

ARTICLE OPEN



Atomistic simulation assisted error-inclusive Bayesian machine learning for probabilistically unraveling the mechanical properties of solidified metals

A. Mahata^{1,2}, T. Mukhopadhyay³✉, S. Chakraborty⁴ and M. Asle Zaeem⁵✉

Solidification phenomenon has been an integral part of the manufacturing processes of metals, where the quantification of stochastic variations and manufacturing uncertainties is critically important. Accurate molecular dynamics (MD) simulations of metal solidification and the resulting properties require excessive computational expenses for probabilistic stochastic analyses where thousands of random realizations are necessary. The adoption of inadequate model sizes and time scales in MD simulations leads to inaccuracies in each random realization, causing a large cumulative statistical error in the probabilistic results obtained through Monte Carlo (MC) simulations. In this work, we present a machine learning (ML) approach, as a data-driven surrogate to MD simulations, which only needs a few MD simulations. This efficient yet high-fidelity ML approach enables MC simulations for full-scale probabilistic characterization of solidified metal properties considering stochasticity in influencing factors like temperature and strain rate. Unlike conventional ML models, the proposed hybrid polynomial correlated function expansion here, being a Bayesian ML approach, is data efficient. Further, it can account for the effect of uncertainty in training data by exploiting mean and standard deviation of the MD simulations, which in principle addresses the issue of repeatability in stochastic simulations with low variance. Stochastic numerical results for solidified aluminum are presented here based on complete probabilistic uncertainty quantification of mechanical properties like Young's modulus, yield strength and ultimate strength, illustrating that the proposed error-inclusive data-driven framework can reasonably predict the properties with a significant level of computational efficiency.

npj Computational Materials (2024)10:22; <https://doi.org/10.1038/s41524-024-01200-1>

INTRODUCTION

Solidification is recognized as one of the most critical phenomena in both conventional and advanced manufacturing processes of metals and alloys¹. The mechanical properties and deformation behavior of these materials depend on the nano/microstructures that form during the solidification process. Various experimental techniques have been used to study the nano/microstructures and tensile/compression/torsion deformation properties of solidified metals and alloys^{2,3}, but experimental procedures are expensive and require more time and resources to generate reliable data. To perform controlled melting and solidification of metals and alloys, special melting and solidification equipment is needed. In addition, the nano/microscale characterization and temperature-dependent mechanical testing of solidified samples are expensive parts of understanding the effects of solidification parameters on nano/microstructures and mechanical properties. Besides, carrying out stochastic analysis (e.g., by Monte Carlo simulations) based on thousands of solidification experiments is practically impossible. A viable alternative is to use numerical simulation tools to study the solidification process and its effect on mechanical properties. It should be noted that the computational and simulation-based approaches can not replace the experiments, but they can provide insights and recommendations for experiments, which can lead to more efficient and effective experiments with more accurate results.

State-of-the-art simulation techniques at different length and time scales for the study of solidification include classical

molecular dynamics (MD)⁴ and phase-field modeling (PFM)⁵, among others. Models at larger length scales (microscale and larger), like PFM, require input parameters and data from atomistic models (nanoscale) like MD and experiments^{4,6,7}. Quantification of stochastic variations and manufacturing uncertainties during solidification is critically important. Accurate atomistic simulations of metal solidification and the resulting properties require excessive computational expenses for probabilistic analyses where thousands of random realizations are necessary. Inadequate model sizes and time scales lead to inaccuracy in each random realization, causing a large cumulative statistical error in the probabilistic results. In this work, we present a machine learning (ML) approach, as a surrogate to MD simulations, for carrying out efficient and accurate stochastic computations.

Classical MD simulations have been used for predicting both nano-scale and bulk properties. The interatomic potentials for the classical MD simulations can be designed in such a way that they can predict from low-temperature properties such as different physical constants and stacking faults to high-temperature properties like melting point or solid-liquid interface free energy^{8–10}. Although MD simulations have the capability of predicting bulk scale properties, they can be computationally demanding depending on the system size, required level of accuracy and efficiency of interatomic potentials. The situation further aggravates in the case of stochastic simulations that need thousands of realizations for carrying out Monte Carlo (MC) simulation. Since quantifying uncertainty in the mechanical properties of solidified materials is getting significant traction

¹Department of Engineering, King's College, Wilkes-Barre, PA 18711, USA. ²School of Engineering, Brown University, Providence, RI 02912, USA. ³Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK. ⁴Department of Applied Mechanics, Indian Institute of Technology Delhi, Delhi, India. ⁵Department of Mechanical Engineering, Colorado School of Mines, Golden, CO 80401, USA. ✉email: t.mukhopadhyay@soton.ac.uk; zaeem@mines.edu

with the rapid adoption of additive manufacturing technologies, there is a strong rationale to develop efficient stochastic computational frameworks. In this context, ML-based approaches can be integrated with MD simulations for improving the feasibility of performing full-scale probabilistic characterization. Conventional ML models require a large number of consistent training data points that can be generated based on MD simulations^{11,12}. Further, since the MD simulation involves random fluctuations of atoms, there is always a notion of inherent variability involved with the repeatability, even with the same simulation parameters. This leads to a mean and standard deviation corresponding to each of the training data-point. In this article, we propose a Bayesian ML approach coupled with MD simulations to overcome the aforementioned critical computational issues. This ML approach, referred to as the hybrid polynomial correlated function expansion (H-PCFE), is data efficient and accounts for the effect of uncertainty in training data by exploiting mean and standard deviation of the MD simulations.

As the goal of this work is to integrate ML and MD methods to accelerate probabilistic characterization of solidified metal properties, here we briefly review the previous research on ML algorithms coupled with MD simulations for the determination of material properties and discuss how our work contributes to this field. ML algorithms available in the literature can be classified into two categories: (a) unsupervised learning, and (b) supervised learning¹³. In unsupervised learning, we have access to unlabeled data. The goal of unsupervised learning is to find some pattern in the data available. On the other hand, in supervised learning, one has access to labeled data and the objective is to predict labels or outputs for a new set of input data. The unlabeled data indicates a scenario when the features, characteristics, classifications and attributes of the input and output parameters are not known a priori. This is unlike the current scenario of supervised learning where we know the inputs (temperature and strain rate) and outputs of the model (Young's modulus, yield stress and ultimate stress). Over the last decade, a significant interest has evolved to exploit the capabilities of ML algorithms to enhance the MD-based simulation outputs for material characterization and innovation^{14–17}. In this paper, we are concerned with supervised learning for predicting outputs corresponding to unseen input parameter combinations. Popular supervised learning techniques available in the literature include deep neural networks^{18–20}, Gaussian process^{21,22} and support vector regression²³. It is worthwhile to note that each of the available techniques has its own merits and shortcomings depending on the nature of the problem or physical model under consideration. In this paper, we integrate a form of Gaussian process called H-PCFE²⁴ with MD simulations. H-PCFE is a probabilistic ML algorithm that provides error estimates (predictive variance), and therefore it is suitable for cases having sparse datasets.

In this work, we would focus on the mechanical properties of solidified aluminum^{4,25} in deterministic and stochastic domains. This work has three different interlinked components with an overarching goal of efficient uncertainty quantification. First, we construct a dataset from MD simulations of tensile deformation of solidified aluminum. During this stage, along with MD simulations of algorithmically generated training data points, we investigate the mechanical properties and structure-property relationship comprehensively by presenting various stress–strain and strength plots versus solidification conditions and strain rates. In the second stage, based on the physics-based training data for mechanical properties such as Young's modulus, yield stress and ultimate strength, we form the ML models for predicting scenarios that have not been simulated (i.e., unseen combination of input parameters). In the final stage, we perform sensitivity analysis for the input variables (deformation temperature and strain rate) and output variables (Young's modulus, yield stress and ultimate

strength) along with complete probabilistic characterization for quantifying uncertainty.

RESULTS

Mechanical properties of solidified polycrystalline aluminum

In this section, we discuss the mechanical properties of polycrystalline aluminum created by solidification at different isothermal temperatures and quench rates. The stress–strain plots are obtained from the simulated polycrystalline aluminum under uniaxial tension for different solidified structures. Figure 1 shows an example of such a stress–strain curve where detwinning of solidification twins happens during tensile deformation²⁵. From each stress–strain curve, the mechanical properties such as Young's modulus, yield stress and ultimate tensile strength are derived. Young's modulus is the slope of the linear portion of the stress–strain plot. The elastic part of the curve where Young's modulus is calculated acts linearly, and no permanent deformation happens. The yield stress is found from the linear regression of the stress data with 0.2% offset on strain. The intersecting point of the linear fit from the regression and the actual stress–strain curve is the yield stress. The maximum tensile stress on the stress–strain plot is defined as the ultimate tensile strength (UTS).

