01 Oct 2022

Predicting Defects in Laser Powder Bed Fusion using In-Situ Thermal Imaging Data and Machine Learning

Sina Malakpour Estalaki

Cody S. Lough

Robert G. Landers
Missouri University of Science and Technology, landersr@mst.edu

Edward C. Kinzel
Missouri University of Science and Technology, kinzele@mst.edu

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/mec_aereng_facwork/4925

Recommended Citation

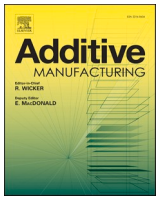
S. M. Estalaki et al., "Predicting Defects in Laser Powder Bed Fusion using In-Situ Thermal Imaging Data and Machine Learning," *Additive Manufacturing*, vol. 58, article no. 103008, Elsevier, Oct 2022.
The definitive version is available at https://doi.org/10.1016/j.addma.2022.103008

Additive Manufacturing 58 (2022) 103008

Contents lists available at ScienceDirect

Additive Manufacturing

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



Predicting defects in laser powder bed fusion using *in-situ* thermal imaging data and machine learning

Sina Malakpour Estalaki [a], Cody S. Lough [b], Robert G. Landers [a], Edward C. Kinzel [a,c], Tengfei Luo [a,d,*]

[a] Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[b] Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO, 65409, USA
[c] Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[d] Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

ARTICLE INFO

Keywords:
LPBF
Additive manufacturing
Machine learning
Binary classification
Confusion matrix

ABSTRACT

Variation in the local thermal history during the Laser Powder Bed Fusion (LPBF) process in Additive Manufacturing (AM) can cause micropore defects, which add to the uncertainty of the mechanical properties (*e.g.*, fatigue life, tensile strength) of the built materials. *In-situ* sensing has been proposed for monitoring the AM process to minimize defects, but successful minimization requires establishing a quantitative relationship between the sensing data and the porosity, which is particularly challenging with a large number of variables (*e.g.*, laser speed, power, scan path, powder property). Physics-based modeling can simulate such an *in-situ* sensing-porosity relationship, but it is computationally costly. In this work, we develop Machine Learning (ML) models that can use *in-situ* thermographic data to predict the micropore of LPBF stainless steel materials. This work considers two identified key features from the thermal histories: the time above the apparent melting threshold (τ) and the maximum radiance (T_{max}). These features are computed, stored for each voxel in the built material, and then used as inputs. The binary state of each voxel, either defective or normal, is the output. Different ML models are trained and tested for the binary classification task. In addition to using the thermal features of each voxel to predict its own state, the thermal features of neighboring voxels are also included as inputs. This is shown to improve the prediction accuracy, which is consistent with thermal transport physics around each voxel contributing to its final state. Among the models trained, the F1 scores on test sets reach above 0.96 for Random Forests. Feature importance analysis based on the ML models shows that T_{max} is more important to the voxel state than τ. The analysis also finds that the thermal history of the voxels above the present voxel is more influential than those beneath it. Our study significantly extends the capability of using *in-situ* thermographic data to predict porosity in LPBF materials. Since ML models are fast, they may play integral roles in the optimization and control of such AM technologies.

1. Introduction

Additive Manufacturing (AM) is recognized as a new paradigm for the manufacturing industry. It stands out due to its capability of creating complex, multi-material, and multi-functional designs and its unique position in advancing manufacturing through data and machine intelligence [1–8]. Recently, data-driven machine learning (ML) techniques are being applied to various AM applications to monitor building processes, detect defects or anomalies, and enhance decision-making leveraging data collected through different sensors [9–18]. Jin et al. reviewed different ML methods that are used to systematically optimize different stages of AM processes, ranging from geometrical design and process parameter configuration to *in-situ* anomaly detection [1]. Wang et al. also provided a comprehensive review on the state-of-the-art of ML applications in a variety of AM domains [19].

Detection of the defects due to variation in process conditions is of prime importance for quality control of AM. Kadam et al. used different ML algorithms in combination with pre-trained convolutional neural networks (CNNs) for fault detection in a fused deposition modeling-based 3D printing process utilizing a layer-by-layer manner, where

* Corresponding author at: Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.
E-mail address: tluo@nd.edu (T. Luo).

https://doi.org/10.1016/j.addma.2022.103008
Received 15 March 2022; Received in revised form 31 May 2022; Accepted 27 June 2022
Available online 1 July 2022

layers were treated as images [20]. Li et al. proposed a ML scheme to detect bumps on the surfaces of AM samples by describing them using point clouds and detecting anomalies in the distances among the points [21]. Jin et al. used layer-by-layer optical images and CNNs to distinguish imperfections for transparent hydrogel-based bio-printed materials [22]. These prior studies focused on obvious defects that are visible in optical images where the optical images are also inputs for the ML models. However, in some other AM techniques, like Laser Powder Bed Fusion (LPBF), defects like lack of fusion pores can be in the microscale which *in-situ* optical sensors cannot detect easily. In addition, using images for defect detection does not directly reveal what process variables are causing the defect and thus cannot be used for informing process optimization or feedback control.

Micropores in LPBF materials are usually detected using high-resolution micro-Computed Tomography (μ-CT) [23], but each scan can take hours, making μ-CT infeasible for *in-situ* defect detection. More advanced techniques such as the synchrotron radiation imaging [24,25] are capable of high-resolution imaging of pores in LPBF materials, but they require highly specialized facilities that are not easily applicable to wide-spread development. High-speed x-ray [26,27] can image pore dynamics *in-situ* during the LPBF process, but it also needs specialized facilities and can only monitor a relatively small volume of a few hundreds of micrometers in size.

