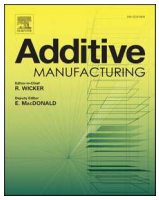
Additive Manufacturing 68 (2023) 103500

Contents lists available at ScienceDirect

Additive Manufacturing

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



Linking process parameters with lack-of-fusion porosity for laser powder bed fusion metal additive manufacturing

Satyajit Mojumder [a], Zhengtao Gan [b], Yangfan Li [c], Abdullah Al Amin [c], Wing Kam Liu [c,*]

[a] Theoretical and Applied Mechanics Program, Northwestern University, Evanston 60208, IL, USA
[b] Department of Aerospace and Mechanical Engineering, The University of Texas at El Paso, El Paso 79968, TX, USA
[c] Department of Mechanical Engineering, Northwestern University, Evanston 60208, IL, USA

ARTICLE INFO

Keywords:
Active learning
Laser powder bed fusion
Additive manufacturing
Lack-of-Fusion porosity
Symbolic regression

ABSTRACT

Structural defects such as porosity have detrimental effects on additively manufactured parts which can be reduced by choosing optimal process conditions. In this work, the relationship between process parameters and lack-of-fusion (LOF) porosity has been studied for the laser powder bed fusion (L-PBF) process of the Ti-6Al-4V alloy (Ti64). A physics-based thermo-fluid model is used to predict LOF porosity in the multilayer multitrack PBF process. To effectively map the high-dimensional processing parameters with porosity, an active learning framework has been adopted for the optimal design of experiments. Furthermore, a customized neural network-based symbolic regression tool has been utilized to identify a mechanistic relationship between processing conditions and LOF porosity. Results indicate that combining the physics-based thermo-fluid model for PBF porosity prediction with active learning and symbolic regression can find an appropriate mechanistic relationship of LOF porosity that is predictive for a wide range of processing conditions. This mechanistic relationship was further tested for other metal AM materials systems (IN718, SS316L) through non-dimensional numbers. The presented workflow effectively explores the high-dimensional process design space for different additive manufacturing materials systems.

1. Introduction

Additive manufacturing (AM), also known as 3D printing, is a highly promising technology for making complex parts by layer-wise deposition of materials [1]. A wide range of materials, including metals and alloys [2], ceramics [3], and polymers and their composites [4], have been tested for use in this disruptive manufacturing technology. Different techniques have been developed over the years to print metals and polymers, such as selective laser melting (SLM) [5–7], laser powder bed fusion (L-PBF) [8,9], direct energy deposition (DED) [10], fused deposition modeling (FDM) [11], etc. While the technology is becoming increasingly robust, the key challenge remains in qualifying and certifying parts, as the manufacturing process can lead to significant variability in material performance, such as fatigue [12–14]. Metal 3D printing processes such as L-PBF [15], SLM [16], and DED [17] often encounter defects such as porosity, spattering, and gas trapping, making it challenging to ensure the desired level of build quality necessary to be used in sophisticated applications like aviation and medical science [18].

Porosity is one of the common defects in additively manufactured parts due to the inherent nature of AM process. Porosity defects are common both in metal and polymer additive manufacturing [9,19]. Porosity can be attributed to several mechanisms, such as lack-of-fusion (LOF), vaporization of the materials, keyhole instability, powder shape, and size distribution. The L-PBF process is often plagued by LOF porosity, which occurs when powder particles fail to fully fuse together due to factors such as poor choice of processing conditions, low energy density, fast cooling rate, thick layers, and poor powder bed quality [20]. This lack of fusion is caused by a portion of the powder bed materials not reaching the melting temperature when considering a single layer thickness and applied laser power. Various methods can be implemented to combat this issue, such as optimizing process conditions, using high-quality powders, controlling layer thickness, developing advanced control systems, and post-processing [15]. In experiments, it can be difficult to separate LOF porosity from other mechanisms. However, it can be identified in thermo-fluid simulations of the additive manufacturing process [18] by identifying the regions under the melting temperature. These simulations, however, can be expensive to perform on a full-scale AM part. Quantifying the porosity

* Corresponding author.
E-mail address: w-liu@northwestern.edu (W.K. Liu).

https://doi.org/10.1016/j.addma.2023.103500
Received 22 November 2022; Received in revised form 6 March 2023; Accepted 13 March 2023
Available online 17 March 2023

Nomenclature

Symbols

h_c	convective heat transfer coefficient ($W{\cdot}m^{-2}{\cdot}K^{-1}$).
H_f	latent enthalpy of fusion ($kJ{\cdot}kg^{-1}$).
T_m	melting temperature (K).
c_p	specific heat($J{\cdot}kg^{-1}{\cdot}K^{-1}$).
f_l	volume fraction of liquid.
r_b	laser beam radius (m).
z_{top}	z-coordinate of the top surface (m).
$\hat{y}$	predicted output feature.
B	numerical parameter.
d	depth of the heat source (m).
g	gravitational acceleration ($m{\cdot}s^{-2}$).
h	enthalpy (J).
H	scan spacing (m).
k	thermal conductivity ($W{\cdot}m^{-1}{\cdot}K^{-1}$).
Ke	keyhole number.
L	layer thickness (m).
NED	normalized energy density.
p	dimensional pressure (Pa).
P	laser power (W).
Q	laser heat (W).
$RMSE$	root mean square error.
t	dimensional time (s).
T	temperature (K).
u	velocity ($m{\cdot}s^{-1}$).
Vs	scan speed ($m{\cdot}s^{-1}$).
x	coordinate system (m).
$\boldsymbol{X}$	input features vector (processing parameters).
$\boldsymbol{y}$	output feature (porosity).

Greek symbols

$\mathscr{L}$	labeled pool.
$\mathscr{U}$	unlabeled pool.
α	consolidation factor.
β	thermal expansion coefficient (K^{-1}).
γ	surface tension ($N{\cdot}m^{-1}$).
δ	dendritic spacing (m).
ε	emissivity.
η	absorptivity.
μ	dynamic viscosity (Pa·s).
ξ	volume fraction of consolidated domain.
ρ	density ($kg{\cdot}m^{-3}$).
σ	Stefan-Boltzmann constant ($W{\cdot}m^{-2}{\cdot}K^{-4}$).
ϕ	porosity.

Subscripts

bulk	bulk material properties.
i	variable index.
j	variable index.
k	variable index.
l	liquid phase.
powder	powder material properties.
ref	reference.
s	solid phase.

through simulation is challenging as it requires multilayer multitrack scanning of the laser over the powder bed. Experiments are also costly, and varying different process parameters (laser speed, laser power, hatch spacing, layer thickness, scan orientations, etc.) to understand their effects on the structure properties, such as LOF porosity, can become infeasible. While identifying the relationship between process parameters and LOF porosity is useful, it is a challenging problem due to the high dimensionality of the process design space and the large number of simulations or experiments required to evaluate all processing conditions [18,21]. In the current work, we aim to address these challenges by:

i. obtaining reliable LOF porosity data through compute-intensive physics-based thermo-fluid models that vary different process parameters,
ii. developing an optimal design of experiment (DOE) strategy to reduce the high-dimensional process design space of numerical and physical experiments,
iii. establishing a mechanistic relationship among process parameters and LOF porosity from fewer experiments and numerical simulations.

The thermo-fluid-based simulation data can be used as an alternative to experimental data if it is validated against experiments. With the advancement of the AM field, quite a few experimental studies are now available for several AM materials systems, such as Inconel 718 (IN718) [21,22], Stainless Steel (SS) [23–25], Ti64 alloy [20,26], and so on. These experimental study results can be used to calibrate the numerical simulation model. Recently, Gan et al. [27,28] developed an AM-CFD simulation tool that successfully predicted the melt pool dynamics of metal AM materials. AM-CFD has been validated against AFRL [28] and NIST [27] benchmark studies for single-layer and multilayer scanning in the PBF process. However, predicting LOF porosity using a thermo-fluid-based model requires calibration of the laser parameters. In this work, we extended the capability of the previously developed AM-CFD tool to predict LOF porosity and understand its relationship with processing conditions.

To understand the relationship between process parameters and their effects on the LOF porosity, many experiments and numerical simulations are required to effectively sample the high-dimensional process parameter space. The challenge is to learn this high-dimensional process parameter space relationship with LOF porosity through the optimal design of experiments. Active learning is a branch of machine learning that focuses on reducing the training dataset by sampling only the important instances that contain most of the information. The active learning process learns from previously sampled instances (also known as the labeled pool) and guides the next sampling points from the unlabeled pool [29,30]. Previously, Houtum and Vlesea [31] used an adaptive weighted uncertainty sampling-based active learning for an additive manufacturing image dataset from a DED process and reported more than 85% computational cost reduction. Teichert et al. [32] used an active learning workflow to learn the free energy derivative in a high-dimensional input space for a phase field model. Active learning techniques efficiently sample the process design space, reducing the number of simulations required to effectively map the process conditions to LOF porosity.

Once reliable porosity data is obtained, a regression tool is required to identify the underlying relationship between process parameters and LOF porosity. Regression is a widely used data science tool that uses linear and nonlinear functions to map input and output features. Recently, Xie et al. [33] used Convolutional Neural Networks (CNN) to establish a relationship between the thermal histories of the processing conditions in the AM process and printed materials properties such as ultimate tensile strength. Symbolic regression has an advantage over linear and nonlinear regression in the sense that it provides a closed-form expression to identify the relation between the input and output. Gan et al. [18] used genetic programming to learn a scaling law relationship between energy density and experimental porosity data for

different AM material systems. While many symbolic regression tools are available, deep learning-based symbolic regression has the benefit of using the universal approximation of the nonlinear functions and solving an optimization problem [34]. Neural network-based symbolic regression can be more powerful than genetic programming because neural networks are able to learn complex, nonlinear relationships in data, whereas genetic programming typically relies on brute-force search through a space of possible solutions [34,35]. Additionally, neural networks can be trained on much larger data sets and more easily fine-tuned, whereas genetic programming can struggle with scaling and require significant hand-tuning. Finally, neural networks can be more easily interpretable as they can be visualized as a set of layers and weights, whereas genetic programming solutions can be difficult to understand. However, neural network-based symbolic regression tends to have higher model complexity than genetic programming, which can make them more powerful but also more prone to over fitting. Genetic programming can be adjusted to control the complexity of the models generated but can be difficult to implement. To establish the relationship between processing conditions and LOF porosity, we propose a customized neural network architecture to identify the relationship through symbolic regression.

To address the computational challenges of relating the processing conditions of L-PBF to LOF porosity, the following contributions have been made in this work:

- The capability of a previously developed AM-CFD tool was extended to predict LOF porosity and understand its relationship with the processing conditions. This allows for more accurate predictions and an understanding of the factors that affect porosity in L-PBF, which is crucial for improving the quality and consistency of printed parts.
- An active learning-based sampling method was adopted to reduce the computational effort needed to evaluate the process design space. This approach efficiently samples the design space, which helps to keep the number of simulations required to a reasonable number, effectively mapping the process to LOF porosity while minimizing the computational burden.
- A customized neural network architecture was proposed to establish the relationship between processing conditions and LOF porosity through symbolic regression. This approach utilizes the universal approximation capabilities of neural networks, enabling the efficient modeling of the complex relationship between the process conditions and porosity.

