Computational Materials Science 214 (2022) 111703

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Full Length Article

Knowledge database creation for design of polymer matrix composite

Hannah Huang [a,1], Satyajit Mojumder [b,1], Derick Suarez [a], Abdullah Al Amin [a], Mark Fleming [a], Wing Kam Liu [a,*]

[a] Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA
[b] Theoretical and Applied Mechanics Program, Northwestern University, Evanston, IL 60208, USA

A R T I C L E I N F O

Keywords:
Mechanistic data science
Unidirectional fiber
Polymer composite
Materials design
Mechanistic features
Dimension reduction

A B S T R A C T

We present a mechanistic data science (MDS) framework capable of building a composite knowledge database for composite materials design. The MDS framework systematically leverages data science to extract mechanistic knowledge from composite materials system. The composite response database is first generated for three matrix and four fiber combinations using a physics-based mechanistic reduced-order model. Next, the mechanistic features of the composites are identified by mechanistically analyzing the composites stress–strain responses. A relationship between the composite properties and the constituents' material features are established through a mechanics constrained data science-based learning process after representing materials in latent space, following a dimension reduction technique. We demonstrate the capability of predicting a composite materials system for target properties (material elastic properties, yield strength, resilience, toughness, and density) from the MDS created knowledge database. The MDS model is predictive with reasonable accuracy, and capable of identifying the materials system along with the tuning required to achieve desired composite properties. Development of such MDS framework can be exploited for fast materials system design, creating new opportunity for performance guided materials design.

1. Introduction

Fiber reinforced polymer (FRP) composites have become ubiquitous in industry due to their strong, stiff, and lightweight properties. Ranging from the frame of the Boeing 787 comprising of nearly 50 % advanced composites by weight [1] to utilization in repair and reinforcement of infrastructure [2], structural applications of FRPs [3] are now widespread. However, design of composite materials, including FRPs, remains a challenge. With various parameters to consider – constituent combinations, volume fraction, *meso*/micro/nano structure, temperature – performing in-lab experiments to determine mechanical properties necessary to make design choices calls for expensive equipment and significant time. Even computational methods, such as the Finite Element Method (FEM), are time consuming, computationally costly, and often rely on data from physical experiments for validation.

The heightened interest in Machine Learning (ML) along with the availability of necessary computational hardware has facilitated the widespread use of data science. Recently, ML techniques have gained popularity in the materials design space [4–9]. Application of ML techniques range in scale from molecular [4,5] to macro design [8]. For example, the Accelerated Metallurgy project used ML techniques, including neural network models, to rapidly suggest new alloy compositions, drastically decreasing development time [10].

Similar ML techniques have begun to be applied to composite material systems [11,12]. The applications of the ML techniques for composites vary greatly in scope, some aiming at the prediction of path dependent constitutive response [13] and optimal manufacturing processing parameters [14] through neural networks, and others aiming at reducing the computational cost of multiscale simulations through reduced-order models [3,15–18]. Most applications focus on efficiently predicting specific properties of a composite, such as the compressive strength of concrete [19], elastic stiffness of unidirectional fiber composites [20], effective thermal conductivity [21], interfacial shear strength [22], among others. These predictions largely focus on the forward problem, determining composite properties based on constituent information and micro/nano structure. There exists a gap in attempting to solve inverse problems [23], i.e., determining constituent information and micro/nano structure based on desired composite

* Corresponding author.
E-mail address: w-liu@northwestern.edu (W.K. Liu).
[1] Equal contribution.

https://doi.org/10.1016/j.commatsci.2022.111703
Received 14 May 2022; Received in revised form 27 July 2022; Accepted 30 July 2022
Available online 10 August 2022
0927-0256/

properties. Herein lies the opportunity to extend ML techniques for prediction of the inverse problem.

One of the major pitfalls of ML techniques is their data hungry nature, often requiring the collection or generation of extensive data to train a model. Even when such copious amounts of data are collected, it becomes cumbersome to extract meaningful features from the data. Previously, researchers have demonstrated machine learning models with a lower amount of data and no simulated data at all [24–27]. To ameliorate parts of these data collection challenges, we propose utilizing Self-Consistent Clustering Analysis (SCA) [28–30], a reduced-order modeling approach orders of magnitude faster than FEM, for data generation and Principal Component Analysis (PCA) for data reduction. Coupling data science techniques with scientific knowledge, we propose a Mechanistic Data Science (MDS) [31,32] framework for guiding the materials design process in FRP composites.

The paper is organized as follows: First, the paper will describe the key steps of the MDS framework: multimodal data generation and collection, mechanistic feature extraction, knowledge driven dimension reduction, mechanistic learning through regression, reduced-order surrogate model, and system and design. Then, each step will be expanded upon, particularly addressing how the data was generated and the methods of the data reduction and neural network surrogate modeling for composite materials design. Further, representative examples of the capability of the MDS framework for guiding the design of process of FRP composites for lightweight, high stiffness, and high toughness will be demonstrated. Finally, a conclusion is provided with possible future directions.

2. Overview of mechanistic data science framework

In this work, a mechanistic data science (MDS) framework is used to solve the inverse problem of composite materials design which does not solve the traditional inverse problem; instead, it focuses on "*training the inverse relations*". MDS is a general framework that can deal with three types of problems encountered in science and engineering [32]. Based on the available data and underlying physics, MDS tackles three types of problems: i) problems with inadequate physics but abundant data, ii) problems with limited physical understanding that have sufficient data for a highly accurate solution, iii) problems with well-known physics requiring minimal data for an acceptable solution.

To address these three types of problems, a MDS framework (see Fig. 1) is proposed that combines existing scientific knowledge, data science techniques, and engineering modeling to make informed decisions based on the problem scope [32]. The MDS framework follows six steps: i) multimodal data generation and collection, ii) mechanistic feature extraction, iii) knowledge-driven dimension reduction, iv) reduced-order model, v) mechanistic learning through regression and classification, and vi) system and design. Through these six steps, a systematic approach analyzes the available data to extract mechanistic knowledge of a system using data science tools and build a knowledge database. Not every problem will require all six steps, and the steps need not always follow a specific order. It is expected that the steps will closely interact with each other to build a knowledge database. We briefly introduce the six steps below, which will be further expanded with specific details in section 3 for the composite design problem.

- Multimodal data generation & collection: In this step, data can be gathered from existing databases, conclusive research, or can be produced from experimentation or simulation. This step also involves deciding where and how to sample data. Like other fields (e.g., finance, biology, etc.), data is not available in great extent for science and engineering problems. Experiments generate data in the form of images and sensor readings, which tend to be expensive and incomplete. Modeling and simulations can help generate new data to complement the incomplete dataset from experimentation. After the data is collected, it requires further processing such as formatting, cleansing, and wrangling to use it for feature extraction.
- Mechanistic feature extraction: After the data generation and collection process, mechanistic features [33] are extracted from the data using mathematical tools such as Fast-Fourier Transform (FFT), Short-time Fourier Transform (STFT), Wavelet Transform, etc. These mechanistic feature extraction step provides an efficient way for data management by discarding the unnecessary features of the problem. The feature extraction techniques are problem dependent, and it requires a domain expert knowledge to decide what mechanistic features can be extracted from the available data to solve the problem.
- Knowledge-driven dimension reduction: Design problems are often very high-dimensional as there are many design parameters to play with. Understanding the mechanistic features and further reducing them can reduce the problem size in reasonable way to handle it. Two common ways of dimension reduction are clustering and principal component analysis (PCA). Clustering represents similar datapoints as a single point or value. PCA takes highly correlated features and expresses them as a reduced feature on a linear basis.
- Mechanistic learning: In the mechanistic learning for regression or classification step, the functional relationships of the input and output features are established. Several machine learning techniques can be used for this step such as neural network, support vector machine, and Gaussian process regression to name a few. Neural networks are a versatile tool with universal approximation capability and can be customized to integrate the physics of the problem

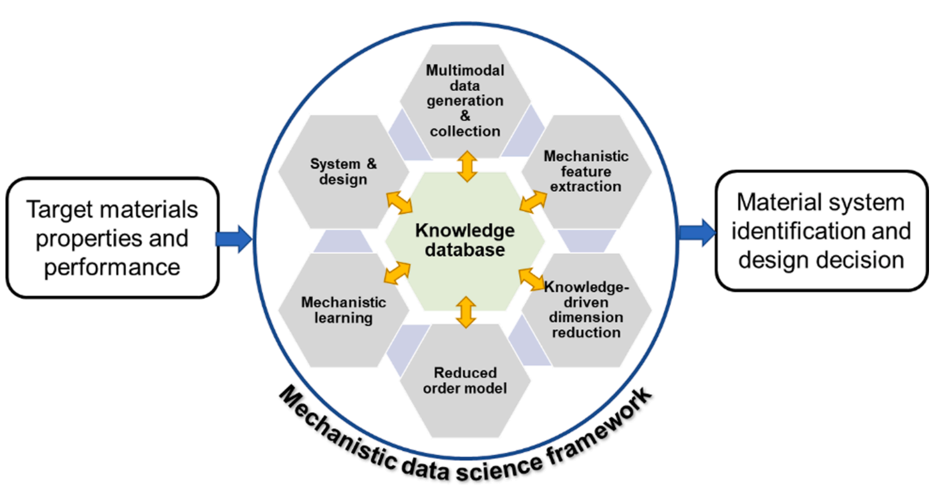

Fig. 1. Mechanistic data science framework with the six steps to create a knowledge database for materials design.