The origin of the micropores is rooted in the thermal history the material has experienced during the LPBF process [24,28]. Using thermal features that can be monitored *in-situ* to predict micropore is thus attractive [3], and several studies have explored their relationships. Scime and Beuth monitored the melt-pool using a high-speed camera in the LPBF of stainless steel and used unsupervised learning and computer vision techniques to distinguish normal and abnormal melt-pools from the optical images [29]. It is desirable to further establish quantitative models to link the melt-pool information to porosity. Gobert et al. established such a relation using the support vector machine ML model to classify if a printed region is defective or nominal by training the model against *in-situ* optical images with *ex-situ* CT data. They achieved a defect detection accuracy, defined as (true positive + true negative)/total population, greater than 80 % as demonstrated using cross-validation [30]. However, since the number of defective regions is usually much smaller than that of normal regions in common LPBF materials, ~80 % accuracy is not particularly high. This was indicated by the true positive and true negative rates, which were both ~60 % at best. That is to say, their model had high probability to predict false positive and false negative labels.

Using visible optical cameras to monitor melt-pool is convenient; however, it offers indirect information of the thermal history of the LPBF process. Directly sensing the thermal features may provide more direct information for defect prediction. Baumgartl et al. used *in-situ* thermographic imaging to detect printing defects by training CNN ML models, achieving an overall accuracy of 97.9 % and true positive and true negative rates both above 96.8 %. However, the detected defects were delamination and spatter, which are much larger defects than micropores and easier to detect in nature [31]. Paulson et al. monitored the surface temperature histories of a Ti64 LPBF process *in-situ* by an infrared (IR) camera and used statistical ML models to correlate the thermal history to the keyhole porosity formation in the subsurface [28]. They extracted different features from the temperature histories and formulated the problem into a binary classification task to determine if subsurface pores would form or not. In this study, the pores were identified in the melt pool as circular regions with lighter intensity than the other solid regions using high-speed x-ray *in-situ* imaging. From the x-ray images, they were able to distinguish small pores from large pores, which can have different origins of temperature history, but they found that ML models perform better when not distinguishing between small and large pores. Trained by 112 temperature histories, the cross-validation accuracy using the leave-one-out scheme was shown to be 83.9–89.3 % depending on the different ML models and features used. On the test set, which contained 8 temperature histories and the corresponding porosity information, the ML model accuracy was 87.5–100.0 % for different models and features used. However, *in-situ* high-speed x-ray is a very specialized tool not widely accessible, and the overall data size was relatively small for ML. Around the same time, Lough et al. used Short-Wave Infrared (SWIR) imaging to monitor the spatial and temporal thermal features of the whole LPBF process of stainless steel, voxel by voxel [32]. They showed that various thermal features extracted from the SWIR data for each voxel could be indicative of whether the voxel contains pore defects, which were characterized by *ex-situ* μ-CT. In a follow-up study, Lough et al. found that two thermal features are most influential for defect detection, including the time above the apparent melting threshold and the maximum radiance [33]. In this study, they used linear regression to correlate these two thermal features to the porosity in a voxel-by-voxel manner and obtained area under curve (AUC) scores up to ~0.86 of the receiver operating characteristic (ROC) curve [33]. However, as shown by Paulson et al. [28], pores can form in the subsurface of the present top layer, and thus including thermal features surrounding a voxel, instead of only those of itself, may be needed to accurately predict the existence of pores.

In this work, we reduce thermal time histories recorded by an *in-situ* short-wave camera to take the two proven important thermographic features extracted from our previous work [33], including the time above apparent melting threshold (τ) and the maximum radiance (T_{max}). We then develop supervised ML models to predict the micropore of LPBF stainless steel materials in a voxel-by-voxel manner. The micropore is characterized using *ex-situ* μ-CT, and voxels are labeled as either defective or healthy, depending on the density of micropores in each voxel. Different supervised ML models, including K-nearest neighbors (KNN), Random Forests (RF), Decision Trees (DT), Multi-Layer Perceptron (MLP), Logistic Regression (LR) and AdaBoost, are trained and tested for this binary classification task. Besides using the thermal features of each voxel to predict its own state, those of neighboring voxels are also included, which is shown to improve the prediction accuracy. Such a finding suggests that the thermal history around a voxel can also impact its state due to heat transfer. Among the models trained, the F1 score on the hold-out test sets reaches above 0.96 for Random Forest (RF) model. Feature importance analysis based on the ML models shows that T_{max} is more important to the voxel state than τ. The analysis also finds that the thermal features of the voxels above the present voxel are more influential than those beneath it, correctly reflecting the influence of thermal history. Our study demonstrates the viability to use *in-situ* thermographic data to predict porosity in LPBF materials in high precision. Since ML models are fast, they may play integral roles in the optimization and control of such AM technologies.

2. Methodology

2.1. Experimental procedure and data registration

A 304 L stainless steel cylindrical part with a diameter of 4 mm and a height of 20 mm is fabricated on a Renishaw AM250 selective laser melting machine. A point-to-point exposure strategy is used to process the part by scanning an SPI fiber laser with a maximum power of 200 W and a wavelength of 1070 nm. In this process, the laser spot moves discretely in a fixed step size of 60 μm along a hatch line. The Renishaw AM250 system uses a step and dwell scan strategy as opposed to continuous scanning. This is motivated by stabilizing the melt pool. The laser and scan stages are coordinated and dwell at a point for an exposure time. The laser is turned off and the galvo scanners rapidly reposition to the next point. A detailed discussion of the effects of the AM250 scan strategy can be found in Ref. [34]. The exposure time at each spot varies from 50 to 125 μs and the laser power varies between 100 and 200 W. These parameters were selected to target different process regimes to enable nominal states and pore defects. The process of laser scanning is repeated with a rotation (67°) of the laser raster pattern from