This paper is organized as follows. Section 2 describes the thermo-fluid modeling and prediction of the LOF porosity. Section 3 shows the optimal design for the number of simulations using the active learning workflow. A customized neural network architecture for symbolic regression is described in Section 4. Section 5 shows the results with necessary interpretation and discussion. Finally, a conclusion section is provided in Section 6.

2. Thermal-fluid modeling for LOF porosity prediction

2.1. Additive manufacturing process modeling

To simulate the L-PBF process, a thermal-fluid model is used for multilayer simulation and prediction of LOF porosity. The thermal-fluid AM process model is a transient three-dimensional model that predicts the velocity field for the melt pool region and the thermal field for the entire part. The governing equations solve the mass, momentum, and energy conservation equations are given as follows:

$$\frac{\partial \rho}{\partial t}+\frac{\partial \rho u_i}{\partial x_i}=0 \tag{1}$$

$$\frac{\partial \rho u_i}{\partial t}+\frac{\partial \rho u_i u_j}{\partial x_j}=\frac{\partial}{\partial x_j}\left(\mu\left(\frac{\partial u_i}{\partial x_j}\right)\right)-\frac{\partial p}{\partial x_i}-\frac{180\mu}{\delta^2}\frac{(1-f_l)^2}{f_l^3+B}u_i+\rho_{ref}g_i\beta(T-T_{ref}) \tag{2}$$

$$\frac{\partial \rho h}{\partial t}+\frac{\partial \rho u_i h}{\partial x_i}=\frac{\partial}{\partial x_i}\left(k\frac{\partial T}{\partial x_i}\right)+Q \tag{3}$$

where ρ is the density, t is the time, u_i is the i th component of the velocity, and x_i is the local coordinates. μ, k, β are the viscosity, thermal conductivity, and thermal expansion coefficient, respectively. g_i is the acceleration of gravity and considered as 9.8 m/s^2 in the vertical direction. For this study, we set μ as a constant value and primary dendritic spacing is set to $1\mu m$. B is set to a value of 0.001 to avoid division by zero. The enthalpy (h) and temperature (T) are related by the following equation

$$\rho h=\int_0^T \rho c_p(T')dT'+\rho H_f f_l \tag{4}$$

where $c_p$is temperature-dependent heat capacity, H_f is the latent enthalpy of fusion, and f_l is the volume fraction of the liquid phase.

The heat transfer by laser (Q) into the materials system is modeled using a cylindrical shape heat source with Gaussian distribution as shown below:

$$Q=\left\{\begin{array}{l}\frac{2P\eta}{\pi r_b^2 d}exp\left(-\frac{2(x^2+y^2)}{r_b^2}\right);\ z_{top}-z\le d\\ 0;\ z_{top}-z>d\end{array}\right\} \tag{5}$$

where P is the laser power, η is the absorptivity, r_b is the laser beam radius, d is the depth of the heat source, and z_{top} is the z-coordinate of the top surface of the computational domain. In Eq. 5, we assume that the laser power is an important process parameter that varies from process to process, and η, r_b are parameters that remains fixed for a material system during printing. However, these parameters require calibration with the physical experiment described in Section 2.3.

The thermal boundary conditions are specified as

$$q_{ener}=-h_c(T-T_\infty)-\sigma\varepsilon(T^4-T_{ref}^4) \tag{6}$$

where $h_c, \sigma, \varepsilon, T_{ref}$ are the convective heat transfer coefficient, Stefan-Boltzmann constant, emissivity, and reference temperature, respectively. The boundary conditions for the momentum conservation equation (Eq. 2) are set as follows for the top surface:

$$\mu\frac{\partial u_1}{\partial z}=f_l\frac{d\gamma}{dT}\frac{\partial T}{\partial x} \tag{7a}$$

$$\mu\frac{\partial u_2}{\partial z}=f_l\frac{d\gamma}{dT}\frac{\partial T}{\partial y} \tag{7b}$$

$$u_3=0 \tag{7c}$$

where γ is the surface tension, $\frac{d\gamma}{dT}$ is the Marangoni coefficient, and u_i is the i th component of the velocity. To distinguish between the powder layer and consolidated layer, a consolidation factor (α) is defined as given below, and its values vary in the range between 0 (no consolidation) to 1 (fully consolidated).

$$\alpha=\frac{T_{peak}-T_s}{T_l-T_s} \tag{8}$$

The materials properties depend on this consolidated factor (state variable), and we consider a linear dependence of it on the effective material properties, shown in the following expression.

$$\lambda=\lambda_{bulk}\alpha+\lambda_{powder}(1-\alpha) \tag{9}$$

For this study, we considered the material properties of the Ti64 as

Table 1
Ti64 materials properties used for LOF porosity prediction [28].

Property	Ti 64
Solid density ($kg{\cdot}m^{-3}$)	4420
Liquid density ($kg{\cdot}m^{-3}$)	3920
Powder density ($kg{\cdot}m^{-3}$)	2210
Solidus temperature (K)	1878
Liquidus temperature (K)	1923
Boiling temperature (K)	3000
Solid specific heat capacity ($J{\cdot}kg^{-1}{\cdot}K^{-1}$)	0.2012 + 410.97 T
Liquid specific heat capacity ($J{\cdot}kg^{-1}{\cdot}K^{-1}$)	830
Solid thermal conductivity ($W{\cdot}m^{-1}{\cdot}K^{-1}$)	0.0118 + 4.243 T
Liquid thermal conductivity ($W{\cdot}m^{-1}{\cdot}K^{-1}$)	34.6
Powder thermal conductivity ($W{\cdot}m^{-1}{\cdot}K^{-1}$)	0.995 T
Latent heat of fusion ($kJ{\cdot}kg^{-1}$)	1126.6
Dynamic viscosity (Pa·s)	3.0×10^{-3}
Thermal expansion (K^{-1})	5.85×10^{-5}
Marangoni coefficient ($N{\cdot}m^{-1}{\cdot}K^{-1}$)	-2.6×10^{-4}
Convection coefficient ($W{\cdot}m^{-2}{\cdot}K^{-1}$)	5
Absorptivity	0.527
Emissivity	0.3

summarized in Table 1 [28].

2.2. Porosity prediction

LOF porosity in the L-PBF process can be attributed to interaction among different processing conditions such as lower power, higher scanning speed, higher layer thickness, and hatch spacing which does not allow the materials to absorb sufficient energy to fuse the powders. Here, we considered the powder layers as continuous media and tracked the region where the temperature history is below the melting temperature. Based on that, we identified the region for possible LOF porosity and quantified that using the following equation.

$$\phi = 1 - \xi \tag{10}$$

where ξ denotes the volume fraction of the domain with a temperature history above the melting temperature($T \geq T_m$).

In the thermo-fluid simulation model, multitrack and multilayer materials are printed. Fig. 1(a) and (c) show a prediction of the LOF porosity for a multitrack, single-layer printing using the AM-CFD tool. An experimental observation (see Fig. 1b) also shows similar LOF porosity for the unmelted powder in the bed. To reduce the computational effort, two layers are printed and then repeated to obtain the printed part height of 2.1 mm, which is of the same size as the experimental observation performed by du Plessis [20]. It is worth noting that the dwell time considered in printing each layer is sufficient to ensure the thermal stability of the melt pool, and printing more than two layers and repeating them will not change the results significantly. To ensure solidification of the deposited materials, a dwell time is provided following each laser scan to change the position of the laser. The temperature of the different layers is plotted against the dwell time in Fig. 1 (d) and a temperature evolution movie during the scan is provided as a supplementary material for the simulation set up of Fig. 1. The results indicate that the effect of residual heat on the subsequent layer can be disregarded if the dwell time is greater than 1 s. This is consistent with our model, which only considers two layers, as a sufficient dwell time of 8–10 s between scans allows for cooling of the substrate. Using two layers instead of one accommodates for variations in toolpaths between even and odd layers.

2.3. Model calibration and validation

As mentioned earlier, several printing parameters, such as absorptivity and laser spot radius, need to be calibrated to predict the experimental LOF porosity for a specific materials system. In this study, we calibrated our model for the Ti64 materials. The calibration has been performed against the experimental results published by du Plessis [20] for the Ti64 materials system. Note that the knowledge of stochastic calibration is applied in this work [37,38], in which the process parameters are not deterministic but follow a statistical distribution. The stochastic scheme can add uncertain information into melt pool

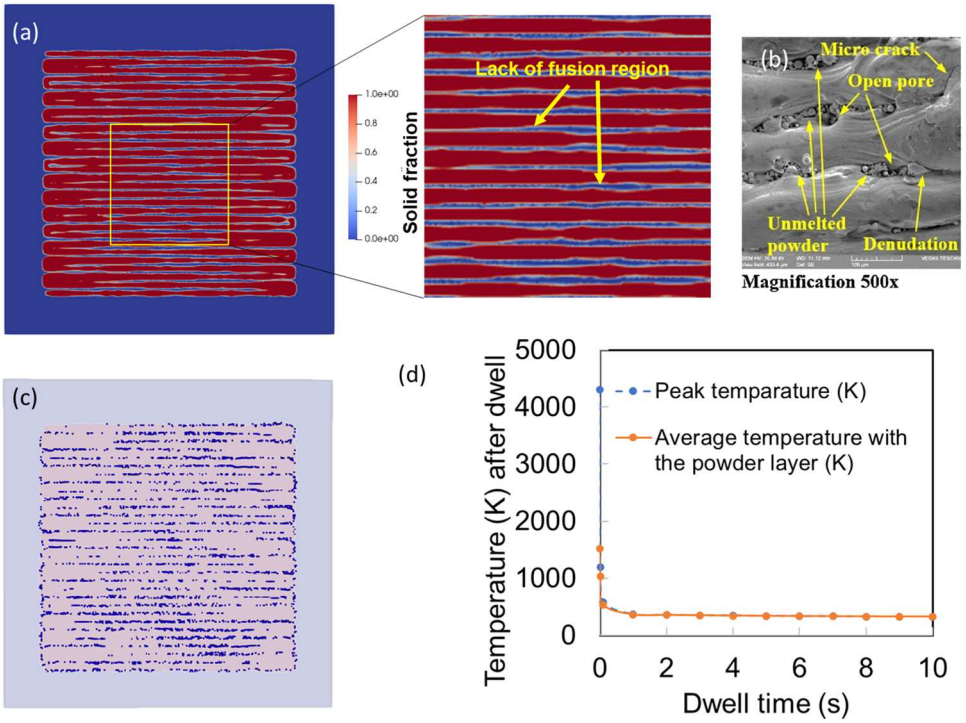


Fig. 1. (a) LOF porosity predicted using thermo-fluid modeling for Ti64 materials with laser power of 300 W, scan speed of 1230 mm/s, layer thickness of 0.04 mm, hatch spacing of 0.1 mm for 30 tracks, (b) experiment showing the unmelted powder region which turns into LOF porosity [36], (c) LOF porosity (on blue color) predicted using the multitrack simulations, (d) Variation of part temperature with dwell time.