Young's modulus of the solidified polycrystalline samples is compared with that of the single crystal aluminum in Fig. 2a–d. The expected decline of Young's modulus is observed with increasing the deformation temperature when the strain rate is kept constant between 10^8 s^{-1} to 10^{10} s^{-1} . A higher strain rate and/or a lower deformation temperature produce a stiffer sample (Fig. 2a). The single crystal has Young's modulus of ~60–65 GPa at 300 K under the strain rate of 10^{10} s^{-1} . The samples prepared by isothermal solidification at 500 K can reach up to 58 GPa of Young's modulus (Fig. 2a) for a deformation temperature of 300 K and strain rate of 10^8 s^{-1} . Young's modulus remains much lower for all other samples prepared by isothermal and quenching condition. The samples solidified at the slowest quench rate of 10^{11} K s^{-1} , have a Young's modulus of ~59 GPa (Fig. 2b). At higher strain rates the dislocation movements are affected by the velocity of the atoms. When the strain rate is reduced, it is observed that the single crystal and samples with larger grain sizes (annealing at 500 K and quenching at 10^{11} K s^{-1}) have closer Young's modulus. As shown in Fig. 2c, d, the single crystal has Young's modulus of ~58 GPa at 300 K compared to ~57 GPa of the sample prepared by isothermal solidification at 500 K at the slowest quench rate of 10^{11} K s^{-1} .

Table 1 shows a detailed comparison of Young's modulus obtained in this study with the single crystal and nanocrystalline experiments. The results from this MD calculation are close to previous experimental results. The margin of error is less than 5% for single crystal including Young's modulus determined by indentation tests^{26,27}. Similarly, for nanocrystalline aluminum, the difference of Young's modulus predicted by MD simulation and the one from experiments is below 10%. However, it should be noted that the experiments are normally done with pristine single crystal aluminum with preexisting evenly distributed grain boundaries (GBs), whereas the samples produced in all our solidification simulations have several defects (such as dislocations, twins, vacancies etc.) and randomly distributed GBs.

Unlike Young's modulus, UTS largely differs between single crystal and polycrystals (see Fig. 2e, f). At deformation temperature of 300 K and strain rate of 10^{10} s^{-1} , (Fig. 2e), the UTS of single crystal aluminum is ~7.25 GPa whereas the sample solidified isothermally at 500 K has the UTS of ~3.75 GPa. So, the UTS of the single crystal is almost twice as the polycrystalline for different strain rates and deformation temperatures. UTS of isothermally produced samples and quenched samples is similar and remains between 2.5 and

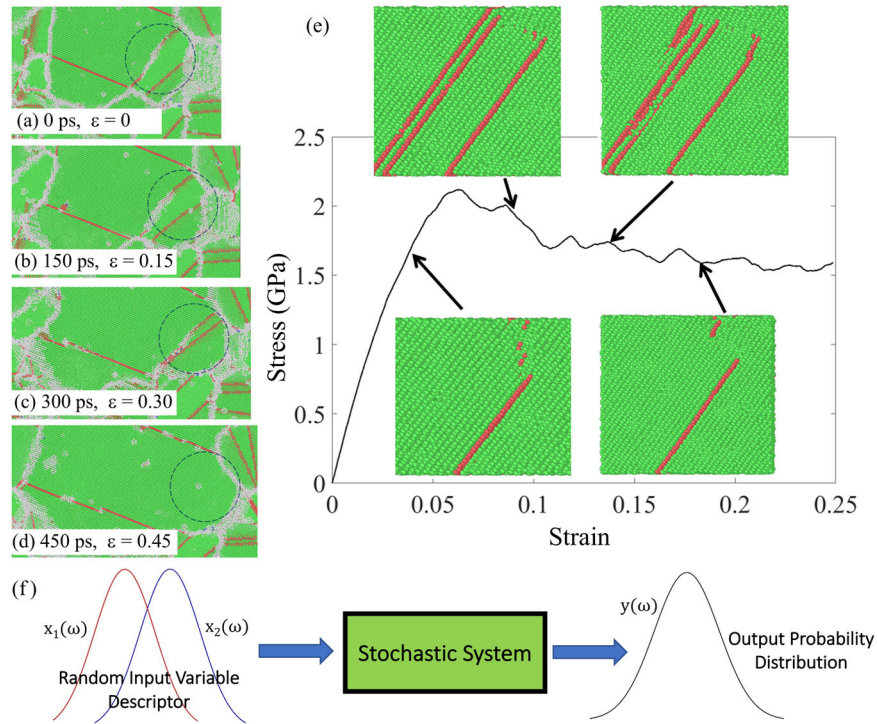


Fig. 1 Tensile deformation of an isothermally solidified aluminum sample under stochastic environment. **a** The solidified polycrystalline aluminum before applying deformation, **(b–d)** Snapshots of the polycrystalline aluminum during deformation under the strain rate of 10^9 K s^{-1} at 300 K, **(e)** Stress–strain plot of the polycrystalline aluminum. The inset images show detwinning during plastic deformation. **f** Schematic representation of the stochastic analysis. The random input parameters x_1 and x_2 are temperature and strain rate respectively, while the outputs y are Young's modulus, yield stress and ultimate stress. The parameter ω is used to represent the stochastic nature of the input and output quantities.

4 GPa. The temperature acts as the major factor in reducing the strength for both single crystal and polycrystals. UTS of a single crystal goes down from 7.25 GPa to 5.5 GPa (Fig. 2e, f) with the increase in deformation temperature from 300 K to 500 K which is a decrease of strength by almost 24%. However, in samples isothermally solidified, UTS reduces only by 10%; for example, for the sample isothermally solidified at 500 K in Fig. 2e, UTS goes down from 3.5 GPa to 3.25 GPa in this deformation temperature range. The strength also reduces by a similar margin for quenched aluminum. The values of the UTS for aluminum are relatively higher in comparison to experimental data. However, our results of 2–7 GPa are well-aligned with existing MD simulations of aluminum nanocrystalline structures^{28–30}. We also studied the effects of defects on solidified polycrystalline aluminum under similar strain rates, where the strength of aluminum remains within 2–7 GPa range²⁵. The higher strength of nanocrystalline aluminum utilizing MD simulation generally occurs due to the smaller size of simulation box and shorter time scale of pico to nano seconds. Such differences between MD simulation and bulk results are well-accepted by the computational materials science community. In theory, we could build a micro or millimeter-scale simulation box for achieving better experimental agreement, but that will contain trillions of atoms and will be unrealistic to simulate by using classical MD simulations.

It can be noted from Fig. 2e, f that the samples prepared at higher annealing temperatures or slower cooling rate have a higher UTS. As shown in Fig. 2e, the UTS for the samples isothermally solidified at 500 K is much higher than those solidified at 300 K. And the slowest quench rate of 10^{11} K s^{-1} produces the highest strength among all the polycrystalline samples prepared by quenching (Fig. 2d). Classically, one would expect an increase in yield stress (σ_y) for smaller grain sizes according to the Hall–Petch equation^{31,32} given by

$$\sigma_y = \sigma_0 + kd^{-1/2}, \quad (1)$$

where σ_0 is the friction stress in the absence of grain boundaries, k is a strengthening coefficient and d is the grain size. In general yield stress increases as grain size decreases because pile-ups in fine-grained materials contain fewer dislocations, the stress at the tip of the pile-up decreases and, thus, a larger applied stress is required to generate dislocations in the adjacent grain. In very small grains, this mechanism breaks down as grains are unable to support dislocation pile-ups. As a consequence, a threshold value is expected at which a maximum yield stress can be achieved. Experimental studies on nano-crystalline aluminum have revealed that the Hall–Petch slope (k value in Eq. (1)) is reduced or becomes negative below a certain grain size^{33–35}. This is known as the inverse Hall–Petch effect. Furthermore, the transition in the Hall–Petch slope normally occurs for grain sizes above 25 nm, where dislocation pile-ups are still possible³⁶. The maximum grain size achieved in this study is $\sim 14 \text{ nm}$, so we don't observe the transition in the Hall–Petch effect in this study. The inverse Hall–Petch relationship was discussed in detail for all the polycrystalline samples in our previous work³⁷.

In this subsection, we have discussed the critical deterministic physics involved behind the mechanical analysis of solidified aluminum that leads to the current output quantities of interest such as Young's modulus, yield stress and ultimate stress. The ML models are created here considering the influencing factors like strain rate and temperature for mapping the output quantities based on a data-driven Bayesian framework with the aim of generating stochastic insights (refer to the following subsection).