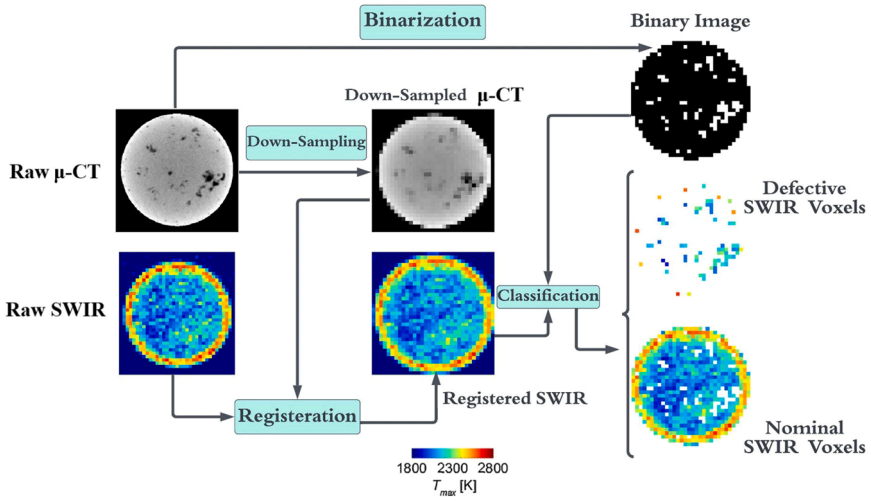


Fig. 1. A schematic flowchart for voxelization and registration of SWIR thermal feature and the μ-CT data. In the binary image in the upper right corner, black and white colors correspond to nominal and defective voxels, respectively.

layer-to-layer with a hatch spacing of 85 μm [32]. The commanded powder elevator step size is 50 μm. It should be noted that the true powder layer thickness may be greater due to consolidation over the first few layers. This does not affect the voxelized registration because the layer height is still the commanded 50 μm.

For the SWIR imaging measurements, a FLIR SC6201 camera (sensitive: 0.9–1.7 μm, filtered: 1.45 ± 0.05 μm) is installed in a staring configuration above the build chamber to observe the build plate through a custom window. A high frame rate with a frequency of 2500 Hz is enabled through dimension reduction from 640×512 camera pixel array to an 80×80 pixel window. The x- and y-direction instantaneous fields of view of the SWIR camera are ~ 130 μm/pixel and ~ 135 μm/pixel, respectively. The observation angle of the SWIR camera (θ) is 15°.

Based on the SWIR data, thermal features are extracted. Meaningful quantitative relationships can be developed with the information corresponding to the qualitative differences that are observed in the time series radiance data and the LPBF part properties through thermal features extraction [32]. The feature extraction allows the multi-frame recording SWIR raw data to reduce to a single image representation of each layer, with information stored in each pixel in the layer. Then, 2D thermal feature data of each layer are concatenated to build the 3D reconstruction of the feature space. Values of the voxel thermal feature retain information directly related to the local thermal history for the development of thermal feature-porosity correlations with a significantly smaller amount of feature data [32]. The two proven important thermal features (*i.e.*, τ and T_{max}) [33] for each voxel of the built material are used as inputs for the supervised ML models.

The binary labels characterizing whether the voxel is defective (*i.e.*, porous) or normal (*i.e.*, fully dense) are used as outputs for the ML models. The state of the voxel (defective or normal) is determined by processing the μ-CT scanning data, which establishes the ground truth data for our ML model training and testing. The sliced x-ray images from the μ-CT were collected as a stack of .tiff files, which offer gray scale intensities due to beam attenuation, and pores in the sample reduce the x-ray attenuation, providing information of a relative density decrease. These slices are reconstructed using a home-built Matlab code to form the complete 3D voxel data for the sample. We note that there were edge effects that produce grayscale features on the boundaries, which were corrected using a non-uniformity beam hardening. The resolution of our μ-CT data is 15 μm/pixel perpendicular to the build direction and 10 μm/pixel in the build direction.

Since there is a resolution mismatch between the voxels from the thermal feature data and the μ-CT data, proper registration needs to be performed before they can be used for ML model training. The voxel-wise binary state labels are obtained by registering the part's thermal feature voxel reconstruction with its post-processed μ-CT data. In the post-processing, the μ-CT data is first down-sampled from $15 \times 15 \times 10$ μm^3 per voxel to the thermal features' $130 \times 135 \times 50$ μm^3 voxel size. The down-sampling is performed by defining a grid with the larger voxel size and then using the grayscale intensity values of the smaller voxels encompassed by a larger voxel. This determines both a new grayscale voxel value by averaging and a binary state by the percentage of porosity present. Such a process produces both a grayscale intensity μ-CT-based reconstruction and a binary μ-CT-based reconstruction for the part, where the binary states are defective and normal. A binary voxel is considered as porosity if more than 5 % of the original smaller voxels within the new voxel volume correspond to porosity. This small percentage allows the lower resolution binary map to capture the smaller porosity features contained in the full resolution μ-CT data. Specifically, the 5 % threshold is set to allow us to retain a flag for the fine keyholing porosity features in the down-sampled μ-CT data. We have run the down-sampling at several thresholds and the values around 5 % yield the best results in capturing the fine porosity features. Setting too high a threshold will lead to less defective voxels, preventing the ML models from capturing fine defects. Next, the down-sampled grayscale μ-CT reconstruction is registered with the thermal feature reconstruction. Image space registration was performed using a home-built Matlab code using the imregistar function (intensity based multimodal based registration). The data sets are registered in the z-direction (*i.e.*, build direction) by a manual translation and then registered automatically in the x-y plane by translations determined through a multimodal intensity-based algorithm. Fiducial features were included for registration and the change between zones processed with different parameters provided additional references for registration in the z-direction. The translations from the registration step are then applied to the binary μ-CT-based reconstructions. This finally provides the binary labels for the thermal feature voxels because the porosity state is now known from the corresponding point in the binary reconstruction for a given (x, y, z) coordinate. Fig. 1 shows the workflow of the SWIR thermal feature and μ-CT data processing procedure for voxelization and registration.

