Data-driven modeling of process, structure and property in additive manufacturing: a review and future directions

Zhuo Wang[1], Wenhua Yang[2], Qingyang Liu[3], Yingjie Zhao[4], Pengwei Liu[4], Dazhong Wu[3], Mihaela Banu[5*], Lei Chen[1*]

[1] *Department of Mechanical Engineering, University of Michigan-Dearborn, Dearborn, MI 48128, USA*
[2] *Department of Mechanical Engineering, Mississippi State University, Starkville, MS 39762, USA*
[3] *Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL 32816, USA*
[4] *School of Mechanical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China*
[5] *Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI48109, USA*

Abstract

A thorough understanding of the complex process-structure-property (P-S-P) relationship in additive manufacturing (AM) has long been pursued, due to its paramount importance in achieving AM process optimization and quality control. The physical modeling and experimental approaches are usually time-consuming and/or costly. With the increasing availability of digital AM data and rapid development of data-driven modeling techniques, especially machine learning (ML), data-driven AM modeling is emerging as an effective approach towards this end. It allows for automatic exploration of pattern and trend in the data, construction of quantitative P-S-P relationship over the parameter space and prediction at unseen points without having to perform new physical modeling or experiments. A proliferation of researches on data-driven modeling of process, structure and property in AM have been witnessed in recent years. In this context, this paper aims to provide a systematic review of existing data-driven AM modeling with respect to different quantities of interest (QoI) along the process-structure-property chain. Specifically, this paper offers a summary of important information (i.e., input features, QoI-related output, data source and data-driven model used) of existing data-driven AM modeling, as well as an in-depth analysis on relevant success achieved so far. Based on the comprehensive review, this paper also critically discusses the major limitations faced today and brings up some research directions that are promising for significantly advancing data-driven AM modeling tomorrow.

Keywords: Data-driven modeling, Machine Learning, Additive Manufacturing, Process-structure-property.

* mbanu@umich.edu (M. Banu), leichn@umich.edu (L. Chen).

1. Introduction

Additive manufacturing (AM) [1-3], which builds a part directly from a 3D CAD model in a layer-by-layer manner, can fabricate components with complex geometry in a time- and cost-saving manner, which makes it potent and increasingly used in many industries. While the unique layer-wise building endows AM unrivaled manufacturing capabilities, those benefits come at the cost of complicated AM process and miscellaneous physics involved. A plethora of complex physical events, of which some [4-7] rarely occur in conventional manufacturing process, exist at different stages of AM process. This would result in extremely non-linear and complex process-structure-property (P-S-P) relationship [8]. Meanwhile, (multi-)physical AM modeling and AM experiment are notoriously time-consuming and/or costly. Those two facts pose great challenges in fully exploring and understanding AM P-S-P linkage that, however, is vital to efficient AM process optimization and quality control.

In the context of the fourth research paradigm of data-intensive discovery [9], research methodologies across science and engineering have seen a shift to data-driven and informatics approaches. This is in part due to the increasing availability of digital scientific data or data deluge [10], both theoretical and experimental, with the fast advancement of computational power and experimental instrumentation. On the other hand, with the revival of artificial intelligence (AI) researches from AI winter [11], the rapid development of machine learning (ML), a disruptive data-driven modeling technique, has especially accelerated such paradigm shift sweeping through many scientific disciplines [12-18]. It should be stressed that, although becoming more and more popular, data-driven modeling is not a standalone approach and, instead, augments physics-based modeling and experiments by making best use of the generated data thereof. Specifically, data-driven modeling can have different types for diagnostic, descriptive, predictive and prescriptive purposes, respectively [19,20].

Among others, additive manufacturing (AM) is arguably one of the most affected domains in the age of data. For example, data-driven model has appeared frequently as a key building block of various AM design and management frameworks/strategies, such as design for additive manufacturing (DfAM) [21,22], digital twin (DT) for AM [23,24], smart additive manufacturing [25], cloud additive manufacturing [26]. While applications of data-driven model in AM are diverse, data-driven predictive modeling is especially useful for studying P-S-P relationship in AM. It usually discovers the P-S-P relationship in a mathematical form (or the predictive function) via regression analysis. More specifically, data-driven predictive modeling allows for automatic exploration of pattern and trend in the AM data, construction of P-S-P relationship over the parameter space and prediction at unseen points without having to perform new physics-based modeling or experiments. This data-driven attempt of training a cheap relationship model to replace original physics-based modeling or experiment is sometimes also called surrogate modeling or metamodeling [27]. Such a data-driven modeling approach is crucial for achieving AM process optimization based on a complete, quantitative understanding of P-S-P relationship. AM community has witnessed an upsurge of data-driven AM researches on this topic and produced a wealth of relevant literature.

Although existing related review efforts [28-35], more or less, touched those literature, they usually focus neither specifically nor fully on the topic of data-driven modeling of P-S-P relationships. Most of those reviews cover the comprehensive applications of data-driven modeling techniques and ML in AM. They thus have quite different emphasis from this review. For example, the review paper by Wang et al. [28] is organized based on existing applications of ML for AM design, AM production, and AM process. The review by Razvi et al. [29] is focused on four applications - AM design, process and performance optimization, in-situ monitoring and control, and AM part inspection and validation. Similarly, the survey by Jin et al. [30] concentrates on geometrical design, process parameter configuration, and in-situ anomaly detection. Baumann et al. [31] focuses on process control, process monitoring, and quality enhancement of manufactured objects and so on. Goh et al. [32] have broadly reviewed applications of ML in AM including design for 3D printing, process optimization, in-situ monitoring, cloud 3D printing, and even security of attach detection. Therefore, the specific topic of data-driven predictive modeling of P-S-P relationships has not been explicitly addressed in those reviews. While the reviews by Qi et al.[33], Kouraytem et al. [34] and Meng et al.[35] have clearly surveyed existing data-driven modeling of P-S-P relationships, this topic constitutes only a portion of the entire article in their reviews. In conclusion, a systematic and fine-grained review exclusively devoted to the important topic of data-driven modeling of process, structure and property in AM (simply referred to as data-driven AM modeling hereinafter) is still absent.

To fill this gap, this paper provides an extensive review and detailed analysis of existing researches on data-driven modeling of process, structure and property in metal, polymer, ceramics and composite-based AM. For a systematic review, this research classifies concerned data-driven AM modeling into three categories - process modeling, structure modeling and property modeling (Fig. 1). The three main categories are further subdivided based on the detailed modeling quantity of interest or output of data-driven modeling. More specifically, process modeling broadly involves modeling of AM machine activities and its thermal and physical interactions with AM building during manufacturing process. Typical process modeling is thermally related modeling, e.g., modeling of process temperature development and melt pool formation; structure modeling precisely refers to those modeling interested in various structures across multiple scales, ranging from microstructures to mesoscale geometries and macroscopic shapes; property modeling includes modeling of different mechanical properties of AM-fabricated parts.

In this paper, the review and analysis of those different types of data-driven AM modeling will contain three main parts: 1) tabulated summary that gives a quick overview of existing data-driven AM modeling; see Table 1, 2 and 3 for data-driven process, structure and property modeling respectively. We summarize basic information of existing data-driven AM modeling, namely the four components - input features, QoI-related output and data-driven model used to link them, as well as the data source. Note that, data source (simulation or experiment) largely implies if the related data-driven modeling is more in the stage of proof-of-concept or rather practically significant. Besides the four components, data-driven modeling result is briefly summarized in the table, in terms of predictive accuracy or error. It gives readers a general sense of the performance of existing data-driven models and may provide baselines for future studies. 2) further analysis that provides deeper insights into existing data-driven AM modeling; see subsections of Data-driven Process Modeling, Data-driven Structure Modeling and Data-driven Property Modeling. In those three subsections, we further analyze in depth the success achieved by existing researches. 3) detailed discussion that offers guidance on future data-driven AM modeling; see Future Directions section. Based on the two review parts, we finally discuss major limitations faced today and raise some promising research directions for significantly advancing data-driven AM modeling tomorrow.

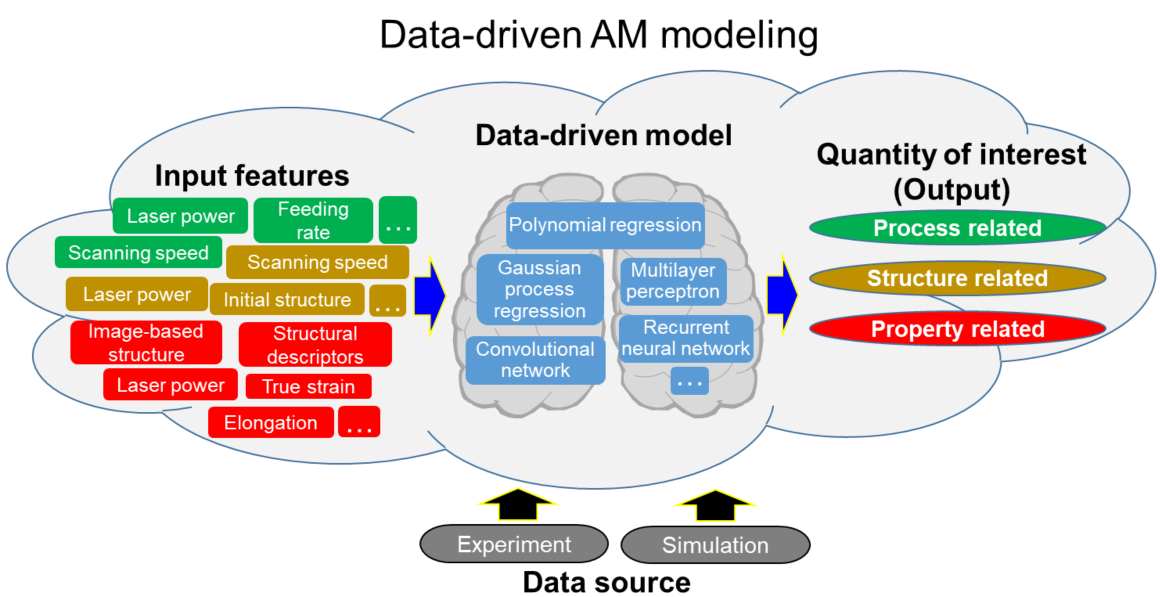

Fig. 1 Schematic illustration of data-driven AM modeling. Data-driven AM modeling usually consists of four basic components - input features, quantity of interest (output) and data-driven model used to link them, as well as the data source that fuels data-driven modeling. Note that, this review divides data-driven AM modeling into three groups - data-driven process modeling, structure modeling and property modeling, according to the modeling quantity of interest or output of data-driven modeling.

2. Data-driven model

2.1 Common data-driven models in AM

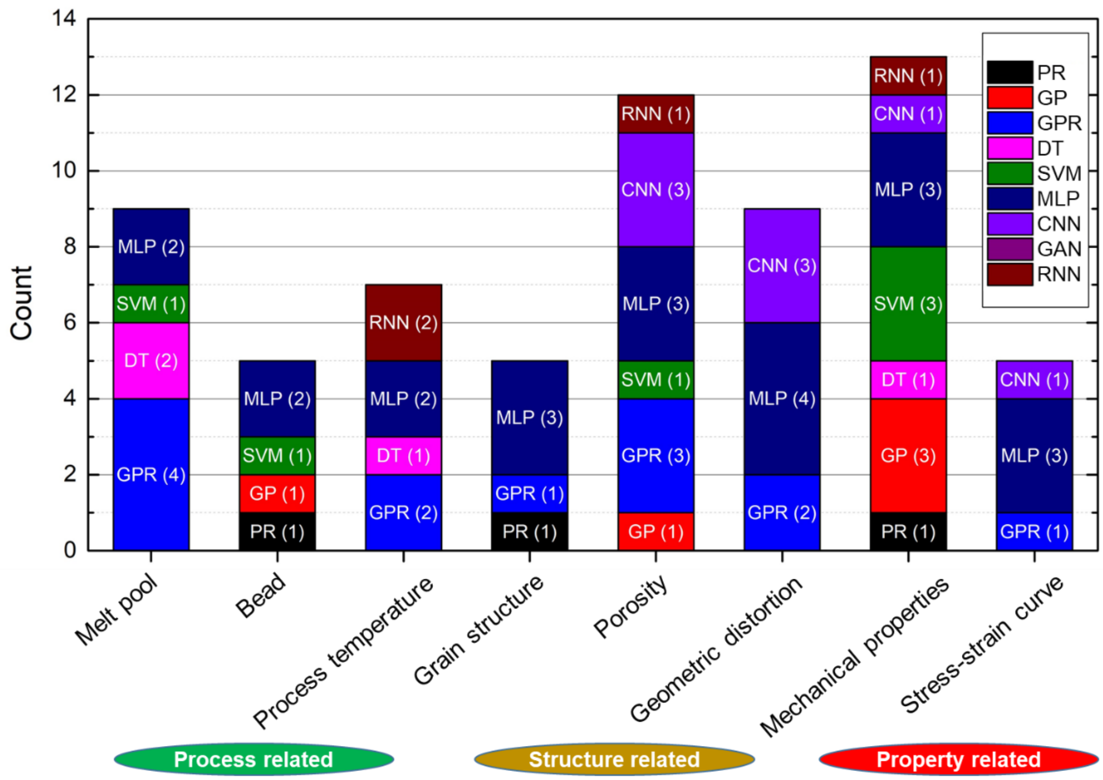

Fig. 2 Frequency of common data-driven models used in different data-driven AM modeling. Even for a certain data-driven AM modeling, researchers may use quite different data-driven models.

This subsection gives a brief introduction to those common data-driven AM models involved in this review. Readers interested in more detail are referred to related references in this paper; see Table 1-3 as a guide. Additionally, Fig. 2 summarizes the frequency of those different data-driven models used in previous researches. It may serve as a quick reference for model selection or model novelty evaluation against existing ones in future studies.

1. **Polynomial Regression (PR):** PR is a rudimentary data-driven modeling technique by using the polynomial function. PR is easy to implement but usually limited to approximating simple or moderately complex relationship.
2. **Genetic Programming (GP):** GP and its variants are based on combining a more varied set of essential elements and functions, such as arithmetic operators (+, −, ×, /), non-linear functions (sin, cos, tan etc.) and/or Boolean operators. The satisfactory structure and parameters of GP model are iteratively found through genetic operations on a population of initial configurations. GP is thus a more flexible data-driven modeling technique, but the proper setup of GP model (e.g., genetic operation details) requires much expert knowledge.
3. **Gaussian Process Regression (GPR):** GPR is instead a "non-parametric" (i.e., without a rigid regression model) and more universal approximation method. The input-response relationship is formulated within covariance kernel. GPR has the outstanding capability of measuring uncertainty on the prediction, as GPR essentially considers the whole data as a sample from a multivariate Gaussian distribution and therefore the prediction as a conditional distribution. GPR would become painfully slow in face of large dataset, as whole samples or features information are used to perform the prediction. Note that in the literature, Gaussian process regression is often called GP for short. To avoid confusion with genetic programming, we refer to Gaussian process regression as GPR throughout this article.
4. **Decision Tree (DT):** DT for either classification or regression is conducted by recursively and discriminatingly partitioning data into branches until sufficiency or user-defined depth. DT is easy to understand and known as a white box. It does not require much data preparation such as data normalization, dummy variables creation and can handle both numerical and categorical data. To mitigate the overfitting issue of deep DT, a common practice is to use ensemble of decision trees i.e., Random Forests (RF), which randomly selects observations and features to build several decision trees and then averages the results.