Machine learning model formation and stochastic analysis

As discussed in the preceding section, the mechanical properties vary with changes in temperature and strain rate considerably. To understand how the mechanical properties are influenced by these two factors, ML-assisted mapping has been performed

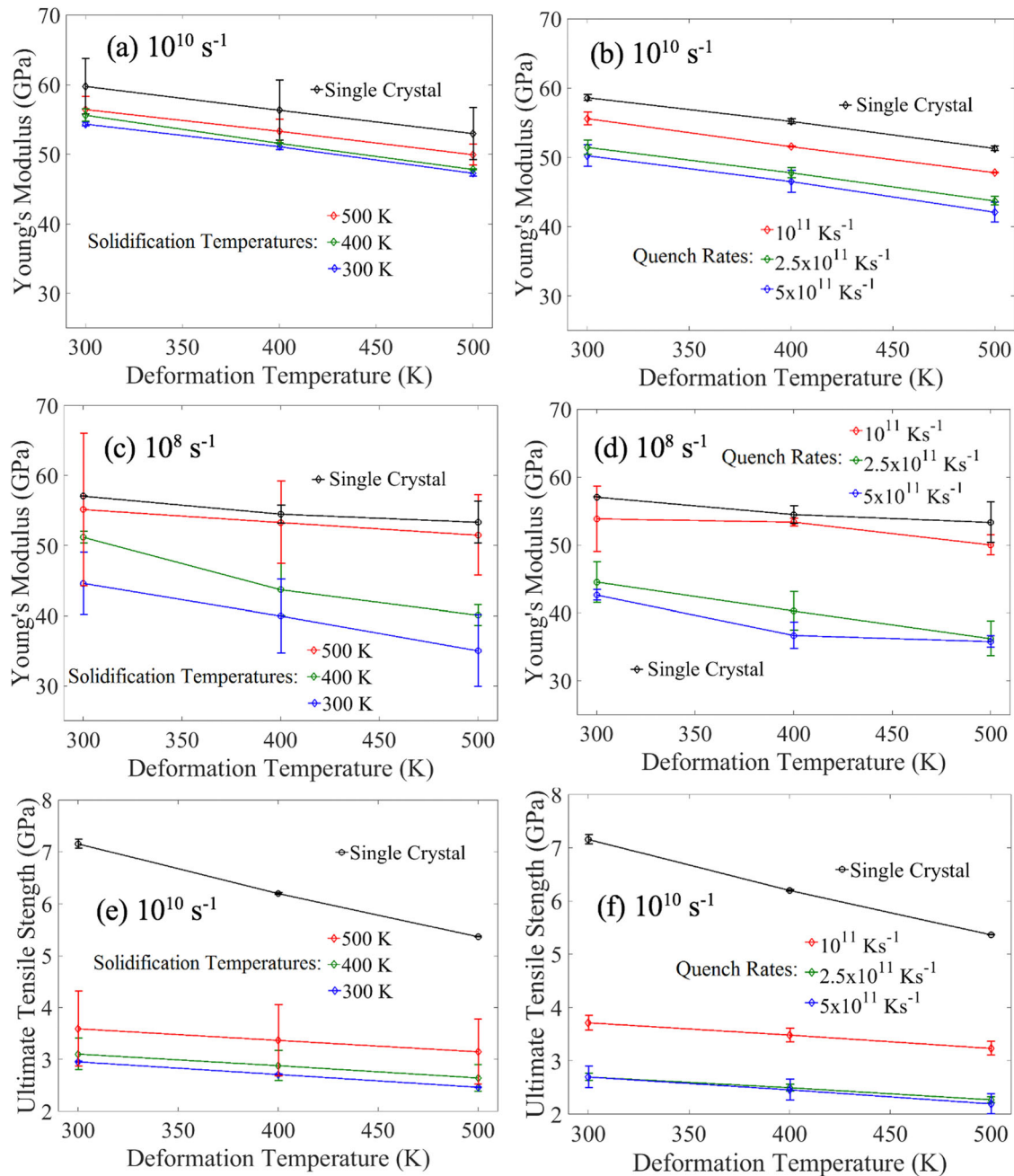


Fig. 2 Young's modulus and ultimate tensile strength for isothermally solidified polycrystalline samples. Young's modulus of (a, c) isothermally solidified polycrystalline samples at the strain rate of 10^{10} s^{-1} and 10^8 s^{-1} , and (b, d) solidified polycrystalline samples prepared by quenching at the strain rate of 10^{10} s^{-1} and 10^8 s^{-1} . UTS of (e) isothermally solidified polycrystalline at the strain rate of 10^{10} s^{-1} , and (f) solidified polycrystalline samples prepared by quenching at the strain rate of 10^{10} s^{-1} . The error bars represent the deviation originating from the three different simulations for three different deformation directions.

considering the individual and combined methods of training (refer to section 2.2.1 and 2.2.2). First, simulations have been performed considering the individual method (refer to section 2.2.1). As shown in Fig. 3a the margin of error remains as high as 37% for UTS. The prediction of Young's modulus remains close to MD simulation data (i.e., ground truth). However, as we calculate the yield stress or UTS, the ML predictions deviate considerably from MD simulations. When we follow the combined method (refer to section 2.2.2), the ML predictions improve significantly with respect to the MD-based ground truths (15% or more), Fig. 3b. The reduction in error can be attributed to the fact that the ML model for the combined method also learns from other datasets

while predicting for a particular case. To visualize the consistency and prediction accuracy between the ML model and MD results, we show parity plots of all the output quantities (i.e., Young's modulus, yield stress and UTS) for isothermally solidified sample at 400 K (Case 3) in Fig. 3c–e. Given the superior accuracy of the combined method, the same has been used for generating these plots. To illustrate the generalization of the proposed approach, a test dataset (dataset not used during training, i.e., unseen data points) has been used for the parity plots. The agreement between the MD and ML results is quite evident and within acceptable limits, suggesting the validity of ML model for further analyses.

There are multiple ways to check the prediction accuracy of ML models^{13,38}. Though R^2 values are a popular way to have some preliminary sense of accuracy, it is widely accepted that checking only R^2 is not adequate and can sometimes be misleading. For this reason, we have resorted to parity plots or scatter plots, which provide a way for direct comparison between predicted results and ground. In such a plot, the closer the points are to the diagonal line, the lesser the prediction error is. From the presented results, we notice that the points are quite close to the diagonal line for the combined method, indicating its high prediction accuracy. We have further reported the maximum error values which show reasonably less prediction error for the converged cases, ascertaining further confidence in ML models. Note that the maximum error is a stricter indicator of the prediction quality compared to the mean error or the root mean square error. For comprehensiveness, we have also checked the R^2 values, wherein we note that the values are above 0.97. This further indicates a high prediction accuracy as the values are very close to 1. Based

Table 1. Young’s modulus values for deformation at 300 K for bulk aluminum in the form of single crystal and nanocrystalline, obtained by MD calculations and/or experimental measurements.

Bulk Aluminim	Young’s Modulus (GPa)	
	Current MD work	Previous Results (Methods)
Single Crystal	58-64	67.2-69.5 (Expt.) ²⁶ 64 (MD) ²⁵ 62.3 ± 3.1 (Expt.) ²⁷
Nanocrystalline	40-56 (Average grain size: 5–14 nm)	67 (MD, Grain size 11.1 nm) ⁵⁹ 60.2 (Expt., Average grain size: 11.1 nm) ⁶⁰

The average grain size for polycrystalline cases is given in the parenthesis.

on these three forms of validation or accuracy analysis, we conclude that ML models have a high prediction capability.

As the combined method produces a much lower error margin, we have obtained the probability density functions based on these surrogate models by performing MC simulations. Unlike conventional frequentist ML approaches, the proposed H-PCFE based approach accounts for the modeling uncertainty by considering both mean and standard deviation of the MD simulation results. Additionally, H-PCFE being a Bayesian surrogate model, also captures the epistemic uncertainty due to limited and noisy training data. The error bounds (shaded portion) shown in Fig. 4 represent the overall uncertainty captured using the proposed approach. The probabilistic distribution of Young’s modulus for single crystal and isothermally solidified samples at different temperatures follows the normal distribution. The effect of quench rate on the probabilistic descriptions of Young’s modulus, yield stress and UTS is presented in Fig. 5. Interestingly, all the probability distribution plots roughly follow normal distribution with a single peak under the compound effect of source uncertainties. In this context, it should be noted that we have presented the probability density function (PDF) plots in their original form based on raw data rather than fitting them to a particular predefined distribution. This allows us to show the probabilistic characteristics of the data without any bias. An asymmetric distribution of Young’s modulus, yield stress or ultimate stress can be advantageous or disadvantageous from a mechanical design perspective depending on the nature of asymmetry. For example, if the PDF is asymmetric towards higher values (i.e., positively skewed distribution) of Young’s modulus, yield stress or ultimate stress, this means the probability of having a higher value of these quantities is higher and can be treated as advantageous.