[26,34,35]. Using such customized neural networks can reduce both the data required and training time, while providing better predictions.
- Reduced-order model: Instead of solving the expensive full-scale model, a reduced-order model can be developed that can provide a faster solution with less computation effort. Such reduced-order models can be either mechanistic or surrogate type, which takes different design parameters as input and predicts the system response as output. In this work, we demonstrated the use of mechanistic reduced-order model Self-consistent Clustering Analysis (SCA) to generate the data for further analysis.
- System and design: Lastly, system and design is the real-world application of the model. In this step, the efficacy of the model in meeting the original design objectives is tested [36]. This may include testing the model against known data or using the model to make a prediction that can be tested experimentally. As a system and design problem, we aim to solve an inverse design problem where composite materials system is predicted for a given set of target properties. Previously, Thomas et al. [37,38] and Chuaqui et al. [39] have studied machine learning methods such as Bayesian framework for inverse design of the short reinforced composite materials for prediction of elastic properties with good accuracy.

3. Knowledge database creation using mechanistic data science steps

The six steps of the MDS framework described in the previous section build a knowledge database that can be used for design of composite materials system. In this section, the steps are described in detail to show how the framework works for a composite design problem.

3.1. System and design problem description

For a set of desired properties of polymer matrix composite, choosing the right combination of the fiber and polymer matrix is a challenging problem, as the design space or choice is unbounded. There is no unique solution of the material combinations; however, the most important consideration is whether the combinations closely satisfy the design requirement. If a relationship between the composite property space (e.g., elastic modulus, yield strength, resilience, toughness) and the matrix and fiber individual properties and microstructure design space (e.g., fiber volume fractions) can be established or "trained" using a deep-learning neural network, it can guide the designer to choose appropriate materials for a set of composite target properties. For instance, football helmets are made of composite materials that reduce weight significantly. However, for safety purposes, a designer needs to ensure the stiffness and toughness of the materials so that it can absorb the energy during impacts and reduce injury from a concussion [40]. In this work, we will use this example as a motivation to find materials that are lightweight, but have high stiffness, and high toughness properties. Using the steps of MDS, a knowledge database will be created to find the appropriate combination of three matrices—Polyamide-imides (PAI), Polycarbonate (PC), Polymethyl methacrylate (PMMA)—and four fibers—Carbon, E-Glass, Kevlar, S-Glass—at different fiber volume fractions and temperatures to satisfy a design requirement. Through this MDS framework, a hidden relationship of the composite properties and the constituent's matrix and fiber will be established for designing a new materials system with target properties. This idea will significantly reduce the experimental trial-and-error process to choose a materials system with less intuition.

3.2. Multimodal data generation and collection

As described in the problem description, the primary goal of this MDS demonstration is to find a composite material system having low weight, high stiffness, and high toughness. Thus, we need to build a composite materials response database by varying different fibers, matrices, volume fractions, and temperatures. The materials response database consists of the tensile stress–strain response along the transverse of fiber directions of unidirectional composite microstructures. For the design of experiment, three matrices (PAI, PC, PMMA) and four fibers (Carbon, E-Glass, Kevlar, and S-Glass) are chosen with a volume fraction varying from 1 to 50 % at five different temperatures ranging from 213 to 393 K for different matrices. Five different realizations of each microstructure descriptor (volume fraction of fiber) are considered to study the statistical variations in the response. This design of experiments gives us a total of 15,000 tensile simulations that have been performed using Self-consistent Clustering Analysis (SCA).

A schematic of the data generation process through SCA has been depicted in Fig. 2. Material microstructure and the properties of the fiber and matrix are taken as inputs of the SCA model to calculate the transverse mechanical response of the UD composites. Transverse loading is a critical loading mode for the fiber reinforced composites, where the matrix damage plays a critical role and very important to consider for the practical design purpose [41,42]. To model the temperature-dependent stress–strain behavior, the following assumptions have been made on the constituent's materials properties: i) matrix properties (elastic modulus, Poisson's ratio, and the hardening properties) change with temperature, ii) fiber properties variation are negligible as temperature changes and remains mostly constant [43], iii) matrix-fiber interface is perfectly bonded (no interphase) and temperature effects are not included. The fibers and temperature-dependent matrices properties are collected from the literatures [44,45] and the properties are given in supplementary information (Table S1).

SCA is a mechanistic reduced-order model that has offline database computation and an online prediction stage as shown in the Fig. 2. In the offline stage, the material microstructure is elastically loaded, and the local strain response is collected. Based on this strain distribution the materials are clustered using an unsupervised learning algorithm (K-means) [46]. This clustering reduces the problem domain degree of freedom from millions to hundreds. Once the clustering is done, the fiber–fiber, matrix–matrix, and fiber–matrix interaction tensor is computed and stored in an offline database. In this work, 32 clusters are used for the matrix phase, and 32 clusters are used for the fiber phase, which reduces the 4 million voxels of the original microstructure domain to 64 clusters. In the online stage, the offline database is used to solve Lipmann-Schwinger equation [47] for each cluster and their interactions, which gives the local stress–strain response for any arbitrary loading. The local stress–strain field is further homogenized, and an effective stress–strain response is recorded. A continuum damage model [29] has been adopted within the SCA model to capture the strain softening behavior of the materials (see section S2 of the supplementary information). Here, we loaded the microstructure in the fiber transverse direction only to obtain the stress–strain response curve as shown in Fig. 2. The details of the SCA method [28], damage model [29] and the material properties are provided in sections S1, S2, and S3 of the supplementary information.

A previous study showed that SCA is very efficient and accurate in predicting the homogenized response [3], with a speed up of around 5,000 times in the online stage. It is to be noted that the offline clustering does not change when the temperature is varied with the same microstructure. Therefore, the online steps only needed to be performed to enrich our database for different temperatures. This greatly accelerates the database generation process and allows us to study the parametric space of the design problem. These SCA-computed stress–strain data for different fiber and matrix combinations are stored in the knowledge database.

3.3. Mechanistic feature extraction

From the SCA generated stress–strain curves, several mechanistic features such as elastic modulus, yield strength, resilience, and modulus

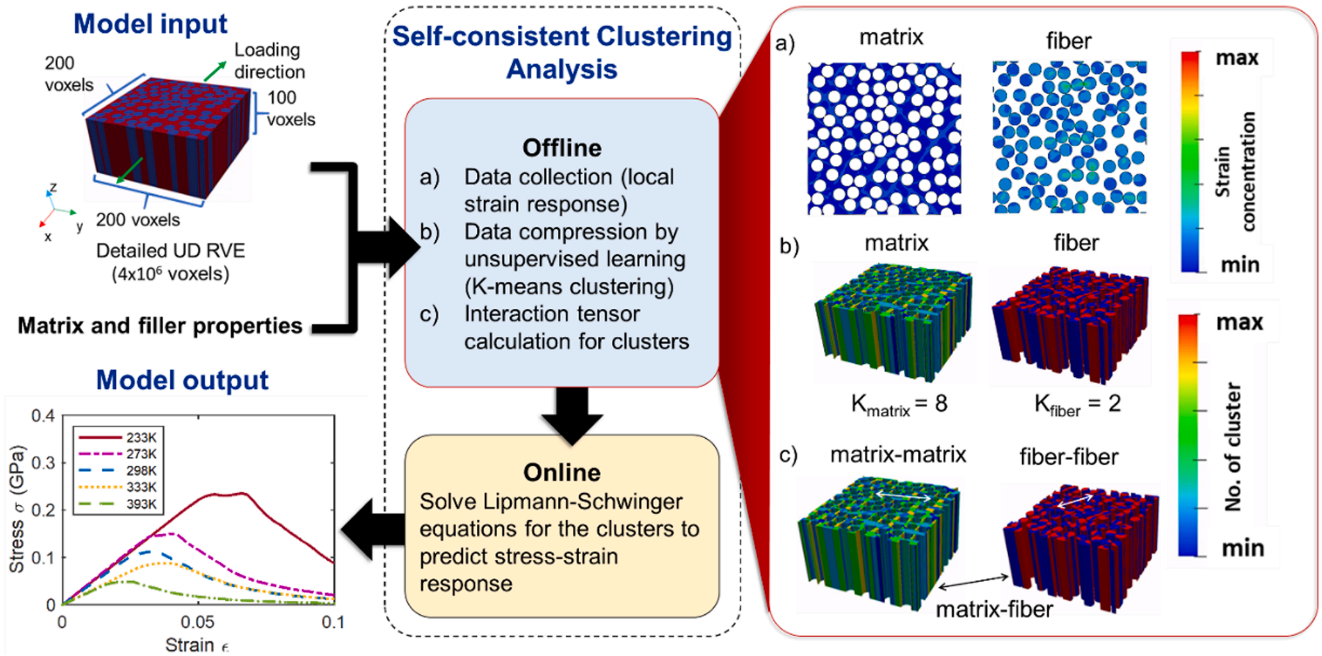

Fig. 2. Unidirectional fiber composites tensile data generation and collection process using the SCA method. (a) Local strain response, (b) clustering of fiber and matrix phase, (c) interaction tensor calculation steps of the offline stage are shown in the right panel of the figure. SCA outputs the stress–strain response of the composites as shown in bottom left.