5. **Support Vector Machine (SVM):** SVM for regression analysis is a close variant of classification-purposed SVM. They follow similar principles to use Kernel trick to transform data into a more amenable high-dimensional space, which then permits simply finding a N-dimensional hyperplane that has best fit to all data points. Contrary to the intuitive DT, SVM is more difficult to interpret as a black-box model.
6. **Multilayer Perceptron (MLP):** MLP as the classical type of neural network consists of multiple layers of fully connected neurons, which include input layer, hidden layer(s), and output layer. MLP can learn a mapping from inputs to outputs of tabular datasets. For other types of data such as images and words of document, they have to be converted to one long row of data as the input, which however usually results in a parameter-intensive MLP with inferior performance.
7. **Convolutional Neural Networks (CNNs)**: CNNs on the other hand were designed to make predictions by directly using image data as an input. More generally, CNNs perform well on array data that has a spatial relationship, because they can develop an internal representation of both one-dimensional sequence and high-dimensional matrix by using stacks of convolutional layers to extract salient patterns. However, CNNs usually demand large amounts of training data.
8. **Generative Adversarial Networks (GANs):** GANs as another important image-based neural networks are significantly promising deep learning models. GAN and its variants can generate sharp images from latent variables and/or conditional information, and the distribution of generated samples usually matches the true data distribution well. Especially, they can be employed to generate deterministic results when complicated conditional information is provided. This fact brings them huge potential in a wide range of engineering applications, including microstructure reconstruction, microstructure synthesis and image-based microstructure prediction.

9. **Recurrent Neural Networks (RNNs):** RNNs were designed to handle sequential data, by permitting output from previous step to be fed to the current step during processing long sequences. RNNs perform well on processing and predicting sequences with even variable length, e.g., words and spoken language in natural language processing (NLP) applications. However, the looping structure of RNNs for dealing with sequences can increase training complexity due to the more complicated signal movements. Long short-term memory (LSTM) [36], which can learn the very long dependence in sequences, is the most commonly used RNN variant in data-driven AM modeling currently.

2.2 Model selection

Selecting the suitable data-driven models is an essential step towards successful data-driven AM modeling. However, as shown in Fig. 2, different research groups adopted different data-driven models even for the same data-driven AM modeling problem (up to 7 in data-driven modeling of mechanical properties). In part, this is attributable to different simplification and formulation of a data-driven modeling problem. For instance, the detailed modeling quantity of data-driven porosity modeling could be simple porosity percentage or real porosity structure; see Table 2. Also, different researchers may be interested in using distinct input features. On the other hand, even for a well-specified data-driven AM modeling problem, e.g., building the relationship between the melt pool area and AM parameters in Table 1, there is still no widely accepted criterion for choosing the best data-driven model. Regarding this, one of the best strategies is ensemble learning [37,38], which use the weighted summation of predictions by data-driven models of different types to achieve better prediction accuracy than any constituent model.

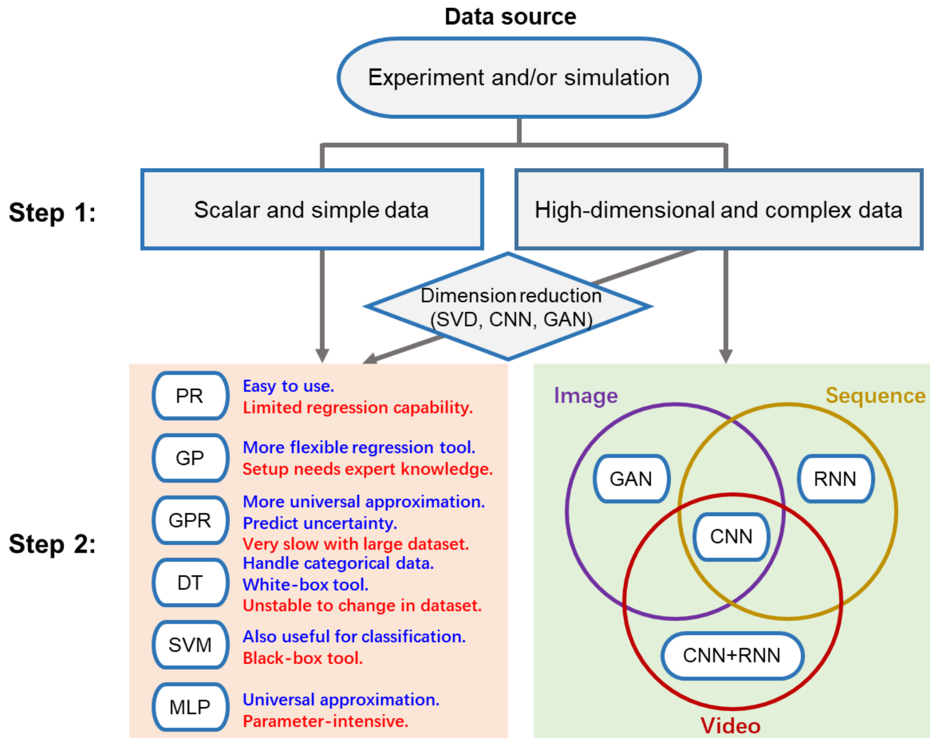

Fig. 3 Flowchart of the two-step model selection.

While model selection is still an open research problem in AM, we introduce a general two-step model selection method for different problems in AM as a guideline. From previous practices, one can first narrow down data-driven models based on data involved in the data-driven modeling, followed by determining a suitable model according to the unique characteristics of different data-driven models and properties of the modeling problem; see flowchart in Fig. 3.

For step 1, all the data-driven model candidates may be classified into two groups, based on their capability in tackling data of different types and complexities. The first group (PR, GP, GPR, DT, SVM, MLP) usually deals with scalar quantities and sometimes categorical/label data, while the second group (CNN, GAN, RNN, and CNN+RNN) handles more complex data (e.g., time series, image and video). For step 2, a proper data-driven model can be further determined depending on the model's specialties and modeling problem. When only scalar and simple quantities are involved in the data-driven modeling, the modeling task lies more within conventional regression analysis. Generally, all of those six data-driven models would be useful, although MLP tends to be the first choice nowadays as reflected in Fig. 2. Meanwhile, as discussed in Subsection 2.1, those data-driven models also have their own characteristics in terms of friendliness to non-expert, interpretability, regression power, categorical prediction capability and so on. This may make a particular data-driven model favored in some special situations where those factors carry weight. Besides the above common data-driven AM modeling, AM modelers often encounter high-dimensional and complex quantities, e.g., acoustic signal, thermal history curve, melt pool image and 3D temperature field. Those complex quantities can arise from the input feature, QoI output, or both sides. Sometimes the high-dimensional data might be further processed before performing data-driven modeling by using data dimension reduction techniques, such as singular value decomposition (SVD)[39], principal component analysis (PCA)[40], and GAN-based encoder[41]. Additionally, CNN, though not yet used for dimension reduction in AM-related applications, is effective in learning latent representations of high-dimensional data[42,43]. Dimension reduction will then transform the original modeling problem to the common regression analysis (i.e., left-hand side in Fig. 3) in the latent space. In other cases, some capable ML models are leveraged to directly deal with complex data. RNN is preferred when modeling time series and other sequences; GAN is usually selected for image (generation) modeling; CNN is able to handle images, 3D array (e.g., stacked images, video and 3D structure[44]), as well as long sequences (e.g., thermal history profile[45]). Furthermore, integration of CNN and RNN for processing image streams is also a common practice [46,47].

Table 1 Summary of data-driven process modeling in AM

Quantity of interest (QoI)	Input features	QoI-related output	Data-driven model	Data source (Dataset size)	Modeling results*	Ref.
Melt pool	Laser power, scanning speed.	Pool depth	GPR	Exp. (139) Simu. (26)	MAPE = 10.91 μm MAPE = 6 μm**	48
	Laser power, scanning speed.	Pool depth	GPR, DT	Simu. (462)	ReMSE = 3.6%***	49
	Laser power, scanning speed.	Remelting depth	GPR	Simu. (24)	MAPE = 1.4 μm	50
	Chemistry of powders, materials thermal property, powder bed information and laser parameters.	Pool width, depth, area within substrate, height, area based on height	RF, SVM, etc.	Exp. (472)	R^2=0.75-0.9	51
	Laser power, scanning speed, neighboring effect factors.	Pool area	PR	Exp. (4,957)	NRMSE = 0.08	52
	Laser power, scanning speed, neighboring effect factors, layer-wise effect factors.	Pool area	MLP	Exp. (118,928)	AREM = 12.21%	53
	Laser power, scanning speed and five uncertainty sources.	Pool length, width and depth	GPR	Exp. (7) Simu. (300)	AREM = 5.74/3.26/10.08%	54
Bead	Laser parameters and feeding rate.	Bead width, depth and height	MLP	Exp. (90)	RMSE=0.59/0.53/0.14mm	55
	Laser processing parameters.	Bead width	PR	Simu. (70)	MAE = 0.2997 μm	56
	Layer thickness, laser power and scanning speed.	Bead width	GP, MLP	Exp. (54)	RMSE=65.53 μm	57
	Power, speed, feeding rate.	Bead height	SVM	Exp. (180)	MSE = 2.89E-8 mm	58
Process temperature	Long-term, short-term memory descriptors and temperature feedback from previous step.	Maximum temperature within heat-affected zone	MLP	Simu. (54,450)	RMSE=108.87 °C	59
	Layer power, scan speed, layer index, time index, average height, average width.	Maximum temperature within pool boundary	LSTM	Exp. (~17,000)	RMSE = 20.1 °C	60
	Thermal and spatial information of the voxel and its neighboring elements.	Thermal history of voxel	DT	Simu. (9.05E6)	MAE = 0.21	61
	Tool-path feature, time of deposition, location information of the point, etc.	Thermal history of point	RNN	Simu. (>2.5E5)	MSE = 3.84e-5	62
	A set of relative distances from the cooling surfaces & the heat sources and a set of deposition times influencing thermal behavior.	Thermal history of point	MLP	Simu. (2.6E5)	NRMSE < 2%	63
	Printing settings, index variables and calibration parameters.	Layer-to-layer thermal field	GPR	Simu. & Exp. (N/A****)	RMSE = 5.71 K	64
	AM process conditions and materials properties.	3D high-temperature	GPR	Simu. (280)	N/A	39

		field				

Note: * Metrics used to measure modeling performance in previous researches include mean absolute prediction error (MAPE), relative mean-squared error (ReMSE), residuals (R^2), normalized root-mean-square error (NRMSE), average relative error magnitude (AREM), root mean-squared error (RMSE), mean absolute error (MAE), mean-squared error (MSE), mean-squared prediction error (MSPE), accuracy (Acc), global accuracy (GAcc), intersection over union (IoU), $F1$ score ($F1$) and mean error (ME). Detailed definition of each metric can be found in corresponding referred literature.

** When simulation and experiment data are used separately to train two individual data-driven models, both of their predictive results are listed.

*** When more than one type of data-driven models have been used in a research, only the best predictive result is listed.

**** In a few researches, related information was not directly reported.

Table 2 Summary of data-driven structure modeling in AM

Quantity of interest (QoI)	Input features	QoI-related output	Data-driven model	Data source (Dataset size)	Modeling results	Ref.
Grain structure	AM manufacturing conditions and materials properties.	Mean and variance of grain aspect ratio	GPR	Simu. (150)	N/A	39
	Seven processing conditions including scan pattern, molten zone width, velocity, molten zone depth, etc.	Principle component (PC) scores of the microstructure	PR	Simu. (1,799)	MAE = 0.0168	65
Porosity	Seven process parameters including layer thickness, laser power, hatch spacing, scanning speed, etc.	Density	MLP	Exp. (26)	N/A	66
	Laser power and scanning speed.	Density	GPR	Exp. (42)	MSPE = 0.2593	67
	Laser power, laser velocity and hatching space.	Relative density	MLP	Exp. (60)	RMSE = 0.3249%	68
	Laser power and scan speed.	Relative density	GPR	Exp. (82)	MAE < 0.3%	69
	Volumetric energy density.	Density	GPR	Exp. (N/A)	MAPE = 1%	70
	Layer thickness, laser power, laser scan speed.	Open porosity (%)	GP, MLP, SVM	Exp. (36)	MAPE = 8.82%	71
	Thermal images by pyrometer and IR camera.	"Bad" (porous) or "good" (neglectable porosity) label	CNN+RNN	Exp. (840)	Acc = 99.29 %	46
	Surface images by DSLR camera.	"With flaw" or "without flaw" label	CNN	Exp. (1708)	Acc = 92.50 %	72
	Laser power, scanning speed, initial powder-bed structure.	Porous structure after sintering	CNN	Simu. (130,500)	GAcc = 99.13%	73
Geometrical distortion	Nodal coordinates of designed geometry.	Nodal coordinates of actual geometry	MLP	Simu. (N/A)	MSE = 1.12E-05	74
	Five process parameters including part bed temperature, laser power, scan speed, scan spacing, scan length.	Shrinkages (%) along length, width and thickness	MLP	Exp. (50)	$R^2 = 0.54$	75
	Polar angle, polar radii of point on the designed shape.	Polar radii of point on the actual shape	MLP	Exp. (20,000-40,000)	MSE < 10E-7	76
	Polar angle of point on the contour of the designed shape, transformation parameter set.	Polar radii of point on the contour of the actual shape	GPR	Exp. (360)	$R^2 = 0.9022$	77
	Polar angle of point on the contour of the designed shape, infill parameters.	Polar radii of point on the contour of the actual shape	GPR	Exp. (N/A)	MSE = 0.21%	78
	Designed shape.	Actual shape	CNN	Exp. (18,500)	IoU > 0.90	79
	Designed shape.	Actual shape	CNN	Simu. (39,424)	$F1 > 0.93$	80
	Thermal image and process/design parameters including laser	Pointwise distortion	MLP+CNN	Exp. (21,818)	RMSE = 56	81

	power, scan speed, location, print angle and material.				µm	

Table 3 Summary of data-driven property modeling in AM

Quantity of interest (QoI)	Input features	QoI-related output	Data-driven model	Data source (Dataset size)	Modeling results	Ref.
Mechanical property	Structure descriptors including volume fraction of inclusions and average inclusion size.	Yield strength, strain hardening coefficient, etc.	PR	Simu. (900)	MAE = 0.0038	[82]
	3D grain structure.	Yield strength	CNN	Simu. (7680)	RMSE=9.23 MPa	[44]
	Materials compositions and microstructural features including mean equiaxed alpha size, volume fraction of equiaxed alpha, etc.	Yield strength	GP, MLP	Exp.(N/A)	N/A	[83]
	Layer thickness, orientation, raster angle, raster width, air gap.	Compressive strength	GP	Exp.(32)	MAPE=3.93	[84]
	Layer-wise temperature and vibration information, materials property, AM process conditions.	Tensile strength	MLP+RNN	Exp. (144)	RMSE=0.59MPa	[85]
	Melt temperatures, layer thickness, raster pattern orientation.	Tensile strength	MLP	Exp. (108)	RMSE=0.040756	[86]
	Layer thickness, orientation, raster angle, raster width, air gap.	Compressive strength	GP, SVM, etc.	Exp.(32)	MAE=0.0558	[87]
	Angle of incline, overlapping length, number of specimen shells	Tensile strength	SVM, etc.	Exp. (192)	RMSE=2.648MPa	[88]
	Number of fiber layers and fiber rings as well as polymer infill patterns of AM-fabricated carbon fiber-reinforced polymer	Flexural strength	SVM, DT, etc.	Exp. (162)	RMSE=7.75MPa	[89]
Stress-strain curve	Grain structure descriptors including mean and variance of grain aspect ratio.	Stress-strain curve	GPR	Simu. (150)	N/A	[90]
	Strain rate, strain and temperature.	Flow stress	MLP	Exp. (128)	ME = -0.3%	[91]
	Temperature, strain rate, uniform elongation.	Flow stress	MLP	Exp. (720)	RMSE = 2.056%	[92]
	True strain and 12 microstructural descriptors.	True stress	MLP	Exp. (111)	RMSE < 0.2	[93]
	Raw image-based structure.	Stress-strain curve	CNN	Simu. (100,000)	N/A	[40]