Sensitivity analysis

The relative sensitivity of the mechanical properties is quantified here for both strain rate and temperature. Moment-independent

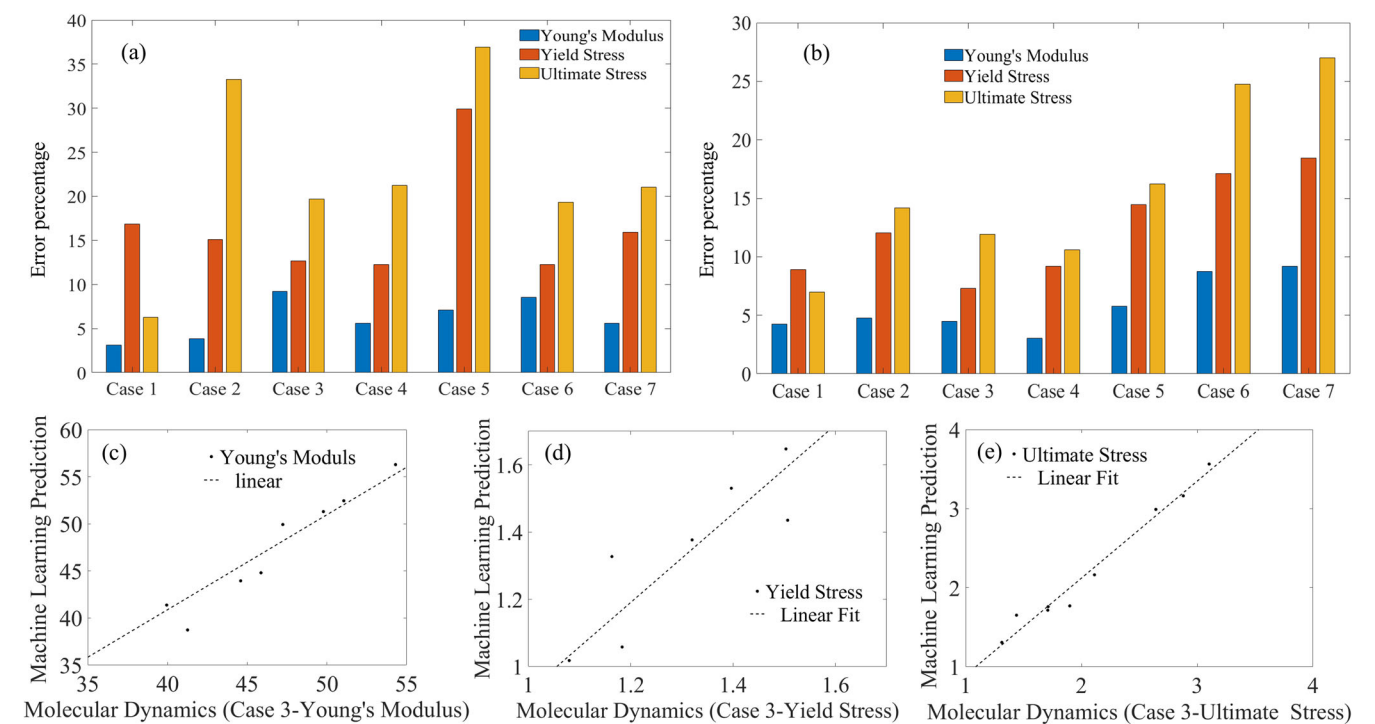


Fig. 3 Error margins in H-PCFE predictions based on individual and combined methods. Error margin for different cases in H-PCFE predictions, obtained using (a) individual method and (b) combined method. Representative scatter plots for the ML model and MD prediction considering (c) Young’s modulus, (d) yield stress and (e) UTS for all the case 3 or isothermally solidified samples at 400 K. Young’s modulus, yield stress and ultimate stress are presented here in GPa unit.

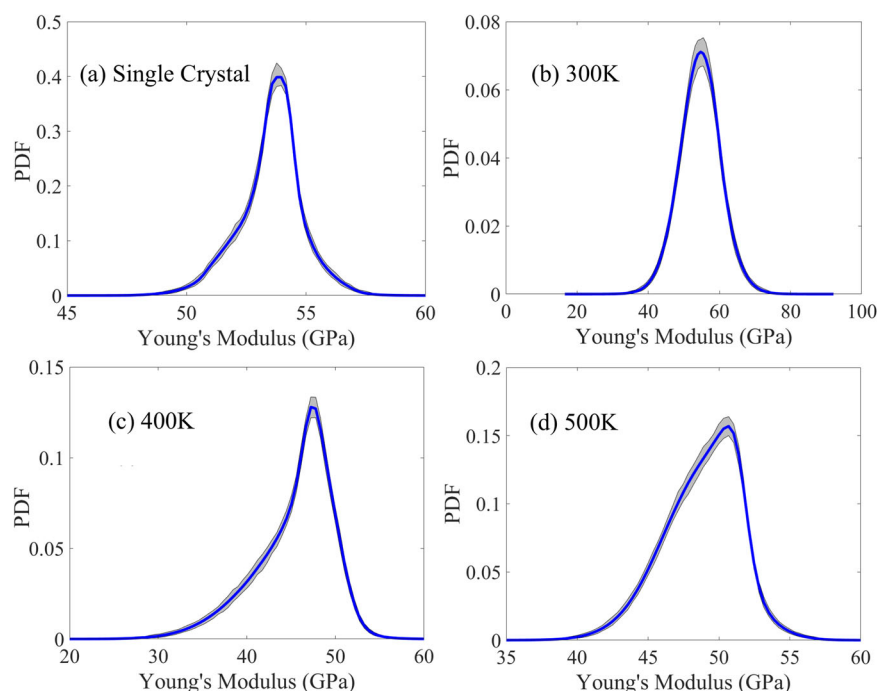


Fig. 4 Probabilistic descriptions of Young's modulus for various aluminum samples. Probability distribution function plots of Young's modulus for different aluminum samples of (a) single crystal, isothermally solidified at (b) 300 K, (c) 400 K and (d) 500 K.

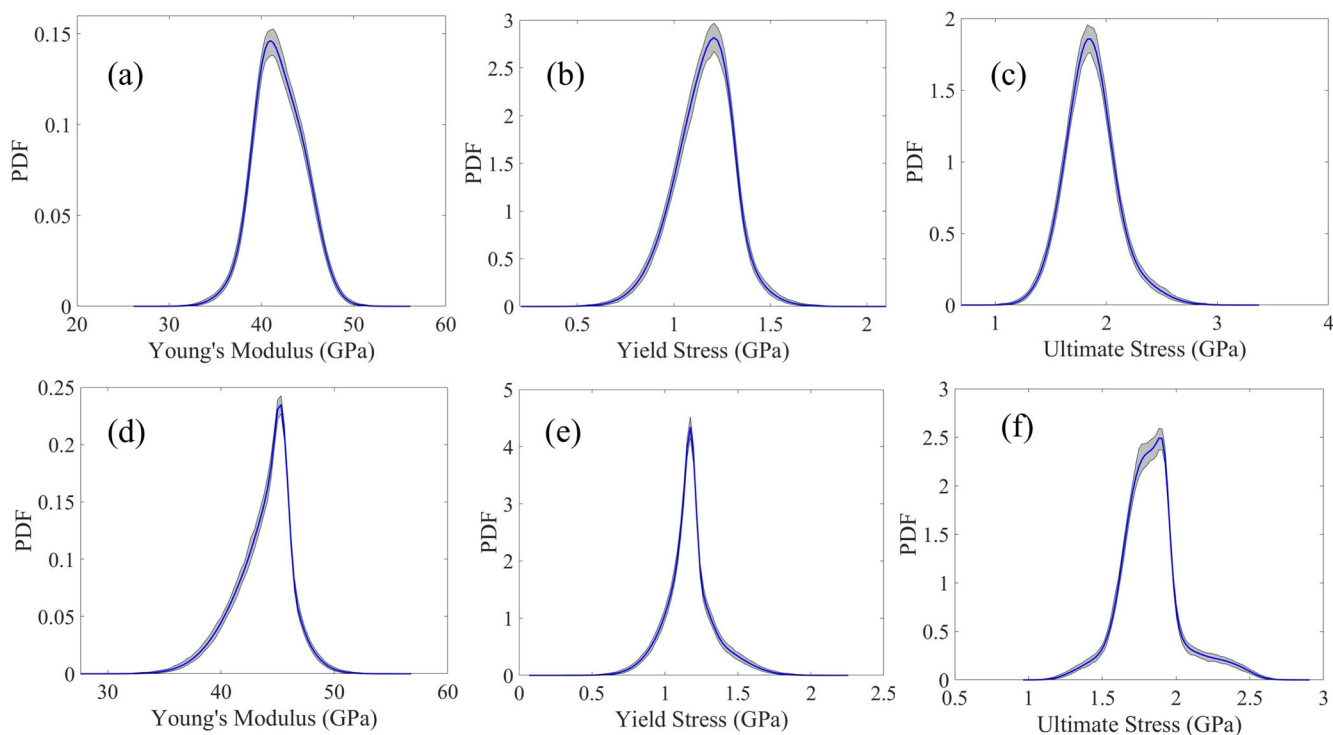


Fig. 5 Probabilistic descriptions depending on quench rate. Probability distribution function plots of Young's modulus, yield stress and ultimate stress for different Al samples, solidified at quench rate of (a–c) $2.5 \times 10^{11} \text{ K s}^{-1}$ and (d–f) $5 \times 10^{11} \text{ K s}^{-1}$.

sensitivity analysis is performed for obtaining the sensitivity plots^{39,40}. The advantage of moment-independent sensitivity measure resides in the fact that it utilizes the complete information of the probability distribution and not just the second-order moments; detailed descriptions are provided in

Supplementary Material (Section III). However, moment-independent sensitivity analyses are expensive from a computational point of view. To address this issue, the trained H-PCFE models based on the combined method (refer to section 2.2.2) are exploited in the current analysis.

