

RESEARCH ARTICLE | FEBRUARY 22 2024

## Integrating uncertainty into deep learning models for enhanced prediction of nanocomposite materials' mechanical properties

Yuheng Wang; Guang Lin ; Shengfeng Yang  



*APL Mach. Learn.* 2, 016112 (2024)

<https://doi.org/10.1063/5.0177062>



CrossMark



### APL Quantum

Bridging fundamental quantum research with technological applications

## Now Open for Submissions

No Article Processing Charges (APCs) through 2024

**Submit Today**

 AIP  
Publishing

# Integrating uncertainty into deep learning models for enhanced prediction of nanocomposite materials' mechanical properties

Cite as: APL Mach. Learn. 2, 016112 (2024); doi: 10.1063/5.0177062

Submitted: 19 September 2023 • Accepted: 29 January 2024 •

Published Online: 22 February 2024



Yuheng Wang,<sup>1,2</sup> Guang Lin,<sup>2,3</sup>  and Shengfeng Yang,<sup>1,2,a)</sup> 

## AFFILIATIONS

<sup>1</sup> School of Mechanical Engineering, Purdue University in Indianapolis, Indianapolis, Indiana 46202, USA

<sup>2</sup> School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907, USA

<sup>3</sup> Department of Mathematics, Purdue University, West Lafayette, Indiana 47907, USA

<sup>a)</sup> Author to whom correspondence should be addressed: [shengfengyang@purdue.edu](mailto:shengfengyang@purdue.edu)

## ABSTRACT

In this paper, we present a novel deep-learning framework that incorporates quantified uncertainty for predicting the mechanical properties of nanocomposite materials, specifically taking into account their morphology and composition. Due to the intricate microstructures of nanocomposites and their dynamic changes under diverse conditions, traditional methods, such as molecular dynamics simulations, often impose significant computational burdens. Our machine learning models, trained on comprehensive material datasets, provide a lower computational cost alternative, facilitating rapid exploration of design spaces and more reliable predictions. We employ both convolutional neural networks and feedforward neural networks for our predictions, training separate models for yield strength and ultimate tensile strength. Furthermore, we integrate uncertainty quantification into our models, thereby providing confidence intervals for our predictions and making them more reliable. This study paves the way for advancements in predicting the properties of nanocomposite materials and could potentially be expanded to cover a broad spectrum of materials in the future.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0177062>

## I. INTRODUCTION

Nanocomposite materials, which exhibit unique properties, have become prevalent in a wide range of engineering applications, including but not limited to the biomedical and aerospace sectors.<sup>1–4</sup> The task of understanding and predicting the properties and behaviors of these materials poses a substantial challenge to materials research due to the intricate microstructures of nanocomposites and their dynamic changes under diverse conditions. Traditionally, the prediction of mechanical properties of nanomaterials has relied heavily on empirical equations. A notable example is the Hall–Petch relation,<sup>5</sup> which correlates yield stress with the average grain size, indicating that smaller grains typically strengthen a material by obstructing dislocation motion. However, these types of empirical models do not consider crucial aspects such as material composition and internal defects. This gap in considering a broader range

of factors affecting mechanical properties can be addressed by high-fidelity modeling, which offers a more comprehensive approach by incorporating these additional factors into the prediction process. High-fidelity modeling techniques,<sup>6</sup> such as Molecular Dynamics (MD) simulations and first-principle Density Functional Theory (DFT) calculations,<sup>7</sup> while effective, also impose a significant computational burden when attempting to predict structure–property relationships in nanocomposite materials. This immense computational demand often limits the exploration of the design space inherent to these materials. As a result, Machine Learning (ML) methods,<sup>8</sup> with their capability for predicting material properties at a markedly lower computational cost, are becoming increasingly significant in materials research.<sup>9–11</sup> Consequently, ML models,<sup>12</sup> which are trained on extensive material datasets to anticipate structure–property relationships, are providing new avenues for

the rapid exploration of large design spaces.<sup>13</sup> These methodologies facilitate the development of advanced nanocomposite materials with highly tailored properties.<sup>8,14</sup>

ML models, utilizing training data obtained from continuum-level Finite Element (FE) simulations, have been developed to predict a broad range of mechanical properties for composite materials.<sup>15–22</sup> Yang *et al.*, for instance, applied conditional generative adversarial networks (cGANs) to anticipate stress and strain fields in hierarchical composites.<sup>18,19</sup> Likewise, Bhaduri *et al.* utilized Convolutional Neural Networks (CNN) to predict the stress field of fiber-reinforced composites.<sup>15</sup> Drawing inspiration from the Modified National Institute of Standards and Technology (MNIST) dataset widely used in the computer vision research community,<sup>23</sup> Lejeune *et al.* developed a mechanically relevant equivalent, “Mechanical MNIST,”<sup>16,17</sup> for heterogeneous materials using the finite element method. Leveraging these datasets, they used various deep-learning models, including CNN, to predict total changes in strain energy<sup>16</sup> and full-field displacement.<sup>17</sup>

However, ML models trained on continuum-level datasets may not accurately characterize nanocomposite materials’ behavior as these simulations often overlook the discreteness of matter and defects predominantly found at the nanoscale. In response, researchers have trained ML models on datasets derived from atomistic simulations, like MD simulations. These models have been employed to predict atomic-level material quantities, such as grain boundary energy<sup>24,25</sup> and segregation energy.<sup>26</sup> Beyond predicting individual quantities, ML models have been further developed to predict full-field atomic scale quantities like stress distribution<sup>27,28</sup> and fracture behavior.<sup>29–32</sup> For instance, Dewapriya *et al.* utilized cGANs to predict the behavior of defects in graphene materials and forecast the corresponding mechanical stress distribution, leveraging training data from MD simulations.<sup>28</sup>

Owing to the various sources of uncertainty inherent in atomistic simulations at the nanoscale, considering these uncertainties becomes essential for providing more reliable predictions of nanocomposite materials’ properties.<sup>33,34</sup> Previous work by Winovich *et al.* integrated quantified uncertainty into the development of CNN models for solving partial differential equations on different domains,<sup>35</sup> and the Gaussian Process method has been integrated into CNN models to obtain the model uncertainty by Yin and Du.<sup>36</sup> Nonetheless, limited research has been conducted concerning the inclusion of uncertainty in predicting structure–property relations for materials at the nanoscale. Tavaza *et al.*,<sup>34</sup> Tran *et al.*,<sup>33</sup> and Gruich *et al.*<sup>37</sup> have previously employed machine learning models to predict material properties such as bulk modulus, formation energy, and adsorption energies, incorporating uncertainty quantification (UQ). These studies primarily focused on parameters or small-scale representations of atomic or molecular structures. In contrast, our research extends this scope by predicting material properties at a larger scale, utilizing images to capture the microstructural complexity inherent at the nanoscale. This approach allows for a more comprehensive representation of the materials being studied.

In this study, we aim to develop a deep learning framework that incorporates quantified uncertainty to predict the properties of nanocomposite materials under mechanical stress, taking into account the morphology and composition of these materials. While our focus primarily lies on predicting the mechanical properties

of a model nanocomposite material, the framework we develop holds potential applicability for predicting diverse properties across a broad spectrum of materials.

## II. METHODOLOGY

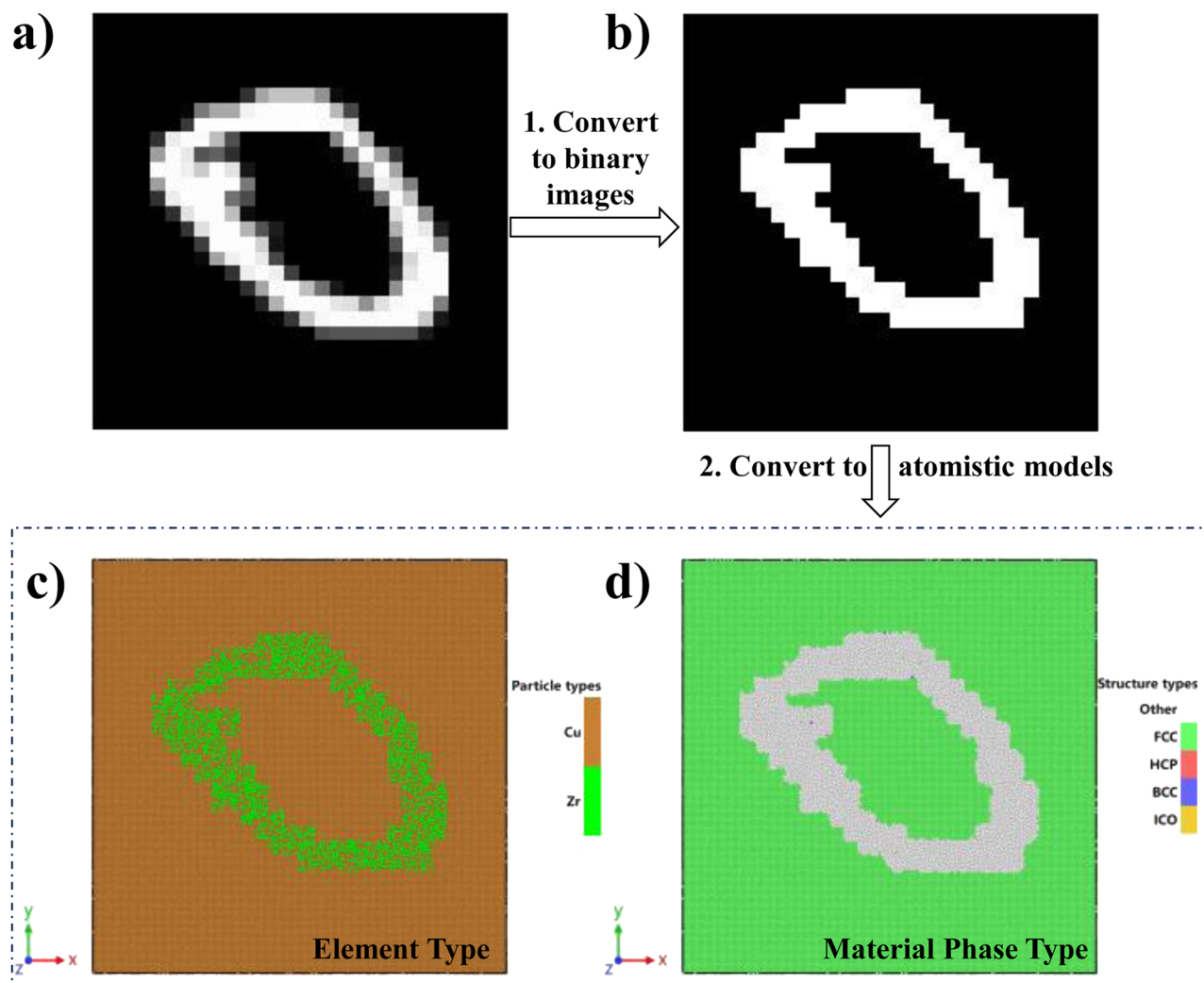
### A. Atomistic models and simulations

The two-phase nanocomposite material of Cu–Zr, comprising the crystalline Cu phase and the amorphous Cu<sub>64</sub>Zr<sub>36</sub> phase, is chosen as the model material system in this study due to its engineering applications.<sup>38,39</sup> To construct these two-phase nanocomposite models, we divide the whole region into two distinct phases using the following approach. Our approach utilizes images from the MNIST dataset,<sup>23</sup> a resource widely recognized not only in computer vision research but also in mechanical research on heterogeneous materials, as exemplified by the mechanical MNIST framework developed by Lejeune *et al.*<sup>16,17,40</sup> In this context, mechanical MNIST employs the MNIST dataset for pattern generation, facilitating finite element simulations and the construction of machine learning models that calculate mechanical properties, such as force fields and crack paths. Drawing inspiration from this framework, our methodology repurposes the MNIST dataset using its hand-written digits not as numerical representations but as diverse microstructural patterns. This innovative use of MNIST enables us to efficiently generate a broad range of simple yet varied images, which are integral in training our machine learning models for simulating mechanical properties of heterogeneous materials.