3. Data-driven modeling of additive manufacturing

Complementary to Table 1-3, this section gives detailed analysis on the achievement made so far for different data-driven AM modeling. Unless otherwise pointed out, throughout this article we survey the relevant literature indiscriminately in terms of AM techniques and AM materials. The literature is grouped based merely on the modeling quantity of interest, which determines the type and name of data-driven modeling.

3.1 Data-driven process modeling

3.1.1 Data-driven modeling of melt pool

Data-driven melt pool (size) modeling is one of the most widely studied data-driven AM modeling for achieving AM process control. This is because the melt pool size, although a simple quantity, is an effective indicator of overall manufacturing quality, intimately associated with the development of columnar grain structure[4], solidification textures at the sub-grain scale [94] and lack-of-fusion porosity [95], etc. A robust data-driven melt pool model is of huge value for optimizing AM process on-the-fly and ensuring part quality. Moreover, unlike many other AM quantities, melt pool is readily measurable during AM process by using pyrometer, thermal camera or high-speed camera. This fact facilitates experimental data-based data-driven modeling that can generate actionable insights for AM practice. Therefore, melt pool is a prevailing QoI in data-driven AM modeling nowadays.

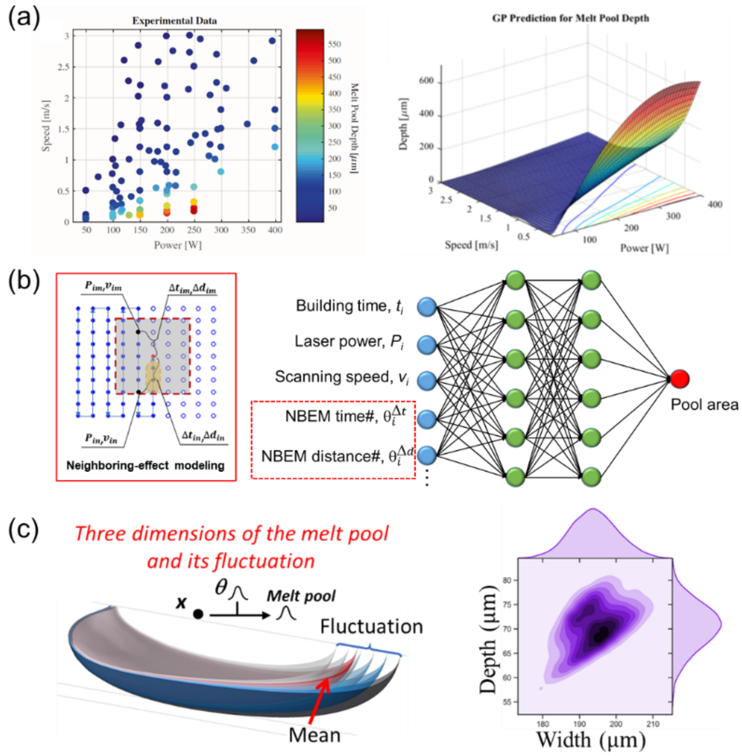

Fig. 4 Data driven melt pool modeling. a A common data-driven melt pool modeling that builds melt pool depth simply as a function of primary manufacturing conditions (i.e., laser power and speed), based on pool depth data obtained under different conditions [48]. **b** An improved data-driven melt pool modeling designs neighboring-effect modeling (NBEM) factors, $\theta_i^{\Delta t}$ and $\theta_i^{\Delta d}$, to account for inter-track heating effect or scanning history on melt pool development [52]. **c** Uncertainty-incorporated data-driven melt pool modeling includes uncertainty sources, $\boldsymbol{\theta}$, as inputs to enable prediction of caused fluctuation of melt pool [54].

Since melt pool size (e.g., pool width, length, depth or area) is a scalar quantity, data-driven melt pool modeling itself is technically not difficult, generally falling within the task of common regression analysis like curve fitting. The key challenge and research interest today lie in the rational selection and even design of input features correlated with melt pool development. Most of existing data-driven melt pool modeling [48-50] just build melt pool as a simple function of common AM process parameters, typically laser power and scanning speed (Fig. 4a). Lee et al. [51] have adopted as many as 23 input features, including various process parameters and materials parameters, to enhance data-driven melt pool modeling. However, those data-driven melt pool modeling by incorporating simple input features, even broadly, cannot adequately account for the complexity of AM process physics, including but not limited to layer-by-layer heat accumulation, inter-track heating effect and, not less importantly, many aleatory uncertainties (i.e., natural variabilities) present in AM process. Those factors can greatly complicate melt pool dynamics. In this sense, more advanced input features, such as cumulative time- and distance-neighboring effect factors [52] (Fig. 4b) as well as various layer-wise effect factors [53], have been designed to improve the predictive accuracy. Wang et al. [54] have also introduced five experimentally calibrated uncertainty sources as inputs in the data-driven model and permits reasonable prediction of melt pool fluctuations (Fig. 4c).

3.1.2 Data-driven modeling of bead

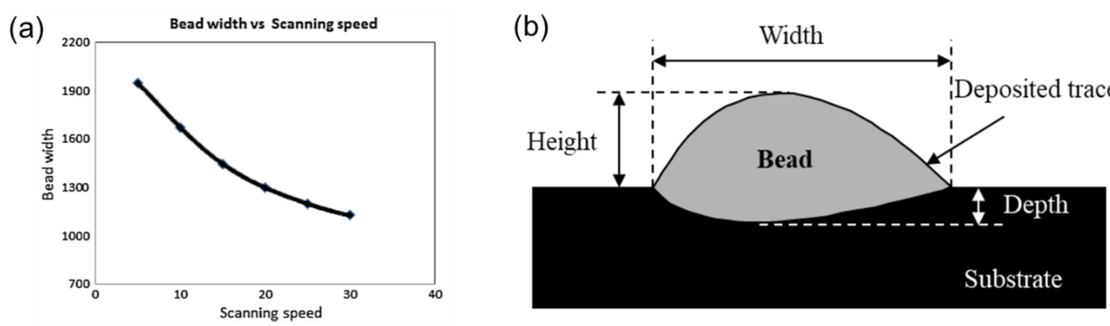

Fig. 5 Data-driven bead modeling. a A typical data-driven bead modeling that builds bead with as a function of scanning speed [57]. **b** Geometrical quantities of interest in data-driven bead modeling, adapted from [55]. Data-driven bead modeling shares the same modeling philosophy and similar modeling quantities with data-driven melt pool modeling.

Another type of data-driven AM modeling very similar to melt pool modeling is the modeling of bead geometry and size (Fig. 5a). Bead formation is a remarkable phenomenon especially in extrusion/feeding/deposition-based AM. Melt pool and bead are highly relevant AM quantities, both essentially concerned with the high-temperature molten part, but bead is just more related to molten materials upon solidification. In fact, their geometrical characteristics, e.g., depth and width between melt pool and bead, are sometimes not really distinguished [51,56]. The only big difference might be that solidification would give rise to a new geometrical quantity - bead height (Fig. 5b), data-driven modeling of which is important for building precision control along z-axis [58]. However, since the overall significance and philosophy of data-driven bead modeling are almost the same to that of data-driven melt pool modeling, this paper will not separately analyze data-driven bead modeling in detail.

3.1.3 Data-driven modeling of process temperature

Data-driven thermal modeling of AM process is a more advanced data-driven process modeling. It can give general thermal characteristics of AM process that includes melt pool informed by high-temperature region. Temperature field formed during AM process is closely related to later development of various structures across many scales and, in turn, largely dictate the properties of final AM parts. Thermal modeling is thus of fundamental importance as a starting point of exploring P-S-P relationship, as evidenced by many multi-stage AM studies of dendrite structure [96,97], grain structure [98,99], porosity [100] and mechanical property [101].

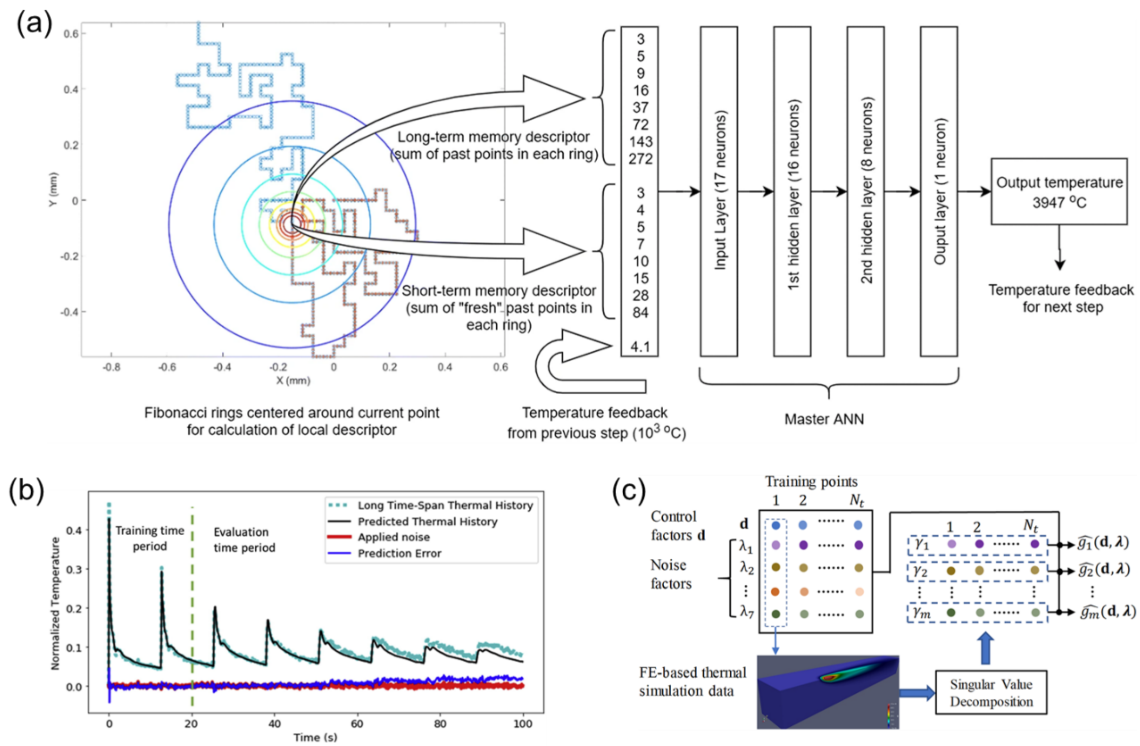

Fig. 6 Data-driven thermal modeling. a Localized data-driven thermal modeling of maximum temperature under the beam [59]. **b** Another type of localized data-driven thermal modeling that allows for modeling the real-time temperature or thermal history for any spatial point/node in the component [62]. **c** Field-level data-driven thermal modeling by using singular value decomposition (SVD) for dimension reduction of original 3D temperature field, followed by data-driven modeling in the latent space [39].

However, unlike data-driven melt pool modeling, data-driven modeling of temperature field is challenging due to its high-dimensional nature. In light of this difficulty, a widely adopted strategy is data-driven thermal modeling of local temperature, instead of the entire temperature field. There exists two such types of data-driven thermal models. The first type focus on a specific, local thermal quantity like peak temperature within the current heating zone (Fig. 6a) [59] or melt pool boundary[60]. The second type is general to all spatial points; that is, the modeling subject is the real-time temperature or thermal history for any given point within the building [61-63] (Fig. 6b). In this way temperature field can be implicitly obtained by aggregating separate modeling results of many points [63]. Besides the mainstream point-/voxel-level data-driven thermal modeling, Wang et al. [39] have utilized dimension reduction technique - singular value decomposition (SVD), thus facilitating the data-driven modeling of the entire 3D temperature field in the latent space (Fig. 6c). But the above SVD-assisted approach is limited to data-driven modeling of temperature field with predefined dimensions and, therefore, focus only on a cuboid of high-temperature part in their work.

3.1.4 Other data-driven process modeling

Other relatively less studied data-driven process modeling include modeling of powder spreading behavior (measured in powder bed surface roughness [102] or represented by six categorical anomalies [103]) and thermal stress field developed in AM process [104]. Apart from powder- and wire-based 3D printing, data-driven methods have been also used in other printing techniques such as the inkjet-based printing. In inkjet-based AM, a special and important process phenomenon is droplet formation determined by the fluid flow pattern, which is critical to printing quality. Wu and Xu [105] developed a supervised data-driven method to predict droplet formation process including droplet volume and velocity based on the AM process parameters (e.g., polymer concentration, dwell time, rise time, and excitation voltage) in the inkjet-based printing process. The proposed method was able to predict droplet volume and velocity in high accuracy. Data-driven method has been also applied to achieve the in-situ process optimization of the process parameters in the inkjet-based printing through a neural network based PID control system [106].

3.2 Data-driven structure modeling

3.2.1 Data-driven modeling of grain structure

Data-driven modeling of grain structure is of particular significance for AM process, since complex and unconventional grain structure is one of the signatures of AM-fabricated components [107]. Grain structure development within AM part has always been an active research topic in AM community [108]. Data-driven grain structure modeling will greatly benefit process plan for tailoring grain structure within AM parts.