From the sensitivity analysis results presented in Fig. 6, we observe that the mechanical properties are less influenced by temperature than strain rate for all the solidified samples. As there are no grain boundaries present in single crystals, the motion of atoms with the increase in temperature is much more significant than the variation in strain rate. Unlike the single crystal, the solidified polycrystalline samples have grain boundaries, which may either resist or facilitate the process of deformation. As shown

in Fig. 6b–g the strain rate is a dominant factor in determining the mechanical properties of the aluminum samples. In polycrystalline samples, the position and orientations of the grain boundaries are also influencing factors that can determine the sensitivity of the two parameters. As shown in Fig. 6e, Young's modulus is equally sensitive to the temperature and strain rate for the sample solidified at 10^{11} K s^{-1} . Due to the slower quench rate, the grain boundaries are larger than any other solidified samples, and as a

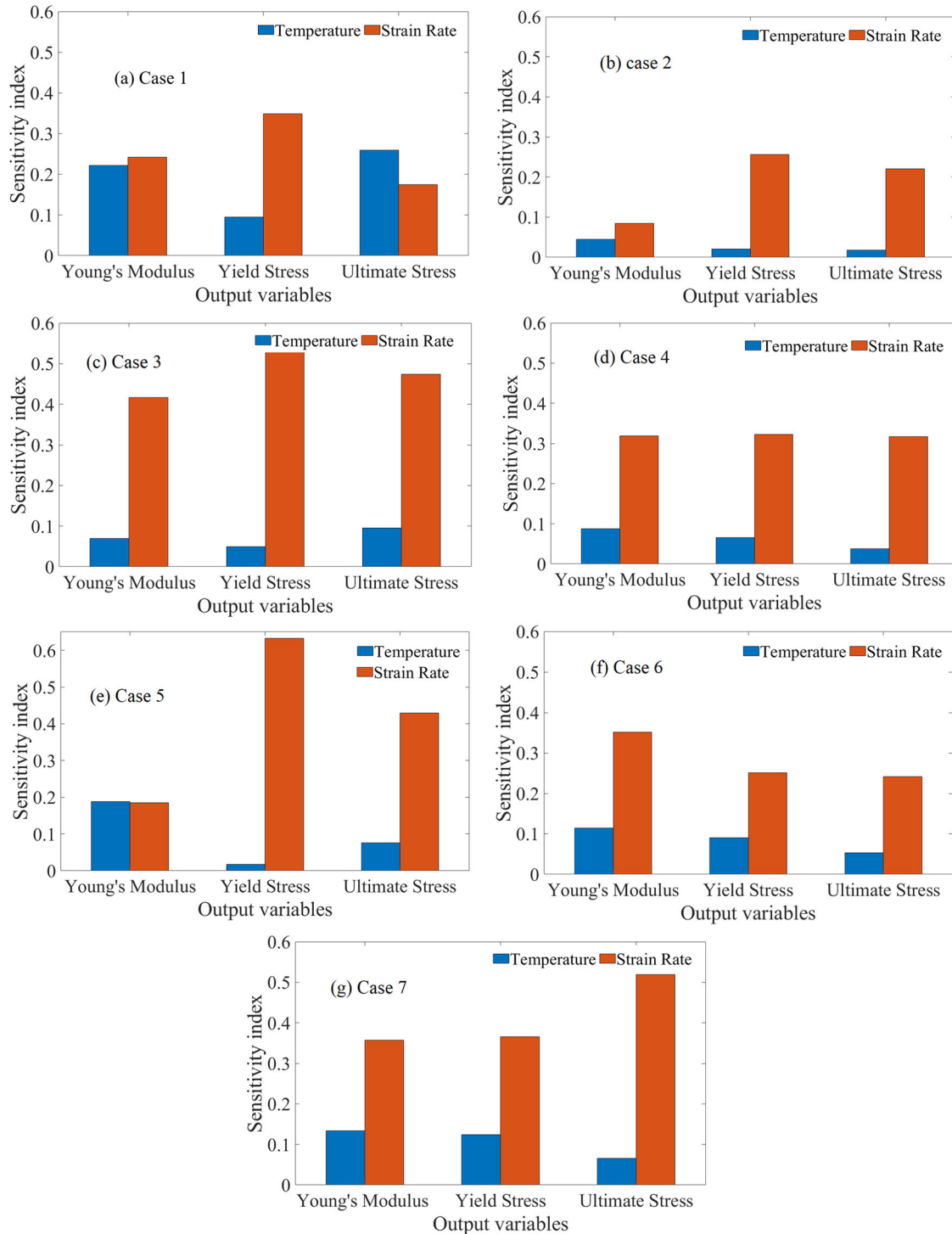


Fig. 6 Sensitivity quantification for temperature and strain rate. Sensitivity to temperature and strain rate calculated for (a) single crystal, isothermal temperatures at (b) 300 K, (c) 400 K, (d) 500 K and the samples prepared by quenching rates of (e) 10^{11} , (f) 2.5×10^{11} and (g) $5 \times 10^{11} \text{ K s}^{-1}$.

result it behaves similarly to the single crystal one for Young's modulus. The yield and ultimate stresses, however, are more sensitive to strain rate, which shows the polycrystalline nature of the samples. An important general observation made from all the results is that the presence of grain boundaries and twin boundaries makes the samples more sensitive to the strain rates applied to perform tensile deformation than the deformation temperature.

DISCUSSION

Solidification has been at the heart of the most important manufacturing processes, including rapidly evolving techniques like additive manufacturing, where the quantifications of stochastic variations and manufacturing uncertainties are critically important. Efficient Bayesian ML-based schemes are proposed in this work for full-scale probabilistic characterization and sensitivity analysis of the mechanical properties of solidified metals.

Although MD simulations have the capability of predicting mechanical properties, they can be computationally demanding depending on the system size, the efficiency of interatomic potentials, and the required level of accuracy. The situation further aggravates in the case of stochastic simulations that need thousands of realizations for carrying out MC simulations. Since quantifying uncertainty in the mechanical properties of solidified materials is getting significant traction with the rapid adoption of additive manufacturing technologies, there is a strong rationale to develop efficient stochastic computational frameworks. In this context, ML-based approaches can be integrated with MD simulations for improving the feasibility of performing full-scale probabilistic characterization. Conventional ML models require a large number of consistent training data points that can be generated based on MD simulations. Further, since the MD simulation itself is a random process, there is always a notion of inherent variability involved with the repeatability, even with the same simulation parameters. This leads to a mean and standard deviation corresponding to each of the training data points. In this article, we have proposed a data-efficient Bayesian ML approach coupled with MD simulations, referred to as the hybrid polynomial correlated function expansion, for establishing a computational mapping between the crucial influencing parameters (like strain rate and temperature) and mechanical properties (like Young's modulus, yield stress and ultimate strength). The proposed approach accounts for the effect of uncertainty in training data by exploiting mean and standard deviation of the MD simulations, which in principle can address the issue of repeatability in stochastic simulations with low variance.

The MD simulations for studying the mechanical deformation of single crystal and polycrystalline aluminum were carried out based on 2NN MEAM interatomic potential. The samples were prepared by quenching or isothermal solidification and were deformed under uniaxial tension in three orthogonal directions to determine the mechanical properties while taking into account the errors. The errors or variations originate in solidified polycrystalline aluminum samples where the grain boundaries, vacancies and twin boundaries form spontaneously in arbitrary directions. We have coupled the current MD simulation approach with Bayesian ML models by incorporating the mean and standard deviation of such variations. Two different approaches have been proposed in the context of training the ML model, individual and combined methods, wherein the latter is found to perform significantly better in terms of prediction. Complete probabilistic descriptions are presented based on the ML models showing a significant degree of stochastic response bounds (including symmetric, positively and negatively skewed Gaussian distributions) in the mechanical properties due to the inherent uncertainties involved in temperature and strain rate. Such quantifiable numerical outcomes lead to the realization of the importance of adopting

an inclusive analysis and large stochastic data-informed design framework considering the inevitable random variabilities. The sensitivity analysis results reveal that Young's modulus, yield stress and ultimate strength are relatively more sensitive to strain rate than temperature. Thus, relatively more operational control needs to be ensured in the strain rate to avoid undesirable stochastic variations in the mechanical properties.

It should be noted that a wide range of input parameters is considered in this study for forming ML models. This covers a large MD simulation domain of influencing factors. In principle, ML models are able to predict the outputs accurately even beyond the training domain as long as the governing equations are continuous with no drastic change in their nature and no additional parameters come into play. In such cases, more training data would be needed, and following the same computational framework as proposed here, one can expand the prediction range. Another way of dealing with such cases and to capture the physics more accurately is to involve physics-informed ML models^{41,42}. Since the current focus of this paper is data-driven ML models^{13,43}, such an analysis is out of scope here.

In Summary, efficient Bayesian ML-based schemes are proposed in this work for full-scale probabilistic characterization and sensitivity analysis of the mechanical properties of solidified metals. We have developed a data-efficient approach coupled with MD simulations, referred to as the hybrid polynomial correlated function expansion, for establishing a computational mapping between strain rate, temperature and mechanical properties like Young's modulus, yield stress and ultimate strength. The physics concerning the mechanical deformation of single crystal and polycrystalline aluminum is captured through training based on MD simulations considering the 2NN MEAM interatomic potential. The proposed approach accounts for the effect of uncertainty in training data by exploiting mean and standard deviation of the MD simulations concerning solidified polycrystalline aluminum samples where the grain boundaries, vacancies and twin boundaries form spontaneously in arbitrary directions. This, in principle, can address the issue of repeatability in stochastic simulations with low variance. The complete probabilistic descriptions reveal a significant degree of stochastic variations in the mechanical properties due to the inherent uncertainties involved in temperature and strain rate.