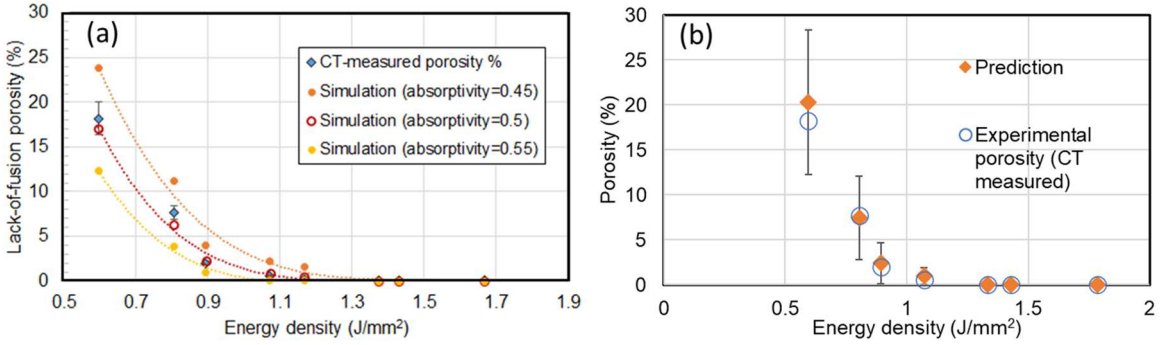


Fig. 2. (a) Calibration of the laser absorptivity parameter with the experimental data[20] (b) Comparison of the experimentally measured porosity with the predicted value.

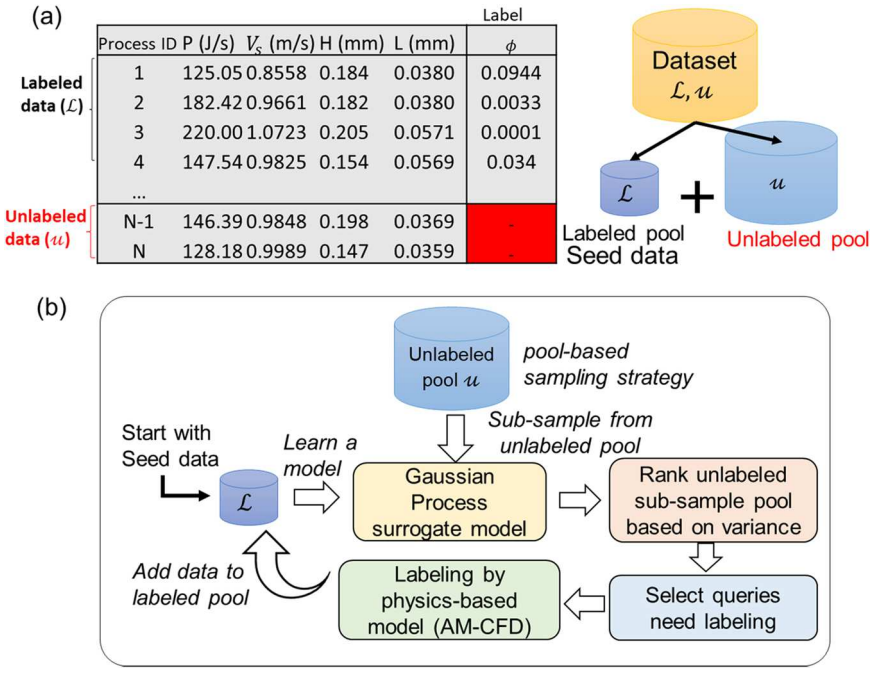


Fig. 3. (a) A sample design of experiment for the process parameters where the dataset is categorized as labeled and unlabeled pool. (b) Active learning workflow.

Table 2
DOE range of the process parameters.

Process parameters	Value
Laser power, P [J/s]	100–220
Scan speed, Vs [m/s]	0.8–1.2
Scan spacing, H [μm]	140–200
Layer thickness, L [μm]	30–60

geometry and help in the prediction of defects like porosity. The details of the stochastic calibration process are presented in Appendix A. The laser spot size (diameter) is considered as 50μm. We varied the absorptivity as shown in Fig. 2(a) and a value of 0.527 is chosen for the rest of the simulations.

It is important to note that the current numerical model has limitations. Specifically, this work uses a cylindrical heat source with gaussian distribution to model the heat source. While various heat source models, such as ellipsoid, double ellipsoid, parabolic, and cylindrical, have been proposed in the literature [39,40], none have been found to be entirely accurate without extensive calibration with experimental data. Our model, however, has undergone extensive calibration against the NIST AM bench [27,41] and AFRL AM modeling [28] challenges for predicting melt-pool shape in L-PBF processes. Additionally, the model does not consider multiphase or multi-species flow and assumes that materials are fully melted when their temperature exceeds the melting point, leading to a simplified mechanism of porosity formation. Furthermore, the materials properties used in the research are taken from literature and may vary considerably with temperature, which is assumed to be constant.

3. Active learning for optimal design of processing parameters

The PBF process of additive manufacturing involves several processing parameters that make the problem high dimensional. Obtaining the LOF porosity data by varying all these process parameters is quite expensive. Active learning can adaptively sample the process parameters from the DOE space and significantly reduce the required number of simulations or experiments for the porosity calculation. The details of the active learning workflow are shown in Fig. 3 and the steps are described below.

3.1. Dataset preparation

For the dataset preparation, important process parameters such as

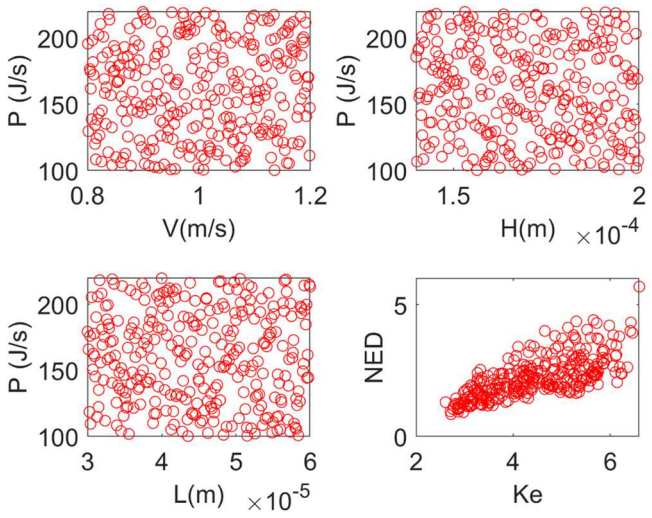


Fig. 4. Latin hypercube sampling of the process design space.

laser power, scan speed, scan spacing, and layer thickness are considered, and they are varied within the ranges shown in Table 2. Two non-dimensional numbers, normalized energy density (NED) and keyhole number (Ke), can be obtained using these processing conditions. The normalized energy density and keyhole number relate processing parameters and materials parameters, such as materials density, heat capacity, and liquidus temperature, and are expressed as follows:

$$NED = \frac{\eta_m P}{V_s H L}\frac{1}{\rho C_p (T_l - T_0)}, Ke = \frac{\eta_m P}{(T_l - T_0)\pi\rho C_p \sqrt{(\beta V_s r_0^3)}} \tag{11}$$

To generate the DOE for the process design space, the Latin hypercube sampling strategy is adopted. Latin hypercube sampling [42,43] is a constrained version of the Monte Carlo sampling scheme that yields more precise estimates than the Monte Carlo random sampling approach [38]. In Latin hypercube sampling, n different values from each of k variables $(X_1, \ldots, X_k)$ are sampled by dividing the range of each variable into n nonoverlapping intervals based on equal probability. From each interval, one value is selected randomly with respect to the probability mass of the interval. Then the n values selected for the variable X_1 are randomly combined with n values of variable X_2 and continued for the k number of variables. This process gives n k-dimensional input vectors, which can be considered as the $(n \times k)$ matrix shown in Fig. 3(a). For the current work, 300 sampling points are generated for the input process parameters as shown in Fig. 4. Initially, all these sampling points are part of our unlabeled pool of the dataset. Labeling means that the LOF porosity for the corresponding process parameters are known either by experiment or simulation. Ten experimental porosity data [20] are validated using our thermo-fluid model; the porosity data for those sampling points are known and included in the dataset as seed data. For these process parameters, the porosity data is known and termed the labeled pool. Therefore, the whole data set consists of two pools: the labeled pool $(\mathscr{L}(X_i, y))$, where the input process parameters and porosity are known, and the unlabeled pool $\mathscr{U}(X_i, ?)$ of input process parameters for which the label or porosity is unknown. For the unlabeled pool, AM-CFD simulations are performed to obtain the porosity label for a given set of processing parameters (from DOE). The DOE dataset and

Box 1
Algorithm of active learning for the process design space.

1: Start with $\mathcal{U}$, a pool of unlabeled samples $\mathcal{U}\{\boldsymbol{X}^{(\mathcal{U})}\}_{u=1}^{U}$

2: Find $\mathcal{L}\{\boldsymbol{X}, \boldsymbol{y}\}$, set of labeled (seed) samples $\mathcal{L}\{< \boldsymbol{X}, \boldsymbol{y} >^{l}\}_{l=1}^{n}$

3: **for** $i = 1{:}\mathcal{L}$ **do**

4: $\theta_i = \boldsymbol{train}(\mathcal{L})$;

5: **end for**

6: Choose a sub-sample $u\{\boldsymbol{X}^{(\mathcal{U})}\}_{u=1}^{m} \in \mathcal{U}$

7: Calculate variance $(\tilde{\sigma}_u^2)$ of sub-samples $u\{\boldsymbol{X}^{(\mathcal{U})}\}_{u=1}^{m}$ using model θ_i

8: Select p number of queries $\{\boldsymbol{X}^{*(\mathcal{U})}\}_{u=1}^{p} \in \mathcal{U}$, based on sorted variance (max to min), where $p \ll m$

9: Send queries $\{\boldsymbol{X}^{*(\mathcal{U})}\}_{u=1}^{p}$ to oracle for the labeling $\{\boldsymbol{y}^{*(\mathcal{U})}\}_{u=1}^{p}$

10: Add $< \{\boldsymbol{X}^{*(\mathcal{U})}\}_{u=1}^{p}, \{\boldsymbol{y}^{*(\mathcal{U})}\}_{u=1}^{p} >$ to $\mathcal{L}$

11: Evaluate performance metrics RMSE

12: Remove $\{\boldsymbol{X}^{*(\mathcal{U})}\}_{u=1}^{p}$ from $\mathcal{U}$

13: Repeat steps 3:12 till RMSE$\leq$ error limit

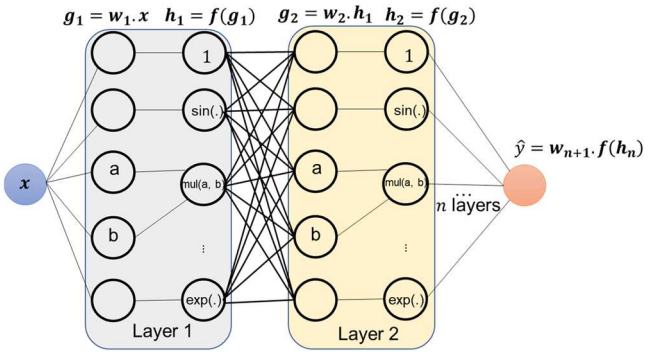


Fig. 5. Neural network architecture for the symbolic regression.