The MNIST dataset of images contains handwritten digits ranging from 0 to 9. Its suitability for this study stems from its ability to encapsulate both substantial changes in digit representation and minor variations in geometries and orientations for the same digit. Figure 1(a) is an example of an image from the MNIST database, which measures 28 × 28 pixels. The original images in the dataset are converted into binary images, each containing two phases [Fig. 1(b)], using a threshold of 128 to differentiate between these two phases. The binary images from the MNIST dataset are used as input for the ML model.

To generate the atomic model for Cu/Cu<sub>64</sub>Zr<sub>36</sub> corresponding to the binary images, we used the following procedure. First, we generated an initial, pure Cu model. The unit cell length for Cu is 3.615 Å. We constructed a Cu atomic model with the dimensions 36.15 × 36.15 × 1.445 nm<sup>3</sup>. Subsequently, the atoms were categorized into two groups based on their respective positions in the corresponding binary images. The atomic model was divided into 28 × 28 small squares corresponding to the binary image. The atoms in the black regions remained Cu, while the atoms in the white regions (representing the digits) were replaced by Cu<sub>64</sub>Zr<sub>36</sub>. In the Cu<sub>64</sub>Zr<sub>36</sub> region, 36% of the Cu atoms were replaced by Zr. Furthermore, to construct a model with the correct density for the Cu<sub>64</sub>Zr<sub>36</sub> region, 24% of atoms were randomly deleted to account for the different densities of Cu and Cu<sub>64</sub>Zr<sub>36</sub>. Figures 1(c) and 1(d) clearly illustrate that the digit region is a mixture of Cu and Zr atoms, while the remaining area consists of pure, crystalline Cu.

After constructing the atomistic models for the two-phase nanocomposite, we performed molecular dynamics (MD) simulations to relax the models at a temperature of 300 K. This enabled each atom to stabilize in an appropriate position. Following relaxation, we applied uniaxial tension with a strain of 0.2 to the atomic



**FIG. 1.** Constructing atomistic models for nanocomposites utilizing the MNIST dataset of handwritten digits commonly employed within the computer vision community: (a) depiction of images extracted from the MNIST dataset; (b) conversion of the images into binary bitmap format; (c) visualization of the atomic nanocomposite model illustrating the spatial arrangement of Cu and Zr elements; (d) representation of the atomic nanocomposite model, with distinct coloring of the FCC and amorphous phases facilitated by common neighbor analysis (CNA).

models. The tensile simulations of the nanocomposite materials produced the corresponding stress–strain curves, examples of which are presented in Fig. 3. We obtained the values of both yield strength and ultimate tensile strength (UTS) from these stress–strain curves. For modeling the two-phase nanocomposite materials, we used the Embedded Atom Method (EAM) potential. This interatomic potential can accurately predict the structure and energy for various ratios of Cu and Zr atoms.<sup>41</sup> We performed all the atomistic simulations using the open-source package, Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).<sup>42</sup>

## B. Machine learning models for predicting mechanical properties

In this study, our primary objective was to establish a foundational framework for integrating uncertainty quantification into

neural network models for predicting material properties, specifically focusing on the mechanical properties of nanocomposites. To this end, we employed basic Feedforward Neural Networks (FNN) and CNN models due to their proven efficacy and simplicity. FNN models comprise different sizes of fully connected layers, leaky rectified linear unit (ReLU) activated layers, and dropout layers. The input images are first transformed from a 2D matrix to a 1D vector. Then, through iterative adjustments, the weights in the fully connected layers calculate an accurate output. The detailed architecture of the FNN model is shown in Table V in the Appendix.

On the other hand, CNN models are constructed using convolution layers, max-pooling layers, activation layers, and dropout layers. Unlike FNN models, the input is not expanded initially; convolution aids in feature extraction from the input, and pooling layers are employed to reduce the size. Finally, fully connected



layers reduce the size to  $1 \times 1 \times 1$  and provide the prediction. The detailed architecture of the CNN model is presented in Table VI in the Appendix.

In our study, we have allocated the training samples into training, validation, and testing sets, with 8000 samples for training, 2000 for validation, and 2000 for testing. This split was conducted randomly to ensure an unbiased distribution of data across all sets. For both FNN and CNN models, we trained two separate models to predict the yield strength and ultimate tensile strength, respectively. Each deep learning (DL) model was trained for at least 100 epochs with a batch size of 10 to ensure convergence. We used the Mean Squared Error (MSE) as the loss function in our work during the training process. In our study, we systematically tested various combinations of learning rate, dropout probability, and neuron count, with the learning rate decreasing to one-tenth of its previous value and dropout probability increasing in 5% increments. The optimal combination was selected based on achieving the lowest Mean Squared Error (MSE), ensuring an effective balance between learning efficiency and model generalizability.

### C. Development of uncertainty-integrated deep-learning models

In this study, our primary objective is to incorporate the uncertainty of ML predictions into the DL models. To achieve this, we aim to calculate the probability of the DL predictions, enabling us to provide confidence intervals for the predictions. This is based on the ConvPDE-UQ framework developed by Winovich *et al.*<sup>35</sup> By assuming that the prediction follows a Gaussian distribution, we can ascertain the distribution of our prediction using the mean and the corresponding standard deviation. Here, we design two uncertainty-integrated models named FNN-Prob and CNN-Prob. These models predict both the mean and the standard deviation simultaneously. The detailed architecture for the FNN-Prob and CNN-Prob models, including uncertainties, is outlined in the Appendix, specifically in Tables VII and VIII.

In this study, the framework for integrating uncertainty quantification into machine learning models primarily revolves around leveraging the inherent capabilities of neural networks, like FNN and CNN models, to estimate uncertainty through their predictive outputs. This approach differs from existing methods, such as Gaussian Processes (GPs). While GPs are well-known for their natural incorporation of uncertainty, our neural network-based method allows for more flexibility and scalability, which is particularly beneficial for complex datasets like those in material science. The computational cost for GPs is scaled with  $N^3$ , where  $N$  is the number of samples, while our neural network-based method is only linearly scaled with  $N$ . Hence, the proposed method can handle machine learning tasks with large datasets.

The probabilistic model, including CNN-Prob and FNN-Prob models, can be divided into two components: the first part predicts the mean of predictions, akin to the classic CNN and FNN models, and the second part calculates the standard deviation of predictions. The latter shares the initial few layers with the mean part and incorporates several additional fully connected layers for predicting the standard deviation.

The CNN-Prob and FNN-Prob models were developed based on the classic CNN and FNN architectures, each tailored to enhance

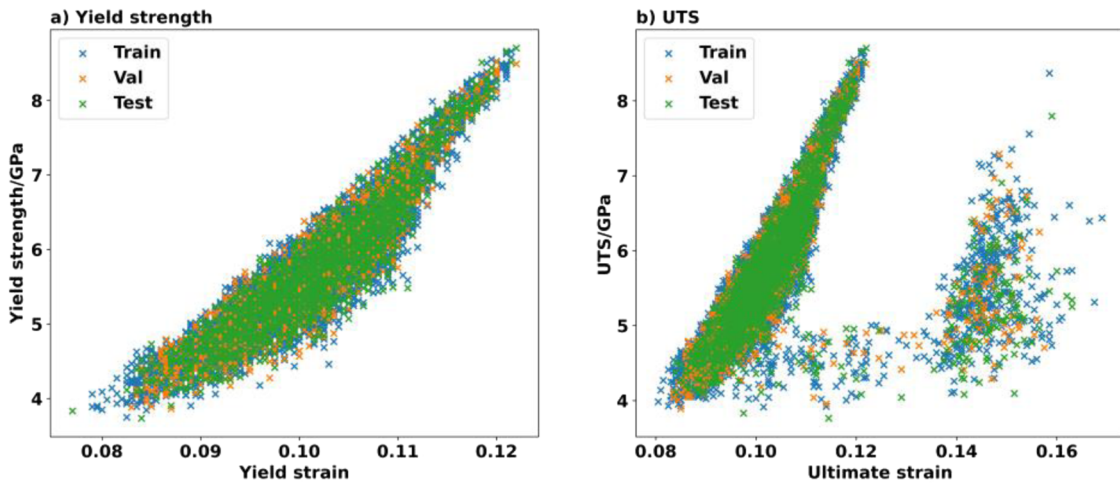
prediction accuracy while integrating uncertainty quantification. For example, the CNN-Prob model is a convolutional neural network designed for probabilistic predictions, where it not only forecasts an outcome but also estimates the associated uncertainty. It starts with convolutional layers that extract features from input data, using filters that capture various aspects of the data through convolution operations. These are followed by activation functions like ReLU to add non-linearity, and pooling layers that reduce the spatial dimensions of the data, simplifying the network and reducing computational load. The network also includes dropout layers to mitigate overfitting by randomly omitting units during training. After convolutional and pooling layers, the architecture transitions to fully connected layers that interpret the features and make predictions. In the CNN-Prob model, the final layer is designed to output two values per prediction: one for the mean or expected value of the prediction and another for the standard deviation, representing the prediction's uncertainty. This dual output allows the model to provide a confidence interval for each prediction, making it particularly useful for applications where understanding the reliability of predictions is crucial. For a thorough understanding of their architecture of FNN-Prob and CNN-Prob models, the detailed configurations and activation functions of these models are presented in Tables VII and VIII in the Appendix.

A probabilistic loss function is utilized during training. Specifically, the negative log estimated likelihood, as illustrated in Eq. (1). Here,  $y$  represents the ground truth value, which is the corresponding stress obtained from MD simulations,  $\mu$  symbolizes the predicted mean value, and  $\sigma$  signifies the predicted standard deviation. Analogous to training the classic DL models detailed in Sec. II B, for both FNN-Prob and CNN-Prob models, we trained two distinct models to predict the yield strength and ultimate tensile strength, respectively. Each model underwent at least 100 epochs of training with a batch size of 10 to ensure convergence. We employed the Adam optimizer during the training phase, and the learning rate decayed each parameter group by 0.2 every 50 epochs,

$$Loss_{prob} = -\log p(y) = \frac{(y - \hat{\mu})^2}{2\hat{\sigma}^2} + \frac{1}{2} \log(2\pi\hat{\sigma}^2). \quad (1)$$

Our approach to uncertainty prediction primarily utilizes probability loss, where the loss diminishes as the predicted mean  $\hat{\mu}$  approximates the actual value  $y$ , and the predicted variance  $\hat{\sigma}^2$  reaches an optimum level minimizing this loss. This optimal variance is directly influenced by the discrepancy between the prediction and the ground truth; a smaller discrepancy leads to a lower critical variance value. Initially, our models generate predictions with larger standard deviations, but as training progress, these predictions gradually align more closely with the ground truth. Concurrently, the uncertainty range narrows. During testing, samples with characteristics similar to those in the training set exhibit smaller UQ, while those with less similarity display comparatively larger UQ, reflecting the model's confidence in its predictions based on familiarity with the data.

Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and  $R^2$  are employed to evaluate the model's performance in predicting the mechanical properties of nanocomposites. The formulations for these evaluation metrics are outlined in Eqs. (2)–(5). In these equations,  $\hat{y}$  represents



**FIG. 2.** Results from tensile simulations conducted on nanocomposite samples in the dataset: (a) distribution of yield strain and yield strength values; (b) distribution of ultimate tensile strength and ultimate strain values.

the prediction,  $y$  is the ground truth, and  $\bar{y}$  is the average of all the ground truth values,

$$MAE = \text{abs}(y - \hat{y}), \quad (2)$$

$$MSE = (y - \hat{y})^2, \quad (3)$$

$$MAPE = \text{abs}\left(\frac{y - \hat{y}}{y}\right), \quad (4)$$

$$R^2 = 1 - \frac{\sum_{l=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2}. \quad (5)$$

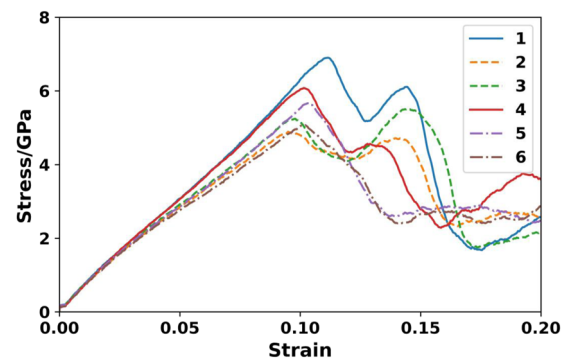
### III. RESULTS

#### A. Atomistic simulations of nanocomposite materials

Tensile simulations were performed on 10 000 atomistic nanocomposite samples with varying geometries. The distribution of both yield strength and ultimate tensile strength (UTS) obtained from these samples is depicted in Fig. 2. Generally, as can be observed from Fig. 2(a), a higher yield strength in a nanocomposite sample corresponds to a higher yield strain. The yield strain ranges from 0.08 to 0.12, while the yield strength varies from 3.8 to 8.5 GPa. Additionally, Fig. 2(a) shows that all the data points fall below a particular upper-bound straight line. Turning to Fig. 2(b), which presents the distribution of ultimate tensile strength and ultimate strain, the data points appear to cluster into two regions: left and right. The majority of the data points in the left region exhibit a similar pattern to the yield strength, but the data points in the right region register much higher ultimate strains than those in the left. The ultimate tensile strength ranges from 3.8 to 8.5 GPa, whereas

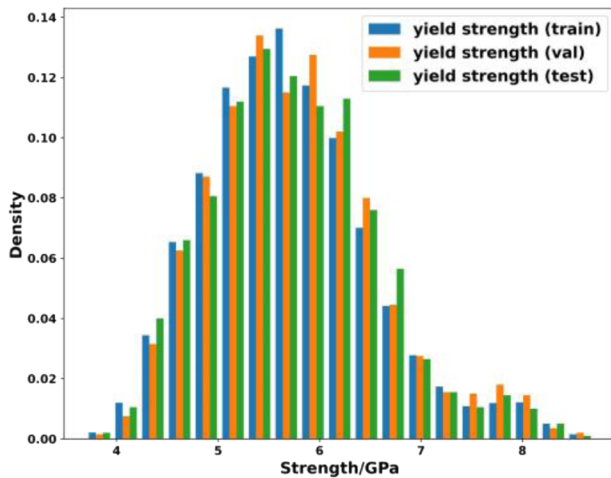
the ultimate strain fluctuates between 0.08 and 0.16. This range is larger compared to the yielding strain observed in Fig. 2(a).

We further calculated the stress-strain curves for all the nanocomposite samples and noticed two different major types of these curves. The first type of stress-strain curve has only one maximum value, as illustrated by dashed-dotted curves 5 and 6 in Fig. 3. The other type features two peaks, as demonstrated by solid and dashed curves 1–4 in Fig. 3. We also observed that when the first peak is higher than the second, as in solid curves 1 and 4, the yield strength and ultimate tensile strength are identical. However, if the second peak is higher, like in dashed curves 2 and 3, then the yield strength and ultimate tensile strength differ. The appearance of these two different types of stress-strain curves is tied to the occurrence of twinning deformations, which create nano-sized twinning boundaries in the samples. Twinning, a common strengthening mechanism under loading, significantly increases the ultimate strain. Upon examining the atomistic structure of samples during



**FIG. 3.** Comparative stress-strain curves of six nanocomposite samples within the training set during tensile simulations.

a) Yield strength



b) UTS

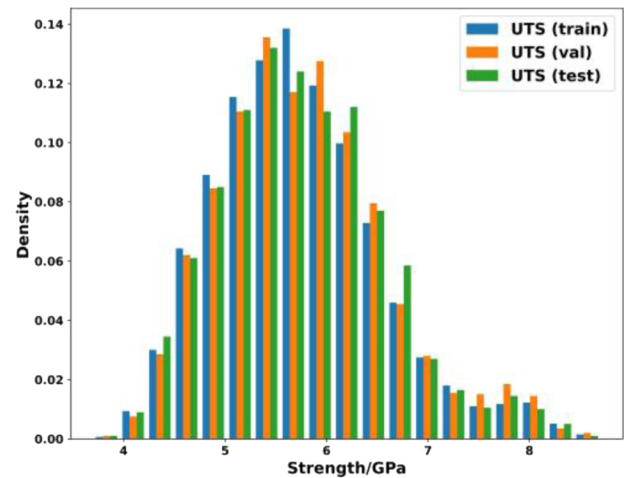


FIG. 4. The density distribution of (a) yield strength and (b) ultimate tensile strength in the dataset.

the simulation, we found that a significantly greater number of twinning boundaries are formed in the samples with two peaks in their stress-strain curves compared to the samples with only one peak. As twinning formation can strengthen the material, a peak typically appears on the stress-strain curves. This mechanism also explains why the data points in the ultimate tensile strength plot in Fig. 2 are divided into two regions.

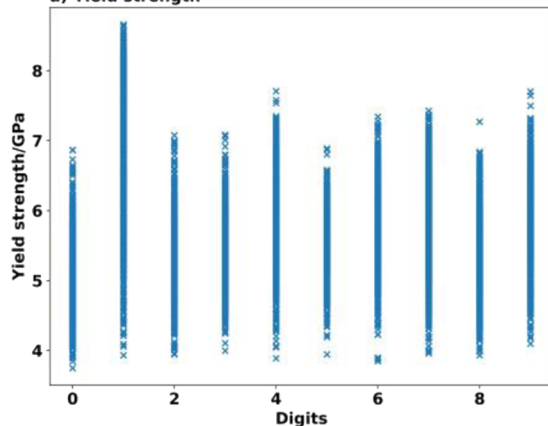
## B. Factors affecting the mechanical properties of nanocomposite materials

In this section, we aim to study how different factors affect the mechanical properties of nanocomposite materials. Figure 4 plots the density distribution of both yielding and ultimate tensile strength

in the training and test sets. Both yield strength and ultimate tensile strength register their highest frequency between 5 and 5.5 GPa.

The digit depicted in the original images within the samples is an important feature that directly influences the geometry of nanocomposite models. Therefore, it is worthwhile to study how this digit can affect both the yield strength and ultimate tensile strength. To that end, we plotted the ultimate tensile strength and yield strength across digits ranging from 0 to 9. As illustrated in Figs. 5(a) and 5(b), both the yield strength and ultimate tensile strength peak for digit 1, while all other digits demonstrate yielding and ultimate tensile strength lower than 8 GPa. A possible explanation for this could be that samples with the digit 1 exhibit the simplest geometrical structure. However, based on the distribution plotted in Fig. 5, it is not feasible to estimate stress solely through the

a) Yield strength



b) UTS

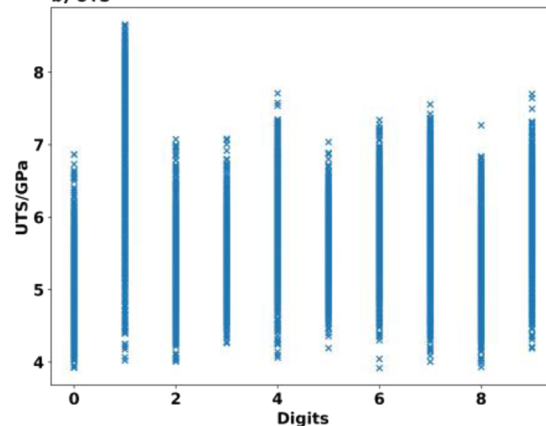
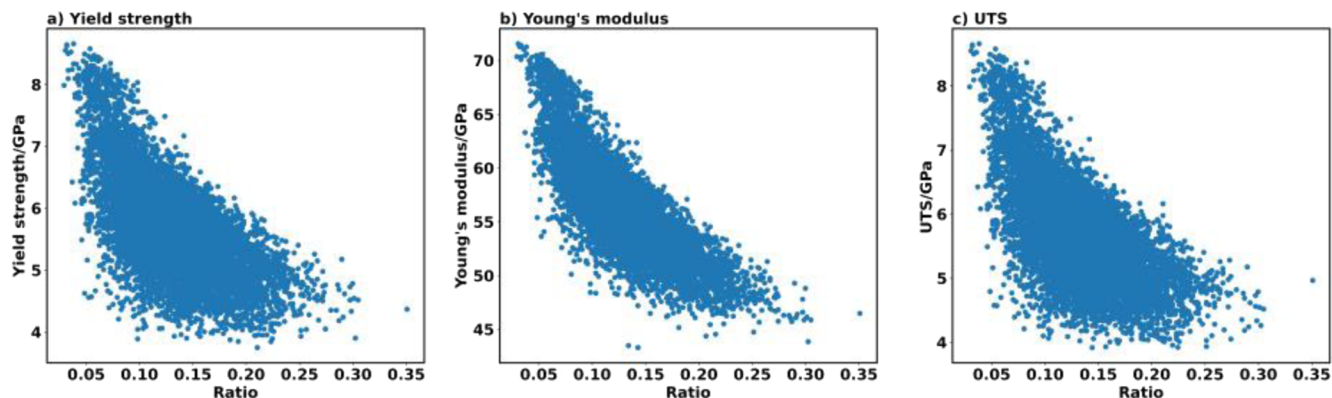


FIG. 5. Distribution of (a) yield strength and (b) ultimate tensile strength correlated with the numerical values from handwritten digits.