of toughness (up to strain of 0.1) of the composite response are extracted. The feature extraction process from a sample transverse loading stress–strain curve is shown in Fig. 3. The modulus of elasticity is the stiffness of a material and is defined by the slope of the linear region. The 0.2 % strain offset method is used to calculate the yield strength of the composites. Resilience is the total energy a material can absorb before experiencing permanent deformation and comprises of the area under the elastic curve. The modulus of toughness is the energy absorbed by a material, and it is determined as the area under the stress–strain curve to a strain value of 0.1 for this work. The transverse modulus of elasticity (E_T), yield-strength (σ_y), resilience (U_R), and modulus of toughness (U_T) are the composite properties as a function of their matrix and fibers combinations, and their ranges for the present work are given in Table 1.

3.4. Knowledge-driven dimension reduction

The data generation and feature extraction resulted in a 15,000-ten-

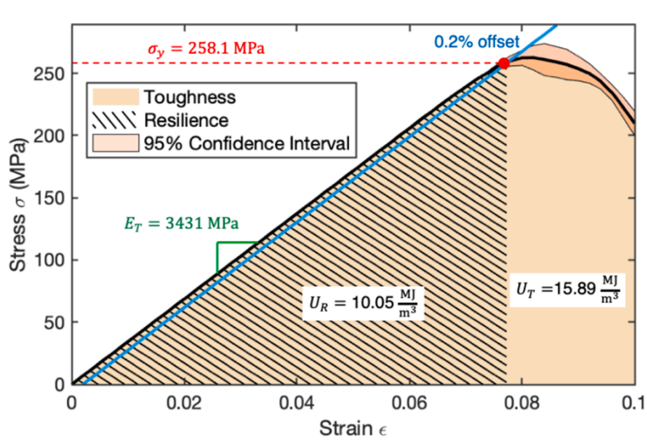

Fig. 3. An illustrative example of mechanistic feature extraction from the stress–strain response of 40 % volume fraction of E-Glass in PMMA at 233 K. Confidence intervals are calculated from the five realizations of the same microstructure volume fractions.

Table 1
Range of matrix and fiber properties taken as an input for SCA simulations and composite properties features extracted in mechanistic feature extraction step.

Materials structure features	Min	Max	Composite property features	Min	Max
Volume fraction ϕ	0.01	0.50	Composite density ρ_c (kg/m^3)	1183	2015
Matrix density ρ_m (kg/m^3)	1180	1490	Composite elastic modulus (MPa)	1089	8035
Fiber density ρ_f (kg/m^3)	1470	2540	Composite yield strength (MPa)	51	302
Matrix modulus of elasticity (kg/m^3)	1076	2743	Resilience (MJ/m^3)	0.40	13.17
Matrix Poisson's ratio	0.36	0.38	Toughness (MJ/m^3)	1.42	16.14
Matrix yield strength (MPa)	43	252			
Matrix hardening parameter	0.23	0.55			
Fiber modulus of elasticity 1 (MPa)	75,000	220,000			
Fiber modulus of elasticity 2 (MPa)	4200	8500			
Fiber modulus of elasticity 3 (MPa)	4200	8500			
Fiber Poisson's ratio 1	0.20	0.35			
Fiber Poisson's ratio 2	0.20	0.35			
Fiber Poisson's ratio 3	0.22	0.35			
Fiber modulus of shear 1 (MPa)	2900	36,000			
Fiber modulus of shear 2 (MPa)	2900	36,000			
Fiber modulus of shear 3 (MPa)	1500	36,000			
Temperature T (K)	213	393			

sile-test-sample database. This quantity was reduced by taking the average results (mean) of all five realizations, leading to a database with only 3,000 tensile samples cases. The averaging of the five realizations is done to reduce the data points having same microstructural descriptors. To define the input of matrix and fiber properties used for SCA simulations, 17 features are identified as inputs (see Table 1). Transverse modulus of elasticity, yield strength, resilience, and modulus of toughness are the output features extracted by analyzing the mechanistic features of SCA-generated stress–strain curves. Composite density (ρ_c) is another important property of the composite that is derived by applying the simple rule of mixture using the following expression,

$\rho_c = \rho_f \phi + \rho_m (1 - \phi), \# (1).$

where ϕ is the volume fractions of the fiber, and ρ_m and ρ_f are matrix and fiber densities, respectively.

The matrix and fiber features as shown in Fig. 4 can be reduced further to represent the materials combination in a reduced feature space using Principal Component Analysis (PCA). The steps of the PCA analysis are described in Supplementary section S4. Fig. 4 summarizes how the fourteen materials features were reduced using principal component analysis [46]. Fig. 4a show that 91.0 % of the dataset can be explained with just four principal components (following an elbow method to select number of PCA components); therefore, the original 14 materials features could be reduced to four principal components. These principal components will be referred to as L_1, L_2, L_3, and L_4 in which L_1 explains the component with the greatest eigenvalue. The eigenvalues of the principal components are summarized in Fig. 4b, in which it is apparent that a couple of principal components dominate the dataset. The space created by these four principal components is identified as the latent material feature space (see Fig. 5).

Once PCA is performed, we visualize the latent materials structure space and interpret what different principal components physically represent. The visualization identifies some materials cluster in the reduced space as shown in Fig. 5. Clearly, the problem dimension reduces significantly in the latent space with known (three matrices and four fibers we have chosen) and unknown material systems. The latent space shown in the figure manifests that L_1 and L_2 depend on the fibers,

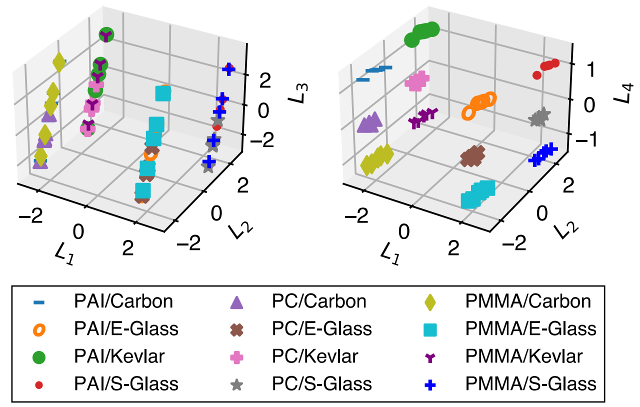

Fig. 5. Clustering pattern of the materials combination in the latent property space.

L_3 represents the temperature variations and L_4 represent the matrix materials in reduced space. The clustering nature and interpretability of the latent space variables enable us with new materials design capability by altering the fiber and matrix properties.