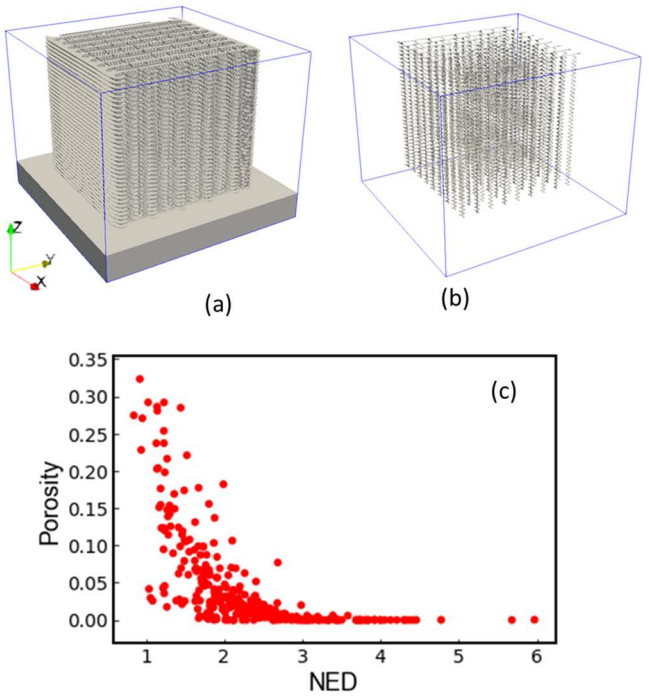


Fig. 6. (a) Printed part structure with laser power 120 W, scan speed 0.8 m/s, scan spacing 0.140 mm, and layer thickness of 0.03 mm. The predicted porosity of 5.8% volume fraction is shown in (b). (c) Variation of LOF porosity with NED for the Ti64 alloy.

their corresponding predicted porosity data is available in the GitHub link in the Data Availability section.

3.2. Gaussian process surrogate model

In the active learning framework, a Gaussian process (GP) surrogate model is fitted on the labeled dataset iteratively. The trained GP model is then utilized to predict the label for a sub-sample of the unlabeled data pool (sampled in a pool-based sampling strategy). The GP surrogate model has the inherent prediction of the variance of the unlabeled dataset along with its label, which is further used for the ranking and query selection of the unlabeled samples. More details on the GP surrogate model and prediction variance are provided in Appendix B.

The GP surrogate model is trained using the Gaussian process regression module [44] available in the scikit-learn machine learning tool using Python for the present work. The input dataset is four dimensional with four process parameters (laser power, scan speed, scan spacing, and layer thickness), and the output is the porosity. The hyperparameters of the GP surrogate model are chosen as $\sigma_f = 1$, and $l_1 = l_2 = l_3 = l_4 = 0.1$ initially and optimized during the training process for a Matern class covariance function. For an input set the GP surrogate model predicts the porosity label along with the variance of porosity prediction. The trained surrogate models for different seed data, sub-sample sizes, and queries are available in the GitHub link in the Data Availability section.

3.3. Pool-based sampling strategy, ranking of sub-sample, and quires selection

Based on the classification of the sampling strategy of the active learning process, we have used the pool-based sampling strategy. In pool-based sampling, a sub-sample from the unlabeled data pool (drawn randomly from the input space generated using Latin hypercube sampling) is used for the prediction of variance. This prediction is done using the GP surrogate model, trained on previously labeled samples as described in the earlier section. We select a sub-sample from the unlabeled pool of m samples and pass it to the GP surrogate model. The GP surrogate model predicts the porosity and associated variance for the input process parameters. Based on the variance of the predicted porosity of the sub-sample, they are ranked in descending order based on the predicted variance using Eq. B7.

3.4. Queries selection and labeling process

Based on the variance associated with the sub-sample of the unlabeled pool, a predefined number of samples (p) are selected based on the ranking of the variance

$$\{\boldsymbol{X}^{*(\mathscr{U})}\}_{\alpha=1}^{p} = X_k, k = \underset{k}{^{argmax}} \left\{ \tilde{\sigma}_{max,k}, k = 1, \ldots, p \right\} \tag{12}$$

This p number of unlabeled sub-samples with the most variance ($\{\boldsymbol{X}^{*(\mathscr{U})}\}_{\alpha=1}^{p} \in \mathscr{U}$) are the most promising candidate samples for the simulations of experiments that can add useful information on the process parameters and LOF porosity relationship.

Once the queries are selected for the labeling, these samples' process parameters are simulated using the thermo-fluid model described in Section 2. From the post-processing of the simulation results, the LOF porosity is predicted, and the unlabeled query sample is labeled for porosity. The labeled data is fed into the GP surrogate model in subsequent iteration and trained to have better predictability for the next sub-sample from the unlabeled pool.

3.5. Stopping criteria and metrics

The stopping criterion depends on the number of seed data available and how many instances are taken for each sub-sampling of the unlabeled pool. Based on that, the active learning process is iterated for the whole design of experiments covering the input process parameter space. We used the root mean square error (RMSE) as a metric to evaluate the performance of the active learning process in each subsequent iteration for a validation dataset of 50 samples. The RMSE is defined based on the true labels of porosity from experiments or simulations and the predicted labels from the GP surrogate model as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{f}_i - f_i\right)^2} \tag{13}$$

The details of the active learning algorithm are presented in Box 1.

4. Neural network-based symbolic regression

In this work, a customized neural network is used for symbolic regression to identify the relationship between process parameters and porosity. Instead of using the individual process parameters as input, the

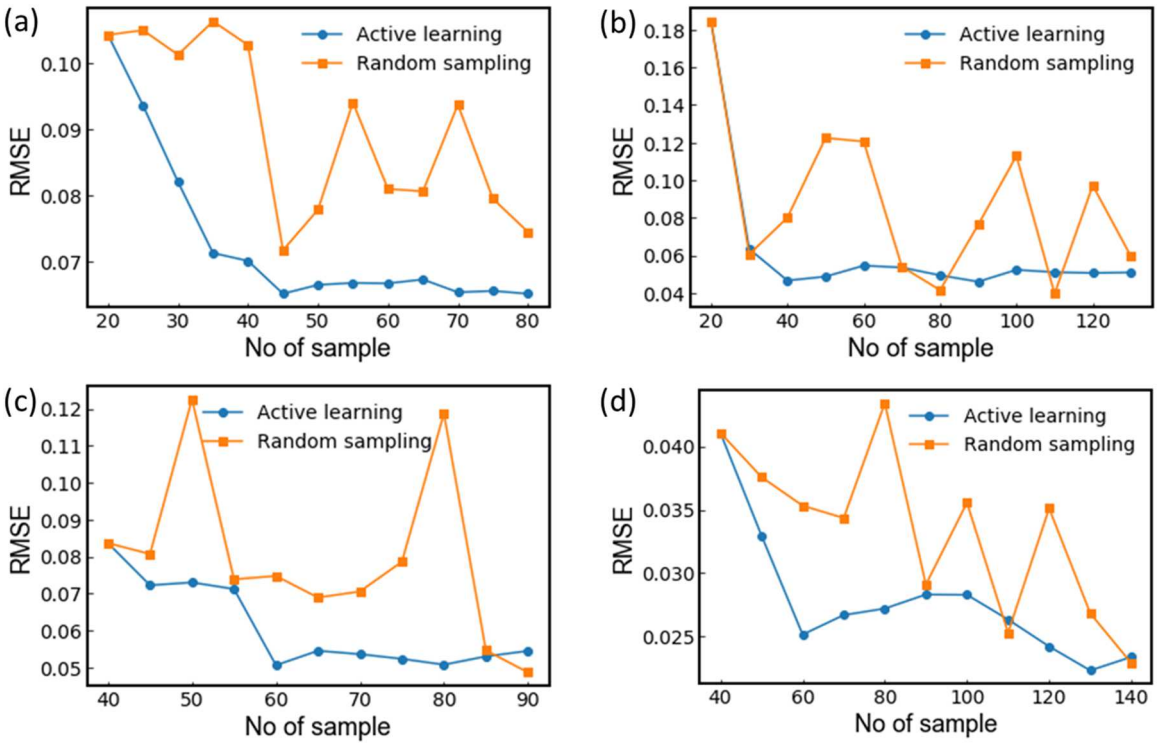


Fig. 7. GP model RMSE prediction on the validation set trained on (a) 20 seed data, 5 queries, (b) 20 seed data, 10 queries, (c) 40 seed data, 5 queries, (d) 40 seed data, 10 queries for a subsample size of 20 in each iteration.

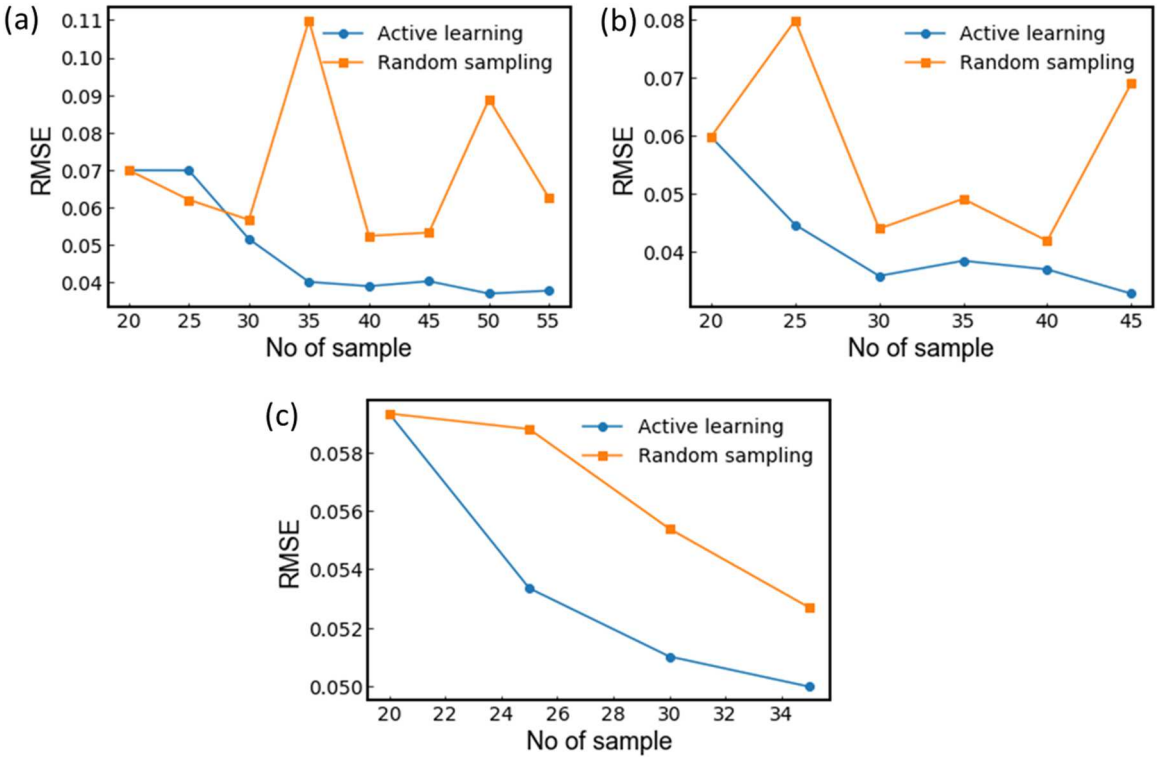


Fig. 8. GP model RMSE prediction on the validation set trained on 20 seed data, 5 queries for a subsample size of (a) 30, (b) 40, and (c) 60 in each iteration.

normalized energy density [18] is used to map the relationship. Using the normalized energy density (NED), the obtained process parameters to porosity relationship can be material independent as such materials parameters are captured in this non-dimensional number.