**FIG. 6.** The impact of the amorphous phase fraction on various mechanical properties: (a) yield strength, (b) Young's modulus, and (c) ultimate tensile strength.

digit in the image due to the extensive variance in stress prediction for each digit.

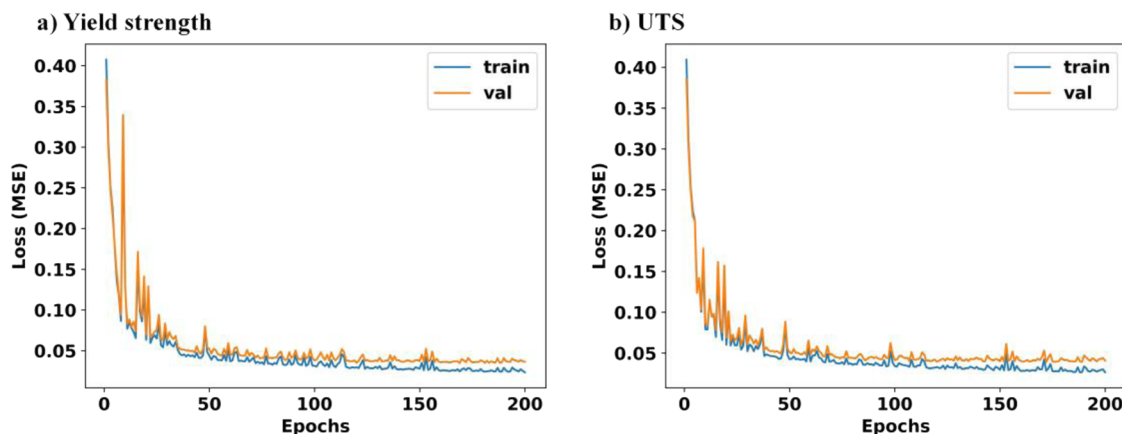
We also studied the effect of the fraction of each phase in the composite materials on mechanical performance. Different geometrical patterns exhibit different ratios between these two phases in the samples. In Fig. 6, we plot the distribution of yield strength, Young's modulus, and ultimate tensile strength in relation to the fraction of the amorphous phase in the nanocomposite samples. From Fig. 6, it can be observed that both yield strength and ultimate tensile strength decrease as the fraction of the amorphous phase increases. Most of the data points fall beneath a certain upper-bound straight line. Regarding the distribution of Young's modulus in Fig. 6(b), a clear boundary is easily discernible at the top right of the data cluster, although the range of distribution is smaller compared to the plots of stresses. Furthermore, in all three plots in Fig. 6, the lower bounds are unclear, making it difficult to predict stresses or Young's modulus based solely on the fraction of each component. These results underscore the challenge of predicting mechanical properties based on a few simple factors. Instead, it is necessary to consider the more complex geometry of the samples to

accurately predict their mechanical properties. To this end, we will develop more intricate ML models to incorporate more information regarding the geometry of two-phase nanocomposite materials in predicting their mechanical properties.

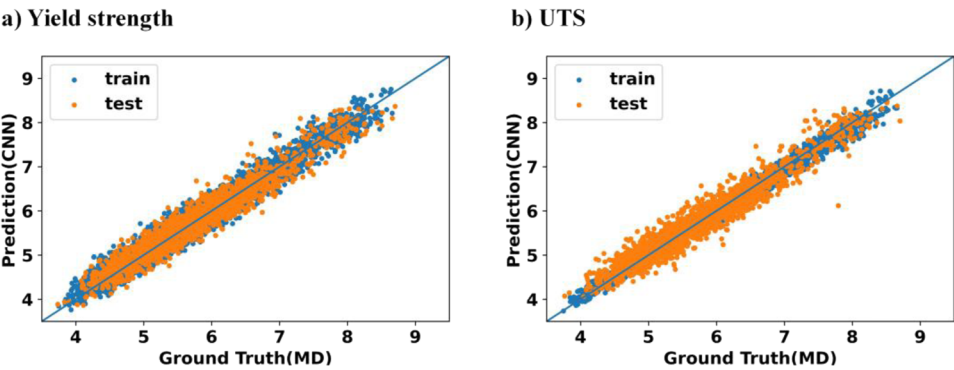
### C. Predicting mechanical properties with classic DL models

Based on the MD simulation results from 10 000 nanocomposite samples, we trained both classic CNN and FNN models to predict their yield strength and ultimate tensile strength. The architecture of these DL models and the training procedure have been discussed in Sec. II B. As demonstrated in Fig. 7, the MSE loss for training the classic CNN models diminishes with an increasing number of training epochs. The training of CNN models for predicting both yield strength and ultimate tensile strength reaches convergence after ~100 epochs.

Figure 8 presents a comparison between the ground truth of yield strength from MD simulations and predictions from the classic



**FIG. 7.** MSE loss of training classical CNN models evolving with the number of epochs for predicting (a) yield strength and (b) ultimate tensile strength.



**FIG. 8.** Comparison of classic CNN model predictions with ground-truth MD simulation results for (a) yield strength and (b) ultimate tensile strength in the training and test sets.

**TABLE I.** Comparative analysis of predictions from classic CNN and FNN models. Values in boldface represent better outcomes comparing different ML models.

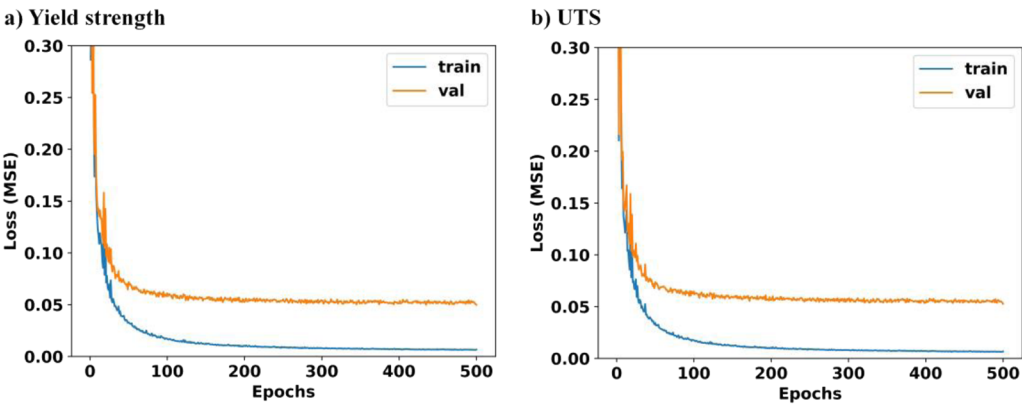
		Train				Test			
		MAE	MSE	MAPE	R <sup>2</sup>	MAE	MSE	MAPE	R <sup>2</sup>
CNN	Yield strength	0.1384	0.0307	0.0247	0.9551	<b>0.1736</b>	0.0482	<b>0.0305</b>	<b>0.9326</b>
	UTS	0.1412	0.0322	0.0251	0.9518	<b>0.1788</b>	0.0521	<b>0.0315</b>	<b>0.9264</b>
FNN	Yield strength	<b>0.0675</b>	<b>0.0079</b>	<b>0.0116</b>	<b>0.9885</b>	0.1811	0.0539	0.0316	0.9246
	UTS	<b>0.0667</b>	<b>0.0076</b>	<b>0.0115</b>	<b>0.9886</b>	0.1860	0.0582	0.0326	0.9178

CNN model. It can be observed that most data points are located near the center line, indicating high prediction accuracy.

To further evaluate the models' performance quantitatively, we employed four evaluation metrics: MSE, MAE, MAPE, and  $R^2$ . The evaluation results are listed in Table I. The MAE for the training set is 0.1384 for yield strength and 0.1412 for ultimate tensile strength. For the test set, the MAE is 0.1736 and 0.1788, respectively. The MAPE values for these parameters in the training set are 2.47% and 2.51%, respectively. Meanwhile, the MAPE for the test set is slightly higher, at 3.05% and 3.15%, for the predictions of yield strength and ultimate tensile strength, respectively.

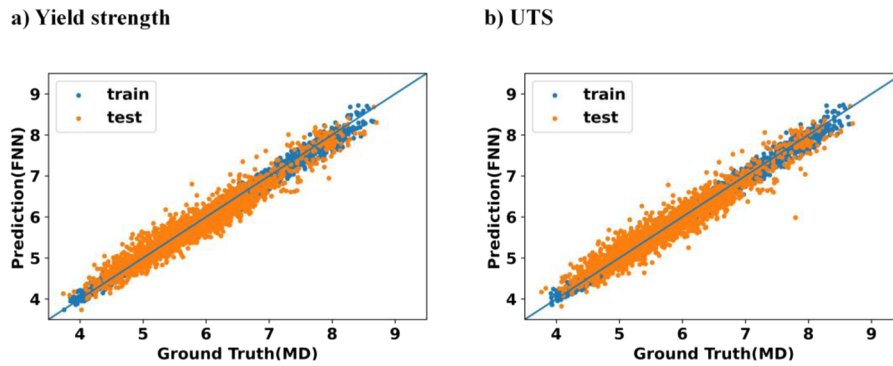
Furthermore, we trained classic FNN models to predict the yield strength and ultimate tensile strength. Compared to the classic CNN models, the classic FNN model converges more slowly, as shown in Fig. 9, taking ~200 epochs to converge.

The MAE, MSE, MAPE, and  $R^2$  for predicting the yield strength in the training set are 0.0675, 0.0079, 1.16%, and 0.9885, respectively, which outperform the CNN model. A similar trend can also be observed in UTS prediction. In the training set, the MAE, MSE, MAPE, and  $R^2$  are 0.0667, 0.0076, 1.15%, and 0.9886, respectively. However, the FNN model does not perform as well as the classic CNN model for the test set. The MAE, MSE, MAPE, and  $R^2$  in the



**FIG. 9.** Variation of MSE loss during the training of classic FNN models as a function of the number of epochs for predicting (a) yield strength and (b) ultimate tensile strength.