3.5. Mechanistic learning through regression

With the reduced representation of the matrix and fibers features, a relationship with the composite property space can be established through a neural network. The mapping of the latent material features and the microstructure features to the composite properties features is done here using a feed forward neural network (FFNN) [48]. The inputs of the FFNN are the composite properties feature (transverse modulus of elasticity, yield strength, resilience, modulus of toughness, and composite density), and the targets are the latent space materials features (L_1, L_2, L_3, and L_4), along with volume fraction and the matrix and fiber density as shown in the Fig. 6a. FFNN are deep neural networks (DNN)

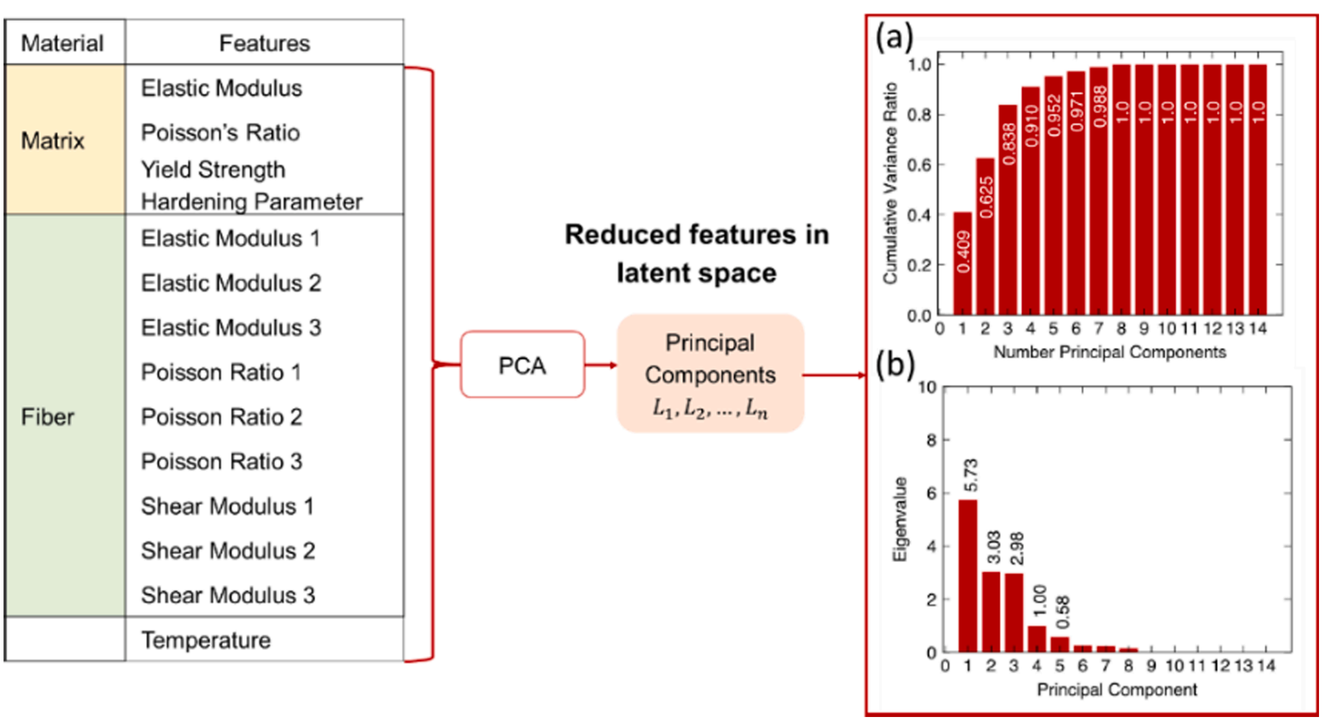

Fig. 4. Dimension reduction of the matrix and fiber features using PCA analysis. (a) shows that four principal components can capture 91.0% variance of the data, and their corresponding eigenvalues are shown in (b).

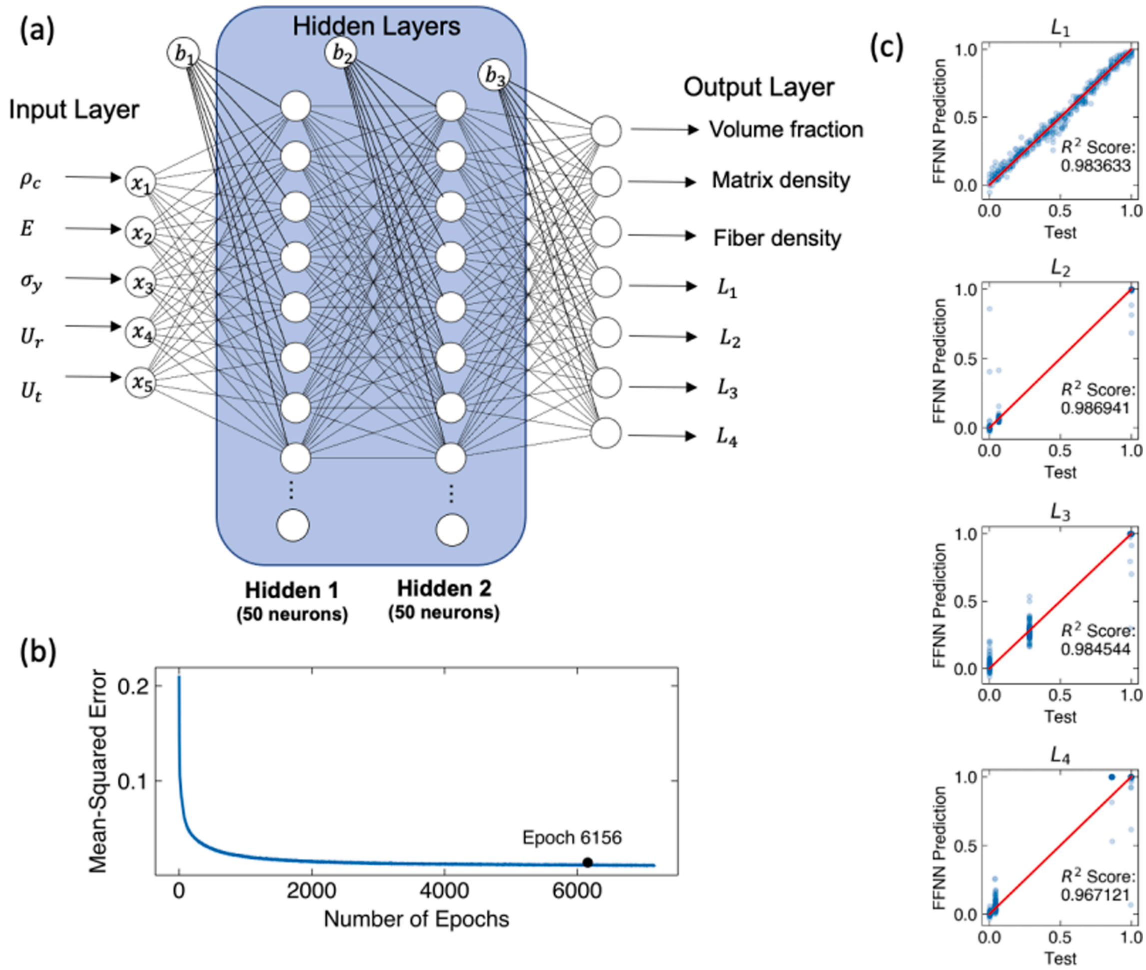

Fig. 6. (a) FFNN architectures, (b) training loss convergence, (c) prediction accuracy to relate composite properties with latent materials space using neural network regression.

with n hidden layers and can be described with the following standard equations.

$$z_l = \boldsymbol{b}_l + \boldsymbol{W}_l \boldsymbol{a}_{l-1}$$
$$\boldsymbol{a}_l = g(z_l)$$
$$\boldsymbol{Y} = \boldsymbol{b}_{n+1} + \boldsymbol{W}_{n+1}\boldsymbol{a}_n, \quad (2)$$

where, $\boldsymbol{b}_l$ and $\boldsymbol{W}_l$ are the bias vector and weight matrix of hidden layer l, g is the activation function, a_l and z_l are intermediate vector values at each layer, and Y is the output feature—in our case, the latent space material features: fiber volume fraction, fiber density, and matrix density.

Several neural network architectures with different activation functions are tested for the dataset and the details of these tests are summarized in the supplementary materials S5 of the paper. Based on our analysis, a final model having 2 hidden layers comprising 45 neurons in each layer, sigmoid as the hidden activation function, tangent sigmoid as the output activation function, and a dropout rate of 0.1 is chosen for the final neural network model. Coefficient of determination, R^2 is used as an accuracy measure of the neural network. The model uses 70 % of the data for training, 15 % for validation and 15 % for testing. Mean squared error has been set as the loss function of the neural network, and Adam optimizer is used as the optimization algorithm. No overfitting is ensured considering a simple neural network architecture with only two hidden layers. We test the R^2 value for the individual output and overall output vector. A plot of the neural network training loss and R^2 of the individual components of the reduced structure feature is shown in Fig. 6b. An overall root mean squared error of the testing set has been obtained as 0.0142 and value of R^2 predicted is 0.90, which means the model is good enough to predict the latent material features for a given new property set. For the individual component of the reduced space structure features we obtained R^2 as 0.98, 0.99, 0.98, 0.97 for L_1,L_2,L_3, and L_4 respectively (see Fig. 6c). The neural network model is a part of the knowledge database.

As mentioned earlier, the outputs of the FFNN are the latent materials structure features obtained through the PCA. Thus, inverse PCA needs to be performed on the FFNN outputs to get the materials features in the original space. However, this makes finding a specific material that matches all the properties exactly in the original space with 14 features difficult. Therefore, the designer might need to engineer the existing matrix and fibers to tune their properties or add more materials in the database to match closely with an existing material.