4.1. Neural network architecture

The neural network-based symbolic regression tool is based on a similar architecture proposed for the equation learner network [45,46].

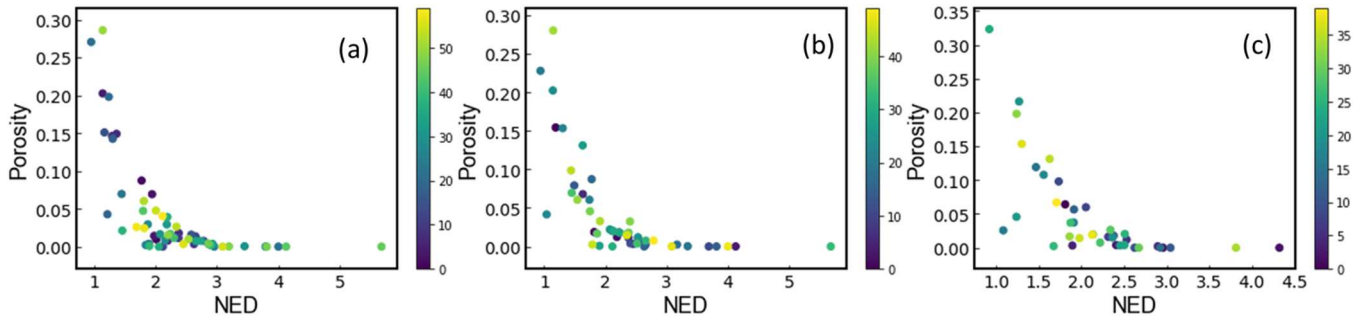


Fig. 9. Process design space learning pattern through active learning using 20 seed data, 5 queries for a subsample size of (a) 30, (b)40, and (c) 60 in each iteration. Legend on the right shows the data index and the first 20 samples are the seed data for each plot.

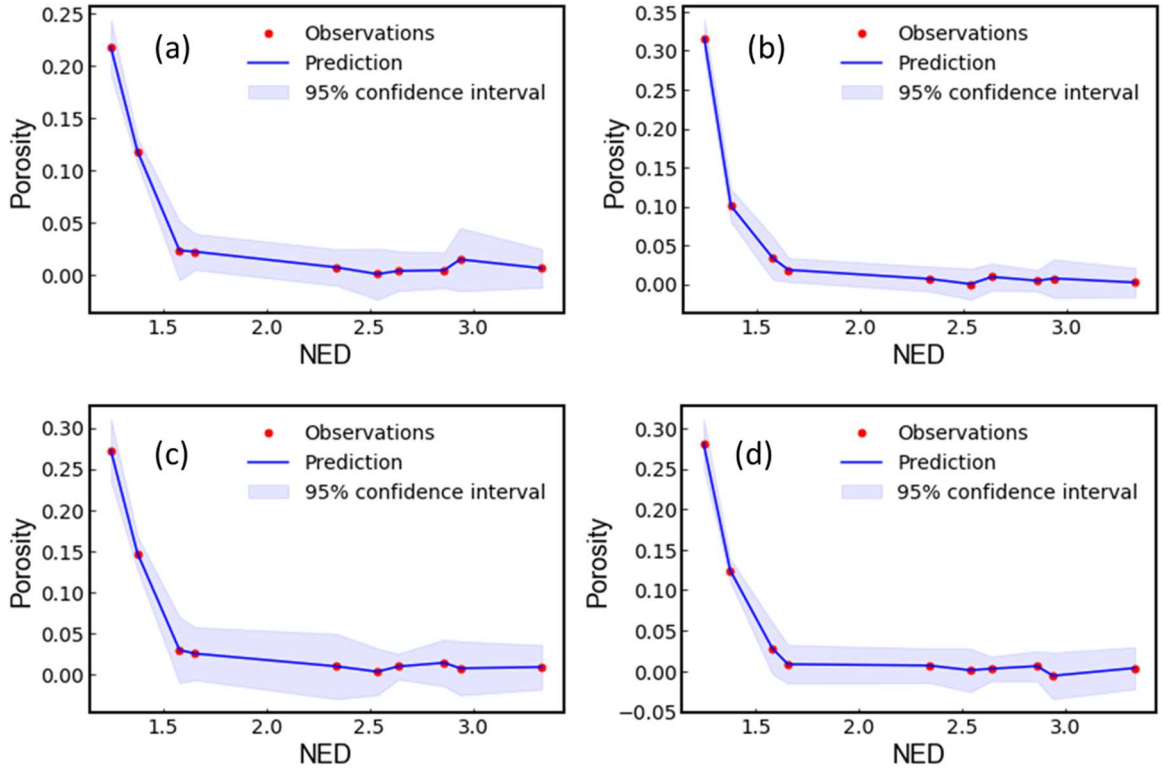


Fig. 10. GP surrogate model prediction with 95% confidence interval for a model trained on (a) 20 seed data, 5 queries, (b) 20 seed data, 10 queries, (c) 40 seed data, 5 queries, (d) 40 seed data, 10 queries for a subsample size of 20 in each iteration.

The general architecture of the neural network is shown in Fig. 5. The ith layer of this fully connected neural network is given by

$$\boldsymbol{g}_i = \boldsymbol{w}_i \boldsymbol{h}_{i-1}$$

$$\boldsymbol{h}_i = \boldsymbol{f}(\boldsymbol{g}_i) \tag{14}$$

where $\boldsymbol{w}_i$ is the weight matrix of the ith layer, and the input layer $\boldsymbol{x}$ can be obtained by $\boldsymbol{h}_0$. $\boldsymbol{f}(\boldsymbol{g})$ denotes the activation function which is different than the conventional machine learning activation functions. The output layers are given by the following relations

$$\widehat{\boldsymbol{y}} = \boldsymbol{h}_{n+1} = \mathbf{w}_{\mathbf{n+1}} \mathbf{h}_{\mathbf{n}} \tag{15}$$

The output layer does not have any activation function and is just a summation of the previous layer's output. In conventional machine learning, activation functions such as rectified linear unit (ReLU), sigmoid, tangent sigmoid, etc., are used. To construct the functional form of the input and output data, we customized the activation functions with several candidate mathematical functions such as sine, square, cubic, exponential, logarithmic, unity, etc. We also define operators, such as multiplication, in an activation function format that take input numbers and output the multiplied value of the two numbers. No additional bias term is defined as they can be absorbed into the $f(g)$, which is given by the following form:

$$f(g) = \begin{bmatrix} f_1(g_1) \\ f_2(g_2) \\ f_3(g_3) \\ \vdots \\ f_{nh}(g_{ng-1}, g_{ng}) \end{bmatrix} = \begin{bmatrix} 1 \\ \sin(.) \\ ()^3 \\ \vdots \\ (a \times b) \end{bmatrix} \tag{16}$$

These activation functions are the primitive functions that are the building blocks of the symbolic regression, and their combinations try to approximate the functions by minimizing the loss functions. Along with the functions shown in Fig. 5, we also considered sigmoid and exponential functions as the activation function for the current work. Same functions are repeated within the layer to reduce the sensitivity of the

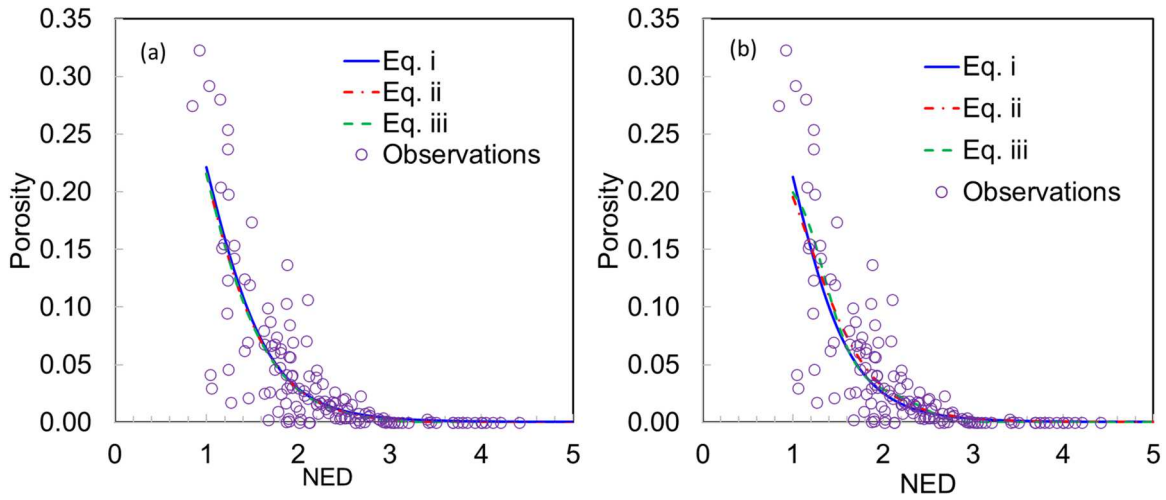


Fig. 11. Plot of the predicted expressions with the original dataset for (a) 20 seed data, (b) 40 seed data with a subsampling of 20, and queries of 10.

Table 3
Identified equations using the symbolic regression tools for the actively learned space with 20 and 40 seed data.

	Identified expression ($NED \leq 5$)	
Eq. no	Seed data= 20, Subsampling data= 20, query data= 10	Seed data= 40, Subsampling data= 20, query data= 10
i	$\phi = \frac{0.47}{0.07e^{9.46*\sin(0.3*NED)}+1}$	$\phi = \frac{0.46}{0.065e^{9.46*\sin(0.32*NED)}+1}$
ii	$\phi = 0.19\left(1-\sin\left(14.21+\frac{7.85}{e^{1.35*NED}+1}\right)\right)$	$\phi = \frac{0.30}{0.02e^{9.72*\sin(0.32*NED)}+1}$
iii	$\phi = 0.19*\sin\left(29.93+\frac{7.85}{e^{1.35*NED}+0.59}\right)$	$\phi = \frac{0.26}{274.23e^{-6.76\sin(1.19*NED)-1.88\sin(2.82*NED)}+1}$

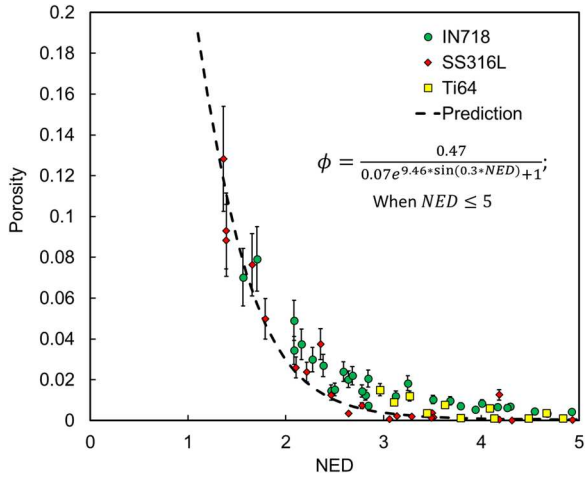


Fig. 12. Predictive ability of the proposed expression for different metal additive materials systems (IN718 [21,22], SS316L [23–25], Ti64 [20,26]).

random initializations. The number of hidden layers considered for this work is two, which can be increased for a more complex system. This is analogous to the tree depth used in the genetic programming approach for symbolic regression. For the given one-dimensional input of the normalized energy density with the porosity, three layers are good enough to predict the relationship. Sparsity is ensured by enforcing a $L_{0.5}$ regularization in the loss function. More details of the sparsity and regularization for the current network are discussed in Appendix C.