In general, the numerical results presented here illustrate that the proposed Bayesian ML-based approach can reliably predict the mechanical properties in a computationally efficient error-inclusive framework for data-intensive analyses such as uncertainty and sensitivity quantification. The proposed data-driven framework is generic in nature and can be extended to other metals and alloys for predicting different multi-physical properties.

METHODS

Molecular dynamics simulation

Deterministic MD simulations of solidification and deformation of solidified aluminum^{4,37,44} are at the core of the current ML-assisted analysis. MD simulations of homogenous nucleation from pure aluminum melt were performed in a simulation box with size of $25 \times 25 \times 25 \text{ nm}^3$ ($64 \times 64 \times 64$ unit cells, with 1000188 atoms) and with the isothermal-isobaric (NPT) ensemble. Periodic boundary conditions were employed in all three directions during the melt preparation. A non-periodic boundary condition was applied in the direction of the solidification. The time step of simulations was 0.003 ps. The adequate model size is determined based on a convergence study and validating the results with existing literature (refer to Tables 1 and 2). The second nearest neighbor-modified embedded atom method potential (2NN-MEAM) for aluminum⁹ was used for the MD simulations. We recently tested this interatomic potential which showed reliable

Table 2. The low-temperature elastic properties and high-temperature melting properties of aluminum predicted by 2NN-MEAM⁹ MD simulations.

Properties	Experimental	2NN-MEAM MD
Bulk modulus (GPa)	76.4 ⁶¹	79.4
C_{11} (GPa)	111.5 ⁶¹	114.3
C_{12} (GPa)	58.8 ⁶¹	61.9
C_{44} (GPa)	29.5 ⁶¹	31.6
Specific Heat ($\text{J mol}^{-1} \text{K}^{-1}$)	26.15 ⁶²	24.70
Thermal expansion coeff. (10^6K^{-1})	17.31 ⁶²	23.50
Melting point (T_m) (K)	934 ⁶³	925
Solid-liquid interface free energy (mJ/m^2)	168.9 \pm 21 to 158 \pm 30 ^{64–66}	172.6

predictions for solid-liquid coexistence properties (i.e. melting point, solid-liquid energy, melting point, specific heat etc.) of aluminum^{4,9}.

The melting and solidification simulations were done in two steps. First, we prepared the aluminum melt and performed equilibration at 1325 K. We obtained a homogenous liquid aluminum in the simulation box at about 100 ps. The simulation was continued to 300 ps to make sure the initial melt was properly equilibrated. Second, after the melt was ready, the next stage of equilibration was performed to create the solid samples by keeping the aluminum melt at a constant temperature of 300, 400 and 500 K for 3,000 ps. Each isothermal simulation was repeated five times to evaluate the possible errors. Further, each isothermal simulation was run for a total of 500 ps to simulate the crystal nucleation and solidification. Additional details about MD simulations of solidification and deformation can be found in the Supplementary Material (Section I) and in our previous works on aluminum^{4,25,45}. Table 2 shows the low-temperature elastic properties and high-temperature melting properties of aluminum determined by 2NN-MEAM MD simulations. A good agreement can be noticed with the results from the literature. This gives us adequate confidence in using the current MD simulation methodology to generate the training data for the development of the subsequent ML model.

The OVITO visualization package was used to monitor the deformation processes⁴⁶. Within OVITO, common neighbor analysis (CNA) was used⁴⁷ to identify the local crystalline structure of atoms. Using CNA, one can distinguish atoms in different crystal structure regions by calculating the statistics of diagrams formed from the nearest neighbors of each atom and comparing them with those previously known for standard crystals. Atoms not identified as face-centered cubic, hexagonal close-packed, or any other crystal type as implemented in OVITO can be identified as amorphous liquid or amorphous solid atoms.

After completing the solidification simulations at different isothermal and quenching conditions, MD simulations of the deformation of solidified single and polycrystalline aluminum samples were completed. The isothermal samples were prepared by keeping the aluminum melt at a constant temperature of 300, 400 and 500 K for 3000 ps. The results concerning the effect of undercooling temperature on the average grain size were reported in our previous work^{4,37,48,49}, showing that the average grain size increases with decreasing the undercooling. In later part of the present article, the effect of grain size on mechanical properties is investigated. For simulating quenching conditions, aluminum melt is quenched from 1325 K to 300 K at cooling rates of 10^{11} , 2.5×10^{11} and $5 \times 10^{11} \text{ K s}^{-1}$. Six polycrystalline samples were created, three for isothermal and three for quenching cases.

Table 3. Single crystal, isothermal and quenching solidification simulations.

Case 1	Single crystal	
Case 2	Isothermal	300 K
Case 3		400 K
Case 4		500 K
Case 5	Quenching	10^{11} K s^{-1}
Case 6		$2.5 \times 10^{11} \text{ K s}^{-1}$
Case 7		$5 \times 10^{11} \text{ K s}^{-1}$

The quenching method was used to mimic the actual experimental procedure for producing undercooling where the temperature decreases from above the melting temperature with a certain cooling rate. More details on the isothermal and quenching simulations are available in our previously published works^{4,25,37,45,48,49}. All the polycrystalline aluminum models were deformed at three different deformation temperatures (300, 400 and 500 K) and three strain rates (10^8 , 10^9 and 10^{10} s^{-1}), as shown in Table 3. During the MD simulation, we fix one end of the simulation box and apply “constant displacement” over a time, which is effectively a “constant engineering strain rate”. This means the box dimension changes linearly with time from its initial to final value. This method can be applied in LAMMPS by utilizing ‘erate’ command⁵⁰. The stress in MD simulations is a measurement of pressure, and it can be done by using standard expressions for virial stress⁵¹, where the further the atoms move the higher the stress becomes. As a result, if we deform in the x directions, the values for the stress in the x-direction are the actual engineering stress, and the stress in the y and z directions remains close to zero.

For the purpose of comparison, we also deformed a single crystal aluminum at the same strain rates and temperatures (9 cases). To calculate the statistical error from all the simulations, each uniaxial tensile simulation was replicated in x, y and z directions. The average stress data were recorded every 1000 steps (or 0.3 ps). Overall, 195 simulations (6 solidification cases and 63 deformation cases each at 3 directions) were performed to analyze the deformation behavior and mechanical properties of solidified aluminum. The results of these simulations are exploited in the next stage for ML model formation.

H-PCFE based machine learning modeling

In general, MD simulations with adequate model sizes can involve unreasonably high computational expenses for probabilistic analyses where thousands of random realizations are necessary. The situation further aggravates in the case of high strain rate simulations. Reducing model size can lead to inaccurate results in each random realization, resulting in a large cumulative statistical error in the probabilistic results obtained through MC simulations. The challenges of running multiple MD simulations at the necessary time and length scales can be addressed by ML-based approaches. In this work, we propose to perform a few MD simulations, with an adequate model size, based on which the ML models are formed. Subsequently, we perform MC simulations using efficient yet high-fidelity ML models (free from inaccuracies due to small model sizes in MD) to probabilistically characterize the properties of solidified metals. The ML-based approach acts as a surrogate to the computationally expensive MD simulations and predicts the macro-scale properties of solidified metals efficiently. Unlike conventional ML models, the proposed H-PCFE, being a Bayesian ML approach, is data efficient. Further, it can account for the effect of uncertainty in training data by exploiting mean and standard deviation of the MD simulations, which in principle

addresses the issue of repeatability in stochastic simulations with low variance.

The H-PCFE based ML approach is integrated with MD simulations^{24,52}. H-PCFE can be viewed as a type of Gaussian process (GP)^{21,22} where the mean function is represented by using polynomial correlated function expansion (PCFE)⁵². PCFE is formulated by combining the well-known analysis-of-variance (ANOVA) decomposition⁵³ with polynomial chaos expansion (PCE)⁵⁴. Detailed formulation of H-PCFE is provided in the Supplementary Material (Section II).

Consider $\mathbf{X} = \{X_1, X_2, \dots, X_N\} \mathbb{R}^N$ to be the input variables and Y to be the output. In H-PCFE, a probabilistic mapping between the output Y and the inputs \mathbf{X} is assumed

$$Y|\mathbf{X} \sim \mathcal{GP}(\mu_{\text{PCFE}}(\cdot; \boldsymbol{\alpha}), C(\cdot, \cdot; \boldsymbol{\theta}_{\text{GP}})), \quad (2)$$

where $\mathcal{GP}(\cdot)$ represents a GP with mean-function μ_{PCFE} and covariance function C . As for the covariance function C , there exists a number of functions in the literature⁵². In this work, the Gaussian covariance function has been used. $\boldsymbol{\alpha}$ in Eq. (2) represents the unknown coefficients in PCFE, and $\boldsymbol{\theta}_{\text{GP}}$ represents the length-scale parameters and process variance of GP. $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\theta}_{\text{GP}}]$ are referred to as the hyperparameters of H-PCFE. The hyperparameters of H-PCFE are computed by using maximum likelihood estimation and homotopy algorithm⁵⁵. The estimation is carried out based on training dataset Ξ_t and \mathbf{Y}_t . For further details on H-PCFE, interested readers may refer to^{24,53,56}. For details on homotopy algorithm, readers may refer to^{55,57,58}.