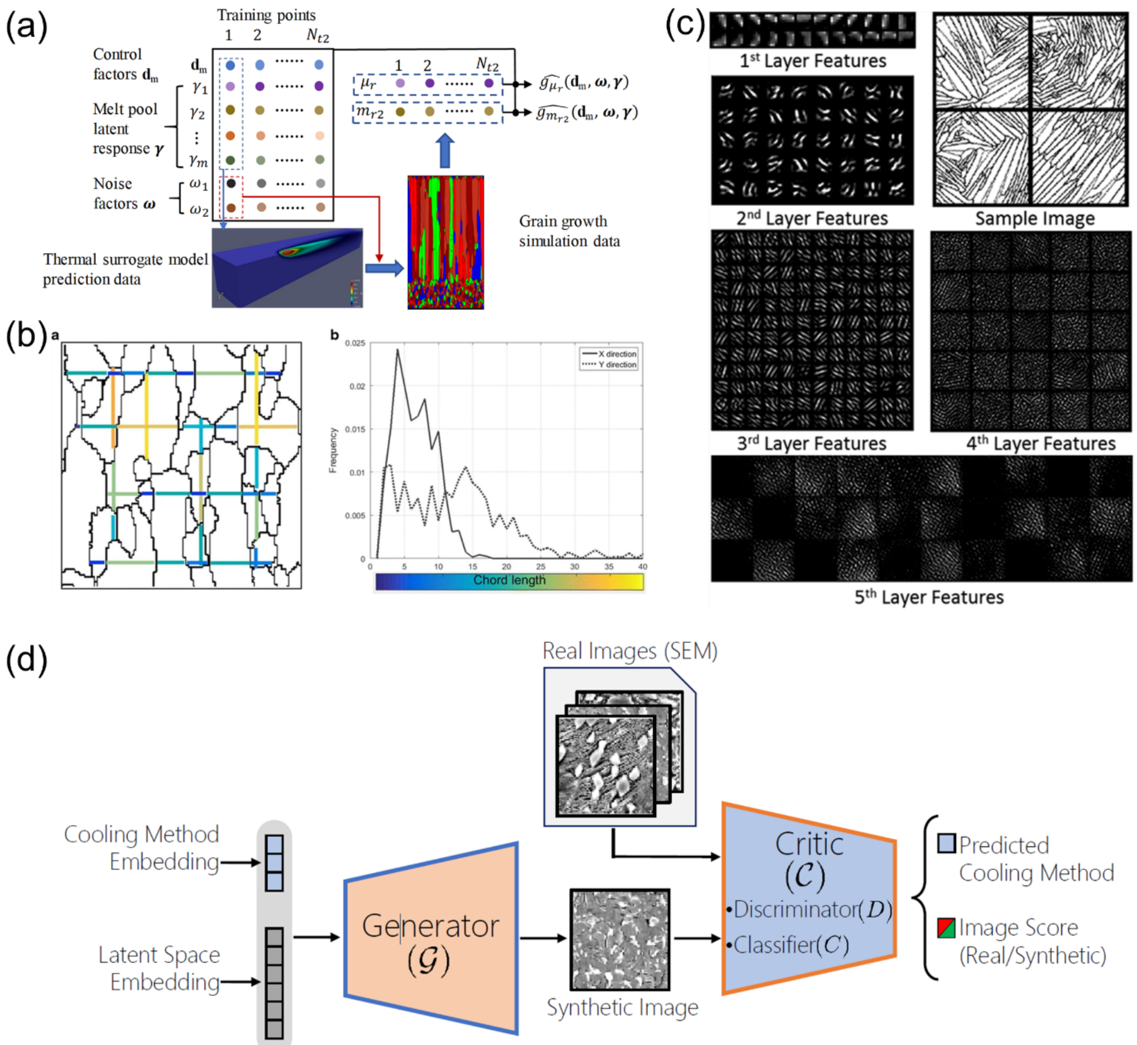

Fig. 7 Data-driven grain structure modeling. a Data-driven grain structure modeling that builds microstructure described by mean and variance of grain aspect ratio, μ_r and m_{r2}, as a function of various input features, such as AM manufacturing conditions and AM materials properties [109]. **b** Data-driven grain structure modeling that chooses to describe complex microstructure using principal component (PC) scores of PC analysis on grain chord length distribution [65]. **c** ML model allows for extraction of highly abstract yet informative features from microstructure through representation learning [110]. **d** GAN allows for direct generation and prediction of microstructure image for a given condition [111].

In contrast with many physics-based modeling and experimental efforts, data-driven grain structure modeling has been surprisingly little explored. Wang et al.[109] have built the formed grain structure, described by the mean and variance of grain aspect ratio, as a function of AM process conditions and other input features (Fig. 7a). Similarly, Popova et al. [65] have correlated grain structure, but quantitatively represented by principal component (PC) scores of PC analysis on grain chord length distribution, with various AM process parameters (Fig. 7b). From the few related researches, it can be seen that data-driven grain structure modeling is still at its early stage, focusing on building mapping between simple microstructure descriptors and input features. This dilemma may be attributed to grain structure as a high-dimensional quantity like temperature field, making explicit modeling of real grain structure difficult. In addition, while the temperature field developed in AM process usually show a tear-drop shape and regular pattern, grain structure is stochastic by nature, thus further preventing the development of advanced data-driven grain structure modeling in AM.

Although related study in AM is scarce, here we would like to provide a glimpse of data-driven grain structure modeling from a wider viewpoint of materials science. In fact, microstructure representation and featurization has long been of great interest in materials community. Early efforts of materials scientist were centered on hand-designing physical and statistical descriptors for various microstructures [112], as did the above AM grain structure modeling. In recent years, ML models, especially CNNs, have substantially advanced representation modeling of microstructure [110,113]. It allows for automatic learning of high-level, abstract features of microstructure (Fig. 7c). Those microstructure representation modeling may be adapted to AM grain structures for extracting more informative and representative microstructure features, thus enabling data-driven construction of high-fidelity process-structure relationship by improving the structure side. Of most relevance, Akshay et al.[111] have even proposed a GAN conditioned on processing conditions - cooling method in processing ultrahigh carbon steel (UHCS) (Fig. 7d). The proposed ML model can directly build process-structure linkage by permitting generation of UHCS microstructure images for given cooling method. Tang et al. [114] has further developed a similar GAN-based approach by conditioning it on numerical conditions, instead of categorical conditions. Although they are not proposed specifically for AM process, but would undoubtedly shed light on data-driven prediction of image-based grain structure under different process conditions in AM.

3.2.2 Data-driven modeling of porosity

Besides grain structure, porosity is another important structural feature of AM-fabricated components. Depending on AM process conditions, various types of porosity can be developed for as-built AM parts, such as lack-of-fusion by insufficient melting [115], keyhole by strong metal evaporation [116], and surface open pore by complex fluid dynamics [5], etc. Porosity modeling might be particularly valuable for the selective laser sintering (SLS), a type of AM process featuring sintering mechanisms for binding raw materials. Unlike laser melting, sintering-based binding is a mild process with insignificant melting and solidification phenomenon, which makes porous structure the prominent feature of SLS components [117]. A data-driven modeling capability of porosity development is necessitated to facilitate process control for not only successfully producing fully-dense parts, but also intentionally fabricating porous structure with desired biocompatibility and permeability [118,119].

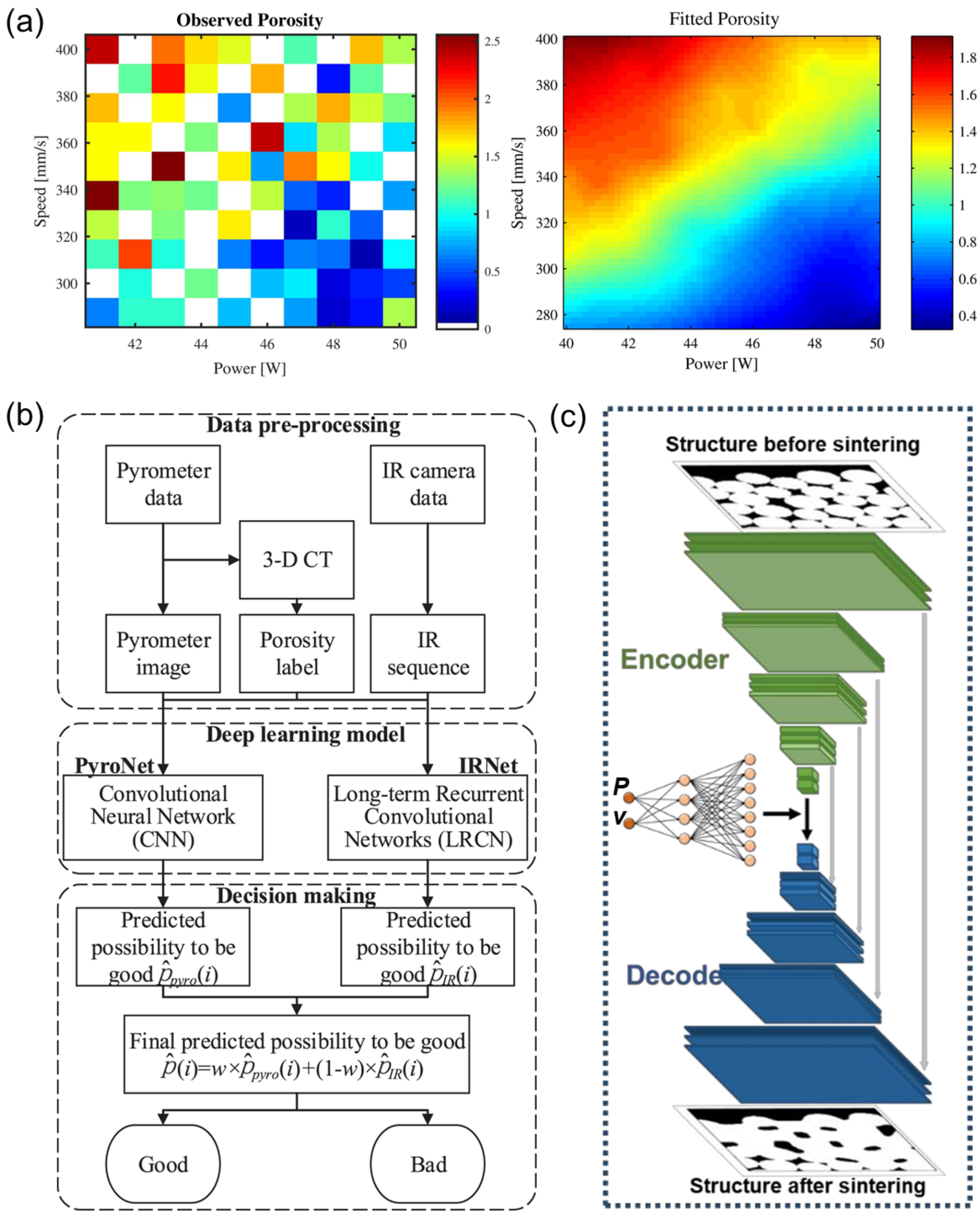

Fig. 8 Data-driven porosity modeling. a A common data-driven porosity modeling that builds porosity simply as a function of manufacturing parameters (i.e., laser power and speed), based on porosity data obtained under different manufacturing conditions [67]. **b** A hybrid CNN and RNN model permits data-driven prediction of porous component by using pyrometer image and IR sequence as inputs [46]. **c** A multi-input CNN enables ultra-fast data-driven simulation of structural evolution and development of porosity for any given manufacturing condition [73].

Data-driven porosity modeling has been more widely studied than data-driven grain structure modeling in AM. The majority of them concentrate on correlating porosity represented in simple porosity-related metrics, typically density [66,67,70] or relative density [68,69], with different AM process parameters (Fig. 8a). The basic idea behind those researches is thus analogous to that of existing data-driven grain structure modeling in AM as aforementioned. Aside from that, there are a few more interesting attempts. Tian et al. [46] have selected some special and complex input features, including pyrometry images and sequential images from infrared camera during AM process, to data-driven predict layer-wise porosity (Fig. 8b) and Imani et al. [72] have taken layer-wise surface images from DSLR camera as inputs. Most recently, Wang et al.[73] have proposed a multi-input CNN for data-driven modeling of structural and morphological evolution of porosity structure for SLS process, based on the input structure of as-deposited powder bed and input parameters of applied laser (Fig. 8c). They have also showed that the proposed data-driven approach permits component-level SLS porosity simulation in minutes.

3.2.3 Data-driven modeling of geometric distortion

Data-driven modeling of geometric distortion is targeted at external shape of AM parts, from mesoscale cross-sectional shape of struts [76] to macroscopic shape of the whole component [77]. Geometrical distortion and dimensional inaccuracy can stem from internal stress associated with the complex thermal behavior of AM process [120], insufficient support for overhangs [76], limited printing resolution of AM machine [79] and inaccurate representation by .stl file (e.g., approximation of curvilinear surface by a mesh of triangles) [121]. They lead to deviation between designed and fabricated geometry/shape. Accurate modeling of geometric distortion is critical for correct compensation of geometrical deviation in the computer-aided design (CAD) phase.