4.2. Training of neural network

Based on the prediction of the neural network, a mean square error loss function can be used with the regularization term as given below

$$\theta^* = \frac{1}{N}\sum(y_i - \hat{y}_i)^2 + \lambda L_{0.5} \tag{17}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and λ is a hyper-parameter that balances the mean squared error and the regularization term.

The training is performed in three stages. In the first stage, a small value of λ is selected, and the network evolves freely to calculate the latent parameter. In the second stage, λ is increased to enforce sparsity. In the third stage, a certain threshold (α) is selected, and the weights below that threshold are set as zero. In the final stage, the $L_{0.5}$ regularization term is dropped to fine-tune the weights of the model.

5. Results and discussion

The focus of this work is to predict the LOF porosity using a physics-based thermo-fluid model and reduce the number of simulations via an active learning framework. Further, a relationship between the processing conditions and porosity is established using a neural network-based symbolic regression tool. First, the porosity is predicted using the thermo-fluid model. Later the predicted porosity data for different processing conditions are used to show the efficacy of the active learning framework to reduce the number of simulations to effectively map the process design space. Finally, the actively learned samples are used to predict the processing conditions and porosity relationship.

5.1. Porosity prediction using physics-based thermo-fluid model

Section 2 presents the details of the thermo-fluid model used for the prediction of LOF porosity for varying processing conditions such as laser power, scan speed, hatch spacing, and layer thickness. Based on the thermo-fluid model, the melt pool due to the laser absorption can be

estimated, which in turn provides the melting region in the powder bed. If the laser power is low, in some scanning region the powder remains unmelted, introducing LOF porosity. In the process modeling, similar phenomena are observed when the process parameters are varied. For different processing conditions, a certain region of the track reaches below the melting point, as shown in Fig. 6(a). The unmelted region from the multilayer multitrack scanning corresponds to the shape of the structure shown in Fig. 6(b). The results shown in Figs. 6a and 6(b) are for laser power 120 W, scan speed 0.8 m/s, scan spacing 0.140 mm, layer thickness 0.03 mm, and a predicted porosity of 5.8% of the 2.1 mm printed cube structure. The porosity data for all samples was generated using a DOE of process parameters, as outlined in Section 3.1. The predicted porosity for each sample is displayed in relation to the non-dimensional number normalized energy density (NED) in Fig. 6(c). At a low NED, the porosity is high, and the primary mode of the porosity formation is a lack of fusion. As the NED increases, the porosity decreases significantly, and the conduction porosity mode becomes active. From the experimental data for various AM materials such as IN718 [21, 22], SS316L [23–25], and Ti64 [20,26], it has been observed that for NED values below 5, the pore formation mechanism primarily follows the lack of fusion mechanism. The predicted porosity from the thermo-fluid simulation also shows similar findings for the Ti64 alloy.

5.2. Active learning for the optimal design of experiments for simulation

The physics-based thermo-fluid simulation is expensive since a high-resolution mesh (~2.5 million finite volume cells) is needed to resolve the melt pool correctly and obtain the porosity data. Each of these simulations has a wall time of ~6 h in a single node with 24 processors using adaptive meshing. Therefore, for a high-dimensional process-design space, it is infeasible to simulate all the samples of the DOE. Using active learning techniques, an optimal DOE can be obtained with a fewer number of samples to simulate while achieving a similar level of confidence in mapping the process to porosity features.

The physics-based simulations have been performed for all the samples to obtain the ground truth, and then we show the efficacy of the active learning techniques based on the RMSE of the GP regression model fitted using the actively learned sample vs. the random sampling from the original DOE. The seed data, subsampling size, and the number of queries are varied to see their effects on the model performance using RMSE metrics. Fig. 7 shows the RMSE of the GP model (fitted on the labeled data after each iteration), varying the seed data from 20 to 40, and queries from 5 to 10 for a subsample size of 20. The active learning shows a lower RMSE value (in the validation set of 50 samples) compared to random sampling indicating that the GP model trained with actively learned samples performs better than with random samples. As the actively learned samples have higher uncertainty variance in the porosity prediction, it provides an effective way to train the model, which reflects in the RMSE of the validation set. Increasing the number of queries from 5 to 10 for the same seed data, sometimes the random sampling may perform better; however, it is not guaranteed. On the other hand, active learning samples always converged better compared to random sampling. As the number of seed data is increased from 20 to 40, the initial RMSE is decreased, which is expected (see Figs. 7c and d).

The effect of choosing different subsampling sizes is presented in Fig. 8. The number of seed data and the queries are fixed at 20 and 5, while the subsample size is varied to 30, 40, and 60. For the subsample sizes of 30 and 40, more iterations are required, and the model learns better from subsequent iterations. On the other hand, for a larger subsample size of 60, a smaller number of iterations are required for sampling, which results in a higher RMSE prediction. From the three parameters (seed data, subsample size, and queries), higher seed data (based on availability), moderate subsampling size, and fewer queries perform better for finding optimal samples of the process design space.

It is clear from the above discussion that active learning effectively samples the design space more efficiently than random sampling. Also, important information about the physics of the model can be captured by analyzing the learning pattern through the active learning process. Fig. 9 shows the learning pattern through active learning for different subsampling. As is evident from Fig. 9, the model effectively identifies the important region for sampling and captures the exponential decrease of the LOF porosity with the normalized energy density. The most uncertainty for the porosity can be found between a NED value of 1–3, and the active learning technique is able to capture the important physics by acquiring samples on these regions.

A Gaussian process (GP) surrogate model is trained as an active learning acquisition function. Following the iteration process of the active learning, the GP surrogate model can be used for predicting the porosity with uncertainty associated with the predictions. In Fig. 10, the predictive capability of the GP surrogate models trained with different seed data and queries is shown for a new set of process parameters. The trained GP surrogate models show similar predictions for these process parameters. One interesting observation is that the predicted uncertainties of all the models show higher error bounds near the elbow region of the NED vs. porosity curves. From Fig. 10(c), the elbow region shows higher uncertainty in the porosity predictions, which clearly reflects in the surrogate model prediction. Also, at higher NED, the error bound is higher due to the lower number of samples available for training. (Fig. 11).

5.3. Symbolic regression for process-structure relationship

Identifying a mechanistic relationship between the processing conditions (combined in NED) and the porosity can provide insight into choosing the right processing conditions for an AM process. Based on the actively learned samples with 20 and 40 seed data, a symbolic regression model is trained, and the predicted equations are presented in Table 3. Symbolic regression has been performed multiple times, and only the top 3 expressions with the lowest errors are shown in the Table. It should be noted that these expressions occurred most frequently over multiple regression trials. A plot of these equations with the original data is visualized in Fig. 12. While all these equations can predict the pattern of the LOF porosity, the first equation gives the simplest expression. We choose this simple expression (Eq. i) as the relationship between the normalized energy density and the porosity as it holds for both trials, each with different numbers of seed data. Further, the expressions are plotted against the observed porosity from thermo-fluid simulations to understand their predictive capabilities. All the expressions show similar predictions and correctly capture the physics of the LOF porosity formation mechanism.

In the identified expression through symbolic regression, the process parameters relate the LOF porosity in terms of the non-dimensional process parameters normalized energy density. The normalized energy density has both the processing conditions with the materials-dependent parameter. Therefore, the identified relationship can be interpreted as a material independent relationship. Fig. 12 shows the porosity data for IN718 [21,22], SS316L [23–25], and Ti64 [20,26] with NED. The experimental data from the literature are compared with our identified expressions for LOF porosity. The porosity prediction of the proposed expression is within the error bar found from those experimental data. Therefore, the proposed mechanistic expression holds for the metal AM system and is predictive of LOF porosity for a wide range of process parameters while $NED \leq 5$.

In this work, only the LOF porosity has been studied; therefore, all the mechanistic equations are only valid for such porosity. As the NED increases, the other mechanisms of porosity formation (e.g., conduction, keyhole porosity) manifest, which cannot be captured in the current physics-based model. Also, the choice of the NED≤5 as a LOF porosity mode is a rough estimate from the experimental observations. The transition of the porosity formation mechanisms with NED needs further investigation to identify a proper range. It should be noted that many mechanistic equations can be identified using the symbolic regression-

based approach and the identified expression is one of many possible cases demonstrated here.

6. Conclusions

The high dimensionality of the process design space for metal AM poses a challenge in determining the optimal processing windows to ensure printed part quality. Both experiment and computation can cause prohibitive costs in evaluating the process design space to identify a feasible processing window. In the present work, a physics-based thermo-fluid modeling approach is demonstrated to predict the LOF porosity and is validated against experiments. Further, the model is used to evaluate a design of experiment for processing conditions to predict porosity. The active learning technique has been utilized to reduce the numerical experiments and effectively map the process design space. Additionally, a neural network-based symbolic regression tool is used to identify a mechanistic relationship between the processing parameters and the porosity. The observations are summarized as follows:

i) The thermo-fluid model is predictive of the LOF porosity; however, at low NED, the uncertainty on the predicted porosity is higher due to the variation of processing conditions.
ii) Active learning effectively samples the process design space and outperforms random sampling of the design space. Seed data, subsampling size, and the number of queries are three important parameters for better sampling through active learning. Higher seed data, lower subsampling, and moderate query selection yield better performance for the active learning process.
iii) The neural network-based symbolic regression tool generates many possible expressions for the process-structure relations. The simpler expression is preferred over a complex expression with good predictive accuracy. The identified mechanistic expression is general enough that it holds for other metal AM systems as well.

Using the proposed mechanistic relation, the LOF porosity for a wide variety of metal AM systems can be predicted, saving a lot of experimental effort and cost. This also identifies the process-structure relationship of the metal AM system, and the framework can be extended to polymer and ceramic materials systems as well to identify the key governing mechanisms of porosity formation.

CRediT authorship contribution statement

Gan Zhengtao: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Mojumder Satyajit:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Li Yangfan:** Writing – review & editing, Writing – original draft, Methodology. **Al Amin Abdullah:** Writing – review & editing, Methodology. **Liu Wing Kam:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

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

Data Availability

The active learning code, relevant data, and trained models can be accessed through the Github link (https://github.com/Satyajit018/AM_active_learning).

Acknowledgments

Authors warmly acknowledge the funding provided by the United States National Science Foundation (NSF) under Grant No. MOMS/CMMI-1762035 and CMMI-1934367. The authors would like to thank the anonymous reviewers for their comments and suggestions to improve the paper. The authors also thank Derick Suarez and Stefan Knapik for their kind help in revising the manuscript.