3.6. Knowledge database

The knowledge database consists of the stress–strain response data generated using SCA, the extracted mechanistic features, reduced representation using PCA in the dimension reduction process, and the trained neural network in the mechanistic regression step. Now, this knowledge database can be used to identify the specific material system for the target properties of composites. A workflow of the use of the knowledge database is shown in Fig. 7. First, the trained neural network will take the composite properties as input and output the latent space material features. Inverse PCA can be applied to the output latent space materials features from the neural network and the materials (matrix, fibers) properties can be identified in the original space. Further, they

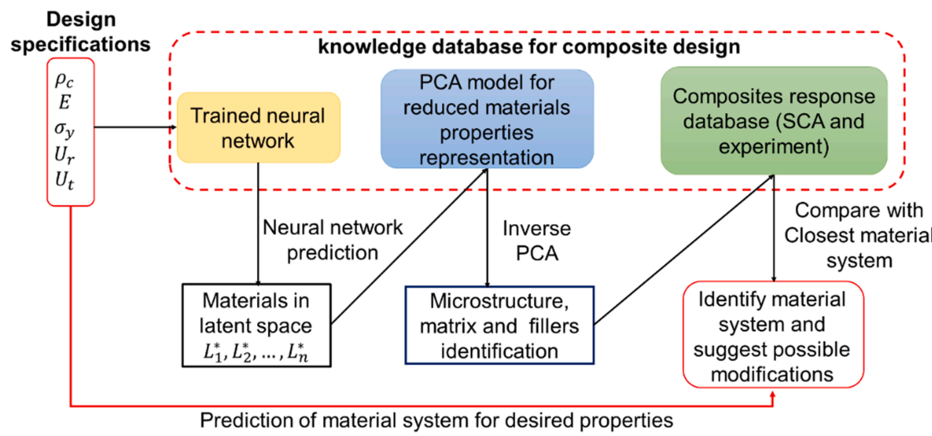

Fig. 7. Flowchart of how to predict a specific material system using the proposed model and a reduced-order surrogate model.

can be searched into the materials response database to find the deviation with the existing materials and propose the closest material systems that show similar properties.

4. Results and discussion

As shown in the previous section, MDS framework can build a knowledge database for the composite materials and further used in the material system prediction. In this section, we first validate the MDS framework predicted materials with the physics-based simulation and further explore design of the new materials system.

4.1. Validation of MDS prediction with physics-based model

As the data used for the MDS framework is generated using the SCA method discussed in section 3.2, it is reasonable to use SCA to validate the output of an MDS-predicted material system. To avoid bias in the MDS prediction, the material system chosen was not used in the training process. First, we choose a set of materials properties as an input for the SCA as given in Table 2. Following the stress–strain data generation and mechanistic feature extraction, the composite properties are calculated (refer to the SCA calculation in Table 3). Later, these composite properties are used as an input to the MDS framework. Using the knowledge database, the materials features developed in the previous section are predicted. However, it is not fair to compare the features of these materials directly with the Table 2 properties, as there is no unique solution to this inverse problem. We use these MDS predicted materials properties as the input properties for a new SCA calculation to compute the composite properties. These properties are referred to as the MDS prediction in Table 3. All the composite properties predicted by MDS are within 8 % difference with the SCA calculation, which demonstrates that the MDS approach has reasonable predictive capability of the inverse problem.

Table 2
Materials properties input to the SCA simulation.

MDS Matrix/Fiber Properties	Values
Volume Fraction	0.45
Matrix density (kg/m^3)	1490.00
Matrix modulus of elasticity (MPa)	1839.66
Matrix Poisson's ratio	0.38
Matrix yield strength (MPa)	85.82
Matrix hardening parameter	0.35
Temperature (K)	322.32
Fiber density (kg/m^3)	1772.83
Fiber modulus of elasticity 1 (MPa)	192122.21
Fiber modulus of elasticity 2 (MPa)	12513.31
Fiber modulus of elasticity 3 (MPa)	12513.31
Fiber Poisson's ratio 1	0.20
Fiber Poisson's ratio 2	0.20
Fiber Poisson's ratio 3	0.24
Fiber modulus of shear 1 (MPa)	7175.45
Fiber modulus of shear 2 (MPa)	7175.45
Fiber modulus of shear 3 (MPa)	3760.30

Table 3
Comparison of the composite properties calculated by SCA and MDS.

Composite Material Properties	SCA calculation	MDS prediction	Difference (%)
Composite density (kg/m^3)	1521.25	1616.98	6.29
Elastic modulus (MPa)	5237.31	4795.92	8.43
Yield strength (MPa)	117.86	117.70	0.14
Resilience (MJ/m^3)	1.53	1.65	7.76
Toughness (MJ/m^3)	3.17	3.09	2.65

4.2. Lightweight, high stiffness and high toughness composite search

As mentioned in section 3.1, the system and design problem considered in this work is to find materials combination for low weight, high strength, and high toughness applications. A correlation between different materials features and the composite properties features has been conducted and shown in Fig. 8. From Fig. 8, we can see the composite density and stiffness are positively correlated with the fiber volume fractions, whereas composite toughness is negatively correlated with the fiber volume fractions. It is also apparent from Fig. 8 that for transverse loading cases, the matrix material is mostly indicative of the composite's stiffness and toughness while the fiber choice is important for making the composite lightweight (Fig. 9).

To understand the interplay between stiffness and toughness, the transverse modulus and the modulus of toughness are plotted against different volume fractions of the fibers (see Fig. 9). All the properties shown in the Fig. 9 are in room temperature conditions (298 K). The property space for the different composites is a complex function and optimizing all the three criteria (lightweight, high strength and toughness) is not straightforward. For example, composites with PAI as the matrix material show higher toughness and stiffness properties than PC; however, composites with PC as the matrix material have lower densities and toughness. The stiffness, toughness, and density can be altered by changing the fiber volume fractions for all materials systems, in which increasing the volume fraction is typically positively correlated with stiffness but negatively with the toughness.

To understand the effect of the temperatures, we identified the materials systems showing lowest composite density, high stiffness, and high toughness. To evaluate the tradeoffs between stiffness, toughness, and density, and define the "best" mechanical properties, we found the material system that maximized the ratio $\frac{E_T U_T}{\rho_c}$. The identified material

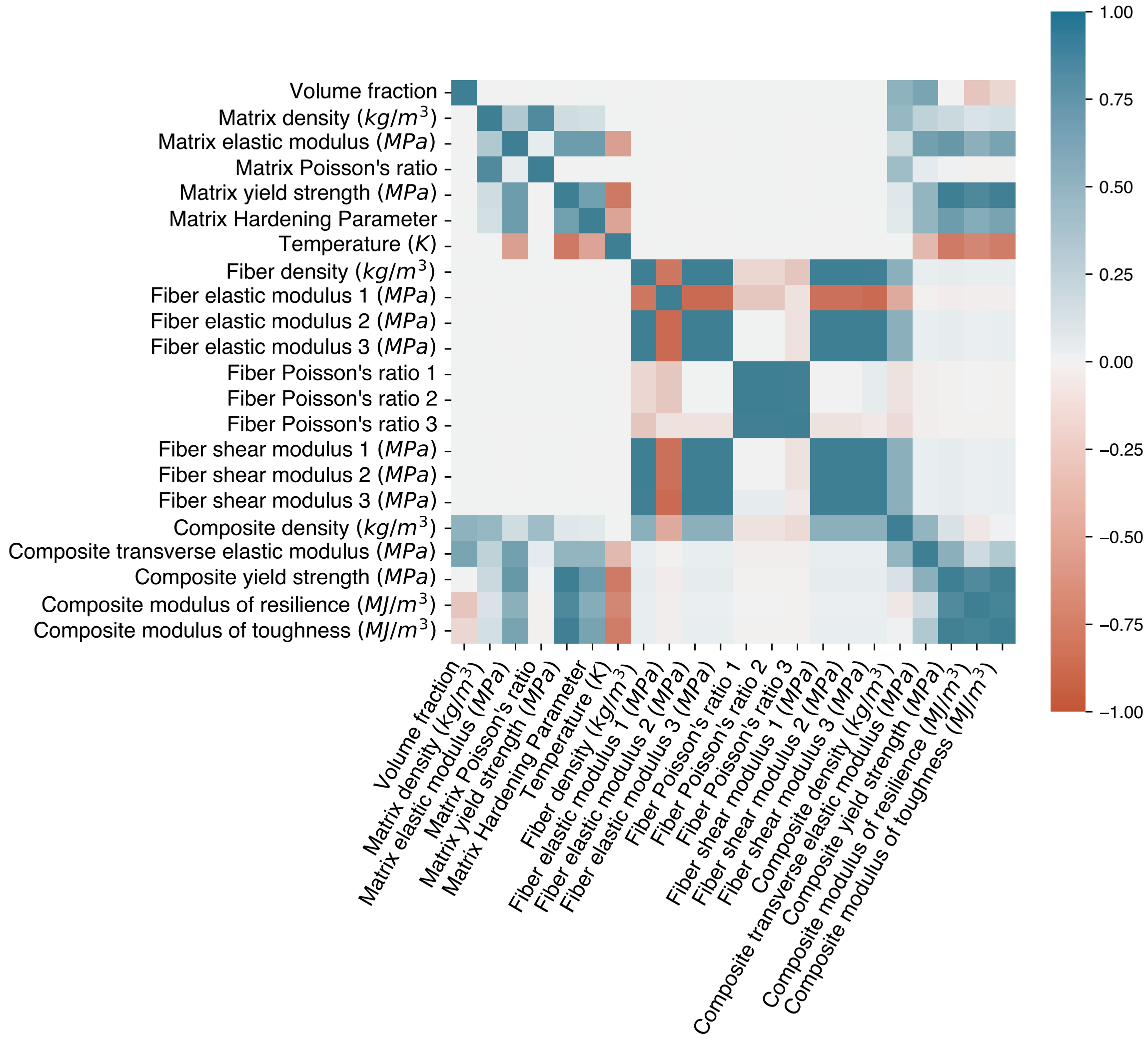

Fig. 8. Correlation plot for the materials features and composite properties features.