In this work, there are two inputs and three outputs. The two inputs are temperature and strain rate while the outputs are Young's modulus, yield stress and ultimate stress. Training data for computing the hyperparameters of H-PCFE are generated by running MD simulations. The MD simulations are run for seven different solidification cases as shown in Table 3 (as discussed in the preceding subsection). For training the ML model, we have proposed two different approaches as discussed below.

Individual method. In this method, we train a different H-PCFE model for each dataset. In other words, we train seven H-PCFE models, each mapping the two inputs and the three outputs. The primary bottleneck of this method resides in the fact that we only have 9 training data (along with the information regarding mean and standard deviation) for training each H-PCFE model. Therefore, it is unlikely that the results obtained using this method will be accurate.

Combined method. To address the issue associated with individual method, we propose a combined method where the idea is to combine datasets from all the seven cases. To that end, a dummy variable Z is introduced. The variable Z is used to indicate the dataset from which the input training sample is taken. The input training matrix takes the following form

$$\Xi_t = \{X_1^i, X_2^i, Z^i\}_{i=1}^{N_s} = \begin{bmatrix} X_1^1 & X_2^1 & 1 \\ X_1^2 & X_2^2 & 1 \\ \vdots & \vdots & \vdots \\ X_1^1 & X_2^1 & 2 \\ X_1^2 & X_2^2 & 2 \\ \vdots & \vdots & \vdots \\ X_1^9 & X_2^9 & 7 \end{bmatrix} \quad (3)$$

where X_1 indicates temperature, X_2 indicates strain rate, and N_s indicates the number of training samples. With this, it is possible to train a single H-PCFE model for all seven cases. The advantage of this setup is two-fold. First, the issue associated with extreme sparsity of data is addressed; instead of 9, the H-PCFE model now has 63 data points to be trained by. Second, with such augmentation, it is

possible to incorporate the correlation among the dataset from different cases. This is expected to enhance the accuracy of the H-PCFE model. Note that the information regarding mean and standard deviation are exploited here for each of the data points.

Stochastic analysis and Monte Carlo simulation

Complete probabilistic description of the mechanical properties, such as Young's Modulus, yield stress and ultimate stress, are obtained based on MC simulations. MC methods (or MC experiments/simulations) are a broad class of computational algorithms that rely on repeated random sampling following a particular probability distribution (such as random uniform distribution) to obtain numerical results. These methods are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use direct mathematical algorithms. Figure 1f presents a schematic representation of a general stochastic system having two random input parameters and one output response. In general, though MC-based analyses are capable of obtaining comprehensive results for a physical/mathematical problem, these are computationally very expensive. These methods normally require thousands of simulations/experiments to be carried out corresponding to random input sets. Thus, the entire process becomes cost intensive, especially for problems where individual simulations/experiments are very costly and time consuming such as MD simulations. To overcome this difficulty, we have employed the ML model as an efficient surrogate of the actual MD code to carry out MD simulation and subsequent sensitivity analyses. Thus, the ML model is capable of obtaining Young's modulus, yield stress and ultimate stress corresponding to any set of values of temperature and strain rate within the considered design space.

The computational efficiency of the proposed ML-assisted framework can be quantified here in terms of the actual number of MD simulations (n_d) for developing the ML models and the number of function evaluations (n_f) for different analyses carried out such as probabilistic characterization and sensitivity analysis (refer to section 3.2 and 3.3). It is noted that the time taken for forming the ML models after having the training dataset ready is negligible compared to a single MD simulation. Thus, the computational efficiency in the current framework can be quantified as $\epsilon = \frac{n_d}{n_f}$. For the combined method proposed here, the computational efficiency can be evaluated as $\epsilon = \frac{63}{10000}$, which leads to efficiency more than 150 times of a purely MD simulation-based analysis without involving ML. This demonstrates that the current probabilistic analysis is practically impossible to conduct using only MD simulations.

DATA AVAILABILITY

All necessary data generated or analyzed during this study are included in this published article and the supplementary material. Other auxiliary data are available from the corresponding author.

Received: 10 August 2023; Accepted: 3 January 2024;

Published online: 24 January 2024

REFERENCES

- Wei, H., Mazumder, J. & DebRoy, T. Evolution of solidification texture during additive manufacturing. *Sci. Rep.* **5**, 1–7 (2015).
- Sreenivasan, S. Nanoscale manufacturing enabled by imprint lithography. *MRS Bull.* **33**, 854–863 (2008).
- Gailevičius, D. et al. Additive-manufacturing of 3D glass-ceramics down to nanoscale resolution. *Nanoscale Horiz.* **4**, 647–651 (2019).
- Mahata, A., Asle Zaeem, M. & Baskes, M. I. Understanding homogeneous nucleation in solidification of aluminum by molecular dynamics simulations. *Model. Simul. Mater. Sci. Eng.* **26**, 025007 (2018).