Appendix A. Stochastic calibration for process parameters

The processing conditions of an AM process can vary during the experiment. To consider these stochastic variations in the processing parameters in physics-based thermo-fluid simulations, extensive calibrations should be performed considering stochasticity. Calibration of simulation model parameters is challenging, especially when the number of unknown parameters becomes important and local minima exist. To have an efficient calibration process, HOPGD [47,48], a non-intrusive data learning and constructing reduced order surrogate models, can be used to handle the model parameter calibration problem. In the thermo-fluid model, the inputs are heat source parameters $P1, P2, P3$ and energy density $e = \frac{P}{V}$, and melt pool width W and the melt pool depth D, are outputs. P *is the l*aser power and V is the scan speed. Based on HOPGD separation idea, surrogate model for stochastic AM simulation can be written as:

$$\boldsymbol{Y}_s = \begin{bmatrix} W_s \\ D_s \end{bmatrix} = \boldsymbol{F}(e, P1, P2, P3) = \sum_{m=1}^{k} \boldsymbol{F}_1^{(\mathrm{m})}(e)\boldsymbol{F}_2^{(\mathrm{m})}(P1)\boldsymbol{F}_3^{(\mathrm{m})}(P2)\boldsymbol{F}_4^{(\mathrm{m})}(P3) \tag{A1}$$

where $\boldsymbol{Y}_s$ represents the simulation output, W_s and D_s are the simulated width and depth, respectively. $\boldsymbol{F}_i$ (i = 1,., 4) are components for the HOPGD method, m is the mode number, and k is the total number of modes.

Difference between experimental and simulation parameters distribution ($f_{We}(\boldsymbol{x}), f_{De}(\boldsymbol{x}), f_{Ws}(\boldsymbol{x}), f_{Ds}(\boldsymbol{x})$) are estimated using kernel density estimate (KDE) [49] in the current model.

$Find f_{Ws(i)}(W), f_{Ds(i)}(D)$

$$\min \sum_{i=1}^{11} \int f_{We(i)}(W) \log \frac{f_{We(i)}(W)}{f_{Ws(i)}(W)} dW + \sum_{i=1}^{11} \int f_{De(i)}(D) \log \frac{f_{De(i)}(D)}{f_{Ds(i)}(D)} dD \tag{A2}$$

The distributions of experiment and simulation can be obtained through KDE:

$$f_{Ws(i)}(W) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{W - W_{sj}}{h}\right), f_{Ds(i)}(D) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{D - D_{sj}}{h}\right) f_{We(i)}(W) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{W - W_{ej}}{h}\right), f_{De(i)}(D) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{D - D_{ej}}{h}\right) \tag{A3}$$

Box B1
Algorithm for the stochastic simulation process.

1. Perform the post-processing of experimental data and find probability density functions (PDFs) of $f_{We}(\boldsymbol{x})$ and $f_{De}(\boldsymbol{x})$ of experimental width W_e and depth D_e.
2. Build HOPGD based surrogate model.
3. Set initial distributions of stochastic heat source input parameters and generate samples.
4. Use sampled $P1, P2, P3$ together with 11 energy density e as input to calculate points of simulation output width W_s and depth D_s using surrogate model.
5. Find PDFs of $f_{Ws}(\boldsymbol{x})$ and $f_{Ds}(\boldsymbol{x})$ of simulated W_s and D_s.
6. Calculate the error of KLD between experimental and simulated PDFs.
7. If the error is minimized, stop; else, go to step 8.
8. Update the input variables through method of moving asymptotes (MMA) optimization method.
9. Go back to step 3.

where i defines the $i-th$ melt pool experiment case $(i = 1, \ldots, 11)$. j is the number of measurement points for each melt pool case, K defines the Gaussian kernel function, n is the number of sample points, and h is the bandwidth. The processes involved in the stochastic simulation model are given in Box B1.

Appendix B. Gaussian process surrogate model

With the small amount of labeled dataset, a Gaussian process (GP) surrogate model is constructed to assess the labels of the unlabeled data pool and their variance. GP is a collection of random variables, and their finite collection has a multivariate normal distribution. Using the labeled data set we can construct a model $f(x)$ as follows:

$$f(x) \sim GP(\mu(x), k(x, x^{'})) \tag{B1}$$

where $\mu(x)$ is a mean function with covariance function $k(x,x^{'})$. We define a nonlinear model $g(x)$ that takes N inputs $\boldsymbol{X} = \{x_1, \ldots, x_N\}$ and predict the outputs $\boldsymbol{g} = \{g(x_1), \ldots, g(x_N)\}$. As the GP model we defined earlier is an interpolation model, there should be $f = g$ where $f(x_i) = g(x_i)$ for each input x_i.

Now, we take M inputs $X^* = \{x_1^*, \ldots, x_M^*\}$ for the predictions using our previously constructed GP surrogate model and the predictions are $f^* = \{f(x_1^*, \ldots, x_M^*\}$. Based on the definition, the finite sampling set and the predictions form a joint Gaussian distribution as follows

$$\mathscr{P}\begin{pmatrix} f \\ f^* \end{pmatrix} = \mathscr{N}\left(\begin{Bmatrix} \mu \\ \mu^* \end{Bmatrix}\begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}\right) \tag{B2}$$

The conditional distribution of the predictive set on the sampling set is a Gaussian distribution and can be obtained using the Bayes' rules as

$$\mathscr{P}(f^*|f) = \mathscr{N}(\widetilde{\mu}, \widetilde{K}) \tag{B3}$$

where $\widetilde{\mu} and \widetilde{K}$ is given using the following relations

$$\widetilde{\mu} = \mu^* + K(X^*,X)K^{-1}(f-\mu),$$

$$\widetilde{K} = K^* - K(X^*,X)K^{-1}K(X,X^*). \tag{B4}$$

The posterior of the GP model can be expressed as

$$f(x)|X, f \sim GP(\widetilde{\mu}(x)\widetilde{k}(x, x^{'})) \tag{B5}$$

where the $\widetilde{\mu}$ and $\widetilde{k}$ are the updated mean and covariance functions and expressed as following

$$\widetilde{\mu} = \mu(x) + K(x,X)K^{-1}(f-\mu),$$

$$\widetilde{k} = k(x,x^{'}) - K(x,X)K^{-1}K(X,x^{'}). \tag{B6}$$

For the prediction of new arbitrary input x^*, the prediction $f(x^*)$ can be shown as

$$\mathscr{P}(f|x^*, X, f) = \mathscr{N}(\widetilde{\mu}(x^*), \widetilde{\sigma}^2(x^*)) \tag{B7}$$

which is a gaussian distribution with variance, $\widetilde{\sigma}^2(x^*) = k(x^*, x^*)$.

For the implementation of the GP model, a constant zero value is assumed for the mean function $\mu(x)$ which gives the whole GP model defined based on the covariance function $k(x, x^{'})$. In practice, different functions are used for estimating the covariance function such as the radial basis function. The radial basis function is a squared exponential function, and it is very smooth. The smoothness of our data is not ensured; therefore, we choose a Matern class covariance function which is defined as the following:

$$k(r) = \sigma_f \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} r\right)^{\nu} K_{\nu}(\sqrt{2\nu} r) \tag{B8}$$

where σ_f is a multiplier, $\Gamma()$ is the Gamma function, K_{ν} is the modified Bessel function for a positive number ν and r is the distance between two inputs x and x' given as

$$r(x, x^{'}) = \sqrt{\sum_{i}^{d} \left(\frac{x_i - x_i^{'}}{l_i}\right)^2} \tag{B9}$$

where d is the number of the input dimension and l_i is the length scale in each dimension. The length scale controls the correlation of the approximation functions.

To train the surrogate model, the hyperparameters $\theta = \{\sigma_f, l_1, \ldots, l_d\}$ are defined and calibrated by maximizing the log marginal likelihood $\log \mathscr{P}(f|X)$.

Appendix C. Sparsity and regularization

The customized neural network activation functions for each node of a layer turn off and on for different combinations of the functions to fit the input and output data. Therefore, sparsity is a key element of this neural network. In the training process, the goal is to set as many zero weights parameters as possible by enforcing the sparsity in the neural network. One of the simplest approaches to enforce sparsity is adding a regularization term in the loss function of the neural network. The regularization term is a function of the weight matrices of the neural network and is given as

$$L_q = \sum_{i=0}^{n+1} \|W_i\|^q \tag{C1}$$

where $\|W_i\|^q$ is the element-wise norm of the matrix and q sets the norm of the matrix. The element-wise norm can be approximated as

$$\|W_i\|^q = \sum_{j,k} \left|w_{j,k}\right|^q \tag{C2}$$

Setting the value of $q = 0$, gives the L_0 regularization which ensures the sparsity by penalizing the non-zero weights regardless of their magnitude. While L_0 regularization pushes the solution towards sparsity, it is not compatible with the gradient descent optimization used in the optimization of the neural networks. Previously, Lampert et al. [45] used L_1 regularization to make the optimization problem convex and push the solution towards sparsity. Recently, Kim et al. [34] proposed a modified $L_{0.5}$ regularization that performs better than L_1 regularization while being compatible with the gradient descent based optimization algorithm. The modified $L_{0.5}$ regularization can be expressed as

$$L_{0.5}(w) = \begin{cases} |w|^{1/2} & |w| \geq a \\ \left(-\frac{w^4}{8a^3} + \frac{3w^2}{4a} + \frac{3a}{8}\right)^{1/2} & |w| < a \end{cases} \tag{C3}$$

where w is the weight and a is a constant value.

Appendix D. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.addma.2023.103500.

References

[1] M. Markl, C. Körner, Powder layer deposition algorithm for additive manufacturing simulations, Powder Technol. 330 (2018) 125–136, https://doi.org/10.1016/j.powtec.2018.02.026.

[2] L.E. Murr, A metallographic review of 3D printing/additive manufacturing of metal and alloy products and components, Metallogr. Microstruct. Anal. 7 (2018) 103–132, https://doi.org/10.1007/s13632-018-0433-6.

[3] Z. Chen, Z. Li, J. Li, C. Liu, C. Lao, Y. Fu, C. Liu, Y. Li, P. Wang, Y. He, 3D printing of ceramics: a review, J. Eur. Ceram. Soc. 39 (2019) 661–687, https://doi.org/10.1016/j.jeurceramsoc.2018.11.013.

[4] S. Yuan, F. Shen, C.K. Chua, K. Zhou, Polymeric composites for powder-based additive manufacturing: materials and applications, Prog. Polym. Sci. 91 (2019) 141–168, https://doi.org/10.1016/j.progpolymsci.2018.11.001.

[5] C. Wei, L. Li, X. Zhang, Y.-H. Chueh, 3D printing of multiple metallic materials via modified selective laser melting, CIRP Ann. 67 (2018) 245–248, https://doi.org/10.1016/j.cirp.2018.04.096.

[6] N.T. Aboulkhair, M. Simonelli, L. Parry, I. Ashcroft, C. Tuck, R. Hague, 3D printing of aluminium alloys: additive manufacturing of aluminium alloys using selective laser melting, Prog. Mater. Sci. 106 (2019), 100578, https://doi.org/10.1016/j.pmatsci.2019.100578.

[7] M. Drexler, M. Lexow, D. Drummer, Selective laser melting of polymer powder – part mechanics as function of exposure speed, Phys. Procedia 78 (2015) 328–336, https://doi.org/10.1016/j.phpro.2015.11.047.