2.2. Data preparation and different ML models

There are a total of 398 layers with each layer consisting of around 465 voxels arranged in a cylindrical shape (Fig. 2**a**). In total, there are 165517 voxels. For each voxel, we use the thermal features of itself and

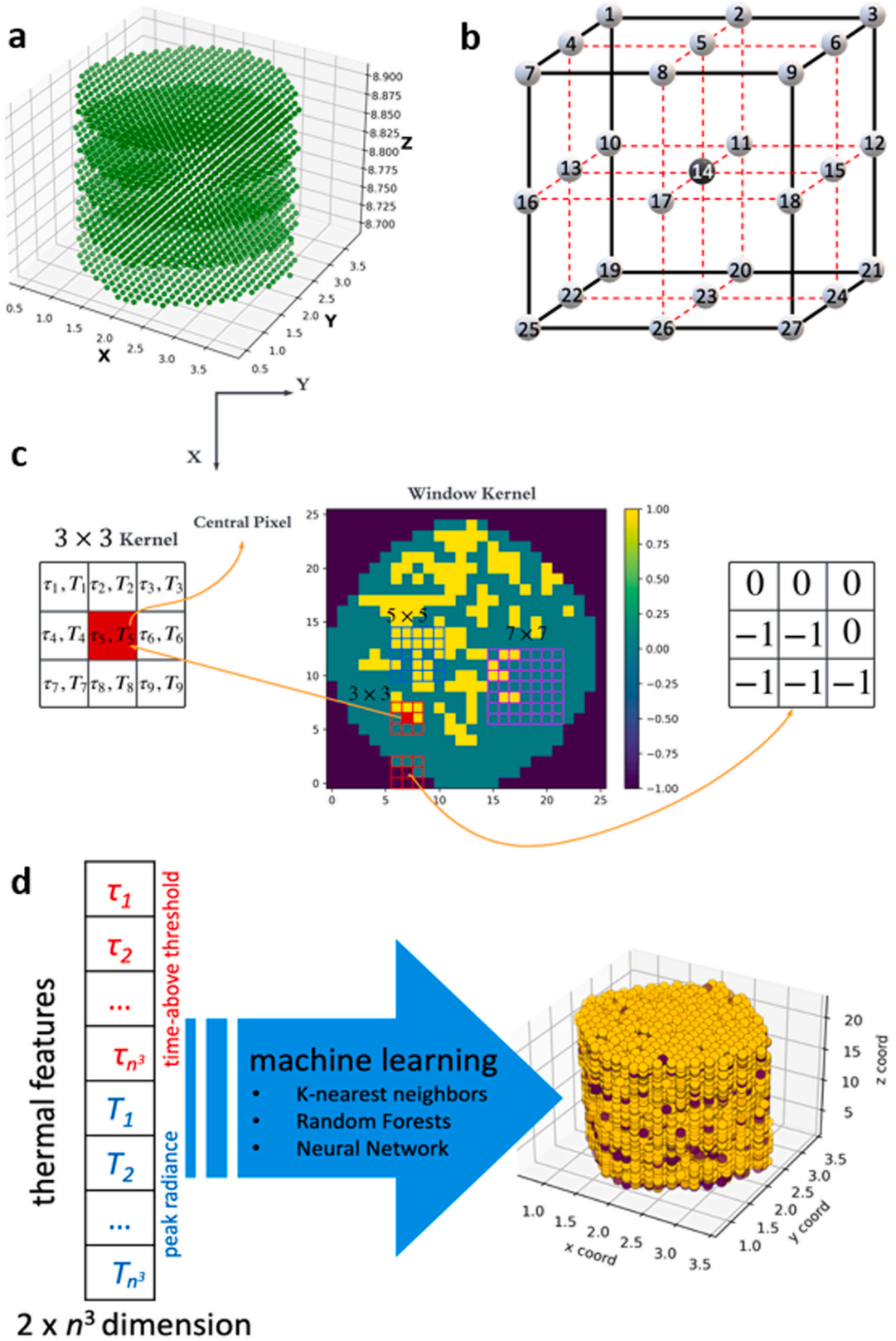


Fig. 2. **(a)** Illustration of the voxelated built volume (for clarification, only five layers are illustrated). **(b)** An example of the neighboring voxels (white) for a central voxel (black), and the indexing convention used. **(c)** Illustration of using sliding window kernels with different dimensions to extract the neighbors for each voxel, and how the surface voxels are detected (right panel), which are excluded in the ML tasks. This was realized by assigning different labels to voxels with "-1" corresponding to exclusion and ["0", "1"] corresponding to inclusion. **(d)** The thermal features of neighboring voxels are put into a one-dimensional vector as inputs for ML models to predict the state (defective – 1; normal – 0) of each voxel.