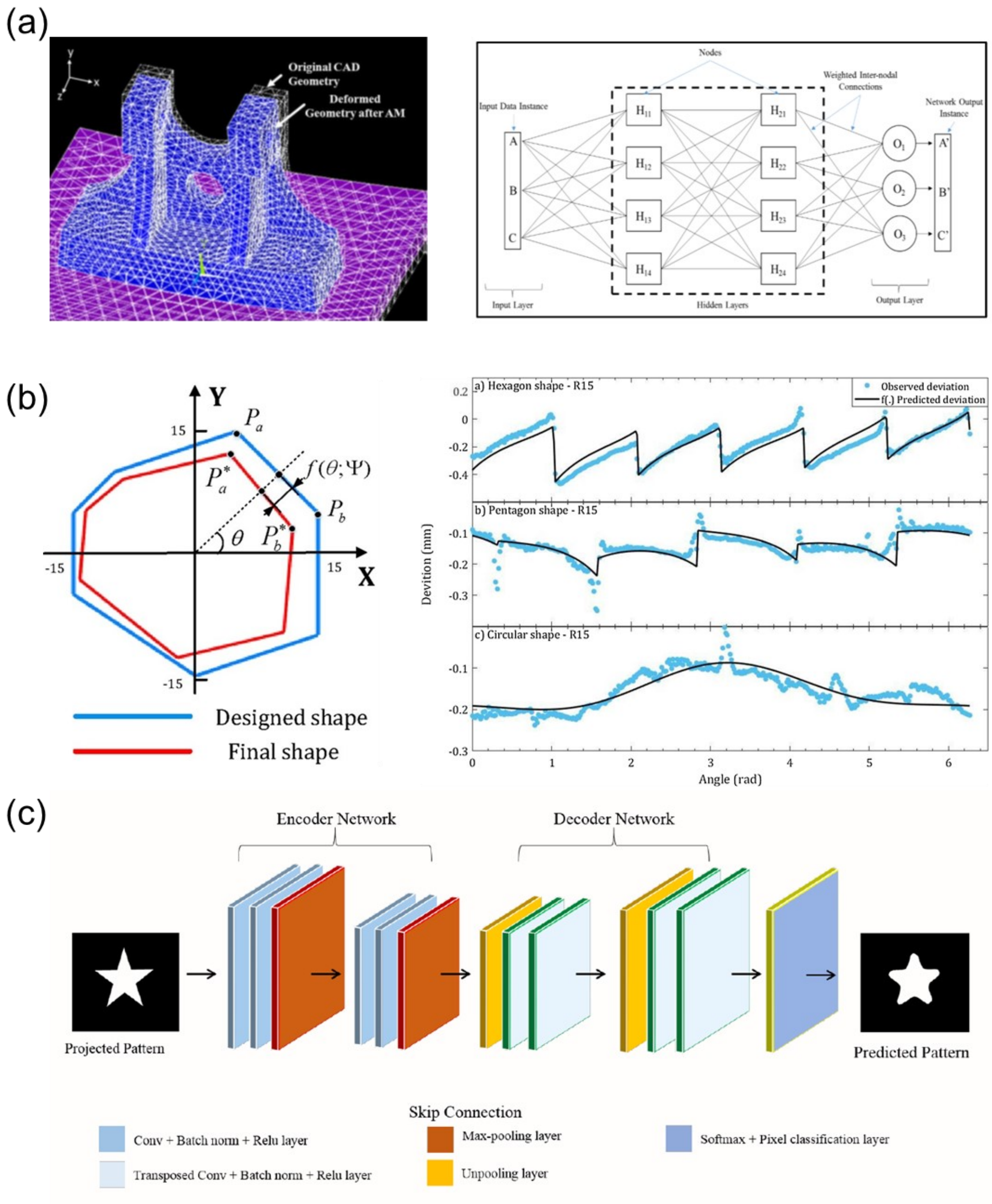

Fig. 9 Data-driven modeling of geometrical distortion. a Localized data-driven modeling that corelates cartesian coordinates of node between theoretical and actual shape [74]. **b**. Localized data-driven modeling that describe deviation of peripheral points as a function of polar angle, θ, in polar coordinate system (PCS) [77]. **c**. CNN with encoder and decoder allows for explicit data-driven modeling of the geometrical distortion of the entire shape [79].

Data-driven modeling of geometric distortion might be the most developed data-driven AM modeling. For example, unlike data-driven grain structure or porosity modeling, the bulk of existing data-driven modeling of geometrical distortion focus on the real geometry[74,76-81], rather than dealing with some simple deviation descriptors (e.g., percentage of shrinkage/expansion[75]). Also, some novel data-driven modeling methodologies based on CNN have been proposed for explicitly handling the 2D or even 3D component geometry. Most of existing data-driven modeling may be categorized into two groups. The first group adopted localized modeling of points that form the geometry/shape. For instance, Chowdhury et al. [122] have attempted to correlate Cartesian coordinates of nodes between the designed and actual model (Fig. 9a). However, a preferred way in this group is to describe points and thus their deviation in polar coordinate system (PCS) [76-78], which facilitates the modeling of complicated deviation mainly as a function of the polar angle (Fig. 9b). The second group mainly leverages CNN for direct data-driven modeling of the whole shape [79,80], thus usually enjoying better generalization to various shapes. The core principle behind them is to use image-to-image regression capability of some CNNs, namely those with an encoder-decoder structure, where encoder reads original shape and decoder reconstructs encoding results to the shape after distortion (Fig. 9c).

3.2.4 Other data-driven structure modeling

There are also some other types of data-driven structure modeling in literature, including surface structure/roughness modeling [37,123] and dendrite structure (arm space) modeling [124]. Although they have not been widely studied yet, data-driven modeling of those structures is of same importance. Besides the above two, other structures, such as sub-grain microstructure and texture associated with solid-state phase transformation [125], seem to have been even less studied. In part this may be attributed to the difficulty in acquiring related data. On the other hand, these AM quantities may have usually been less focused in AM field, even in physics-based modeling or experiment based AM researches.

3.3 Data-driven property modeling

3.3.1 Data-driven modeling of mechanical properties

Stress-strain curve is an informative representation of mechanical property, revealing many scalar properties, such as yield strength, Young's modulus, stiffness, etc. Due to the complex characteristics of stress-strain curve, the bulk of existing data-drive property modeling in AM is centered on those specific scalar properties. Methodologically, data-driven modeling of them are quite similar. They are thus collectively reviewed as data-driven modeling of mechanical properties in this subsection.

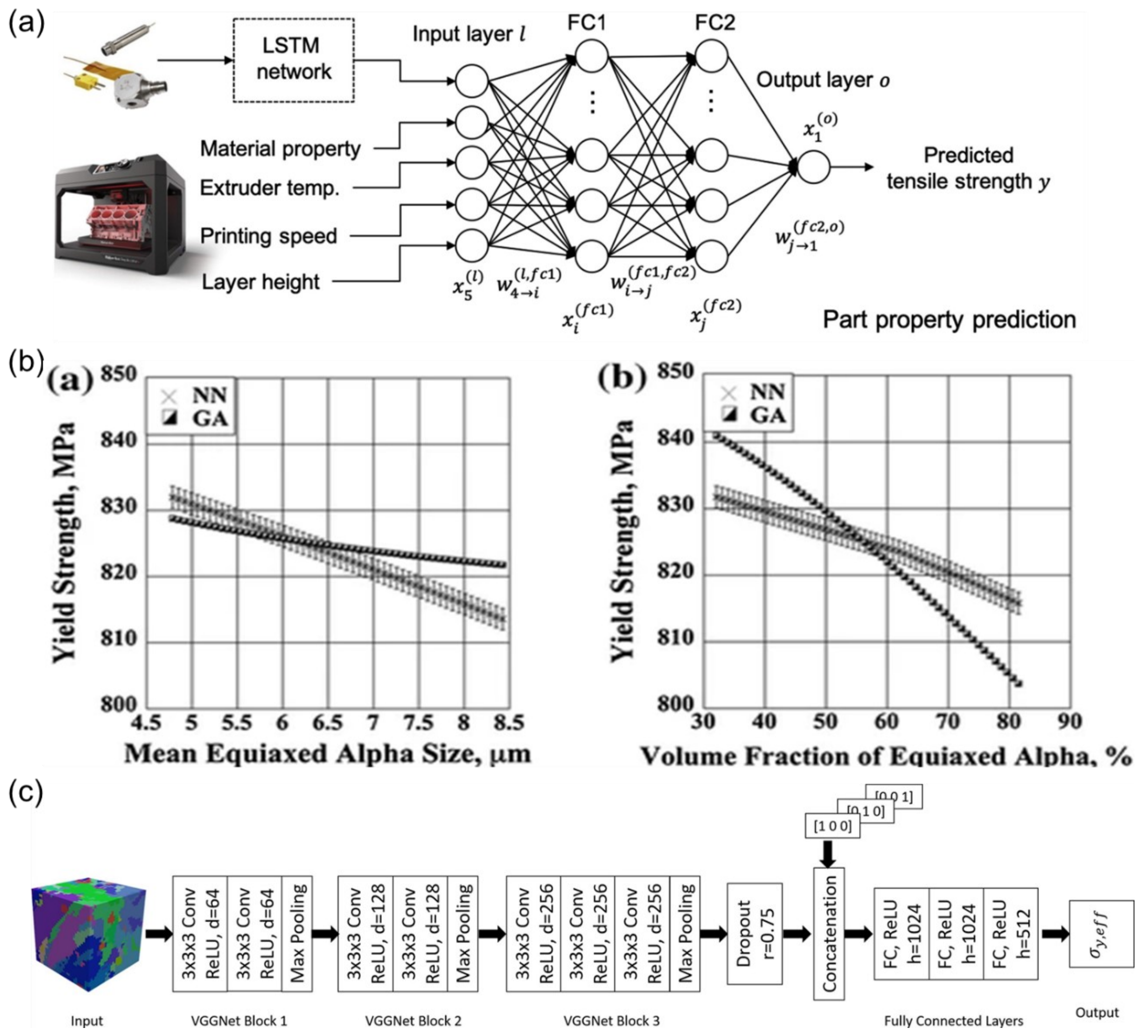

Fig. 10 Data driven modeling of mechanical properties. a Data-driven property modeling that builds process-property (P-P) linkage [85]. **b-c** Data-driven property modeling that builds structure-

property (S-P) linkages by using microstructural descriptors (i.e., mean equiaxed alpha size, volume fraction of equiaxed alphas) [83] and real microstructure [44] as inputs, respectively.

The two most widely studied scalar properties in existing data-driven property modeling are ultimate strength and yield strength. Data-driven modeling of different mechanical properties can be divided into two groups based on the type of input features. The first group directly links AM process (parameters) to properties by skipping the structure stage [84-88], thus essentially building the process-property (P-P) linkage (Fig. 10a). The second group instead builds structure-property (S-P) linkage by using structure as input [44,82,83], and can be more challenging by involving the complex quantity of structure. Among them, some have simplified the modeling problem by using quantitative descriptors to represent structure [82,83] (Fig. 10b), as usually does the data-driven modeling of AM structure; some other data-driven S-P modeling [44,126] have however adopted CNN to explicitly read and process grain structure, and successfully built the relationship between image-based structure and scalar properties (Fig. 10c).

3.3.2 Data-driven modeling of stress-strain curves

As mentioned, it is known that stress-strain curve is a more general representation of mechanical property, indicating many basic properties of a materials. Data-driven modeling of stress-strain curve is thus of prime importance in data-driven property modeling in AM and even materials science.

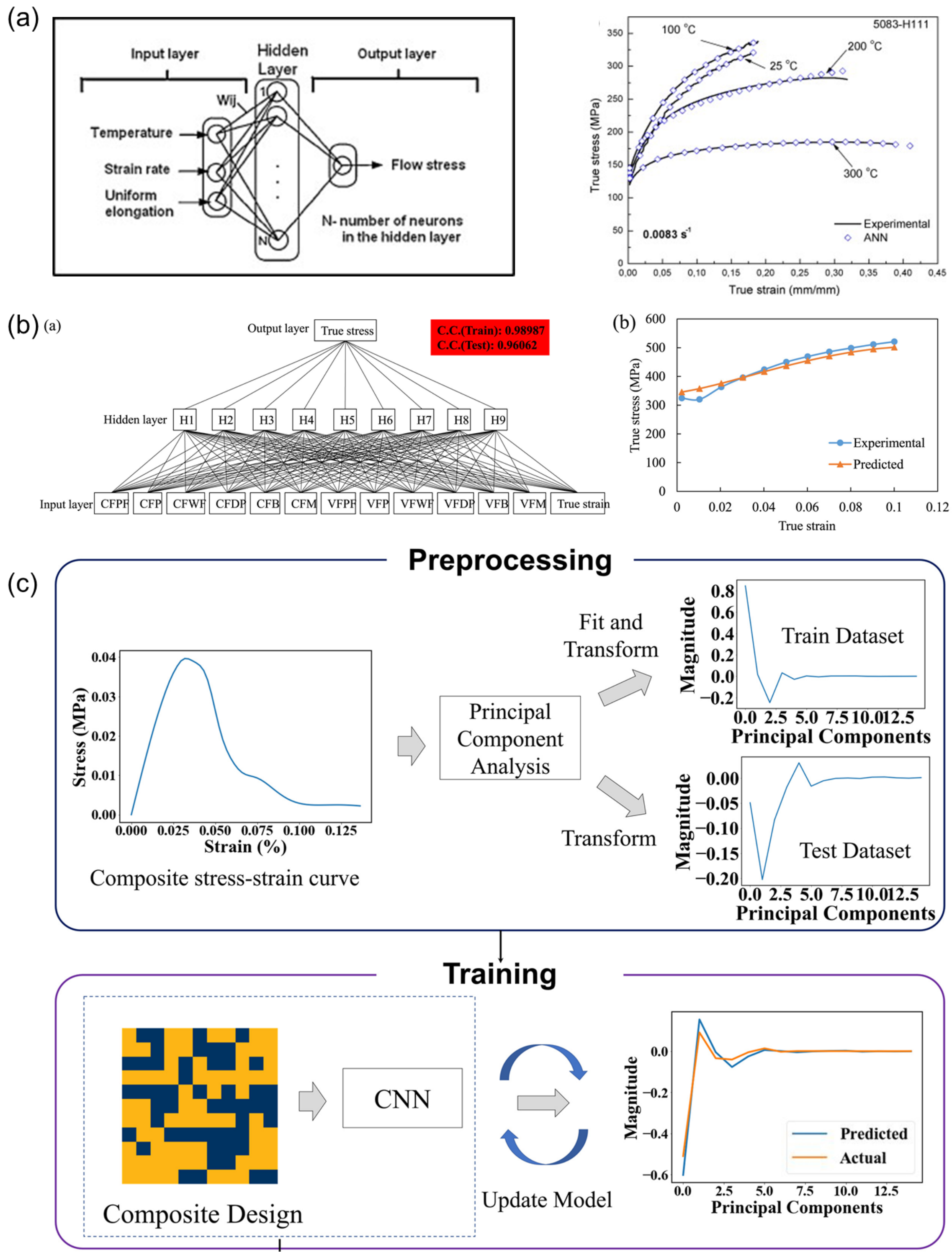

Fig. 11 Data-driven stress-strain curve modeling. a-b Step-wise data-driven modeling of stress-strain curve, where uniform elongation [92] and true strain [93] were respectively used as step indicator. The stress-strain curve was obtained by combining discrete predictions at different step input. **c** Data-driven modeling of the stress-strain curve as a whole, where principal component

analysis (PCA) was utilized to reduce the dimension of original stress-strain curve [40]. It is also noteworthy that 2D image-based structure was just used as the input with the assistance of CNN. In spite of its major significance, data-driven modeling of stress-strain curve in AM has been rather limitedly studied. Wang et al. [90] have proposed to use SVD to reduce dimension of stress-strain curve, followed by data-driven stress-strain curve modeling in the latent space. However, in other fields there are indeed many research efforts on this theme, and those data-driven modeling methodologies can be easily extended to AM. Most of those researches [91-93] adopted a step-wise modeling method, namely building real-time mechanical response as a function of input features (Fig. 11a-b). In this case, elongation or strain as a time-step indicator has to be included as an input. The entire stress-strain curve is then obtained from a series of separate predictions at different time-step inputs, i.e., discrete points in Fig. 11a-b. In addition to step-wise method, Yang et al. [40] have used principal component analysis (PCA) for dimension reduction and facilitated data-driven modeling of stress-strain curve as a whole (Fig. 11c). This idea is similar to the above-mentioned SVD-based approach, but they have further integrated CNN for microstructure representation modeling, instead of using structural descriptors as input features of the data-driven structure-property model.