5. Kavousi, S., Ankudinov, V., Galenko, P. K. & Asle Zaeem, M. Atomistic-informed kinetic phase-field modeling of non-equilibrium crystal growth during rapid solidification. *Acta Mater.* **253**, 118960 (2022).
6. Kavousi, S., Gates, A., Jin, L. & Asle Zaeem, M. A temperature-dependent atomistic-informed phase-field model to study dendritic growth. *J. Cryst. Growth* **579**, 126461 (2022).
7. Kavousi, S., Novak, B. R., Moldovan, D. & Asle Zaeem, M. Quantitative prediction of rapid solidification by integrated atomistic and phase-field modeling. *Acta Mater.* **211**, 116885 (2021).
8. Asadi, E., Asle Zaeem, M., Nouranian, S. & Baskes, M. I. Quantitative modeling of the equilibration of two-phase solid-liquid Fe by atomistic simulations on diffusive time scales. *Phys. Rev. B* **91**, 024105 (2015).
9. Asadi, E., Asle Zaeem, M., Nouranian, S. & Baskes, M. I. Two-phase solid-liquid coexistence of Ni, Cu, and Al by molecular dynamics simulations using the modified embedded-atom method. *Acta Mater.* **86**, 169–181 (2015).
10. Asadi, E., & Asle Zaeem, M. The anisotropy of hexagonal close-packed and liquid interface free energy using molecular dynamics simulations based on modified embedded-atom method. *Acta Mater.* **107**, 337–344 (2016).
11. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
12. Dietterich, T. G. Machine learning for sequential data: A review, Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR) 15–30 (Springer, 2002).
13. Sharma, A., Mukhopadhyay, T., Rangappa, S. M., Siengchin, S. & Kushvaha, V. Advances in computational intelligence of polymer composite materials: machine learning assisted modeling, analysis and design. *Arch. Comput. Methods Eng.* **29**, 3341–3385 (2022).
14. Gupta, K., Mukhopadhyay, T., Roy, A., Roy, L. & Dey, S. Sparse machine learning assisted deep computational insights on the mechanical properties of graphene with intrinsic defects and doping. *J. Phys. Chem. Solids* **155**, 110111 (2021).
15. Mukhopadhyay, T., Mahata, A., Dey, S. & Adhikari, S. Probabilistic analysis and design of HCP nanowires: an efficient surrogate based molecular dynamics simulation approach. *J. Mater. Sci. Technol.* **32**, 1345–1351 (2016).
16. Gupta, K., Roy, A., Mukhopadhyay, T., Roy, L. & Dey, S. Probing the stochastic fracture behavior of twisted bilayer graphene: Efficient ANN based molecular dynamics simulations for complete probabilistic characterization. *Mater. Today Commun.* **32**, 103932 (2022).
17. Gupta, K., Mukhopadhyay, T., Roy, L. & Dey, S. High-velocity ballistics of twisted bilayer graphene under stochastic disorder. *Adv. Nano Res.* **12**, 529–547 (2022).
18. Goswami, S., Anitescu, C., Chakraborty, S. & Rabczuk, T. Transfer learning enhanced physics informed neural network for phase-field modeling of fracture. *Theor. Appl. Fract. Mech.* **106**, 102447 (2020).
19. Chakraborty, S. Simulation free reliability analysis: a physics-informed deep learning based approach. Preprint at <https://arxiv.org/abs/2005.01302> (2020).
20. Chakraborty, S. Transfer learning based multi-fidelity physics informed deep neural network. *J. Comput. Phys.* **426**, 109942 (2021).
21. Nayek, R., Chakraborty, S. & Narasimhan, S. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mech. Syst. Signal Process.* **128**, 497–530 (2019).
22. Chakraborty, S. & Chowdhury, R. Graph-theoretic-approach-assisted Gaussian process for nonlinear stochastic dynamic analysis under generalized loading. *J. Eng. Mech.* **145**, 04019105 (2019).
23. Gunn, S. R. Support vector machines for classification and regression. *ISIS Tech. Rep.* **14**, 5–16 (1998).
24. Chatterjee, T., Chakraborty, S. & Chowdhury, R. A bi-level approximation tool for the computation of FRFs in stochastic dynamic systems. *Mech. Syst. Signal Process.* **70**, 484–505 (2016).
25. Mahata, A. & Asle Zaeem, M. Evolution of solidification defects in deformation of nano-polycrystalline aluminum. *Comput. Mater. Sci.* **163**, 176–185 (2019).
26. Kuo, J.-C. & Huang, I.-H. Extraction of plastic properties of aluminum single crystal using Berkovich indentation. *Mater. Trans.* **51**, 2104–2108 (2010).
27. Kim, S.-H. et al. Deformation twinning of ultrahigh strength aluminum nanowire. *Acta Mater.* **160**, 14–21 (2018).
28. Brandl, C., Derlet, P. M. & Van Swygenhoven, H. Strain rates in molecular dynamics simulations of nanocrystalline metals. *Philos. Mag.* **89**, 3465–3475 (2009).
29. Yuan, L., Shan, D. & Guo, B. Molecular dynamics simulation of tensile deformation of nano-single crystal aluminum. *J. Mater. Process. Technol.* **184**, 1–5 (2007).
30. Shao, J.-L., Wang, P., He, A.-M., Zhang, R. & Qin, C.-S. Spall strength of aluminium single crystals under high strain rates: Molecular dynamics study. *J. Appl. Phys.* **114** (2013).
31. Schiøtz, J., Di Tolla, F. D. & Jacobsen, K. W. Softening of nanocrystalline metals at very small grain sizes. *Nature* **391**, 561 (1998).
32. Yip, S. Nanocrystals: the strongest size. *Nature* **391**, 532 (1998).
33. Mohammadi, A., Enikeev, N. A., Murashkin, M. Y., Arita, M. & Edalati, K. Examination of inverse Hall-Petch relation in nanostructured aluminum alloys by ultra-severe plastic deformation. *J. Mater. Sci. Technol.* **91**, 78–89 (2021).
34. Ito, Y., Edalati, K. & Horita, Z. High-pressure torsion of aluminum with ultrahigh purity (99.9999%) and occurrence of inverse Hall-Petch relationship. *Mater. Sci. Eng.: A* **679**, 428–434 (2017).
35. Haque, M. A. & Saif, M. A. Mechanical behavior of 30–50 nm thick aluminum films under uniaxial tension. *Scr. Mater.* **47**, 863–867 (2002).
36. Xu, W. & Dávila, L. P. Tensile nanomechanics and the Hall-Petch effect in nanocrystalline aluminium. *Mater. Sci. Eng. A* **710**, 413–418 (2018).
37. Mahata, A. & Asle Zaeem, M. Effects of solidification defects on nanoscale mechanical properties of rapid directionally solidified Al-Cu Alloy: A large scale molecular dynamics study. *J. Cryst. Growth* **527**, 125255 (2019).
38. Mukhopadhyay, T., Dey, T. K., Chowdhury, R. & Chakraborty, A. Structural damage identification using response surface-based multi-objective optimization: a comparative study. *Arab. J. Sci. Eng.* **40**, 1027–1044 (2015).
39. Karsh, P., Mukhopadhyay, T., Chakraborty, S., Naskar, S. & Dey, S. A hybrid stochastic sensitivity analysis for low-frequency vibration and low-velocity impact of functionally graded plates. *Compos. B: Eng.* **176**, 107221 (2019).
40. Kushari, S., Mukhopadhyay, T., Chakraborty, A., Maity, S. & Dey, S. Probability-based unified sensitivity analysis for multi-objective performances of composite laminates: a surrogate-assisted approach. *Composite Struct.* **294**, 115559 (2022).
41. Chen, C. T. & Gu, G. X. Physics-informed deep-learning for elasticity: forward, inverse, and mixed problems. *Adv. Sci.* **10**, 2300439 (2023).
42. Chew, A. K. et al. *Advancing Material Property Prediction: Using Physics-informed Machine Learning Models for Viscosity*. (ChemRxiv. Cambridge: Cambridge Open Engage, 2023).
43. Singh, V., Patra, S., Murugan, N. A., Toncu, D.-C. & Tiwari, A. Recent trends in computational tools and data-driven modeling for advanced materials. *Mater. Adv.* **3**, 4069–4087 (2022).
44. Mahata, A. & Asle Zaeem, M. Size effect in molecular dynamics simulation of nucleation process during solidification of pure metals: investigating modified embedded atom method interatomic potentials. *Model. Simul. Mater. Sci. Eng.* **27**, 085015 (2019).
45. Mahata, A. & Asle Zaeem, M. Erratum: Size effect in molecular dynamics simulation of nucleation process during solidification of pure metals: investigating modified embedded atom method interatomic potentials (2019 Modelling Simul. Mater. Sci. Eng. 27 085015). *Model. Simul. Mater. Sci. Eng.* **28**, 019601 (2019).
46. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2009).
47. Tsuzuki, H., Branicio, P. S. & Rino, J. P. Structural characterization of deformed crystals by analysis of common atomic neighborhood. *Comput. Phys. Commun.* **177**, 518–523 (2007).
48. Mahata, A. Mukhopadhyay, T. & Asle Zaeem, M. Modified embedded-atom method interatomic potentials for Al-Cu, Al-Fe and Al-Ni binary alloys: from room temperature to melting point. *Comput. Mater. Sci.* **201**, 110902 (2022).
49. Mahata, A., Mukhopadhyay, T. & Asle Zaeem, M. Liquid ordering induced heterogeneities in homogeneous nucleation during solidification of pure metals. *J. Mater. Sci. Technol.* **106**, 77–89 (2022).
50. LAMMPS. https://docs.lammps.org/fix_deform.html.
51. Thompson, A. P., Plimpton, S. J. & Mattson, W. General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. *J. Chem. Phys.* **131**, 154107 (2009).
52. Biswas, S., Chakraborty, S., Chandra, S. & Ghosh, I. Kriging-based approach for estimation of vehicular speed and passenger car units on an urban arterial. *J. Transp. Eng. A: Syst.* **143**, 04016013 (2017).
53. Chakraborty, S. & Chowdhury, R. Hybrid framework for the estimation of rare failure event probability. *J. Eng. Mech.* **143**, 04017010 (2017).
54. Xiu, D. & Karniadakis, G. E. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24**, 619–644 (2002).
55. Chakraborty, S. & Chowdhury, R. Multivariate function approximations using the D-MORPH algorithm. *Appl. Math. Model.* **39**, 7155–7180 (2015).
56. Chakraborty, S. & Majumder, D. Hybrid reliability analysis framework for reliability analysis of tunnels. *J. Comput. Civ. Eng.* **32**, 04018018 (2018).
57. Li, G. & Rabitz, H. D-morph regression: application to modeling with unknown parameters more than observation data. *J. Math. Chem.* **48**, 1010–1035 (2010).
58. Li, G., Rey-de-Castro, R. & Rabitz, H. D-MORPH regression for modeling with fewer unknown parameters than observation data. *J. Math. Chem.* **50**, 1747–1764 (2012).
59. Xu, W. & Dávila, L. P. Size dependence of elastic mechanical properties of nanocrystalline aluminum. *Mater. Sci. Eng. A* **692**, 90–94 (2017).
60. Haque, M. A. & A Saif, M. T. Mechanical behavior of 30–50 nm thick aluminum films under uniaxial tension. *Scr. Mater.* **47**, 863–867 (2002).
61. Simmons, G. *Single Crystal Elastic Constants and Calculated Aggregate Properties* (Southern Methodist Univ Dallas Tex, 1965).
62. Gale, W. F. & Totemeier, T. C. *Smithells Metals Reference Book* (Elsevier, 2003).
63. James, A. M. & Lord, M. P. *Macmillan's Chemical and Physical Data* (Macmillan, 1992).

64. Jiang, Q. & Lu, H. Size dependent interface energy and its applications. *Surf. Sci. Rep.* **63**, 427–464 (2008).
65. Gránásy, L., Tegze, M. & Ludwig, A. Solid-liquid interfacial free energy. *Mater. Sci. Eng. A* **133**, 577–580 (1991).
66. Gündüz, M. & Hunt, J. The measurement of solid-liquid surface energies in the Al-Cu, Al-Si and Pb-Sn systems. *Acta Metall.* **33**, 1651–1672 (1985).

ACKNOWLEDGEMENTS

This study was supported by the National Science Foundation, CMMI 2031800. The authors are grateful for the supercomputing time allocation provided by the NSF's ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support), Award No. DMR140008 and MAT210018. TM would like to acknowledge the initiation grant received from the University of Southampton.

AUTHOR CONTRIBUTIONS

A.M., S.C., and T.M.: Conceptualization, Methodology, Software, Formal analysis, Writing-Original draft preparation. M.A.Z.: Supervision, Conceptualization, Methodology, Formal analysis, Writing- Reviewing and editing, Funding acquisition.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-024-01200-1>.

Correspondence and requests for materials should be addressed to T. Mukhopadhyay or M. Asle Zaeem.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024