system at room temperature is PMMA/Kevlar at 50 % volume fraction. The result of varying temperature and volume fraction are shown in Fig. 10. Because the correlation matrix in Fig. 8 shows that the composite properties have a high correlation with the matrix material but low correlation with the fiber material (as the loading direction is transverse), the effect of temperature variance are also shown for PAI/Kevlar and PC/Kevlar in Fig. 10.

4.3. *Materials system identification from a design specification*

One of the major capabilities of the MDS built knowledge database is the inverse problem solving to identify the materials system from a given set of target properties. Here, we demonstrate this capability by setting the target properties as the system with the maximal combination of low composite density with high stiffness and toughness. A set of target properties with such attributes are shown in Table 4. For these set of target properties, the MDS predicted materials features are shown in Table 5. We also compare the MDS prediction with the existing database which shows that PMMA/Kevlar composites with 47 % volume fraction at 298 K temperature is the closest materials system to the MDS prediction. Using these PMMA/Kevlar properties we calculated the materials target properties as closest materials system using SCA as shown in Table 4 with the percentage difference in target properties if the predicted PMMA/Kevlar combinations are used. Notice that while the filler properties and even some of the matrix properties are very different from the target properties, the corresponding composite properties (obtained by running SCA on the closest matric/filler properties) are less than 10 % different than the target composite properties. If Kevlar is chosen as the fiber material, its materials properties need to tune up to attain the target properties of the composites. Therefore, materials designers can take Kevlar as an initial guess from the MDS analysis and engineer its properties to achieve the desired target properties. It is evident that there is not just one "best" material system. Other materials in the database can also be compared to the neural network output to find a material that has the desired performance. The next two "closest" materials are shown in Table 6, which are PMMA-Carbon and PMMA E-Glass, both at 298 K and 47% volume fraction.

The MDS framework opens new avenues for materials design (not limited to composites). Currently, it can predict only the materials system, within its database, but increasing the materials system in the database will improve its capability and make it further predictive. Also, the accuracy of the prediction depends on the accuracy of generated and collected data, extracted features, dimension reduction accuracy, neural network predictability. Having accurate individual step is important and it needs some prior experience to set all those different tools together. The knowledge database can be easily interfaced with commercial codes. That makes the MDS approach a powerful tool to solve other science and engineering problems.

5. Conclusion

A composite knowledge database creation process using an MDS

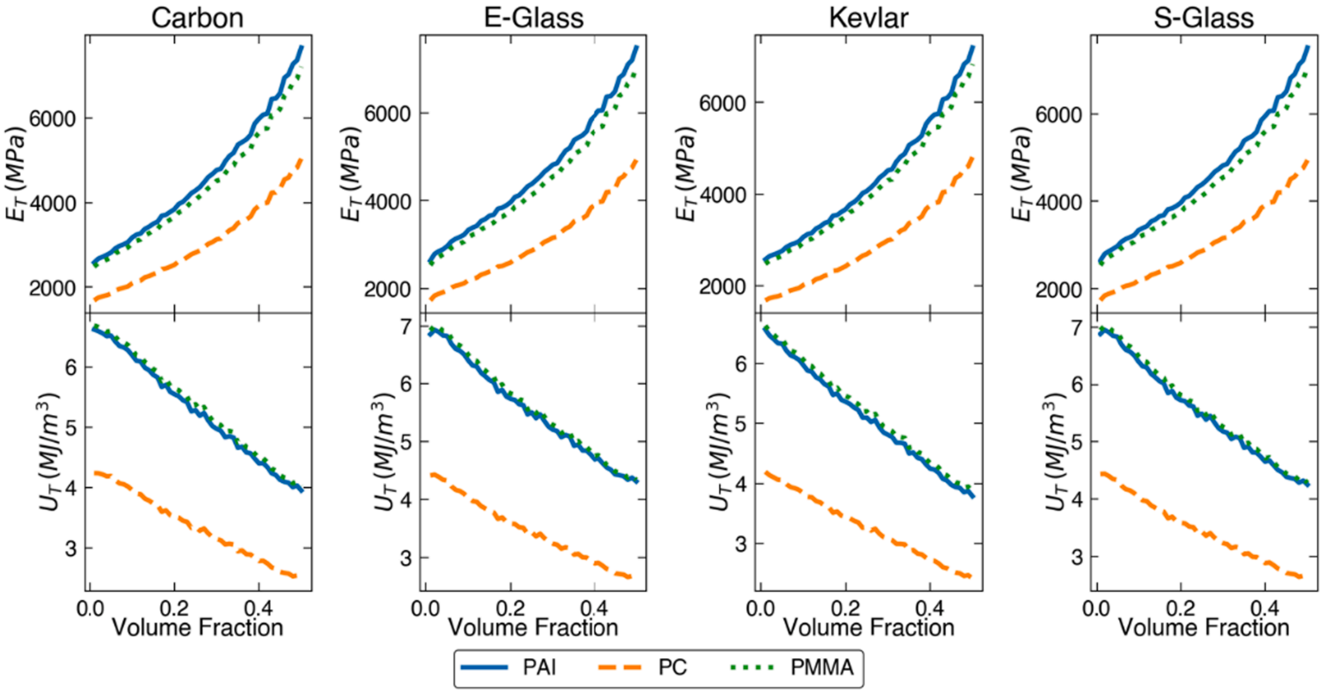

Fig. 9. Variation of composites transverse elastic modulus and modulus of toughness with fiber density for all matrix and fiber combinations in the dataset.

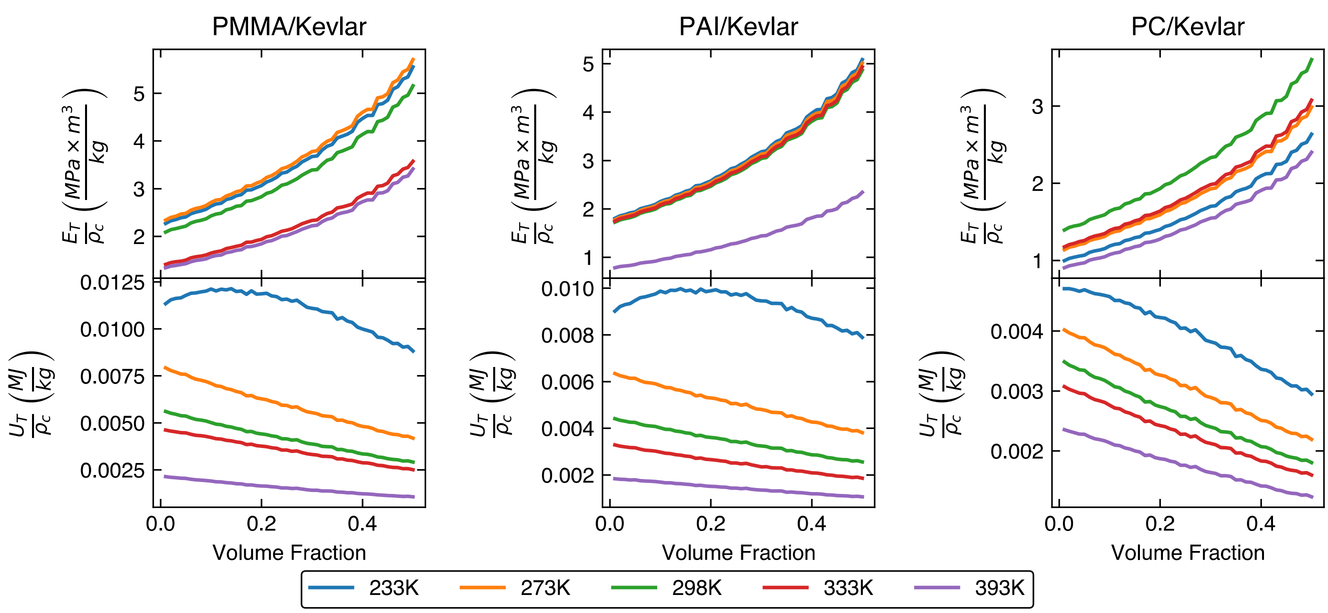

Fig. 10. Effect of temperature on transverse modulus and the modulus of toughness of composite materials varying matrix materials with Kevlar as the fiber material.