3.3.3 Other data-driven property modeling

Other existing data-driven property modeling in AM includes data-driven modeling of fatigue life [127], dynamic properties including storage compliance and loss compliance [128], maximum stress development [129] and even full-field strain development [130]. There is also other data-driven property modeling, which is not performed for but readily extendable to AM-fabricated materials and components, such as data-driven modeling of fatigue crack development [131], stress hotspot formation [132] and full-field stress response [133]. Like many underexplored quantities in data-driven structure modeling, those properties are worth more attention from AM researchers. For instance, maximum stress and even stress hotspot development are important to study considering the main application of AM for fabricating complex structures with intricate geometry, wherein stress concentration and caused fracture may be an outstanding issue. A data-driven modeling capability is thus of immense help for optimizing structure design and improving service lifetime of AM parts.

4. Future directions and conclusion

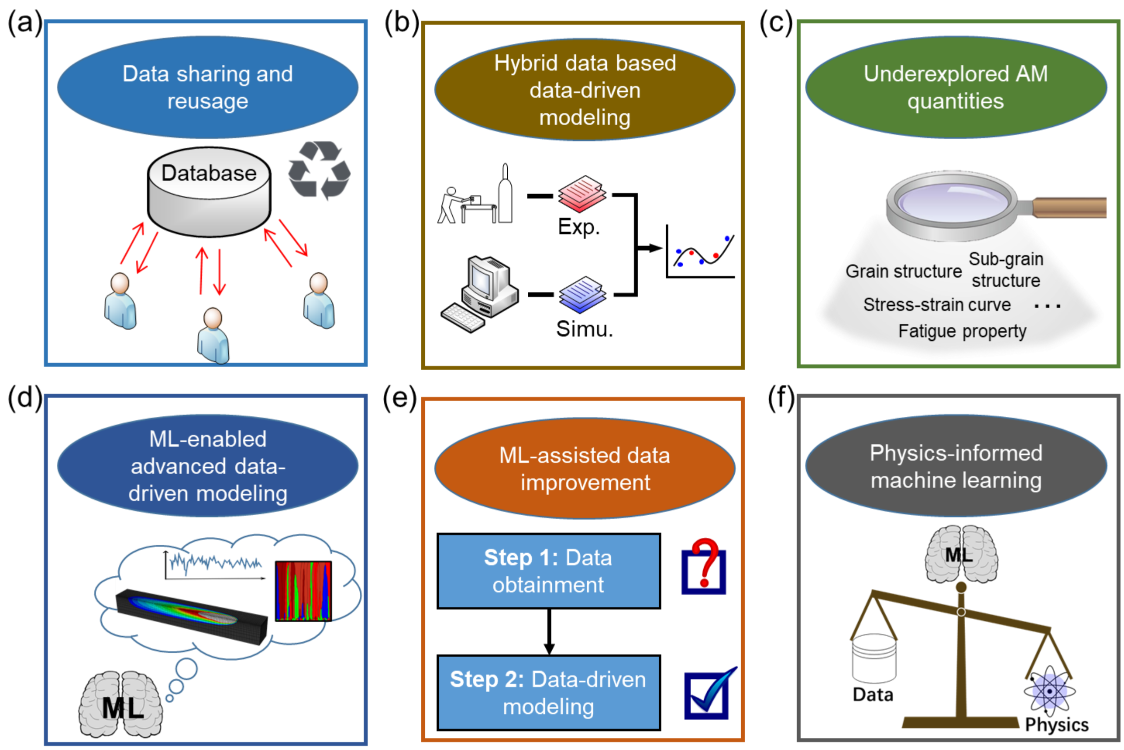

Fig. 12 Potential future research directions for further advancing data-driven AM modeling. **a** Building professional AM database with easy access will greatly advance data-driven AM modeling by improving data efficiency. **b** Using hybrid experimental and simulation will enable to build trustworthy data-driven AM models at minimal cost. **c** In future, more attention can be paid to data-driven modeling of those less studied AM quantities. **d** The distinct capability of machine learning (ML) in explicitly handling high-dimensional and complex data will open enormous opportunities for advanced data-driven AM modeling. **e** ML-assisted data collection, processing and acquisition in Step 1 will hugely benefit data-driven modeling by improving both the quality and quantity of data. **f** Physics-informed ML will potentially revolutionize data-driven modeling by alleviating its dependence on AM data, with the incorporation of physical knowledge into ML model.

4.1 Data sharing and reusage

In future, fostering data-sharing culture and increasing data reusage will be critical for substantially advancing data-driven AM modeling from the data perspective. Data sharing and reusage is a common issue extensively implied in data-driven studies across many scientific domains, from geoscience [16] to agriculture [134] and materials science [135]. Establishing professional database with easy access and building supportive data infrastructure have thus been widely recognized, enabling to greatly improve data efficiency (Fig. 12a). This has been proved in computer science (CS) field, where the constant improvement of different ML techniques has greatly relied on many known datasets, from the early MNIST dataset of handwritten digit [136] to the more recent ImageNet dataset of diverse visual objects [137]. Those large datasets are well curated, easily accessible and commonly used by CS community, providing standardized benchmarks for transparent comparison and accelerated evolvement of ML algorithms. Researchers are able to spend maximum efforts on developing and advancing data-driven modeling itself without data worry. In terms of AM, AM-Bench by National Institute of Standards and Technology (NIST) [138] seems to be the only relevant project, which aims to provides highly controlled AM benchmark P-S-P data with permanent public access. On the other hand, with the ever-increasing interest in data-driven AM studies, an abundance of AM data is actually generated every day. There is however currently still a lack of satisfactory practices that archive and manage them in a unified way. This largely prevents the reusage of those scholarly AM data and maximization of their value.

4.2 Hybrid data-based data-driven AM modeling

Using hybrid data possibly allows for defeating respective drawback of data-driven modeling based solely on simulation data or experimental data (Fig. 12b). That is, physic-based modeling is usually inexpensive compared to experiments, but can only take into account partial physics in practice and thus provide imperfect data. Experiments, on the contrary, offers more practical and reliable data, yet at greater expense. Some researchers [54,64,139] have thus advocated the hybrid-data schema to build high-reliability data-driven AM model at minimal cost. In the few related studies, they first build a primitive data-driven model based on simulation data, followed by incorporating and experimentally calibrating a discrepancy term to describe the bias between physical simulation and practice. Nonetheless, the efficacy and feasibility of hybrid-data strategy are still subject to further verification, not only through a more detailed and clearer comparison with those based purely on theoretical or experimental data, but also in various data-driven AM modeling scenarios. In near future, using hybrid data might be the optimal way to economically build fully trustworthy data-driven AM models for practical use.

4.3 Underexplored AM quantities of interest

Based on the current review, it seems that some AM quantities related to structure and property are limitedly studied (e.g., grain structure, stress-strain curve) or even completely unexplored (e.g., sub-grain structure) in existing data-driven AM modeling. For some of them, such as grain structure and stress-strain curve, the complex quantity involved greatly impedes the development of those data-driven modeling. However, as indicated by the early analysis, one can often find related data-driven modeling researches in other fields than AM. The advanced data-driven modeling methodologies developed therein can be borrowed to facilitate data-driven AM modeling in future. However, for some other AM quantities, the root reason of their little-to-no exploration may stem largely from the difficulty in obtaining (sufficient) related data. To overcome this situation, the further advancement of related experimental instrumentations and physical models are critical. In brief, special care can be taken for those underexplored and underdeveloped areas in future data-driven AM modeling (Fig. 12c).

4.4 Machine learning enabled advanced data-driven modeling

Among various ML techniques, MLP is the most frequently used one. Much of existing data-driven AM modeling basically falls within or is purposely simplified to common regression analysis. One of the most distinct advantages of ML models, the capability of explicitly handling high-dimensional and complex data, such as CNN for image-type data and RNN for sequences, is sometimes neglected. For example, for data-driven grain structure modeling, what one nowadays does is usually using ML to build simple microstructure descriptors as a function of different AM process parameters. Under this circumstance, it poses an intriguing question - can we make deeper use of ML model to realistically predict structural/morphological development and even dynamic evolution of grain structure (as does a physical model)? For instance, in fluid community, different CNNs with encoder-decoder architecture[140-142], CNN+RNN[143,144] and GAN[145] have been used for modeling the evolution and development of flow fields. In solid mechanics modeling, CNN [146] and GAN [147] can accurately predict the stress field developed for given solid structure and loading condition. Admittedly, ML is now serving as the main workhorse in various AM data-driven modeling. However, many capable ML models are overlooked for some data-driven AM modeling within their specialties. Therefore, a more general question raised herein is - in face of high-dimensional and complex AM quantities, which are not uncommon scenarios in data-driven AM modeling, can we take more real advantage of those powerful ML techniques (Fig. 12d)?

4.5 Machine learning assisted data improvement

The whole workflow of a data-driven modeling effort actually includes two main steps – data obtainment and data-driven modeling (Fig. 12e). Acquiring a large quantity of high-quality data would be excellent for any type of data-driven modeling. Existing data-driven AM modeling often uses ML model purely as a data-driven relationship modeling tool, while largely ignoring its multi-functionalities and thus enormous potential for improving data in data-obtainment step. For instance, AM data of experimental images are sometimes not ideal (e.g., noisy and blurry), requiring further processing that is too laborious for human, whereas image processing for various purposes are specialized areas of CNNs and GANs. Also, NLP techniques have shown great potential in information extraction, for example, from literature (including tables, figures, and textual paragraphs) [148,149] and medical records [150]. Similarly, NLP techniques may be used to collect useful AM data from their unstructured form in the voluminous AM literature and other digitalized resources. Ma et al. [151] have used GAN for data augmentation by synthesizing experimental images of grain structures. All of those facts show the tremendous usability and versatility of ML in improving the quality and quantity of data. To authors' best knowledge, there is however only one related work[152], where a CNN-based detection module replaced human to classify local flaws in AM parts and thus effortlessly create required data for further data-driven predictive modeling of local flaws. In short, a valuable research direction implied herein is - can we use ML in a more versatile way, and enhance data-driven AM modeling from the pre-data-driven-modeling aspect, i.e., improving data via ML-assisted data collection, processing and acquisition?

4.6 Small data and physics-informed machine learning

Simulation-derived and experimental datasets in AM are usually limited in size and expensive to be collected. Additionally, training datasets have to be recollected for traditional machine learning methods regarding the new manufacturing conditions such as new geometries, materials, and printing techniques since the feature distribution changes [153]. It is critical for data-driven methods to achieve high prediction accuracy when available datasets are small. Transfer learning, which leverages the knowledge from a related task (in the source domain) to facilitate a new task (in the target domain), can be utilized when the feature distribution shifts and training datasets are deficient [154]. Cheng et al. [155] proposed a hybrid transfer learning framework to predict the in-plane shape deviation of new geometries based on the knowledge from a small number of known geometries. Sabbaghi and Huang[156] developed a transfer learning model to predict the in-plane shape deviations of a new stereolithography process based on the knowledge learned from a previous stereolithography process with limited experimental datasets. Although previous studies demonstrated that transfer learning is effective to predict shape deviation across distinct processes and geometries, more studies are still needed to be extended to the other fields in AM, e.g., learning the knowledge from one material having a large amount of training data to predict the properties of another material having limited training data. Moreover, data-driven models which can transfer the knowledge among process, structure, and property in AM have not been fully developed when training data is limited.

In addition to transfer learning, physics-informed machine learning[157] (PIML) especially holds the promise of reducing, or even completely obviating[158,159], the need for training data (Fig. 12f), while improving the generalization and robustness of the trained model. This is done through incorporating prior knowledge and physical laws into the otherwise purely data-driven model. Specifically, Guo et al. [160] have summarizes five types of PIMLs, namely physics-informed model inputs (PIMI), physics-informed model training (PIMT), physics-informed model components (PIMC), physics-informed model architecture (PIMA), and physics-informed model output (PIMO). That said, the prevalent way to impose physics constraint might be formulating physical governing equations into loss function[161] (i.e., the foregoing PIMT). In this case, the prediction of data-driven model would be penalized by its violation of physical laws. The model training is thus guided directly and sometimes completely by the physical principles, thereby potentially free from any training data. As a transformative modeling technique, there is an explosive application of PIML across various fields in last few years[157]. However, PIML in AM is still at its infancy, with only a handful of relevant research efforts on PIMI[162-165] and PIMT[163,166,167]. Physics-informed ML is anticipated to play a revolutionary role in future data-driven AM modeling, by eliminating its dependence on intensive AM data yet without sacrifice of modeling accuracy.

In conclusion, this paper has systematically reviewed the existing data-driven modeling with respect to different AM quantities of interest along process-structure-property chain. It provides a summary of basic information of existing data-driven AM modeling, as well a further analysis on relevant progress made so far. Promising research directions in future are also discussed in detail, which are associated with improving not only data-driven modeling itself but also AM data.

Particularly, many capable ML models possess unparalleled regression power especially in explicitly handling high-dimensional and complex data, as well as much other versatility. They are expected to significantly improve data-driven AM modeling in future, from perspectives of both AM data obtainment and data-driven modeling methodology. In addition, the burgeoning physics-informed ML even has the potential to completely transform data-driven AM modeling, by alleviating or fully obviating the need for AM data.

5. Acknowledgements

The authors acknowledge financial support by startup fund from Department of Mechanical Engineering at University of Michigan-Dearborn, Propelling Original Data Science (PODS) Grants from Michigan Institute of Data Science (MIDAS) and National Science Foundation (NSF) Grant CMMI-2020527.

Author contributions

Z.W. leaded the development of this paper under the supervision of D.W., M.B. and L.C. W.Y. and Q.L. helped survey relevant literature and participated in the paper writing. Y.Z. and P.L. prepared tables and figures. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

References

1 Attaran, M. The rise of 3-D printing: The advantages of additive manufacturing over traditional manufacturing. *Bus. Horiz.* **60**, 677-688 (2017).

2 Frazier, W. E. Metal additive manufacturing: a review. *J. Mater. Eng. Perform.* **23**, 1917-1928 (2014).

3 Wong, K. V. & Hernandez, A. A review of additive manufacturing. *ISRN Mechanical Engineering* **2012** (2012).

4 Antonysamy, A. A., Meyer, J. & Prangnell, P. Effect of build geometry on the β-grain structure and texture in additive manufacture of Ti 6Al 4V by selective electron beam melting. *Mater. Charact.* **84**, 153-168 (2013).

5 Qiu, C. *et al.* On the role of melt flow into the surface structure and porosity development during selective laser melting. *Acta Mater.* **96**, 72-79 (2015).