[8] S. Sun, M. Brandt, M. Easton, 2 - Powder bed fusion processes: an overview, in: M. Brandt (Ed.), Laser Additive Manufacturing, Woodhead Publishing, 2017, pp. 55–77, https://doi.org/10.1016/B978-0-08-100433-3.00002-6.

[9] E.D. Bain, E.J. Garboczi, J.E. Seppala, T.C. Parker, K.B. Migler, AMB2018-04: benchmark physical property measurements for powder bed fusion additive manufacturing of polyamide 12, Integr. Mater. Manuf. Innov. 8 (2019) 335–361, https://doi.org/10.1007/s40192-019-00146-3.

[10] W.J. Oh, W.J. Lee, M.S. Kim, J.B. Jeon, D.S. Shim, Repairing additive-manufactured 316L stainless steel using direct energy deposition, Opt. Laser Technol. 117 (2019) 6–17, https://doi.org/10.1016/j.optlastec.2019.04.012.
[11] D.P. Cole, F. Gardea, T.C. Henry, J.E. Seppala, E.J. Garboczi, K.D. Migler, C. M. Shumeyko, J.R. Westrich, S.V. Orski, J.L. Gair, AMB2018-03: benchmark physical property measurements for material extrusion additive manufacturing of polycarbonate, Integr. Mater. Manuf. Innov. 9 (2020) 358–375, https://doi.org/10.1007/s40192-020-00188-y.
[12] T.H. Becker, P. Kumar, U. Ramamurty, Fracture and fatigue in additively manufactured metals, Acta Mater. 219 (2021), 117240, https://doi.org/10.1016/j.actamat.2021.117240.
[13] P. Kumar, U. Ramamurty, High cycle fatigue in selective laser melted Ti-6Al-4V, Acta Mater. 194 (2020) 305–320, https://doi.org/10.1016/j.actamat.2020.05.041.
[14] P. Kumar, R. Jayaraj, J. Suryawanshi, U.R. Satwik, J. McKinnell, U. Ramamurty, Fatigue strength of additively manufactured 316L austenitic stainless steel, Acta Mater. 199 (2020) 225–239, https://doi.org/10.1016/j.actamat.2020.08.033.
[15] T. Mukherjee, T. DebRoy, Mitigation of lack of fusion defects in powder bed fusion additive manufacturing, J. Manuf. Process. 36 (2018) 442–449, https://doi.org/10.1016/j.jmapro.2018.10.028.
[16] S. Coeck, M. Bisht, J. Plas, F. Verbist, Prediction of lack of fusion porosity in selective laser melting based on melt pool monitoring data, Addit. Manuf. 25 (2019) 347–356, https://doi.org/10.1016/j.addma.2018.11.015.
[17] L.E. dos Santos Paes, M. Pereira, F.A. Xavier, W.L. Weingaertner, L.O. Vilarinho, Lack of fusion mitigation in directed energy deposition with laser (DED-L) additive manufacturing through laser remelting, J. Manuf. Process. 73 (2022) 67–77, https://doi.org/10.1016/j.jmapro.2021.10.052.
[18] Z. Gan, O.L. Kafka, N. Parab, C. Zhao, L. Fang, O. Heinonen, T. Sun, W.K. Liu, Universal scaling laws of keyhole stability and porosity in 3D printing of metals, Nat. Commun. 12 (2021) 2379, https://doi.org/10.1038/s41467-021-22704-0.
[19] J. Ning, W. Wang, B. Zamorano, S.Y. Liang, Analytical modeling of lack-of-fusion porosity in metal additive manufacturing, Appl. Phys. A 125 (2019) 797, https://doi.org/10.1007/s00339-019-3092-9.
[20] A. du Plessis, Effects of process parameters on porosity in laser powder bed fusion revealed by X-ray tomography, Addit. Manuf. 30 (2019), 100871, https://doi.org/10.1016/j.addma.2019.100871.
[21] Z. Wang, M. Liu, Dimensionless analysis on selective laser melting to predict porosity and track morphology, J. Mater. Process. Technol. 273 (2019), 116238, https://doi.org/10.1016/j.jmatprotec.2019.05.019.
[22] P. Kumar, J. Farah, J. Akram, C. Teng, J. Ginn, M. Misra, Influence of laser processing parameters on porosity in Inconel 718 during additive manufacturing, Int J. Adv. Manuf. Technol. 103 (2019) 1497–1507, https://doi.org/10.1007/s00170-019-03655-9.
[23] J.A. Cherry, H.M. Davies, S. Mehmood, N.P. Lavery, S.G.R. Brown, J. Sienz, Investigation into the effect of process parameters on microstructural and physical properties of 316L stainless steel parts by selective laser melting, Int J. Adv. Manuf. Technol. 76 (2015) 869–879, https://doi.org/10.1007/s00170-014-6297-2.
[24] A. Leicht, M. Rashidi, U. Klement, E. Hryha, Effect of process parameters on the microstructure, tensile strength and productivity of 316L parts produced by laser powder bed fusion, Mater. Charact. 159 (2020), 110016, https://doi.org/10.1016/j.matchar.2019.110016.
[25] J.C. Simmons, X. Chen, A. Azizi, M.A. Daeumer, P.Y. Zavalij, G. Zhou, S. N. Schiffres, Influence of processing and microstructure on the local and bulk thermal conductivity of selective laser melted 316L stainless steel, Addit. Manuf. 32 (2020), 100996, https://doi.org/10.1016/j.addma.2019.100996.
[26] G. Kasperovich, J. Haubrich, J. Gussone, G. Requena, Correlation between porosity and processing parameters in TiAl6V4 produced by selective laser melting, Mater. Des. 105 (2016) 160–170, https://doi.org/10.1016/j.matdes.2016.05.070.
[27] Z. Gan, Y. Lian, S.E. Lin, K.K. Jones, W.K. Liu, G.J. Wagner, Benchmark study of thermal behavior, surface topography, and dendritic microstructure in selective laser melting of inconel 625, Integr. Mater. Manuf. Innov. 8 (2019) 178–193, https://doi.org/10.1007/s40192-019-00130-x.
[28] Z. Gan, K.K. Jones, Y. Lu, W.K. Liu, Benchmark study of melted track geometries in laser powder bed fusion of inconel 625, Integr. Mater. Manuf. Innov. 10 (2021) 177–195, https://doi.org/10.1007/s40192-021-00209-4.
[29] P. Kumar, A. Gupta, Active learning query strategies for classification, regression, and clustering: a survey, J. Comput. Sci. Technol. 35 (2020) 913–945, https://doi.org/10.1007/s11390-020-9487-4.
[30] B. Settles, Active Learning Literature Survey, University of Wisconsin-Madison Department of Computer Sciences, 2009. ⟨https://minds.wisconsin.edu/handle/1793/60660⟩ (Accessed 19 December 2021).
[31] G.J.J. van Houtum, M.L. Vlasea, Active learning via adaptive weighted uncertainty sampling applied to additive manufacturing, Addit. Manuf. 48 (2021), 102411, https://doi.org/10.1016/j.addma.2021.102411.
[32] G.H. Teichert, A.R. Natarajan, A. Van der Ven, K. Garikipati, Scale bridging materials physics: active learning workflows and integrable deep neural networks for free energy function representations in alloys, Comput. Methods Appl. Mech. Eng. 371 (2020), 113281, https://doi.org/10.1016/j.cma.2020.113281.
[33] X. Xie, J. Bennett, S. Saha, Y. Lu, J. Cao, W.K. Liu, Z. Gan, Mechanistic data-driven prediction of as-built mechanical properties in metal additive manufacturing, Npj Comput. Mater. 7 (2021) 1–12, https://doi.org/10.1038/s41524-021-00555-z.
[34] S. Kim, P.Y. Lu, S. Mukherjee, M. Gilbert, L. Jing, V. Čeperić, M. Soljačić, Integration of neural network-based symbolic regression in deep learning for scientific discovery, IEEE Trans. Neural Netw. Learn. Syst. 32 (2021) 4166–4177, https://doi.org/10.1109/TNNLS.2020.3017010.
[35] I. Icke, J.C. Bongard, Improving genetic programming based symbolic regression using deterministic machine learning, 2013 IEEE Congr. Evolut. Comput. (2013) 1763–1770, https://doi.org/10.1109/CEC.2013.6557774.
[36] M. Yakout, M.A. Elbestawi, S.C. Veldhuis, A study of thermal expansion coefficients and microstructure during selective laser melting of Invar 36 and stainless steel 316L, Addit. Manuf. 24 (2018) 405–418, https://doi.org/10.1016/j.addma.2018.09.035.
[37] Y. Li, Y. Lu, A.A. Amin, W.K. Liu, Stochastic additive manufacturing simulation: from experiment to surface roughness and porosity prediction, 2022. https://doi.org/10.48550/arXiv.2208.02907.
[38] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (1979) 239–245, https://doi.org/10.2307/1268522.
[39] S. Hocine, H. Van Swygenhoven, S. Van Petegem, Verification of selective laser melting heat source models with operando X-ray diffraction data, Addit. Manuf. 37 (2021), 101747, https://doi.org/10.1016/j.addma.2020.101747.
[40] M.I. Al Hamahmy, I. Deiab, Review and analysis of heat source models for additive manufacturing, Int J. Adv. Manuf. Technol. 106 (2020) 1223–1238, https://doi.org/10.1007/s00170-019-04371-0.
[41] A.A. Amin, Y. Li, Y. Lu, X. Xie, Z. Gan, S. Mojumder, G.J. Wagner, W.K. Liu, Physics guided heat source for quantitative prediction of the laser track measurements of IN718 in 2022 NIST AM benchmark test, Addit. Manuf. Submitt. (2023).
[42] F.A.C. Viana, A tutorial on latin hypercube design of experiments, Qual. Reliab. Eng. Int. 32 (2016) 1975–1985, https://doi.org/10.1002/qre.1924.
[43] F.A.C. Viana, G. Venter, V. Balabanov, An algorithm for fast optimal Latin hypercube design of experiments, Int. J. Numer. Methods Eng. 82 (2010) 135–156, https://doi.org/10.1002/nme.2750.
[44] 1.7. Gaussian Processes, Scikit-Learn. (n.d.). ⟨https://scikit-learn/stable/modules/gaussian_process.html⟩ (Accessed 20 December 2021).
[45] G. Martius, C.H. Lampert, Extrapolation and learning equations, ArXiv:1610.02995 [Cs]. (2016). ⟨http://arxiv.org/abs/1610.02995⟩ (Accessed 20 December 2021).
[46] S.S. Sahoo, C.H. Lampert, G. Martius, Learning Equations for Extrapolation and Control, ArXiv:1806.07259 [Cs, Stat]. (2018). ⟨http://arxiv.org/abs/1806.07259⟩ (accessed December 20, 2021).
[47] Y. Lu, N. Blal, A. Gravouil, Adaptive sparse grid based HOPGD: toward a nonintrusive strategy for constructing space-time welding computational vademecum, Int. J. Numer. Methods Eng. 114 (2018) 1438–1461, https://doi.org/10.1002/nme.5793.
[48] Y. Lu, N. Blal, A. Gravouil, Datadriven HOPGD based computational vademecum for welding parameter identification, Comput. Mech. 64 (2019) 47–62, https://doi.org/10.1007/s00466-018-1656-8.
[49] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Stat. 27 (1956) 832–837, https://doi.org/10.1214/aoms/1177728190.