Table 4
Target properties for application in football helmet materials.

Composite Material Properties	Target properties	Closest properties (SCA)	% Difference
Composite density (kg/m^3)	1325.0	1316.3	0.66
Elastic modulus (MPa)	6833.47	6260.74	8.38
Yield strength (MPa)	123.16	121.90	1.02
Resilience (MJ/m^3)	1.294	1.375	6.26
Toughness (MJ/m^3)	3.860	3.999	3.60

framework is presented in this work. The MDS framework utilized the mechanistic reduced-order model SCA to generate the composites response database for different matrices (PAI, PC, PMMA) and fibers (Carbon, E-Glass, Kevlar, S-glass) combinations at varied temperatures at a significantly reduced computational time compared to experiments or numerical models. Mechanistic features are extracted from the generated composite data and further reduced using mechanistic relations of rule of mixture. A high dimensional material features space has been reduced using principal component analysis. Deep neural networks are further used to learn the relation of the constituent materials features and composite properties through regression. A knowledge database integrates all these mechanistic knowledges obtained through mechanistic feature extraction, dimension reduction, and deep

Table 5
Comparison of the MDS predicted materials features with the closest materials system in our database. The closest dataset material is 47% volume fraction PMMA Kevlar at 298 K.

Materials input properties	MDS Prediction	Closest dataset material
Volume Fraction	0.47	0.47
Matrix density (kg/m^3)	1179.79	1180.00
Matrix modulus of elasticity (MPa)	2102.49	2448.77
Matrix Poisson's ratio	0.36	0.36
Matrix yield strength (MPa)	130.52	106.56
Matrix hardening parameter	0.41	0.44
Temperature (K)	287.79	298.00
Fiber density (kg/m^3)	1482.24	1470.00
Fiber modulus of elasticity 1 (MPa)	171788.20	150000.00
Fiber modulus of elasticity 2 (MPa)	5465.43	4200.00
Fiber modulus of elasticity 3 (MPa)	5465.43	4200.00
Fiber Poisson's ratio 1	0.35	0.35
Fiber Poisson's ratio 2	0.35	0.35
Fiber Poisson's ratio 3	0.36	0.35
Fiber modulus of shear 1 (MPa)	4402.29	2900.00
Fiber modulus of shear 2 (MPa)	4402.29	2900.00
Fiber modulus of shear 3 (MPa)	2243.48	1500.00

Table 6
Additional materials in the dataset that are close to the MDS prediction.

Materials input properties	PMMA Carbon 298 K	PMMA E-Glass 298 K
Volume Fraction	0.47	0.47
Matrix density (kg/m^3)	1180	1180
Matrix modulus of elasticity (MPa)	2448.77	2448.77
Matrix Poisson's ratio	0.36	0.36
Matrix yield strength (MPa)	106.5574	106.5574
Matrix hardening parameter	0.444529	0.444529
Temperature (K)	298	298
Fiber density (kg/m^3)	1770	2540
Fiber modulus of elasticity 1 (MPa)	220,000	75,000
Fiber modulus of elasticity 2 (MPa)	14,000	75,000
Fiber modulus of elasticity 3 (MPa)	14,000	75,000
Fiber Poisson's ratio 1	0.2	0.2
Fiber Poisson's ratio 2	0.2	0.2
Fiber Poisson's ratio 3	0.25	0.22
Fiber modulus of shear 1 (MPa)	9000	30,000
Fiber modulus of shear 2 (MPa)	9000	30,000
Fiber modulus of shear 3 (MPa)	4600	30,000

neural network regression. Through a workout example, the predictive capability of the MDS presented shows reasonable accuracy with the physics-based simulation. Additionally, with a limited number of material systems, a composite materials system with low weight, high stiffness and high toughness has been identified. MDS can also be useful to solve the inverse problem of materials design to find the materials combination for assesment of target properties.

It is expected that MDS can reduce the experimental trial and error process and provide design guidelines to the materials designer. The MDS framework can also be used for design of other materials system and not limited to composite materials design only. MDS framework can open new avenues in the materials design paradigm that will cut down the materials development to deployment time significantly.

CRediT authorship contribution statement

Hannah Huang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft. **Satyajit Mojumder:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Derick Suarez:** Software, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Abdullah Al Amin:** Writing – original draft, Writing – review & editing. **Mark Fleming:** Conceptualization, Writing – review & editing, Supervision. **Wing Kam Liu:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The SCA generated composites stress-strain dataset and the codes are available on: https://github.com/hannahhuang00/MDS_Composite.

Acknowledgments

H.H. would like to acknowledge the NSF REU program under the Grant No. MOMS/CMMI-1762035. S.M. and W. K. L. thankfully acknowledge the support provided by AFOSR (FA9550-18-1-0381). D. S., A.A.A., and W.K.L. acknowledge the support of the United States National Science Foundation (NSF) under Grant No. MOMS/CMMI-1762035. D.S. also gratefully acknowledges the Walter P. Murphy fellowship provided to first-year graduate students at Northwestern University.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.commatsci.2022.111703.

References

[1] V. Giurgiutiu, Structural Health Monitoring of Aerospace Composites, Academic Press, 2015.
[2] H.-C. Wu, C.D. Eamon, 2 - Fiber-reinforced polymer composites, in: H.-C. Wu, C.D. Eamon (Eds.), Strengthening of Concrete Structures using Fiber Reinforced Polymers (FRP), Woodhead Publishing, 2017, pp. 11–17, doi: 10.1016/B978-0-08-100636-8.00002-8.
[3] J. Gao, M. Shakoor, G. Domel, M. Merzkirch, G. Zhou, D. Zeng, X. Su, W.K. Liu, Predictive multiscale modeling for Unidirectional Carbon Fiber Reinforced Polymers, Compos. Sci. Technol. 186 (2020) 107922, https://doi.org/10.1016/j.compscitech.2019.107922.
[4] L. Tao, G. Chen, Y. Li, Machine learning discovery of high-temperature polymers, Patterns 2 (4) (2021) 100225, https://doi.org/10.1016/j.patter.2021.100225.
[5] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547–555, https://doi.org/10.1038/s41586-018-0337-2.
[6] J.E. Gubernatis, T. Lookman, Machine learning in materials design and discovery: examples from the present and suggestions for the future, Phys. Rev. Mater. 2 (2018), 120301, https://doi.org/10.1103/PhysRevMaterials.2.120301.
[7] T. Zhou, Z. Song, K. Sundmacher, Big Data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design, Engineering 5 (2019) 1017–1026, https://doi.org/10.1016/j.eng.2019.02.011.
[8] K. Guo, Z. Yang, C.-H. Yu, M.J. Buehler, Artificial intelligence and machine learning in design of mechanical materials, Mater. Horiz. 8 (2021) 1153–1172, https://doi.org/10.1039/D0MH01451F.
[9] R. Vasudevan, G. Pilania, P.V. Balachandran, Machine learning for materials design and discovery, J. Appl. Phys. 129 (7) (2021) 070401, https://doi.org/10.1063/5.0043300.
[10] Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning, J. Materiomics 3 (2017) 159–177, https://doi.org/10.1016/j.jmat.2017.08.002.
[11] C.-T. Chen, G.X. Gu, Machine learning for composite materials, MRS Commun. 9 (2019) 556–566, https://doi.org/10.1557/mrc.2019.32.
[12] P. Pattnaik, A. Sharma, M. Choudhary, V. Singh, P. Agarwal, V. Kukshal, Role of machine learning in the field of Fiber reinforced polymer composites: a preliminary discussion, Mater. Today:. Proc. 44 (2021) 4703–4708, https://doi.org/10.1016/j.matpr.2020.11.026.