those of its neighbors as inputs to predict its binary state. Fig. 2**b** shows an example schematic of including the 1st nearest neighbors for the voxel state prediction and how the neighboring voxels are indexed. In this example, the total number of voxels involved in predicting the state of the central voxel is 3 × 3 × 3, and since the neighbor information is processed using sliding window kernels, we call this case K3 (*i.e.*, kernel 3 ×3 × 3). Fig. 2**c** shows a two-dimensional illustration of the windowing kernel. We have also tested 1 × 1 × 1 (no neighbors), 5 × 5 × 5 (up to 2nd nearest neighbors) and 7 × 7 × 7 (up to 3rd nearest neighbors) kernel cases in this study, and they are respectively called K1, K5 and K7 cases. Voxels near the top, bottom and side surfaces of the built volume are excluded from the data if they do not have the complete set of neighbors (see right panel in Fig. 2**c**). This was realized by assigning different labels to voxels with "−1" corresponding to exclusion and ["0", "1"] corresponding to inclusion. Excluding these boundary voxels leads the total numbers of voxels used in the ML studies to be 115541, 76551 and 47287 for the K3, K5 and K7 cases, respectively. The thermal features are then put into a one-dimensional vector with the first half of the elements including τ of voxels and the second half containing T_{max} (Fig. 2**d**). The inputs are scaled according to:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_{min} and x_{max} are the min and max of each feature in Eq. (1). It is noted that the current data is highly imbalanced with 92 % of all the voxels being normal (labeled as 0) and only 8 % defective (labeled as 1). To improve the ML model accuracy, we employ the Borderline-Synthetic Minority Oversampling (Borderline-SMOTE) technique [35] to duplicate samples in the minority class so that the positive and negative labels are balanced to 1:1 for the training process. Although oversampling is used for training, we note that all the model accuracies reported are quantified on the imbalanced data.

Six different supervised ML classification models, including K-nearest neighbors (KNN), Random Forests (RF), Decision Trees (DT), Multi-Layer Perceptron (MLP), Logistic Regression (LR) and AdaBoost, are trained and tested to find the best model for more extensive analysis. Certain percentages (7 %, 10 %, 20 %, 30 % and 40 %) of data are held out for testing to evaluate the predictivity of the models. Fig. S1 in Section S1 of the Supporting information (SI) shows the flowchart that includes all the steps from data preparation, training and testing (final evaluation). The hyperparameters used for each model are optimized *via* five-fold cross-validation (CV) using the balanced data, and they can be found in Section S1 of the SI. The results of five-fold CV using the balanced data for RF based on two metrics including ROC-AUC and accuracy are available in Section S1 of the SI. The fact that the performance metrics in the five-fold CV are close to one another (see Section S1 in SI) suggests that our models are statistically reliable. The receiver operating characteristic (ROC) curve, confusion matrix, ROC-area under curve (AUC), Precision, Recall, F1 and accuracy scores are employed to evaluate the performance of the models. Each of these metrics are explained in Section 2.3. We also visually compare the model predictions with the ground truth of the processed μ-CT data for the test set. The Gini importance [36,37] is examined based on the RF model to assess the feature importance for investigating how the neighboring voxels influence the state of a central voxel. MLP is implemented in PyTorch [38], and all the other models are implemented in scikit-learn [39].

2.3. Evaluation metrics for ML classification models

In this section, we present all metrics used to evaluate the ML model performance.

2.3.1. ROC curve

It is a performance plot of the binary classifier when its discriminative threshold is varied. In the ROC plots of this study, the x-axis is the fraction of false positives out of the negatives (*i.e.*, the false positive rate, FPR), and the y-axis is the fraction of true positives out of the positives (*i.e.*, the true positive rate, TPR). The definitions of the FPR and TPR are:

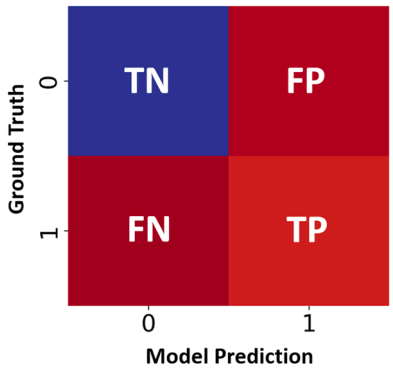


Fig. 3. Confusion matrix with TN, FN, TP and FP quads. 0 and 1 respectively correspond to the normal and defective states of voxels.

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

where TP is true positive, FN is false negative, FP is false positive, and TN is true negative.

2.3.2. ROC-AUC

It is the value of the area under the ROC curve, and it is referred to as the ROC-AUC score in this study. The more ROC curve is towards the top left corner of the plot, the area under it is larger, depicting more accurate classification.

2.3.3. Confusion matrix

The confusion matrix evaluates classification accuracy by computing the matrix with each row and each column corresponding to the true class and prediction, respectively (Fig. 3).

2.3.4. Accuracy score

The accuracy score for our binary classification models is defined as [39]:

$$Accuracy \quad score = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

2.3.5. Precision, recall and F1 scores

When the binary labels are very imbalanced, precision-recall is a useful measure of the success of prediction. In information retrieval, precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. Precision (Pr) is defined as:

$$Pr = \frac{TP}{TP + FP} \tag{5}$$

Recall (Re) is defined as:

$$Re = \frac{TP}{TP + FN} \tag{6}$$

Note that Re equals to TPR (Eq. (3) and Eq. (6)). The precision-recall curve can show the tradeoff between precision and recall for different thresholds. A large area under the curve represents both high recall and high precision, where high precision relates to a low FPR, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as