6 Matthews, M. J. *et al.* Denudation of metal powder layers in laser powder bed fusion processes. *Acta Mater.* **114**, 33-42 (2016).

7 Xiao, Y. *et al.* A gleeble-assisted study of phase evolution of Ti-6Al-4V induced by thermal cycles during additive manufacturing. *J. Alloys Compd.* **860**, 158409 (2021).

8 DebRoy, T. *et al.* Additive manufacturing of metallic components–process, structure and properties. *Prog. Mater Sci.* (2017).

9 Hey, T., Tansley, S. & Tolle, K. M. (2009).

10 Bell, G., Hey, T. & Szalay, A. Beyond the data deluge. *Science* **323**, 1297-1298 (2009).

11 Haenlein, M. & Kaplan, A. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *Calif. Manage. Rev.* **61**, 5-14 (2019).

12 Brunton, S. L., Noack, B. R. & Koumoutsakos, P. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* **52**, 477-508 (2020).

13 Brunton, S. L. & Kutz, J. N. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. (Cambridge University Press, 2019).

14 Schlexer Lamoureux, P. *et al.* Machine learning for computational heterogeneous catalysis. *ChemCatChem* **11**, 3581-3601 (2019).

15 Han, J., Jentzen, A. & Weinan, E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences* **115**, 8505-8510 (2018).

16 Bergen, K. J., Johnson, P. A., Maarten, V. & Beroza, G. C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **363** (2019).

17 Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **3**, 54, doi:10.1038/s41524-017-0056-5 (2017).

18 Li, Y. *et al.* Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renewable and sustainable energy reviews* **113**, 109254 (2019).

19 Bock, F. E. *et al.* A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials* **6**, 110 (2019).

20 Wang, J., Ma, Y., Zhang, L., Gao, R. X. & Wu, D. Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems* **48**, 144-156 (2018).

21 Xiong, Y. *et al.* Data-driven design space exploration and exploitation for design for additive manufacturing. *J. Mech. Des.* **141** (2019).

22 Jiang, J., Xiong, Y., Zhang, Z. & Rosen, D. W. Machine learning integrated design for additive manufacturing. *J. Intell. Manuf.*, 1-14 (2020).

23 Mukherjee, T. & DebRoy, T. A digital twin for rapid qualification of 3D printed metallic components. *Applied Materials Today* **14**, 59-65 (2019).

24 Zhang, L. *et al.* Digital Twins for Additive Manufacturing: A State-of-the-Art Review. *Applied Sciences* **10**, 8350 (2020).

25 Majeed, A. *et al.* A big data-driven framework for sustainable and smart additive manufacturing. *Robotics and Computer-Integrated Manufacturing* **67**, 102026 (2021).

26 Lehmhus, D. *et al.* Cloud-based automated design and additive manufacturing: a usage data-enabled paradigm shift. *Sensors* **15**, 32079-32122 (2015).

27 Wang, G. G. & Shan, S. Review of metamodeling techniques in support of engineering design optimization. (2007).

28 Wang, C., Tan, X., Tor, S. & Lim, C. Machine learning in additive manufacturing: State-of-the-art and perspectives. *Addit. Manuf.*, 101538 (2020).
29 Razvi, S. S., Feng, S., Narayanan, A., Lee, Y.-T. T. & Witherell, P. in *ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* (American Society of Mechanical Engineers Digital Collection).
30 Jin, Z., Zhang, Z., Demir, K. & Gu, G. X. Machine Learning for Advanced Additive Manufacturing. *Matter* **3**, 1541-1556 (2020).
31 Baumann, F. W., Sekulla, A., Hassler, M., Himpel, B. & Pfeil, M. Trends of machine learning in additive manufacturing. *International Journal of Rapid Manufacturing* **7**, 310-336 (2018).
32 Goh, G., Sing, S. & Yeong, W. A review on machine learning in 3D printing: Applications, potential, and challenges. *Artificial Intelligence Review*, 1-32 (2020).
33 Qi, X., Chen, G., Li, Y., Cheng, X. & Li, C. Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives. *Engineering* **5**, 721-729 (2019).
34 Kouraytem, N., Li, X., Tan, W., Kappes, B. & Spear, A. Modeling process-structure-property relationships in metal additive manufacturing: a review on physics-driven versus data-driven approaches. *Journal of Physics: Materials* (2020).
35 Meng, L. *et al.* Machine Learning in Additive Manufacturing: A Review. *JOM*, 1-15 (2020).
36 Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735-1780 (1997).
37 Wu, D., Wei, Y. & Terpenny, J. Predictive modelling of surface roughness in fused deposition modelling using data fusion. *International Journal of Production Research* **57**, 3992-4006 (2019).
38 Wu, D. & Xu, C. Predictive modeling of droplet formation processes in inkjet-based bioprinting. *J. Manuf. Sci. Eng.* **140** (2018).
39 Wang, Z. *et al.* A Data-driven Approach for Process Optimization of Metallic Additive Manufacturing under Uncertainty. *J. Manuf. Sci. Eng.* **141**, 081004 (2019).
40 Yang, C., Kim, Y., Ryu, S. & Gu, G. X. Prediction of composite microstructure stress-strain curves using convolutional neural networks. *Materials & Design* **189**, 108509, doi:https://doi.org/10.1016/j.matdes.2020.108509 (2020).
41 Yang, Z. *et al.* Microstructural materials design via deep adversarial learning methodology. *J. Mech. Des.* **140** (2018).
42 Fukami, K., Nakamura, T. & Fukagata, K. Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Phys. Fluids* **32**, 095110 (2020).
43 Murata, T., Fukami, K. & Fukagata, K. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *J. Fluid Mech.* **882** (2020).
44 Herriott, C. & Spear, A. D. Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine-and deep-learning methods. *Comput. Mater. Sci.* **175**, 109599 (2020).
45 Fang, L. *et al.* Data driven analysis of thermal simulations, microstructure and mechanical properties of Inconel 718 thin walls deposited by metal additive manufacturing. *arXiv preprint arXiv:2110.07108* (2021).
46 Tian, Q., Guo, S., Melder, E., Bian, L. & Guo, W. Deep Learning-Based Data Fusion Method for In Situ Porosity Detection in Laser-Based Additive Manufacturing. *J. Manuf. Sci. Eng.* **143** (2021).
47 Huang, J. *et al.* Unsupervised learning for the droplet evolution prediction and process dynamics understanding in inkjet printing. *Addit. Manuf.* **35**, 101197 (2020).

48 Tapia, G., Khairallah, S., Matthews, M., King, W. E. & Elwany, A. Gaussian process-based surrogate modeling framework for process planning in laser powder-bed fusion additive manufacturing of 316L stainless steel. *Int. J. Adv. Manuf. Technol.* **94**, 3591-3603 (2018).
49 Kamath, C. Data mining and statistical inference in selective laser melting. *Int. J. Adv. Manuf. Technol.* **86**, 1659-1677, doi:10.1007/s00170-015-8289-2 (2016).
50 Meng, L. & Zhang, J. Process Design of Laser Powder Bed Fusion of Stainless Steel Using a Gaussian Process-Based Machine Learning Model. *JOM* **72**, 420-428 (2020).
51 Lee, S., Peng, J., Shin, D. & Choi, Y. S. Data analytics approach for melt-pool geometries in metal additive manufacturing. *Science and technology of advanced materials* **20**, 972-978 (2019).
52 Yang, Z., Lu, Y., Yeung, H. & Krishnamurty, S. From scan strategy to melt pool prediction: A neighboring-effect modeling method. *Journal of Computing and Information Science in Engineering* **20** (2020).
53 Yang, Z., Lu, Y., Yeung, H. & Kirshnamurty, S. in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* V009T009A046 (American Society of Mechanical Engineers).
54 Wang, Z. *et al.* Uncertainty quantification and reduction in metal additive manufacturing. *npj Comput. Mater.* **6**, 175, doi:10.1038/s41524-020-00444-x (2020).
55 Caiazzo, F. & Caggiano, A. Laser direct metal deposition of 2024 Al alloy: trace geometry prediction via machine learning. *Materials* **11**, 444 (2018).
56 Tapia, G. *et al.* Uncertainty Propagation Analysis of Computational Models in Laser Powder Bed Fusion Additive Manufacturing using Polynomial Chaos Expansions. *J. Manuf. Sci. Eng.* **140** (2018).
57 Garg, A., Tai, K. & Savalani, M. Formulation of bead width model of an SLM prototype using modified multi-gene genetic programming approach. *Int. J. Adv. Manuf. Technol.* **73**, 375-388 (2014).
58 Lu, Z. *et al.* The prediction of the building precision in the Laser Engineered Net Shaping process using advanced networks. *Optics and Lasers in Engineering* **48**, 519-525 (2010).
59 Stathatos, E. & Vosniakos, G.-C. Real-time simulation for long paths in laser-based additive manufacturing: a machine learning approach. *Int. J. Adv. Manuf. Technol.* **104**, 1967-1984 (2019).
60 Zhang, Z., Liu, Z. & Wu, D. Prediction of melt pool temperature in directed energy deposition using machine learning. *Addit. Manuf.* **37**, 101692 (2021).
61 Paul, A. *et al.* in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA).* 541-550 (IEEE).
62 Mozaffar, M. *et al.* Data-driven prediction of the high-dimensional thermal history in directed energy deposition processes via recurrent neural networks. *Manufacturing Letters* **18**, 35-39, doi:https://doi.org/10.1016/j.mfglet.2018.10.002 (2018).
63 Roy, M. & Wodo, O. Data-driven modeling of thermal history in additive manufacturing. *Addit. Manuf.* **32**, 101017 (2020).
64 Li, J., Jin, R. & Yu, H. Z. Integration of physically-based and data-driven approaches for thermal field prediction in additive manufacturing. *Materials & Design* **139**, 473-485, doi:https://doi.org/10.1016/j.matdes.2017.11.028 (2018).
65 Popova, E. *et al.* Process-structure linkages using a data science approach: application to simulated additive manufacturing data. *Integr. Mater. Manuf. Innov.* **6**, 54-68 (2017).
66 Wang, R.-J., Li, J., Wang, F., Li, X. & Wu, Q. ANN model for the prediction of density in selective laser sintering. *International Journal of Manufacturing Research* **4**, 362-373 (2009).
67 Tapia, G., Elwany, A. & Sang, H. Prediction of porosity in metal-based additive manufacturing using spatial Gaussian process models. *Addit. Manuf.* **12**, 282-290 (2016).

68 Lee, C. H. *et al.* Optimizing laser powder bed fusion of Ti-5Al-5V-5Mo-3Cr by artificial intelligence. *J. Alloys Compd.*, 158018 (2020).

69 Liu, Q. *et al.* Machine-learning assisted laser powder bed fusion process optimization for AlSi10Mg: New microstructure description indices and fracture mechanisms. *Acta Mater.* **201**, 316-328, doi:https://doi.org/10.1016/j.actamat.2020.10.010 (2020).

70 Rankouhi, B., Jahani, S., Pfefferkorn, F. E. & Thoma, D. J. Compositional grading of a 316L-Cu multi-material part using machine learning for the determination of selective laser melting process parameters. *Addit. Manuf.* **38**, 101836 (2021).

71 Garg, A. & Lam, J. S. L. Measurement of environmental aspect of 3-D printing process using soft computing methods. *Measurement* **75**, 210-217 (2015).

72 Imani, F., Chen, R., Diewald, E., Reutzel, E. & Yang, H. Deep learning of variant geometry in layerwise imaging profiles for additive manufacturing quality control. *J. Manuf. Sci. Eng.* **141** (2019).

73 Wang, Z. *et al.* yNet: a multi-input convolutional network for ultra-fast simulation of field evolvement. *arXiv preprint arXiv:2012.10575* (2020).

74 Chowdhury, S. & Anand, S. in *International Manufacturing Science and Engineering Conference.* V003T008A006 (American Society of Mechanical Engineers).

75 Negi, S. & Sharma, R. K. Study on shrinkage behaviour of laser sintered PA 3200GF specimens using RSM and ANN. *Rapid Prototyping Journal* (2016).

76 Hong, R. *et al.* Artificial neural network-based geometry compensation to improve the printing accuracy of selective laser melting fabricated sub-millimetre overhang trusses. *Addit. Manuf.* **37**, 101594 (2021).

77 Zhu, Z., Anwer, N., Huang, Q. & Mathieu, L. Machine learning in tolerancing for additive manufacturing. *CIRP Annals* **67**, 157-160 (2018).

78 Cheng, L., Wang, A. & Tsung, F. A prediction and compensation scheme for in-plane shape deviation of additive manufacturing with information on process parameters. *IISE Transactions* **50**, 394-406 (2018).

79 He, Y. *et al.* in *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* V01BT02A033-V001BT002A033 (American Society of Mechanical Engineers).

80 Shen, Z. *et al.* A learning-based framework for error compensation in 3D printing. *IEEE transactions on cybernetics* **49**, 4042-4050 (2019).

81 Francis, J. & Bian, L. Deep learning for distortion prediction in laser-based additive manufacturing using big data. *Manufacturing Letters* **20**, 10-14 (2019).

82 Gupta, A., Cecen, A., Goyal, S., Singh, A. K. & Kalidindi, S. R. Structure–property linkages using a data science approach: application to a non-metallic inclusion/steel composite system. *Acta Mater.* **91**, 239-254 (2015).

83 Collins, P. *et al.* Progress toward an integration of process–structure–property–performance models for "three-dimensional (3-D) printing" of titanium alloys. *JOM* **66**, 1299-1309 (2014).

84 Garg, A., Tai, K. & Savalani, M. State-of-the-art in empirical modelling of rapid prototyping processes. *Rapid Prototyping Journal* (2014).

85 Zhang, J., Wang, P. & Gao, R. X. Deep learning-based tensile strength prediction in fused deposition modeling. *Computers in Industry* **107**, 11-21 (2019).

86 Bayraktar, Ö., Uzun, G., Çakiroğlu, R. & Guldas, A. Experimental study on the 3D-printed plastic parts and predicting the mechanical properties using artificial neural networks. *Polym. Adv. Technol.* **28**, 1044-1051 (2017).

87 Garg, A., Tai, K., Lee, C. & Savalani, M. A hybrid $$\text {M} 5^\prime $$-genetic programming approach for ensuring greater trustworthiness of prediction ability in modelling of FDM process. *J. Intell. Manuf.* **25**, 1349-1365 (2014).

88 Zhang, Z., Poudel, L., Sha, Z., Zhou, W. & Wu, D. Data-driven predictive modeling of tensile behavior of parts fabricated by cooperative 3d printing. *Journal of Computing and Information Science in Engineering* **20**, 021002 (2020).

89 Zhang, Z. *et al.* Predicting flexural strength of additively manufactured continuous carbon fiber-reinforced polymer composites using machine learning. *Journal of Computing and Information Science in Engineering* **20**, 061015 (2020).