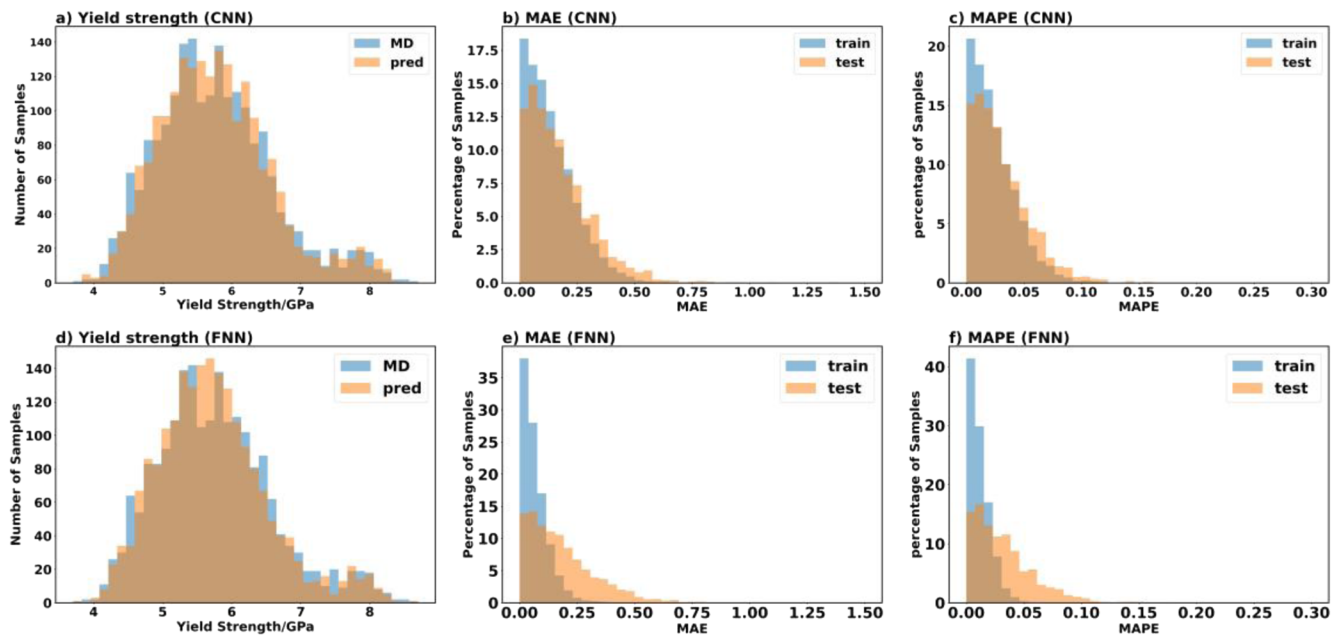


**FIG. 10.** Comparison of prediction results using classic FNN models with MD simulation results in both the training and test sets: (a) yield strength and (b) ultimate tensile strength.

prediction of yield strength are 0.1811, 0.0539, 3.16%, and 0.9246, respectively. In the UTS prediction case, these metrics are 0.1860, 0.0582, 3.26%, and 0.9178, respectively. All these evaluation metrics are slightly worse in the classic FNN model compared to the classic CNN model for the test set. Additionally, we can observe from Fig. 9 that the discrepancy between the training and test sets in the FNN model is much larger than in the classic CNN model. The FNN model exhibits a smoother decrease in the MSE loss and achieves a lower MAE loss in the training set, while the loss in the test set is higher than with the CNN predictions. The result of the FNN prediction, compared to the ground truth values from MD simulations, is shown in Fig. 10. The comparative analysis displayed in Table I

indicates that, compared to the classic CNN model, the classic FNN model performs slightly better for the training set but slightly worse for the test set.

Figure 11 presents a detailed comparison of the distributions of different evaluation metrics. As shown in Fig. 11(a), the CNN model successfully predicts the distribution of yield strength for the test set, and the predicted distribution generally matches the distribution from MD simulations. Figure 11(b) plots the distribution of absolute error, showing that the error in most of the samples is smaller than 0.5 GPa. The percentage error is primarily smaller than 10% for both training and test sets, as shown in Fig. 11(c).



**FIG. 11.** Yield strength prediction comparison between classical CNN [(a)–(c)] and classical FNN [(d)–(f)] models: (a) CNN predictions, [(b) and (c)] absolute error and percentage error in CNN predictions (first row), (d) FNN predictions, and [(e) and (f)] absolute error and percentage error in FNN predictions (second row).

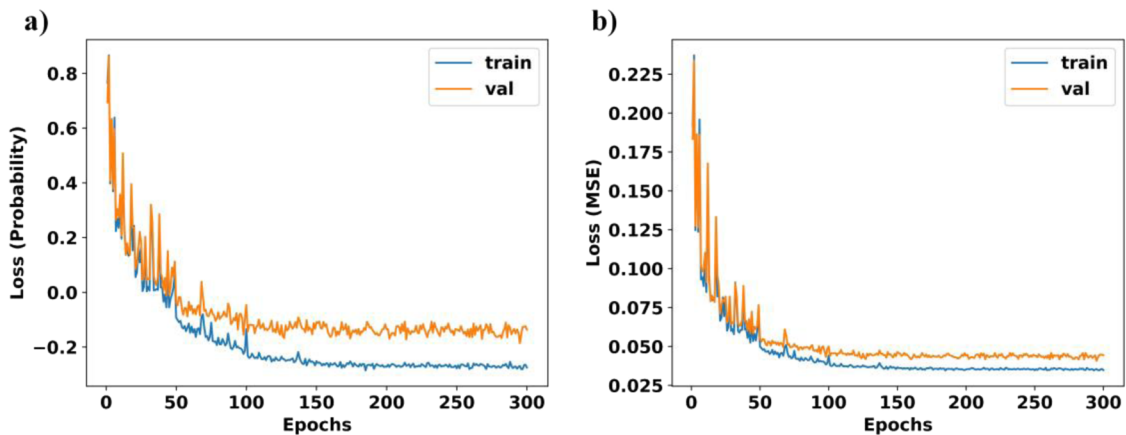


FIG. 12. Evolution of (a) probability loss and (b) MSE loss with the number of epochs in the CNN-Prob model.

In the training set, the maximum MAE and MAPE for the FNN model are 0.437 GPa and 7.28%, respectively, both of which are lower than the values obtained with the CNN model. However, in the test set, the FNN model performs worse than the CNN model, with a maximum MAE and MAPE of 1.059 GPa and 21.3%, respectively. Both these values are higher than those of the CNN model. In Fig. 11, it is clear that the data are more concentrated on the left side, indicating lower error in the CNN result, whereas the FNN results have a more widespread distribution.

In summary, both the CNN and FNN models can accurately predict yield strength and ultimate tensile strength. However, while the FNN model produces better results in the training set, the CNN model shows superior performance in the test set, exhibiting lower error.

#### D. Predictions with uncertainty-integrated DL models

While both CNN and FNN can provide reasonable predictions of yield strength and ultimate tensile strength, the maximum error can still reach 15.7% and 21.3%. Such accuracy may not be sufficient in some applications. Since it is challenging to increase the

prediction accuracy of either CNN or FNN models, providing an uncertainty region simultaneously can indicate which samples may have a large error, which is crucial when selecting samples. For illustration, we take yield strength as an example to display the results of the probabilistic model.

The architecture of the CNN-Prob model with uncertainty is discussed in Sec. II B. The primary part used to predict the mean value is identical to the previously mentioned CNN model. The significant difference here is the addition of three convolution layers to predict the standard deviation simultaneously. Figure 12(a) illustrates the decrease in loss. It takes roughly 300 epochs to reach a steady state, which is significantly longer than with the classic CNN models. This could be because the CNN-Prob probabilistic model requires two variables to be adjusted simultaneously—the mean value and the standard deviation. We also display the decrease in MSE loss in Fig. 12(b). Compared to the classic CNN model, it is evident that this model has a lower MSE for both the training and test sets.

A comparison of predictions of the CNN-Prob model with the ground truth was plotted in Fig. 13(a). This comparison demonstrates the capability of the CNN-Prob model to predict the yield

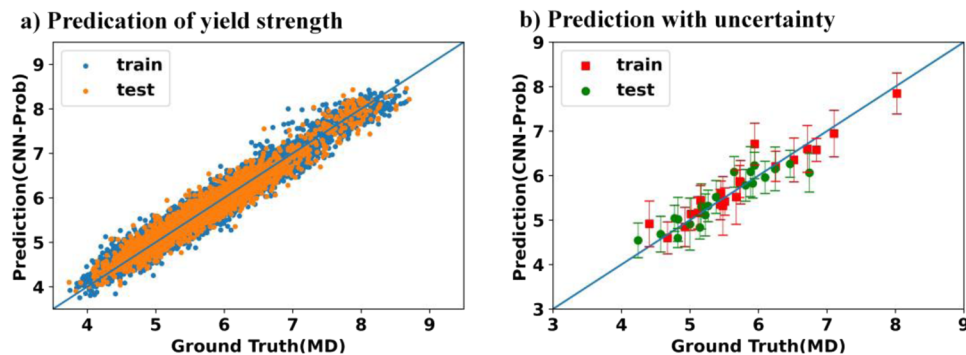
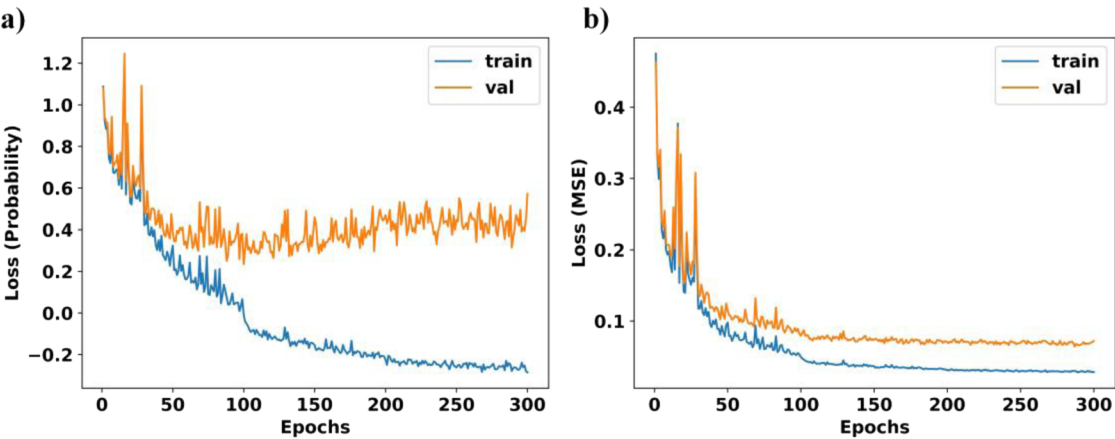


FIG. 13. (a) Prediction results in both the training and test sets; (b) 60 samples with predicted boundaries of two standard deviations in the CNN-Prob model.

**TABLE II.** Comparative analysis of predictions between CNN and CNN-Prob models. Values in boldface represent better outcomes comparing different ML models.

ML models	Train				Test			
	MAE	MSE	MAPE	R <sup>2</sup>	MAE	MSE	MAPE	R <sup>2</sup>
CNN	<b>0.1384</b>	<b>0.0307</b>	<b>0.0247</b>	<b>0.9551</b>	0.1736	0.0482	0.0305	0.9326
CNN-Prob	0.1523	0.0377	0.0270	0.9448	<b>0.1641</b>	<b>0.0435</b>	<b>0.0287</b>	<b>0.9391</b>



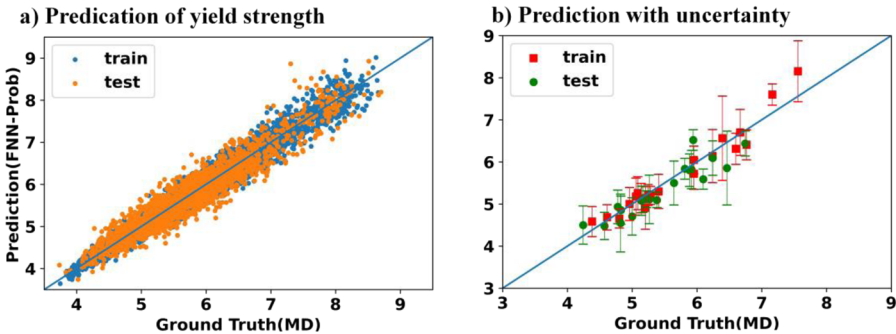
**FIG. 14.** Evolution of (a) MSE loss and (b) probability loss with the number of epochs in the FNN-Prob model.

strength of nanocomposite materials. Moreover, we plot some samples with the two standard deviation plus/minus bounds in both the training and test sets in Fig. 13(b). Most of the samples' confidence intervals cover the ground truth value. 95% of samples in the training set and 90% of samples in the test set fall within two standard deviations of the ground truth, with the mean deviation equal to 0.227 and 0.232 GPa, respectively.