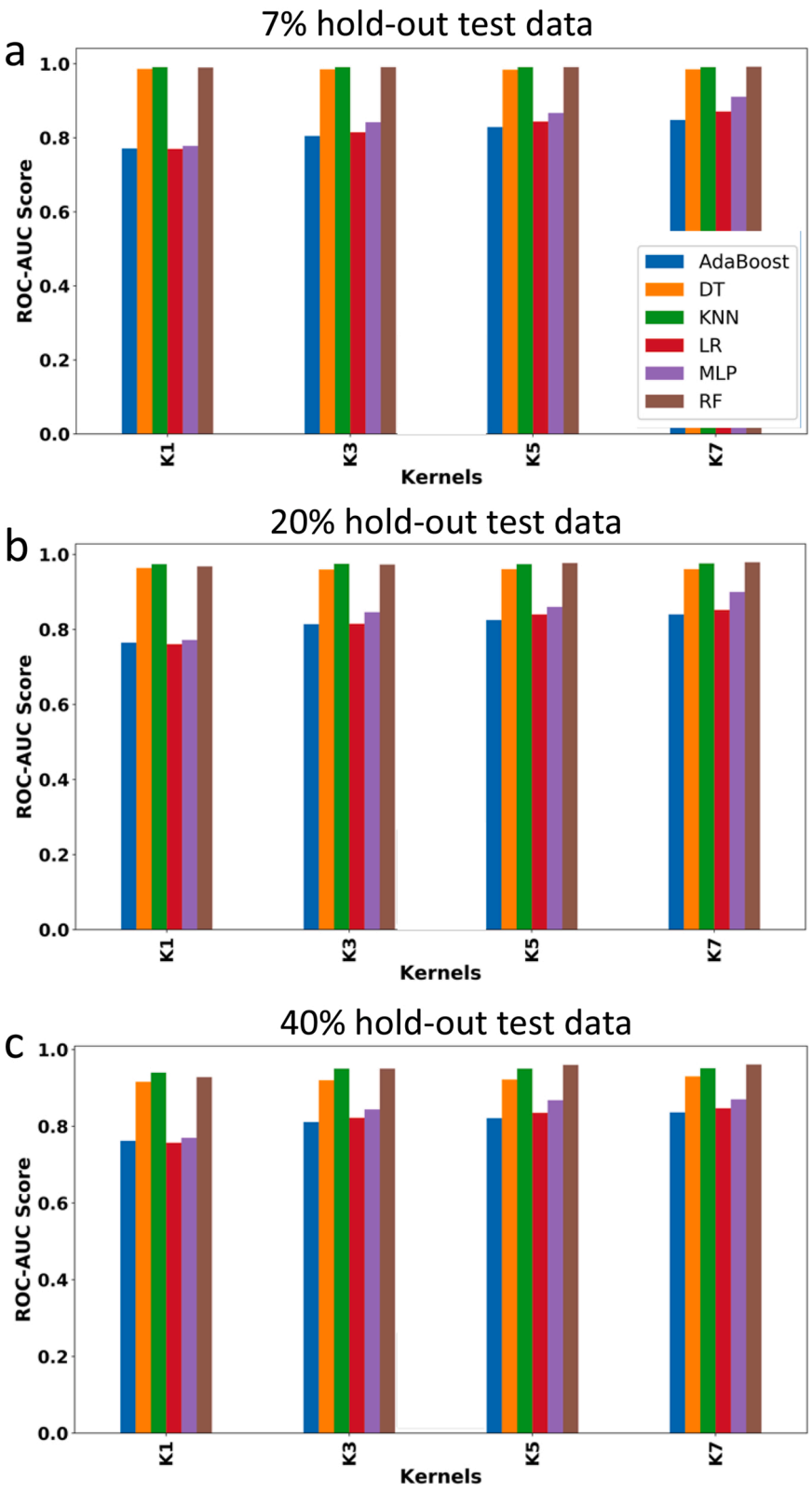


Fig. 4. ROC-AUC score *versus* different kernel sizes using six ML models for **(a)** 7 %, **(b)** 20 % and **(c)** 40 % hold-out data as unseen test set.

returning a majority of all positive results (high recall) [39]. Pr and Re are related to the F1 score which is defined as the harmonic mean of Pr and Re:

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \tag{7}$$

3. Result and discussion

3.1. Results overview

We first compared the ROC-AUC scores of all six ML models and

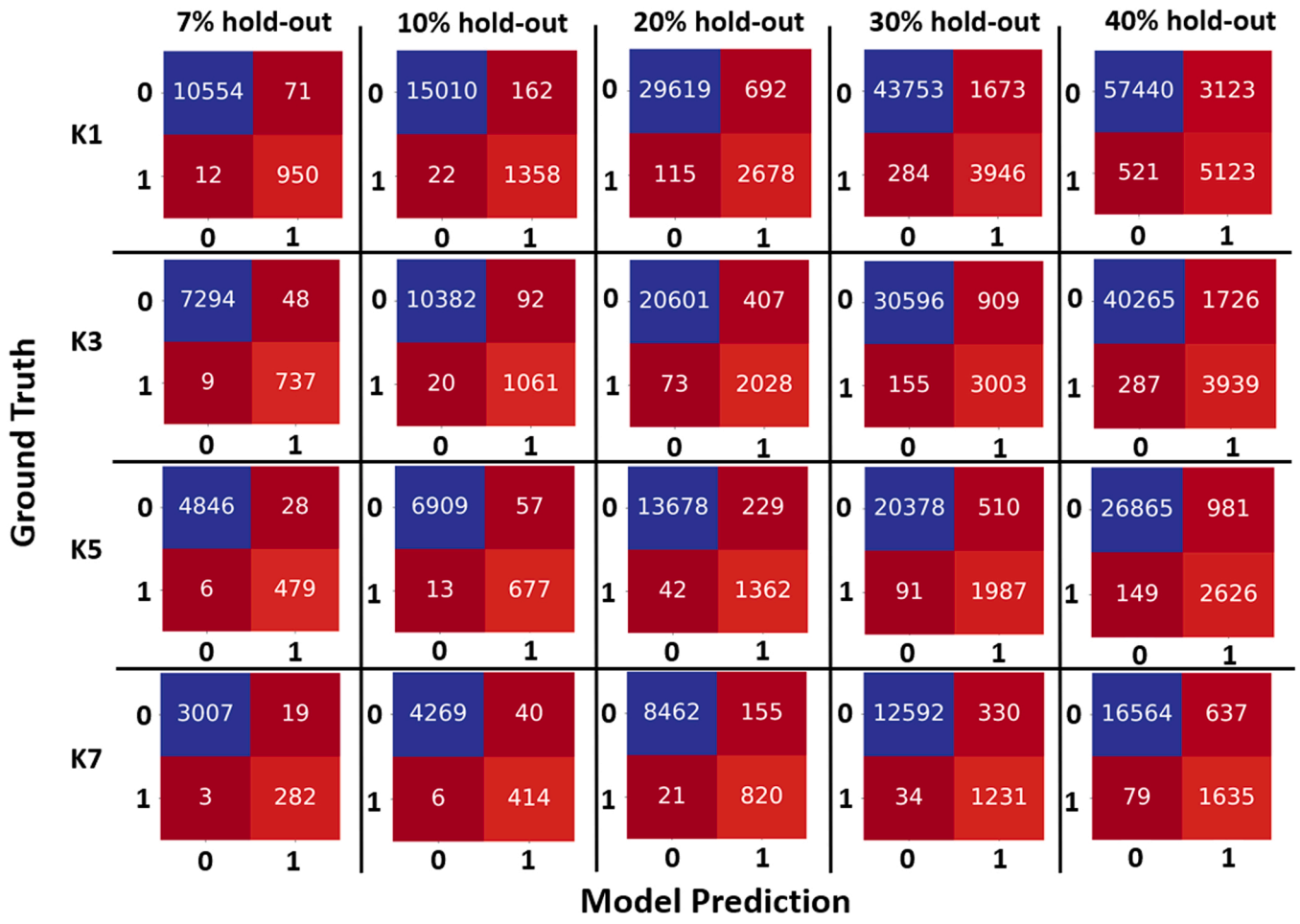


Fig. 5. Confusion matrices for the RF models on different hold-out test data sizes using different kernels.

found that the RF model showed the best prediction performance on the reserved test dataset (Section 3.2.1). We then focused on the RF model to perform more detailed analysis. When using the confusion matrix as the metric, results showed that the RF models predict more FP than FN incidences (Section 3.2.2), meaning that these models were more likely to predict normal voxels as defective ones when they were wrong in prediction, but were less likely to miss defective voxels. Further comparison of Pr, Re, F1 and accuracy scores (Section 3.2.3) and the Pr-Re curves (Section 3.2.4) for the RF models with different kernel sizes (*i.e.*, different neighboring voxel ranges) showed that model performance improved as more neighbors were included but mostly converged at a kernel size of 5, corresponding to including the 2nd nearest neighbors. Image comparison between the RF-predicted voxel states and those derived from the μ-CT demonstrated good model prediction performance (Section 3.3). Based on the RF model, input feature importance was analyzed (Section 3.4). It was found that thermal features of voxels further away from the present voxel were less important than those of the closer ones. It was found that voxels above the present voxel were more influential than those below. These results are discussed in detail in the following sections.

3.2. Detailed results

3.2.1. ROC-AUC analysis and comparison

In Fig. 4, bar plots of the ROC-AUC scores for all ML models are presented for 7 %, 20 % and 40 % of the hold-out test data sizes. The specific values of the ROC-AUC scores for all models and all hold-out data sizes can be found in Table S3. The number of defective voxels in the test set and that predicted from the models are also shown in Table S3. Among all the models tested, KNN and RF consistently outperform other models regardless of the percentage of the hold-out data for testing and kernel size. The DT model is a close follower as reflected by the ROC-AUC scores, while the MLP, LR and AdaBoost models show inferior accuracies. Comparing different test data percentages shows that by decreasing the number of test data, the ROC-AUC generally increases for all ML models, because more data are used for training. For almost all models, by increasing the kernel size, which includes more neighboring voxel thermal features as inputs for prediction, the ROC-AUC scores increase (Fig. 4). For KNN, RF and DT models, which have high ROC-AUC scores, when the percentage of the hold-out test data becomes larger (*e.g.*, 30 % and 40 %), the increasing trend in ROC-AUC as kernel size increases is more obvious (Fig. 4**c** and Table S3). For the hold-out test data size less than 20 %, the ROC-AUC scores for these three models are seen weakly increasing or fluctuating around a high value. For the rest of the models, the increasing trend is evident regardless of the hold-out test data size.

Overall, these findings suggest that neighboring voxels contain useful information that can help determine the state of the voxel of interest. This improvement in ROC-AUC is achieved despite the fact that increasing the kernel size decreases the available data for training and testing since more voxels are considered surface voxels and are excluded (see Fig. 2**c**). Since RF turns out to be the best performing model, we will focus on analyzing the RF model in the rest of the paper. We note that the RF models do not only show the best ROC-AUC scores but also other performance metrics (*e.g.*, the confusion matrix in Fig. 5, S6 and S7, and Precision-Recall curve in Fig. S9).

