complexity of the input – output relationship is expected to increase. In such cases, deeper network architectures or alternative machine learning models may be required to accurately capture the more complex relationships. Moreover, the choice of optimization algorithms for the inverse analysis warrants additional attention. The PSO method employed in this study did not retain information from previous optimization processes, resulting in comparable computation times for each iteration. Adopting algorithms that can leverage accumulated experience, such as reinforcement learning, may offer substantial reductions in the computation time for inverse analyses.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was partially supported by the Grant-in-Aid for Scientific Research [Grant No. (B) 22H01357], the Japan Society for the Promotion of Science (JSPS).

ORCID

Taiyo Maeda http://orcid.org/0009-0007-4552-7251
Toshio Osada http://orcid.org/0000-0003-1539-9264
Shingo Ozaki http://orcid.org/0000-0003-3450-6774

References

[1] de Faoite D, Browne DJ, Chang-Díaz FR, et al. A review of the processing, composition, and temperature-dependent mechanical and thermal properties of dielectric technical ceramics. J Mater Sci. 2012;47:4211–4235. doi: 10.1007/s10853-011-6140-1
[2] Santoliquido O, Bianchi G, Eggenschwiler PD, et al. Additive manufacturing of periodic ceramic substrates for automotive catalyst supports. Int J Appl Ceram Technol. 2017;14(6):1164–1173. doi: 10.1111/ijac.12745
[3] Naslain R. Design, preparation and properties of non-oxide CMCs for application in engines and nuclear reactors: an overview. Compos Sci Technol. 2004;64(2):155–170. doi: 10.1016/S0266-3538(03)00230-6
[4] Padture NP, Gell M, Jordan EH. Thermal barrier coatings for gas-turbine engine applications. Science. 2002;296(5566):280–284. doi: 10.1126/science.1068609
[5] Evans AG, Mumm DR, Hutchinson JW, et al. Mechanisms controlling the durability of thermal barrier coatings. Prog Mater Sci. 2001;46(5):505–553. doi: 10.1016/S0079-6425(00)00020-7
[6] Klemm H. Silicon nitride for high-temperature applications. J Am Ceram Soc. 2010;93(6):1501–1522. doi: 10.1111/j.1551-2916.2010.03839.x
[7] Wang G, Lu Z, Li Y, et al. Electroceramics for high-energy density capacitors: current status and future perspectives. Chem Rev. 2021;121(10):6124–6172. doi: 10.1021/acs.chemrev.0c01264
[8] Goel M. Recent developments in electroceramics: mems applications for energy and environment. Ceram Int. 2004;30(7):1147–1154. doi: 10.1016/j.ceramint.2003.12.012
[9] Huang YH, Wu YJ, Qiu WJ, et al. Enhanced energy storage density of $Ba_{0.4}Sr_{0.6}TiO_3$-MgO composite prepared by spark plasma sintering. J Eur Ceram Soc. 2015;35(5):1469–1476. doi: 10.1016/j.jeurceramsoc.2014.11.022
[10] Evans AG. Structural reliability: a processing-dependent phenomenon. J Am Ceram Soc. 1982;65(3):127–137. doi: 10.1111/j.1151-2916.1982.tb10380.x
[11] Danzer R. A general strength distribution function for brittle materials. J Eur Ceram Soc. 1992;10(6):461–472. doi: 10.1016/0955-2219(92)90021-5
[12] Peterlik H. Relationship of strength and defects of ceramic materials and their treatment by Weibull theory. J Ceram Soc Jpn. 2001;109(8):S121–S126. doi: 10.2109/JCERSJ.109.1272_S121
[13] Quinn GD. Fractography of ceramics and glasses. 3rd ed. National Institute of Standards and Technology Special Publication (NIST SP-960-16e3); 2020. doi: 10.6028/NIST.SP.960-16e3
[14] Danzer R, Lube T, Supancic P, et al. Fracture of ceramics. Adv Eng Mater. 2008;10(4):275–298. doi: 10.1002/adem.200700347
[15] Bažant ZP. Size effect on structural strength: a review. Arch Appl Mech. 1999;69(9–10):703–725. doi: 10.1007/s004190050252
[16] Ozaki S, Aoki Y, Osada T, et al. Finite element analysis of fracture statistics of ceramics: effects of grain size and pore size distributions. J Am Ceram Soc. 2018;101(7):3191–3204. doi: 10.1111/jace.15468
[17] Ozaki S, Yamagata K, Ito C, et al. Finite element analysis of fracture behavior in ceramics: prediction of strength distribution using microstructural features. J Am Ceram Soc. 2022;105(3):2182–2195. doi: 10.1111/jace.18201
[18] Ito C, Maeda T, Higashi R, et al. Application of extreme value statistics to internal pore distribution in ceramics and prediction of size dependency of strength scatter. J Eur Ceram Soc. 2024;44(5):3381–3392. doi: 10.1016/j.jeurceramsoc.2023.12.021
[19] Ramprasad R, Batra R, Pilania G, et al. Machine learning in materials informatics: recent applications and prospects. NPJ Comput Mater. 2017;3:54. doi: 10.1038/s41524-017-0056-5
[20] Yang P, Wu S, Wu H, et al. Prediction of bending strength of Si_3N_4 using machine learning. Ceram Int. 2021;47(17):23919–23926. doi: 10.1016/j.ceramint.2021.05.100
[21] Furushima R, Maruyama Y, Nakashima Y, et al. Fracture toughness evaluation of silicon nitride from microstructures via convolutional neural network. J Am Ceram Soc. 2023;106(2):817–821. doi: 10.1111/jace.18795
[22] Furushima R, Nakashima Y, Maruyama Y, et al. Artificial intelligence-based determination of fracture toughness and bending strength of silicon nitride ceramics. J Am Ceram Soc. 2023;106(8):4944–4954. doi: 10.1111/jace.19147
[23] Zong X, Wu S, Lin K, et al. Advanced ceramics with integrated structures and functions: machine learning prediction and experimental verification. Ceram Int. 2024;50(13):24126–24138. doi: 10.1016/j.ceramint.2024.04.144
[24] Rezasefat M, Hogan JD. Virtual rapid prototyping of materials with deep learning: spatiotemporal stress fields prediction in ceramics employing convolutional neural networks and transfer learning. Virtual Phys Prototyp. 2024;19(1). doi: 10.1080/17452759.2024.2399186
[25] Wang A, He H, Xiong W, et al. Intelligent prediction of Si_3N_4 ceramic strength considering inherent defect characteristics. J Eur Ceram Soc. 2025;45(2):116900. doi: 10.1016/j.jeurceramsoc.2024.116900
[26] Maeda T, Osada T, Ozaki S. Reliability evaluation scheme for ceramics based on defect size distribution inversely estimated from standardized tests. J Am Ceram Soc. 2025;108(9):e20660. doi: 10.1111/jace.20660
[27] Eberhart R, Kennedy J. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, November 27-December 1; 1995. p. 1942–1948. doi: 10.1109/ICNN.1995.488968
[28] Kotz S, Nadarajah S. Extreme value distributions: theory and applications. London: Imperial College Press; 2000. doi: 10.1142/p191
[29] Coles S. An introduction to statistical modeling of extreme values. London: Springer-Verlag; 2001. doi: 10.1007/978-1-4471-3675-0