90 Wang, Z. *et al.* Uncertainty Quantification in Metallic Additive Manufacturing Through Physics-Informed Data-Driven Modeling. *JOM* **71**, 2625-2634 (2019).

91 Liu, J., Chang, H., Hsu, T. & Ruan, X. Prediction of the flow stress of high-speed steel during hot deformation using a BP artificial neural network. *J. Mater. Process. Technol.* **103**, 200-205 (2000).

92 Toros, S. & Ozturk, F. Flow curve prediction of Al–Mg alloys under warm forming conditions at various strain rates by ANN. *Applied Soft Computing* **11**, 1891-1898 (2011).

93 Wang, Z.-L. & Adachi, Y. Property prediction and properties-to-microstructure inverse analysis of steels by a machine-learning approach. *Mater. Sci. Eng., A* **744**, 661-670 (2019).

94 Wei, H., Mazumder, J. & DebRoy, T. Evolution of solidification texture during additive manufacturing. *Sci. Rep.* **5** (2015).

95 Qiu, C. *et al.* Influence of Laser Processing Strategy and Remelting on Surface Structure and Porosity Development during Selective Laser Melting of a Metallic Material. *Metall. Mater. Trans. A* **50**, 4423-4434 (2019).

96 Sahoo, S. & Chou, K. Phase-field simulation of microstructure evolution of Ti–6Al–4V in electron beam additive manufacturing process. *Addit. Manuf.* **9**, 14-24, doi:http://dx.doi.org/10.1016/j.addma.2015.12.005 (2016).

97 Nie, P., Ojo, O. & Li, Z. Numerical modeling of microstructure evolution during laser additive manufacturing of a nickel-based superalloy. *Acta Mater.* **77**, 85-95 (2014).

98 Raghavan, N. *et al.* Numerical modeling of heat-transfer and the influence of process parameters on tailoring the grain morphology of IN718 in electron beam additive manufacturing. *Acta Mater.* **112**, 303-314, doi:10.1016/j.actamat.2016.03.063 (2016).

99 Liu, P. *et al.* Insight into the mechanisms of columnar to equiaxed grain transition during metallic additive manufacturing. *Addit. Manuf.* **26**, 22-29 (2019).

100 Zhang, Z., Yao, X. & Ge, P. Phase-field-model-based analysis of the effects of powder particle on porosities and densities in selective laser sintering additive manufacturing. *International Journal of Mechanical Sciences* **166**, 105230 (2020).

101 Liu, P. *et al.* Integration of phase-field model and crystal plasticity for the prediction of process-structure-property relation of additively manufactured metallic materials. *Int. J. Plast.*, 102670 (2020).

102 Zhang, W., Mehta, A., Desai, P. S. & III, C. F. H. in *2017 Solid Freeform Fabrication Symposium Proceedings.*

103 Scime, L. & Beuth, J. A multi-scale convolutional neural network for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process. *Addit. Manuf.* **24**, 273-286 (2018).

104 Zhou, Z. *et al.* Residual thermal stress prediction for continuous tool-paths in wire-arc additive manufacturing: a three-level data-driven method. *Virtual and Physical Prototyping* **17**, 105-124 (2022).

105 Wu, D., Xu, C. J. J. o. M. S. & Engineering. Predictive modeling of droplet formation processes in inkjet-based bioprinting. **140** (2018).

106 Wang, T., Kwok, T.-H., Zhou, C. & Vader, S. J. J. o. m. s. In-situ droplet inspection and closed-loop control system using machine learning for liquid metal jet printing. **47**, 83-92 (2018).

107 Al-Bermani, S., Blackmore, M., Zhang, W. & Todd, I. The origin of microstructural diversity, texture, and mechanical properties in electron beam melted Ti-6Al-4V. *Metall. Mater. Trans. A* **41**, 3422-3434 (2010).
108 Kok, Y. *et al.* Anisotropy and heterogeneity of microstructure and mechanical properties in metal additive manufacturing: A critical review. *Materials & Design* **139**, 565-586, doi:https://doi.org/10.1016/j.matdes.2017.11.021 (2018).
109 Wang, Z. *et al.* A data-driven approach for process optimization of metallic additive manufacturing under uncertainty. *J. Manuf. Sci. Eng.* **accepted** (2019).
110 Cang, R. *et al.* Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design. *J. Mech. Des.* **139** (2017).
111 Iyer, A., Dey, B., Dasgupta, A., Chen, W. & Chakraborty, A. A Conditional Generative Model for Predicting Material Microstructures from Processing Methods. *arXiv preprint arXiv:1910.02133* (2019).
112 Xu, H., Dikin, D. A., Burkhart, C. & Chen, W. Descriptor-based methodology for statistical characterization and 3D reconstruction of microstructural materials. *Comput. Mater. Sci.* **85**, 206-216 (2014).
113 Li, X. *et al.* A transfer learning approach for microstructure reconstruction and structure-property predictions. *Sci. Rep.* **8** (2018).
114 Tang, J. *et al.* Machine learning-based microstructure prediction during laser sintering of alumina. *Sci. Rep.* **11**, 1-10 (2021).
115 Tang, M., Pistorius, P. C. & Beuth, J. L. Prediction of lack-of-fusion porosity for powder bed fusion. *Addit. Manuf.* **14**, 39-48 (2017).
116 King, W. E. *et al.* Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing. *J. Mater. Process. Technol.* **214**, 2915-2925 (2014).
117 Dürr, H., Pilz, R. & Eleser, N. S. Rapid tooling of EDM electrodes by means of selective laser sintering. *Computers in Industry* **39**, 35-45 (1999).
118 Furumoto, T. *et al.* Permeability and strength of a porous metal structure fabricated by additive manufacturing. *J. Mater. Process. Technol.* **219**, 10-16, doi:https://doi.org/10.1016/j.jmatprotec.2014.11.043 (2015).
119 Basalah, A., Shanjani, Y., Esmaeili, S. & Toyserkani, E. Characterizations of additive manufactured porous titanium implants. *Journal of Biomedical Materials Research Part B: Applied Biomaterials* **100**, 1970-1979 (2012).
120 Paul, R., Anand, S. & Gerner, F. Effect of thermal deformation on part errors in metal powder based additive manufacturing processes. *J. Manuf. Sci. Eng.* **136** (2014).
121 Panda, B. N., Shankhwar, K., Garg, A. & Jian, Z. Performance evaluation of warping characteristic of fused deposition modelling process. *Int. J. Adv. Manuf. Technol.* **88**, 1799-1811 (2017).
122 Chowdhury, S. & Anand, S. in *ASME 2016 11th International Manufacturing Science and Engineering Conference.* (American Society of Mechanical Engineers Digital Collection).
123 Li, Z., Zhang, Z., Shi, J. & Wu, D. Prediction of surface roughness in extrusion-based additive manufacturing with machine learning. *Robotics and Computer-Integrated Manufacturing* **57**, 488-495 (2019).
124 Gan, Z. *et al.* Data-Driven Microstructure and Microhardness Design in Additive Manufacturing Using a Self-Organizing Map. *Engineering* **5**, 730-735, doi:https://doi.org/10.1016/j.eng.2019.03.014 (2019).
125 Ji, Y., Chen, L. & Chen, L.-Q. in *Thermo-Mechanical Modeling of Additive Manufacturing* 93-116 (Elsevier, 2018).
126 Yang, Z. *et al.* Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. *Comput. Mater. Sci.* **151**, 278-287 (2018).

127 Bao, H. *et al.* A machine-learning fatigue life prediction approach of additively manufactured metals. *Eng. Fract. Mech.* **242**, 107508, doi:https://doi.org/10.1016/j.engfracmech.2020.107508 (2021).
128 Mohamed, O. A., Masood, S. H. & Bhowmik, J. L. Analytical modelling and optimization of the temperature-dependent dynamic mechanical properties of fused deposition fabricated parts made of PC-ABS. *Materials* **9**, 895 (2016).
129 Koeppe, A., Padilla, C. A. H., Voshage, M., Schleifenbaum, J. H. & Markert, B. Efficient numerical modeling of 3D-printed lattice-cell structures using neural networks. *Manufacturing Letters* **15**, 147-150 (2018).
130 Muhammad, W., Brahme, A. P., Ibragimova, O., Kang, J. & Inal, K. A machine learning framework to predict local strain distribution and the evolution of plastic anisotropy & fracture in additively manufactured alloys. *Int. J. Plast.* **136**, 102867 (2021).
131 Rovinelli, A., Sangid, M. D., Proudhon, H. & Ludwig, W. Using machine learning and a data-driven approach to identify the small fatigue crack driving force in polycrystalline materials. *npj Comput. Mater.* **4**, 35 (2018).
132 Mangal, A. & Holm, E. A. Applied machine learning to predict stress hotspots I: Face centered cubic materials. *Int. J. Plast.* **111**, 122-134 (2018).
133 Frankel, A., Tachida, K. & Jones, R. Prediction of the evolution of the stress field of polycrystals undergoing elastic-plastic deformation with a hybrid neural network model. *Machine Learning: Science and Technology* **1**, 035005 (2020).
134 Kamble, S. S., Gunasekaran, A. & Gawankar, S. A. Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *International Journal of Production Economics* **219**, 179-194, doi:https://doi.org/10.1016/j.ijpe.2019.05.022 (2020).
135 O'Mara, J., Meredig, B. & Michel, K. Materials data infrastructure: a case study of the citrination platform to examine data import, storage, and access. *JOM* **68**, 2031-2034 (2016).
136 Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine* **29**, 141-142 (2012).
137 Deng, J. *et al.* in *2009 IEEE conference on computer vision and pattern recognition.* 248-255 (Ieee).
138 NIST. *Additive manufacturing benchmark test series (AM-Bench)*, <https://www.nist.gov/ambench> (accessed 2020).
139 Olleak, A. & Xi, Z. Calibration and validation framework for selective laser melting process based on multi-fidelity models and limited experiment data. *J. Mech. Des.* **142**, 081701 (2020).
140 Thuerey, N., Weißenow, K., Prantl, L. & Hu, X. Deep learning methods for Reynolds-averaged Navier–Stokes simulations of airfoil flows. *AIAA J.* **58**, 25-36 (2020).
141 Kashefi, A., Rempe, D. & Guibas, L. J. A Point-Cloud Deep Learning Framework for Prediction of Fluid Flow Fields on Irregular Geometries. *arXiv preprint arXiv:2010.09469* (2020).
142 Bhatnagar, S., Afshar, Y., Pan, S., Duraisamy, K. & Kaushik, S. Prediction of aerodynamic flow fields using convolutional neural networks. *Comput. Mech.* **64**, 525-545 (2019).
143 Hasegawa, K., Fukami, K., Murata, T. & Fukagata, K. Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes. *Theoretical and Computational Fluid Dynamics* **34**, 367-383 (2020).
144 Hasegawa, K., Fukami, K., Murata, T. & Fukagata, K. CNN-LSTM based reduced order modeling of two-dimensional unsteady flows around a circular cylinder at different Reynolds numbers. *Fluid Dynamics Research* **52**, 065501 (2020).

145 Lee, S. & You, D. Prediction of laminar vortex shedding over a cylinder using deep learning. *arXiv preprint arXiv:1712.07854* (2017).
146 Nie, Z., Jiang, H. & Kara, L. B. Stress Field Prediction in Cantilevered Structures Using Convolutional Neural Networks. *Journal of Computing and Information Science in Engineering* **20**, doi:10.1115/1.4044097 (2019).
147 Jiang, H., Nie, Z., Yeo, R., Farimani, A. B. & Kara, L. B. StressGAN: A Generative Deep Learning Model for Two-Dimensional Stress Distribution Prediction. *Journal of Applied Mechanics* **88**, 051005 (2021).
148 Swain, M. C. & Cole, J. M. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. *Journal of chemical information and modeling* **56**, 1894-1904 (2016).
149 Olivetti, E. A. *et al.* Data-driven materials research enabled by natural language processing and information extraction. *Applied Physics Reviews* **7**, 041317 (2020).
150 Sun, W. *et al.* Data processing and text mining technologies on electronic medical records: a review. *Journal of healthcare engineering* **2018** (2018).
151 Ma, B. *et al.* Data augmentation in microscopic images for material data mining. *npj Comput. Mater.* **6**, doi:10.1038/s41524-020-00392-6 (2020).
152 Gardner, J. M. *et al.* Machines as craftsmen: localized parameter setting optimization for fused filament fabrication 3D printing. *Advanced Materials Technologies* **4**, 1800653 (2019).
153 Colosimo, B. M., Huang, Q., Dasgupta, T. & Tsung, F. J. J. o. Q. T. Opportunities and challenges of quality engineering for additive manufacturing. **50**, 233-252 (2018).
154 Pan, S. J., Yang, Q. J. I. T. o. k. & engineering, d. A survey on transfer learning. **22**, 1345-1359 (2009).
155 Cheng, L., Wang, K. & Tsung, F. J. I. T. A hybrid transfer learning framework for in-plane freeform shape accuracy control in additive manufacturing. **53**, 298-312 (2020).
156 Sabbaghi, A. & Huang, Q. Model transfer across additive manufacturing processes via mean effect equivalence of lurking variables. *The Annals of Applied Statistics* **12**, 2409-2429 (2018).
157 Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nature Reviews Physics* **3**, 422-440 (2021).
158 Zhu, Y., Zabaras, N., Koutsourelakis, P.-S. & Perdikaris, P. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* **394**, 56-81 (2019).
159 Sun, L., Gao, H., Pan, S. & Wang, J.-X. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Comput. Methods Appl. Mech. Eng.* **361**, 112732 (2020).
160 Guo, S. *et al.* Machine learning for metal additive manufacturing: Towards a physics-informed data-driven paradigm. *Journal of Manufacturing Systems* **62**, 145-163, doi:https://doi.org/10.1016/j.jmsy.2021.11.003 (2022).
161 Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686-707 (2019).
162 Liu, R., Liu, S. & Zhang, X. A physics-informed machine learning model for porosity analysis in laser powder bed fusion additive manufacturing. *Int. J. Adv. Manuf. Technol.* **113**, 1943-1958 (2021).
163 Kapusuzoglu, B. & Mahadevan, S. Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication. *JOM* **72**, 4695-4705 (2020).

164 Du, Y., Mukherjee, T. & DebRoy, T. Physics-informed machine learning and mechanistic modeling of additive manufacturing to reduce defects. *Applied Materials Today* **24**, 101123 (2021).

165 Ren, Y., Wang, Q. & Michaleris, P. A Physics-Informed Two-Level Machine-Learning Model for Predicting Melt-Pool Size in Laser Powder Bed Fusion. *Journal of Dynamic Systems, Measurement, and Control* **143**, 121006 (2021).

166 Zobeiry, N. & Humfeld, K. D. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence* **101**, 104232 (2021).

167 Zhu, Q., Liu, Z. & Yan, J. Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks. *Comput. Mech.* **67**, 619-635 (2021).