In addition to uncertainty, the model also shows commendable performance in other areas. To further evaluate the models' performance quantitatively, we employed four evaluation metrics: MSE, MAE, MAPE, and R<sup>2</sup>. The evaluation results are listed in Table II.

The MAE, MSE, MAPE, and R<sup>2</sup> in the training set are 0.152, 0.0377, 2.70%, and 0.9448, respectively. In the test set, these metrics are 0.164, 0.0435, 2.88%, and 0.9391, respectively, which are better than those of the classic CNN model.

We have also modified the FNN model to predict uncertainty concurrently. The model's architecture is detailed in Sec. II B. Considering the inclusion of uncertainty, the modified FNN model takes about 300 epochs to run, but the performance disparity between the training and test results is larger than that of the CNN-Prob probabilistic model, particularly. The results are shown in Fig. 14. Furthermore, the probabilistic loss in the test set does



**FIG. 15.** Comparison of predictions of the FNN-Prob model with the ground truth: (a) Prediction results in both the training and test sets and (b) 60 samples with predicted boundaries of two standard deviations in the FNN-Prob model.

not decrease after 100 epochs, and the variation becomes even larger. Meanwhile, the probabilistic loss continues to decrease in the training set. Generally, we need to halt model training to avoid overfitting, but the MSE continues to decrease in both the training and test sets. To maintain the lowest MSE and keep the probabilistic loss at a lower level, we continue running the model for 300 epochs.

In Fig. 15(a), we plot the results of comparing predictions of the FNN-Prob model with the ground truth. Moreover, we plot some samples with the two standard deviation plus/minus bounds in both the training and test sets in Fig. 15(b). The majority of the samples' confidence intervals cover the ground truth value. As shown in the comparative analysis in Table III, the MAE, MSE, MAPE, and  $R^2$  in the training set are 0.1306, 0.0282, 2.26%, and 0.959, respectively. In the test set, these metrics are 0.209, 0.0718, 3.61%, and 0.900, respectively. The FNN-Prob model's performance is inferior to the original FNN model if uncertainty is not considered. When compared with the CNN probabilistic model, both models show similar MAE and MAPE in the training set. However, the test set results differ greatly, with the CNN-Prob probabilistic model outperforming the FNN-Prob probabilistic model. Considering the uncertainty component, the FNN probabilistic model exhibits a higher average standard

deviation and a lower accuracy ratio, where the ground truth is located within the two standard deviations plus/minus bounds of prediction, compared to the CNN-Prob model.

Figures 16(a) and 16(b) show that both the CNN-Prob and FNN-Prob models have a yield strength distribution similar to the ground-truth distribution. Considering the distribution of MAE and MAPE in the training and test sets, the probabilistic models, especially the CNN-Prob, provide better matches. Mainly, the MAE is less than 0.5, and the MAPE is less than 10%. In the case of the FNN-Prob model, the MAE and MAPE distribution widens in the test set compared to the training set.

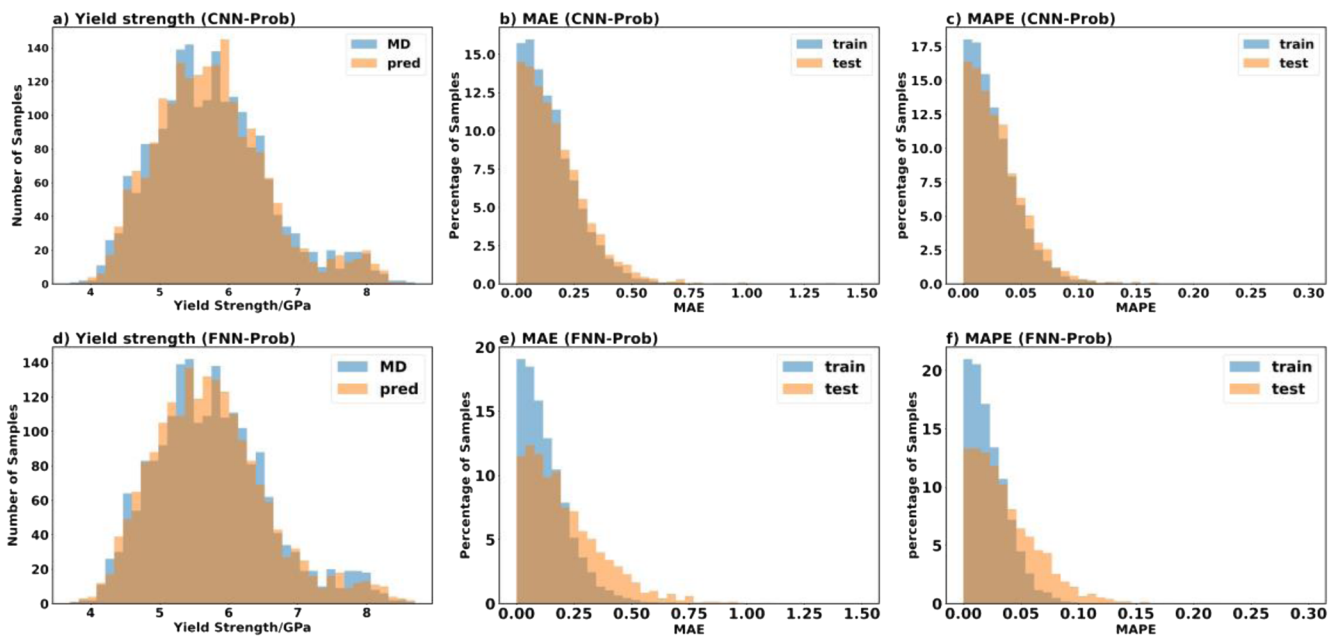
## IV. DISCUSSION

### A. Comparison of DL models for prediction accuracy

Our comparative analysis between the CNN and FNN models reveals that the CNN model performs better overall. Despite FNN demonstrating higher accuracy in the training set, its performance in the test set is weaker, indicating its limitations in feature extraction during training.

**TABLE III.** Comparative analysis of predictions between FNN and FNN-Prob models. Values in boldface represent better outcomes comparing different ML models.

ML models	Train				Test			
	MAE	MSE	MAPE	$R^2$	MAE	MSE	MAPE	$R^2$
FNN	<b>0.0675</b>	<b>0.0079</b>	<b>0.0116</b>	<b>0.9885</b>	<b>0.1811</b>	<b>0.0538</b>	<b>0.0316</b>	<b>0.9246</b>
FNN-Prob	0.1306	0.0282	0.0226	0.9588	0.2089	0.07178	0.0362	0.8995

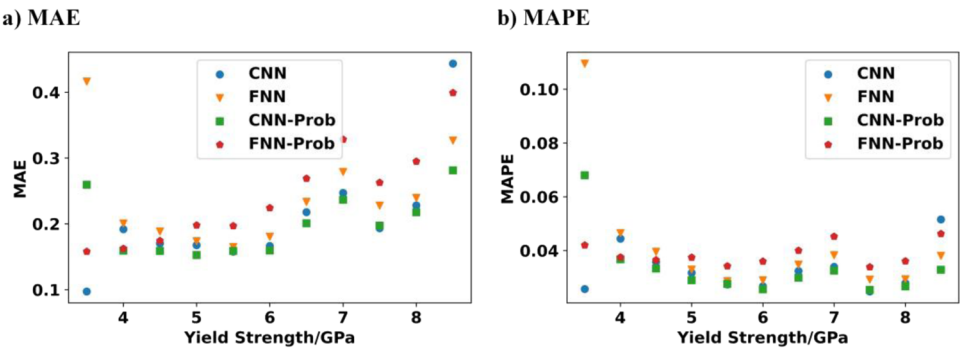


**FIG. 16.** Yield strength prediction comparison between CNN-Prob [(a)–(c)] and FNN-Prob [(d)–(f)] models: (a) CNN-Prob predictions, [(b) and (c)] absolute error and percentage error in CNN-Prob predictions (first row), (d) FNN-Prob predictions, and [(e) and (f)] Absolute error and percentage error in FNN-Prob predictions (second row).

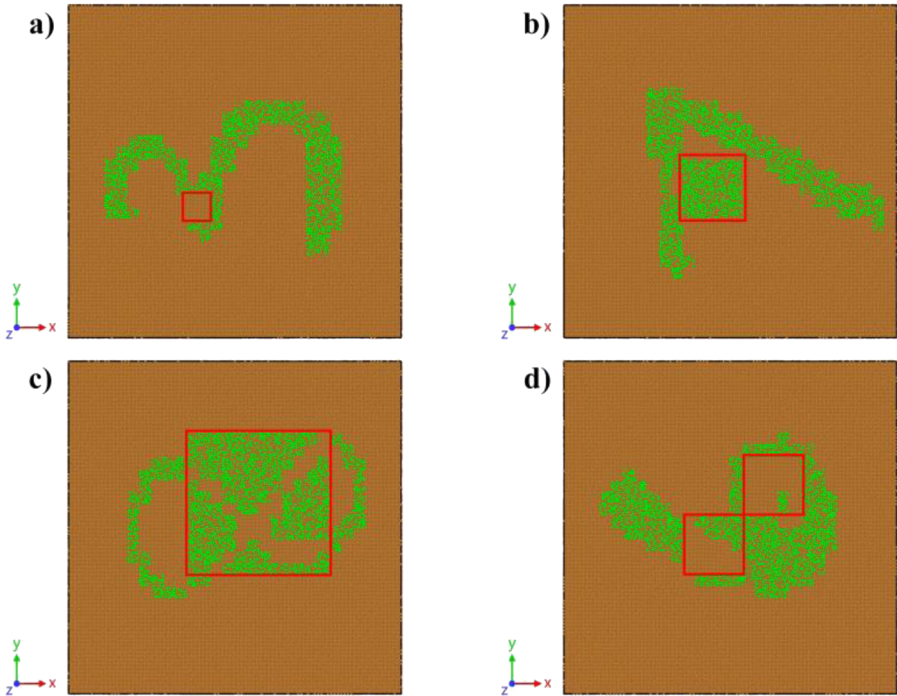
**TABLE IV.** Comparative analysis of Mean Absolute Percentage Error (MAPE) in yield strength predictions among four ML models. Values in boldface represent better outcomes comparing different ML models.