[13] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, J. Cao, M.A. Bessa, Deep learning predicts path-dependent plasticity, PNAS 116 (2019) 26414–26420, https://doi.org/10.1073/pnas.1911815116.
[14] Robbany F, Pramujati B, Suhardjono, Effendi MK, Soepangkat BOP, Norcahyo R. Multi response prediction of cutting force and delamination in carbon fiber reinforced polymer using backpropagation neural network-genetic algorithm. AIP Conf. Proc. 2114 (2019) 030012, doi: 10.1063/1.5112416.
[15] X. Han, J. Gao, M. Fleming, C. Xu, W. Xie, S. Meng, W.K. Liu, Efficient multiscale modeling for woven composites based on self-consistent clustering analysis, Comput. Methods Appl. Mech. Eng. 364 (2020) 112929.
[16] C. He, J. Ge, J. Gao, J. Liu, H. Chen, W.K. Liu, D. Fang, From microscale to mesoscale: the non-linear behavior prediction of 3D braided composites based on the SCA2 concurrent multiscale simulation, Compos. Sci. Technol. 213 (2021) 108947, https://doi.org/10.1016/j.compscitech.2021.108947.
[17] J. Gao, S. Mojumder, W. Zhang, H. Li, D. Suarez, C. He, J. Cao, W.K. Liu, Concurrent n-scale modeling for non-orthogonal woven composite, Comput. Mech. 1 (2022) 14, https://doi.org/10.1007/s00466-022-02199-2.
[18] S. Mojumder, J. Gao, W.K. Liu, Self-consistent clustering analysis for modeling of theromelastic heterogeneous materials, AIP Conf. Proc. 2324 (2021), 030029, https://doi.org/10.1063/5.0038297.
[19] F. Khademi, M. Akbari, S.M. Jamal, M. Nikoo, Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete, Front. Struct. Civ. Eng. 11 (2017) 90–99, https://doi.org/10.1007/s11709-016-0363-9.
[20] M.V. Pathan, S.A. Ponnusami, J. Pathan, R. Pitisongsawat, B. Erice, N. Petrinic, et al., Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning, Sci. Rep. 9 (2019) 13964, https://doi.org/10.1038/s41598-019-50144-w.
[21] Q. Rong, H. Wei, X. Huang, H. Bao, Predicting the effective thermal conductivity of composites from cross sections images using deep learning methods, Compos. Sci. Technol. 184 (2019), 107861, https://doi.org/10.1016/j.compscitech.2019.107861.
[22] A. Rahman, P. Deshpande, M.S. Radue, G.M. Odegard, S. Gowtham, S. Ghosh, A. D. Spear, A machine learning framework for predicting the shear strength of carbon nanotube-polymer interfaces based on molecular dynamics simulation data, Compos. Sci. Technol. 207 (2021) 108627.
[23] L. Lu, M. Dao, P. Kumar, U. Ramamurty, G.E. Karniadakis, S. Suresh, Extraction of mechanical properties of materials through deep learning from instrumented indentation, PNAS 117 (2020) 7052–7062, https://doi.org/10.1073/pnas.1922210117.
[24] R. Eggersmann, L. Stainier, M. Ortiz, S. Reese, Model-free data-driven computational mechanics enhanced by tensor voting, Comput. Methods Appl. Mech. Eng. 373 (2021), 113499, https://doi.org/10.1016/j.cma.2020.113499.
[25] S. Karumuri, R. Tripathy, I. Bilionis, J. Panchal, Simulator-free solution of high-dimensional stochastic elliptic partial differential equations using deep neural networks, J. Comput. Phys. 404 (2020), 109120, https://doi.org/10.1016/j.jcp.2019.109120.
[26] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, https://doi.org/10.1016/j.jcp.2018.10.045.
[27] M.A. Nabian, H. Meidani, A deep learning solution approach for high-dimensional random differential equations, Probab. Eng. Mech. 57 (2019) 14–25, https://doi.org/10.1016/j.probengmech.2019.05.001.
[28] Z. Liu, M.A. Bessa, W.K. Liu, Self-consistent clustering analysis: an efficient multi-scale scheme for inelastic heterogeneous materials, Comput. Methods Appl. Mech. Eng. 306 (2016) 319–341, https://doi.org/10.1016/j.cma.2016.04.004.
[29] Z. Liu, M. Fleming, W.K. Liu, Microstructural material database for self-consistent clustering analysis of elastoplastic strain softening materials, Comput. Methods Appl. Mech. Eng. 330 (2018) 547–577, https://doi.org/10.1016/j.cma.2017.11.005.
[30] M.A. Bessa, R. Bostanabad, Z. Liu, A. Hu, D.W. Apley, C. Brinson, W. Chen, W. Liu, A framework for data-driven analysis of materials under uncertainty: countering the curse of dimensionality, Comput. Methods Appl. Mech. Eng. 320 (2017) 633–667.
[31] S. Mojumder, L. Tao, Y. Li, W.K. Liu, Mechanistic data science for modeling and design of aerospace composite materials, arXiv preprint arXiv:2112.00968.
[32] W.K. Liu, Z. Gan, M. Fleming (Eds.), Mechanistic Data Science for STEM Education and Applications, Springer International Publishing, Cham, 2021.
[33] W.K. Liu, Z. Gan, M. Fleming, in: Mechanistic Data Science for STEM Education and Applications, Springer International Publishing, Cham, 2021, pp. 89–129.
[34] Z. Zhang, G.X. Gu, Physics-informed deep learning for digital materials, Theor. Appl. Mech. Lett. 11 (1) (2021) 100220, https://doi.org/10.1016/j.taml.2021.100220.
[35] S. Saha, Z. Gan, L. Cheng, J. Gao, O.L. Kafka, X. Xie, H. Li, M. Tajdari, H.A. Kim, W. K. Liu, Hierarchical Deep Learning Neural Network (HiDeNN): An artificial intelligence (AI) framework for computational science and engineering, Comput. Methods Appl. Mech. Eng. 373 (2021) 113452, https://doi.org/10.1016/j.cma.2020.113452.
[36] W.K. Liu, Z. Gan, M. Fleming, System and design, in: W.K. Liu, Z. Gan, M. Fleming (Eds.), Mechanistic Data Science for STEM Education and Applications, Springer International Publishing, Cham, 2021, pp. 215–66, doi: 10.1007/978-3-030-87832-0_7.
[37] A.J. Thomas, E. Barocio, R.B. Pipes, A machine learning approach to determine the elastic properties of printed fiber-reinforced polymers, Compos. Sci. Technol. 220 (2022), 109293, https://doi.org/10.1016/j.compscitech.2022.109293.
[38] A.J. Thomas, E. Barocio, I. Bilionis, R.B. Pipes, Bayesian inference of fiber orientation and polymer properties in short fiber-reinforced polymer composites, Compos. Sci. Technol. 228 (2022), 109630, https://doi.org/10.1016/j.compscitech.2022.109630.
[39] T.R.C. Chuaqui, A.T. Rhead, R. Butler, C. Scarth, A data-driven Bayesian optimisation framework for the design and stacking sequence selection of increased notched strength laminates, Compos. B Eng. 226 (2021), 109347, https://doi.org/10.1016/j.compositesb.2021.109347.
[40] D.W. Sproule, E.T. Campolettano, S. Rowson, Football helmet impact standards in relation to on-field impacts, Proc. IMechE 231 (4) (2017) 317–323.
[41] R. Cai, T. Jin, The effect of microstructure of unidirectional fibre-reinforced composites on mechanical properties under transverse loading: A review, J. Reinf. Plast. Compos. 37 (2018) 1360–1377, https://doi.org/10.1177/0731684418796308.
[42] V.I. Kushch, S.V. Shmegera, P. Brøndsted, L. Mishnaevsky, Numerical simulation of progressive debonding in fiber reinforced composite under transverse loading, Int. J. Eng. Sci. 49 (1) (2011) 17–29.
[43] F. Zhou, J. Zhang, S. Song, D. Yang, C. Wang, Effect of temperature on material properties of carbon fiber reinforced polymer (CFRP) tendons: experiments and model assessment, Materials 12 (2019) 1025, https://doi.org/10.3390/ma12071025.
[44] J. Richeton, S. Ahzi, K.S. Vecchio, F.C. Jiang, R.R. Adharapurapu, Influence of temperature and strain rate on the mechanical behavior of three amorphous polymers: characterization and modeling of the compressive yield stress, Int. J. Solids Struct. 43 (2006) 2318–2335, https://doi.org/10.1016/j.ijsolstr.2005.06.040.
[45] Composite Materials and Structures (n.d.). Available from: <http://www.ae.iitkgp.ac.in/ebooks/chapter4.html> (accessed April 29, 2022).
[46] W.K. Liu, Z. Gan, M. Fleming, Knowledge-driven dimension reduction and reduced order surrogate models, in: W.K. Liu, Z. Gan, M. Fleming (Ed.), Mechanistic Data Science for STEM Education and Applications, Springer International Publishing, Cham, 2021, pp. 131–170, doi: 10.1007/978-3-030-87832-0_5.
[47] H. Moulinec, P. Suquet, A numerical method for computing the overall response of nonlinear composites with complex microstructure, Comput. Methods Appl. Mech. Eng. 157 (1998) 69–94, https://doi.org/10.1016/S0045-7825(97)00218-1.
[48] W.K. Liu, Z. Gan, M. Fleming, Deep learning for regression and classification, in: W. K. Liu, Z. Gan, M. Fleming (Eds.), Mechanistic Data Science for STEM Education and Applications, Springer International Publishing, Cham, 2021, pp. 171–214, doi: 10.1007/978-3-030-87832-0_6.