[30] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf. 2003;81(1):23–69. doi: 10.1016/S0951-8320(03)00058-9
[31] ASTM International. ASTM C1239-07 standard practice for reporting uniaxial strength data and estimating Weibull distribution parameters for advanced ceramics. ASTM Annu Book Stand; 2007.
[32] Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation. Departmental technical reports (CS). 2018. p. 1209. Available from: https://scholarworks.utep.edu/cs_techrep/1209/
[33] Joseph VR. Optimal ratio for data splitting. Stat Anal Data Min. 2022;15(4):531–538. doi: 10.1002/sam.11583
[34] Sivakumar M, Parthasarathy S, Padmapriya T. Trade-off between training and testing ratio in machine learning for medical image processing. PeerJ Comput Sci. 2024;10:e2245. doi: 10.7717/peerj-cs.2245
[35] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386–408. doi: 10.1037/h0042519
[36] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–536. doi: 10.1038/323533a0
[37] Werbos PJ. Backpropagation through time: what it does and how to do it. Proc IEEE. 1990;78(10):1550–1560. doi: 10.1109/5.58337
[38] Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16); Georgia (USA); 2016. p. 265–283.
[39] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12(85):2825–2830. doi: 10.5555/1953048.2078195
[40] Haltmeier M, Nguyen L. Regularization of inverse problems by neural networks. Cham: Springer; 2022. p. 1–29. doi: 10.1007/978-3-030-98661-2_81
[41] Jiao Q, Chen Y, Kim JH, et al. A machine learning perspective on the inverse indentation problem: uniqueness, surrogate modeling, and learning elasto-plastic properties from pile-up. J Mech Phys Solids. 2024;185:105557. doi: 10.1016/j.jmps.2024.105557
[42] Ren S, Padilla WJ, Malof J. Benchmarking deep inverse models over time, and the neural-adjoint method. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in neural information processing systems. Vol. 33. Vancouver, Canada: NeurIPS2020; 2020. p. 33–48. https://papers.nips.cc/paper/2020