Percentile	MAPE			
	50	75	90	100 (max)
FNN	0.0265	0.0448	0.0662	0.213
FNN-prob	0.0298	0.0528	0.0755	0.213
CNN	0.0259	0.0434	0.0636	<b>0.157</b>
CNN-Prob	<b>0.0238</b>	<b>0.0413</b>	<b>0.0602</b>	0.167

To provide a detailed comparison, we present the percentile distribution of MAE and MAPE in Table IV. In the test set, the 50th percentile MAE for FNN is 0.151, compared to 0.146 for CNN. The 75th, 90th, and 100th percentiles for FNN are 0.255, 0.391, and 1.059, respectively, whereas for CNN, these values are 0.246, 0.355, and 0.942. Across all percentiles, the FNN model underperforms compared to the CNN model. The same is true for MAPE, with values of 2.65%, 4.48%, and 6.62% for the 50th, 75th, and 90th percentiles in FNN, compared to 2.59%, 4.43%, and 6.36% in CNN. Furthermore, the maximum MAPE for CNN (15.7%) is smaller than that for FNN (21.3%).



**FIG. 17.** Distribution of (a) absolute errors and (b) percentage errors in yield strength predictions across all four machine learning models, with respect to the magnitude of yield strength.



**FIG. 18.** Atomic model modification by swapping crystalline Cu phase and amorphous Cu<sub>64</sub>Zr<sub>36</sub> phase in a limited region enclosed by the red outline: (a) 2 × 2 units, (b) 7 × 7 units, (c) 15 × 15 units, and (d) two 5 × 5 units' areas.



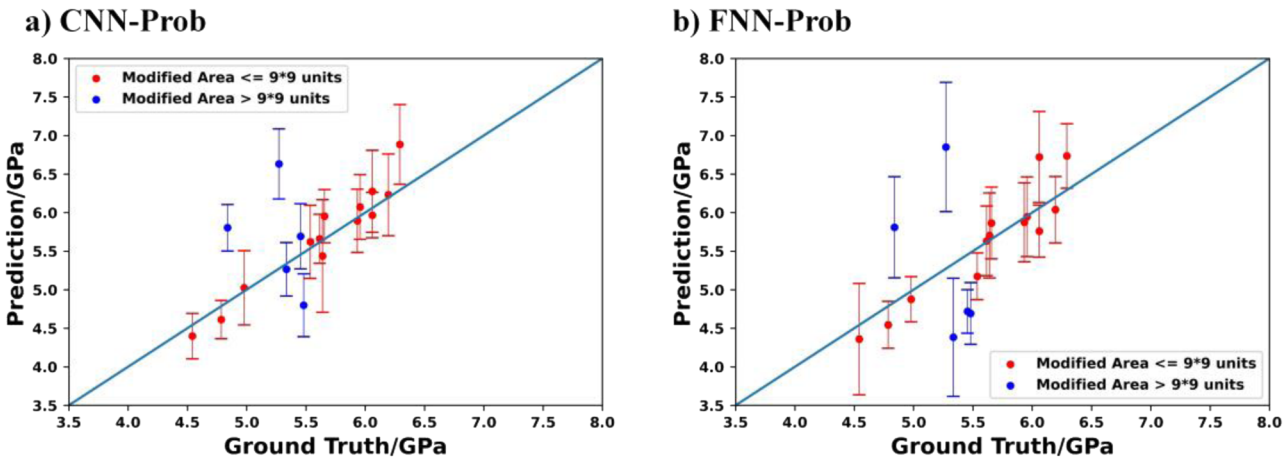


FIG. 19. Comparison of yield strength predictions, including the means and standard deviation on the modified samples using both (a) CNN-Prob and (b) FNN-Prob models.

In terms of the probabilistic models, the MAE in the test set for the FNN-Prob model is 0.169, 0.299, 0.443, and 1.556 for the 50th, 75th, 90th, and 100th percentiles, respectively. In contrast, for the CNN-Prob model, these values are 0.135, 0.235, 0.342, and 0.967. The CNN-Prob model outperforms the FNN-Prob model across all percentiles and presents a significant advantage. A similar trend is observed with MAPE, which is 2.98%, 5.28%, 7.55%, and 21.3% for

the FNN-Prob model and 2.38%, 4.13%, 6.02%, and 16.7% for the CNN-Prob model. In comparison to their non-probabilistic counterparts, the FNN and FNN-Prob models are closely matched, with differences typically under 3%, except for the maximum value where FNN significantly outperforms FNN-Prob. Conversely, the CNN model generally underperforms compared to the CNN-Prob model, with the exception of the maximum value.

TABLE V. The architecture of the FNN model.

Layer	Size
Input	(28, 28, 1)
Full connect	(1, 1, 1568)
Full connect	(1, 1, 1568)
Full connect	(1, 1, 256)
Full connect	(1, 1, 1)

TABLE VI. The architecture of the CNN model.

Layer	Size
Input	(28, 28, 1)
Convolution	(26, 26, 32)
Convolution	(24, 24, 64)
Max-pool	(12, 12, 64)
Convolution	(10, 10, 128)
Convolution	(8, 8, 256)
Max-pool	(4, 4, 256)
Full connect	(1, 1, 4096)
Full connect	(1, 1, 500)
Output	(1, 1, 1)

TABLE VII. The architecture of the FNN-Prob model.

Layer	Size
Input	(28, 28, 1)
Full connect	(1, 1, 6272)
Full connect	(1, 1, 1568)
Full connect	(1, 1, 784)
Full connect	(1,1,1) stress
Full connect	(1, 1, 784)
Full connect	(1,1,64)
Full connect	(1,1,8)
Output	(1,1,1) std

TABLE VIII. The architecture of the CNN-Prob model.

Layer	Size
Input	(28, 28, 1)
Convolution	(26, 26, 32)
Convolution	(24, 24, 64)
Max-pool	(12, 12, 64)
Convolution	(10, 10, 128)
Convolution	(8, 8, 256)
Max-pool	(4, 4, 256)
Full connect	(1, 1, 4096)
Full connect	(1, 1, 512)
Output	(1, 1, 1) stress
Full connect	(1, 1, 4096)
Full connect	(1, 1, 512)
Full connect	(1, 1, 64)
Output	(1, 1, 1) std

Additionally, we plotted the relationship between yield strength and the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) in Fig. 17. In general, the error tends to be larger when the yield strength is either small or large, while samples with medium yield strength exhibit smaller errors. Comparing the various models, it is clear that those incorporating uncertainty (i.e., FNN-Prob and CNN-Prob) yield more uniform results than the conventional CNN and FNN models. For instance, the FNN model typically exhibits smaller errors than FNN-Prob when the yield strength is greater than 4.5 GPa. Conversely, when the yield strength is less than 4.5 GPa, FNN-Prob demonstrates much smaller MAE and MAPE values than the standard FNN model. We also found that CNN-Prob generally has the lowest MAE and MAPE among all four models. However, when the yield strength falls below 4 GPa, CNN-Prob performs worse than the standard CNN model, which exhibits the smallest error at this point. This phenomenon can be attributed to the fewer training data points at both ends of the scale, which can lead to larger errors in the machine learning predictions.

## B. Generalizability of uncertainty-integrated ML models to unseen samples

In addition to the samples in the test set, we also sought to test the generalizability of the model. First, we selected samples from the test set. Then, we selected one or two square regions, with the length of the square ranging from 2 to 15 (given that the length of the atomic model is 28 units). We also tried to use two  $5 \times 5$  units' regions instead of just one. Finally, within this square region, the material was converted. If initially it was a crystalline Cu phase, it was replaced by the amorphous  $\text{Cu}_{64}\text{Zr}_{36}$  phase, and if it was initially an amorphous  $\text{Cu}_{64}\text{Zr}_{36}$  phase, it was changed to a crystalline Cu phase. The new structure is shown in Fig. 18, where the material phase in the red square has been switched.

The predicted results for these modified samples are shown in Fig. 19. We tested both CNN-Prob and FNN-Prob models. Initially, when the area change is small, both models are able to provide relatively accurate predictions, even if the area is two squares instead of one. However, when the length of the square exceeds  $9 \times 9$  units, the prediction accuracy of these two models decreases. It should be noted that the size of the modified region exceeding  $9 \times 9$  units represents a change of about 30% of entire models without additional training for unseen samples. Compared to the FNN-Prob model, the CNN-Prob model behaves better in predicting the yield strength of samples with the modified region exceeding  $9 \times 9$  units. This analysis indicates that while our model maintains accuracy with minor sample disturbances, its uncertainty quantification becomes less reliable under conditions of large disturbance. This contrasts with traditional methods, which typically lack uncertainty quantification, highlighting both the advancements and limitations of our approach.

## V. CONCLUSIONS

In conclusion, this study has successfully presented a deep learning framework capable of predicting various mechanical properties for nanocomposite materials while also integrating quantified uncertainty into the predictions. We have demonstrated the efficacy

of our machine learning models, both FNN and CNN models, in predicting yield strength and ultimate tensile strength, offering a less computationally intensive alternative to traditional methods, such as molecular dynamics simulations.

While MD simulations provide detailed insights, they are computationally intensive. For instance, in our study, conducting a single sample MD simulation using LAMMPS on bigred200 (AMD EPYC 7742) with  $4 \times 48$  central processing unit (CPU) configuration requires  $\sim 1$  h. In contrast, training our neural network model takes about the same time (1 h) but on a single v-100 graphics processing unit (GPU). However, once trained, the neural network model can predict properties for a new sample in under 0.5 s. This represents a drastic reduction in time, particularly beneficial when dealing with large datasets or when rapid predictions are essential.

Our models, trained on comprehensive material datasets, allow for rapid exploration of design spaces and make more reliable predictions by including a measure of uncertainty. The addition of uncertainty quantification not only enhances the confidence in our predictions but also provides a greater understanding of the prediction variance, a vital feature when making decisions based on these predictions in practical engineering applications.

Furthermore, our work exhibits how the standardization and generalization of these models can provide invaluable insights into the structure-property relationships of nanocomposite materials. The deep learning framework developed here has the potential applicability to predict diverse properties across a broad spectrum of materials, thereby holding promise for future research in the field of materials science and engineering.

Ultimately, this research paves the way for advancements in the prediction of nanocomposite materials' properties and sets a precedent for integrating uncertainty quantification in machine learning models within this field, promoting a more robust and reliable paradigm for materials property prediction. Building on the foundations of this study, our future research will focus on advancing deep learning architectures and expanding the material scope beyond nanocomposites. A significant emphasis will be placed on real-world experimental validation to enhance the practical applicability of our models. To achieve this, we can utilize experimental images from advanced characterization techniques like Scanning Electron Microscopy (SEM) and Scanning Transmission Electron Microscopy (STEM). Additionally, mechanical property data obtained from nano-testing methods, particularly nanoindentation, will be crucial for providing property data to validate and refine our developed models. We will also explore model interpretability and engage in industry collaborations to align our research with practical needs. Additionally, efforts will be made to augment and diversify our data sources and integrate our neural network approach with other computational methods, aiming to create a more versatile and robust tool for material property prediction and analysis in material science.

## ACKNOWLEDGMENTS

S.Y. acknowledges the support of BRIDGE and Release Time for Research (RTR) programs at Indiana University Purdue University Indianapolis (IUPUI). The computer simulations used the clusters in the high-performance computing (HPC) at Indiana University, which are supported in part by Lilly Endowment, Inc.,

through its support for the Indiana University Pervasive Technology Institute.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Yuheng Wang:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Guang Lin:** Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Shengfeng Yang:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

### DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub.<sup>43</sup>

## APPENDIX: ARCHITECTURE OF THE MACHINE LEARNING MODELS

Tables V–VIII show the architecture of the FNN model, CNN model, FNN-Prob model, and CNN-Prob model.

## REFERENCES

- <sup>1</sup>B. Ates, S. Koytepe, A. Ulu, C. Gurses, and V. K. Thakur, “Chemistry, structures, and advanced applications of nanocomposites from biorenewable resources,” *Chem. Rev.* **120**(17), 9304–9362 (2020).
- <sup>2</sup>R. Sharma, S. M. Jafari, and S. Sharma, “Antimicrobial bio-nanocomposites and their potential applications in food packaging,” *Food Control* **112**, 107086 (2020).
- <sup>3</sup>X. Shen, Q. Zheng, and J.-K. Kim, “Rational design of two-dimensional nanofillers for polymer nanocomposites toward multifunctional applications,” *Prog. Mater. Sci.* **115**, 100708 (2021).
- <sup>4</sup>A. Bhat, S. Budholiya, S. Aravind Raj, M. T. H. Sultan, D. Hui, A. U. Md Shah, and S. N. A. Safri, “Review on nanocomposites based on aerospace applications,” *Nanotechnol. Rev.* **10**(1), 237–253 (2021).
- <sup>5</sup>S. N. Naik and S. M. Walley, “The Hall–Petch and inverse Hall–Petch relations and the hardness of nanocrystalline metals,” *J. Mater. Sci.* **55**(7), 2661–2681 (2019).
- <sup>6</sup>E. van der Giessen, P. A. Schultz, N. Bertin, V. V. Bulatov, W. Cai, G. Csányi, S. M. Foiles, M. G. D. Geers, C. González, M. Hütter, W. K. Kim, D. M. Kochmann, J. Llorca, A. E. Mattsson, J. Rottler, A. Shluger, R. B. Sills, I. Steinbach, A. Strachan, and E. B. Tadmor, “Roadmap on multiscale materials modeling,” *Modell. Simul. Mater. Sci. Eng.* **28**(4), 043001 (2020).
- <sup>7</sup>B. Yin and W. A. Curtin, “First-principles-based prediction of yield strength in the RhIrPdPtNiCu high-entropy alloy,” *npj Comput. Mater.* **5**(1), 14 (2019).
- <sup>8</sup>K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, and C. Wolverton, “Recent advances and applications of deep learning methods in materials science,” *npj Comput. Mater.* **8**(1), 59 (2022).
- <sup>9</sup>K. T. Butler, F. Oviedo, and P. Canepa, *Machine Learning in Materials Science* (American Chemical Society, 2022).
- <sup>10</sup>J. Damewood, J. Karagunesian, J. R. Lunger, A. R. Tan, M. Xie, J. Peng, and R. Gómez-Bombarelli, “Representations of materials for machine learning,” *Annu. Rev. Mater. Res.* **53**, 399 (2023).
- <sup>11</sup>F. Oviedo, J. L. Ferres, T. Buonassisi, and K. T. Butler, “Interpretable and explainable machine learning for materials science and chemistry,” *Acc. Mater. Res.* **3**(6), 597–607 (2022).
- <sup>12</sup>T. Wen, L. Zhang, H. Wang, W. E. and D. J. Srolovitz, “Deep potentials for materials science,” *Mater. Futures* **1**(2), 022601 (2022).
- <sup>13</sup>W. Jin, J. Pei, P. Xie, J. Chen, and H. Zhao, “Machine learning-based prediction of mechanical properties and performance of nickel-graphene nanocomposites using molecular dynamics simulation data,” *ACS Appl. Nano Mater.* **6**(13), 12190–12199 (2023).
- <sup>14</sup>M. Parsazadeh, S. Sharma, and N. Dahotre, “Towards the next generation of machine learning models in additive manufacturing: A review of process dependent material evolution,” *Prog. Mater. Sci.* **135**, 101102 (2023).
- <sup>15</sup>A. Bhaduri, A. Gupta, and L. Graham-Brady, “Stress field prediction in fiber-reinforced composite materials using a deep learning approach,” *Composites, Part B* **238**, 109879 (2022).
- <sup>16</sup>E. Lejeune, “Mechanical MNIST: A benchmark dataset for mechanical metamodels,” *Extreme Mech. Lett.* **36**, 100659 (2020).
- <sup>17</sup>S. Mohammadzadeh and E. Lejeune, “Predicting mechanically driven full-field quantities of interest with deep learning-based metamodels,” *Extreme Mech. Lett.* **50**, 101566 (2022).
- <sup>18</sup>Z. Yang, C.-H. Yu, and M. J. Buehler, “Deep learning model to predict complex stress and strain fields in hierarchical composites,” *Sci. Adv.* **7**(15), eabd7416 (2021).
- <sup>19</sup>Z. Yang, C.-H. Yu, K. Guo, and M. J. Buehler, “End-to-end deep learning method to predict complete strain and stress tensors for complex hierarchical composite microstructures,” *J. Mech. Phys. Solids* **154**, 104506 (2021).
- <sup>20</sup>Z. Nie, H. Jiang, and L. B. Kara, “Stress field prediction in cantilevered structures using convolutional neural networks,” *J. Comput. Inf. Sci. Eng.* **20**(1), 011002 (2019).
- <sup>21</sup>W. Dai, H. Wang, Q. Guan, D. Li, Y. Peng, and C. N. Tomé, “Studying the micromechanical behaviors of a polycrystalline metal by artificial neural networks,” *Acta Mater.* **214**, 117006 (2021).
- <sup>22</sup>J. R. Mianroodi, N. H. Siboni, and D. Raabe, “Teaching solid mechanics to artificial intelligence—A fast solver for heterogeneous materials,” *npj Comput. Mater.* **7**(1), 99 (2021).
- <sup>23</sup>Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE* **86**(11), 2278–2324 (1998).
- <sup>24</sup>M. Guziewski, D. Montes de Oca Zapiaín, R. Dingreville, and S. P. Coleman, “Microscopic and macroscopic characterization of grain boundary energy and strength in silicon carbide via machine-learning techniques,” *ACS Appl. Mater. Interfaces* **13**(2), 3311–3324 (2021).
- <sup>25</sup>Y. Wang, X. Li, X. Li, Y. Zhang, Y. Zhang, Y. Xu, Y. Lei, C. S. Liu, and X. Wu, “Prediction of vacancy formation energies at tungsten grain boundaries from local structure via machine learning method,” *J. Nucl. Mater.* **559**, 153412 (2022).
- <sup>26</sup>J. Messina, R. Luo, K. Xu, G. Lu, H. Deng, M. A. Tschopp, and F. Gao, “Machine learning to predict aluminum segregation to magnesium grain boundaries,” *Scr. Mater.* **204**, 114150 (2021).
- <sup>27</sup>Y. Cui and H. B. Chew, “Machine-learning prediction of atomistic stress along grain boundaries,” *Acta Mater.* **222**, 117387 (2022).
- <sup>28</sup>M. A. N. Dewapriya, R. K. N. D. Rajapakse, and W. P. S. Dias, “Uncovering stress fields and defects distributions in graphene using deep neural networks,” *Int. J. Fract.* **242**, 107 (2023).
- <sup>29</sup>Y.-C. Hsu, C.-H. Yu, and M. J. Buehler, “Tuning mechanical properties in polycrystalline solids using a deep generative framework,” *Adv. Eng. Mater.* **23**(4), 2001339 (2021).
- <sup>30</sup>A. J. Lew, C.-H. Yu, Y.-C. Hsu, and M. J. Buehler, “Deep learning model to predict fracture mechanisms of graphene,” *npj 2D Mater. Appl.* **5**(1), 48 (2021).
- <sup>31</sup>Y. Wang, D. Oyen, W. Guo, A. Mehta, C. B. Scott, N. Panda, M. G. Fernández-Godino, G. Srinivasan, and X. Yue, “StressNet—Deep learning to predict stress

with fracture propagation in brittle materials,” *npj Mater. Degrad.* **5**(1), 6 (2021).

<sup>32</sup>M. S. R. Elapolu, M. I. R. Shishir, and A. Tabarraei, “A novel approach for studying crack propagation in polycrystalline graphene using machine learning algorithms,” *Comput. Mater. Sci.* **201**, 110878 (2022).

<sup>33</sup>K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, “Methods for comparing uncertainty quantifications for material property predictions,” *Mach. Learn.: Sci. Technol.* **1**(2), 025006 (2020).

<sup>34</sup>F. Tavazza, B. DeCost, and K. Choudhary, “Uncertainty prediction for machine learning models of material properties,” *ACS Omega* **6**(48), 32431–32440 (2021).

<sup>35</sup>N. Winovich, K. Ramani, and G. Lin, “ConvPDE-UQ: Convolutional neural networks with quantified uncertainty for heterogeneous elliptic partial differential equations on varied domains,” *J. Comput. Phys.* **394**, 263–279 (2019).

<sup>36</sup>J. Yin and X. Du, “Uncertainty quantification by convolutional neural network Gaussian process regression with image and numerical data,” *AIAA SciTech 2022 Forum* (AIAA, 2022), p. 1100.

<sup>37</sup>C. J. Gruich, V. Madhavan, Y. Wang, and B. R. Goldsmith, “Clarifying trust of materials property predictions using neural networks with distribution-specific uncertainty quantification,” *Mach. Learn.: Sci. Technol.* **4**(2), 025019 (2023).

<sup>38</sup>T. Phan, J. Rigelesaiyin, Y. Chen, A. Bastawros, and L. Xiong, “Metallic glass instability induced by the continuous dislocation absorption at an amorphous/crystalline interface,” *Acta Mater.* **189**, 10–24 (2020).

<sup>39</sup>A. Khalajhedayati, Z. Pan, and T. J. Rupert, “Manipulating the interfacial structure of nanomaterials to achieve a unique combination of strength and ductility,” *Nat. Commun.* **7**(1), 10802 (2016).

<sup>40</sup>E. Lejeune and B. Zhao, “Exploring the potential of transfer learning for meta-models of heterogeneous material deformation,” *J. Mech. Behav. Biomed. Mater.* **117**, 104276 (2021).

<sup>41</sup>M. I. Mendelev, Y. Sun, F. Zhang, C. Z. Wang, and K. M. Ho, “Development of a semi-empirical potential suitable for molecular dynamics simulation of vitrification in Cu-Zr alloys,” *J. Chem. Phys.* **151**(21), 214502 (2019).

<sup>42</sup>A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in’t Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, “LAMMPS—A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales,” *Comput. Phys. Commun.* **271**, 108171 (2022).

<sup>43</sup>Y. Wang, G. Lin, and S. Yang (2023). “Integrating uncertainty into deep learning models for enhanced prediction of nanocomposite materials’ mechanical properties,” GitHub. <https://github.com/shengfeng-yang/Integrating-Uncertainty-to-DL>