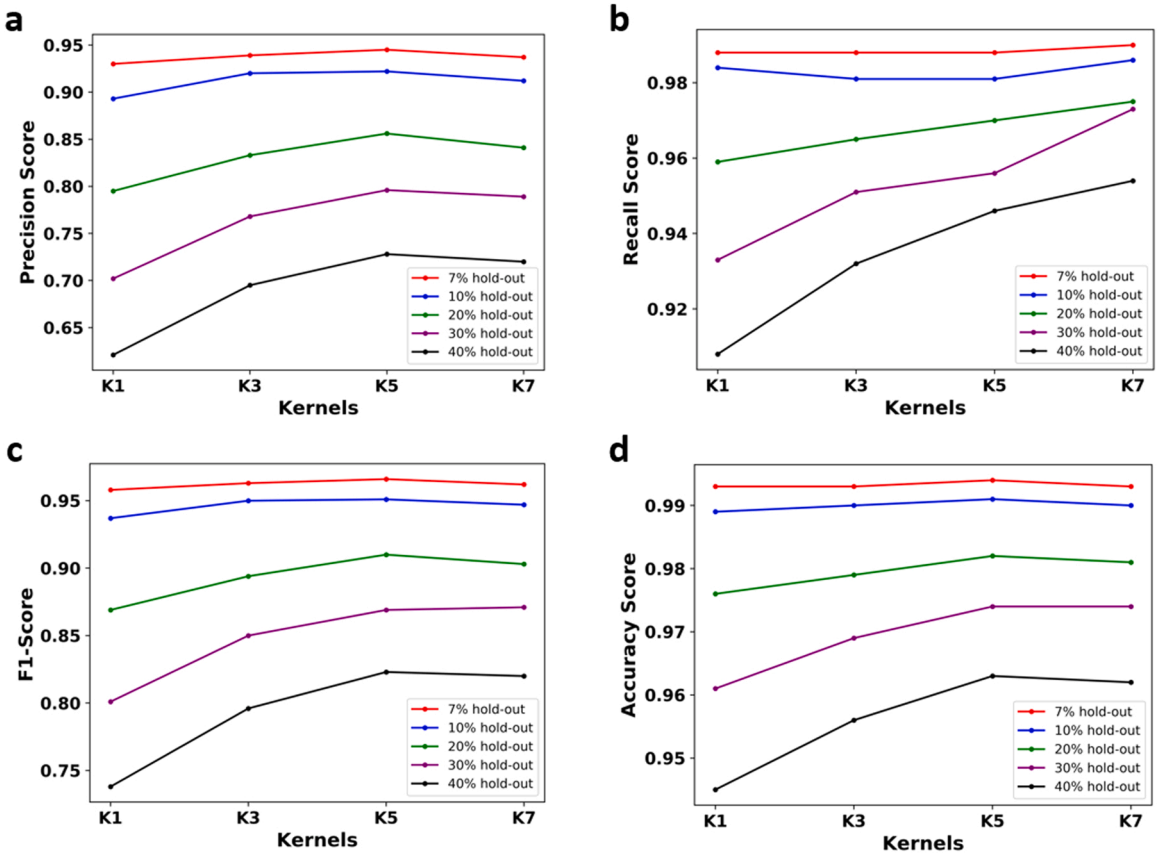


Fig. 6. **(a)** Precision, **(b)** Recall, **(c)** F1 and **(d)** Accuracy scores *versus* kernel size for the RF models with different percentages of hold-out data as unseen test set.

3.2.2. *Confusion matrix analysis and comparison*

Fig. 5 shows the confusion matrices of the RF model as the highest performing ML model with different kernel sizes and different hold-out test data sizes. The confusion matrices show that the RF models predict more FP (upper right quad in the matrix) than false negative (FN, lower left quad) incidences. This means that models tend to predict normal voxels as defective ones when they are wrong in prediction but are less likely to miss defective voxels. This suggests that the models are "safe" but "conservative", which are favorable for practical AM applications by providing a larger margin of safety. The same observation can be made for the other two high-performance ML models (*i.e.*, KNN and DT), see Section S3 of the SI.

3.2.3. *Pr, Re, F1 and accuracy scores analysis*

Fig. 6**a-d** show the Precision, Recall, F1 and accuracy scores, respectively, for the RF models using different kernels for all percentages of the hold-out test data. As shown in Fig. 6, by increasing the size of hold-out test data, Pr, Re, F1 and accuracy scores all decrease. This is because less data is used for training when the hold-out test set size increases. By increasing the kernel size, the Pr score generally increases from K1 to K5, but it sees a slight decrease in some cases when changing from K5 to K7 (Fig. 6**a**). The Re score generally increases when the kernel size increases and this increase is more significant for the larger hold-out test set. One exception is the 10 % hold-out case where the Re score fluctuates without a monotonic trend. The Re scores are uniformly higher than their Pr counterparts because the model predicts more FP than FN as discussed previously (Fig. 6**b**). The F1 score, which is a combined description of Pr and Re, almost always increases from K1 to K5, but it stagnates or decreases slightly when the kernel size increases to K7 (Fig. 6**c**). A similar trend can be found in the accuracy score (Fig. 6**d**). From these quantifications, it can be inferred that there is a gain of useful information by including thermal features from more neighboring voxels up to K5 (*i.e.*, the 2nd nearest neighbors), which leads to the increase in the model quality. However, using the K7 kernel no longer improves model accuracy, suggesting that the thermal history of the 3rd nearest neighboring voxels has little impact on the state of the present voxel. Table S4 in **section S4** of the SI shows the values of Pr, Re, F1 and accuracy scores using RF model for different kernel sizes and all percentages of the hold-out test data. Furthermore, Fig. S8 compares the F1 scores of six ML models for 10 % hold-out test data with kernel K5, and RF has the best F1 score.

3.2.4. *Precision-Recall (Pr-Re) curves analysis*

Fig. 7 further demonstrates Precision-Recall curves for hold-out test data percentages of 10 %, 20 %, 30 % and 40 % to provide a visual comparison between the RF model and a dummy classifier (DC) as a baseline for different kernel sizes. DC makes predictions without considering the input features, and thus it is essentially a random guess [39]. The curve stretched towards the upper right corner in the Precision-Recall indicates better model prediction quality and this is the case for all hold-out test data percentages when the kernel size increases from K1 to K7. However, there is a small difference between curves of K5 and K7 for all percentages and this difference is negligible for 10 % and 20 % hold-out test data.

3.3. *ML model prediction and comparison with μ-CT data*

Fig. 8 visually compares the predictions from the RF model and the ground truth for the 30 % hold-out test data case in the 3D view and from randomly selected layers. The RF model shown uses the K5 kernel and the dots shown are the test data, *i.e.*, training data is not shown. As can be seen, both the 3D view (Figs. 8**a** and **b**) and the randomly selected layers (Fig. 8**c**) show good agreement between ground truth and RF prediction with a relatively small number of false predictions (circled).

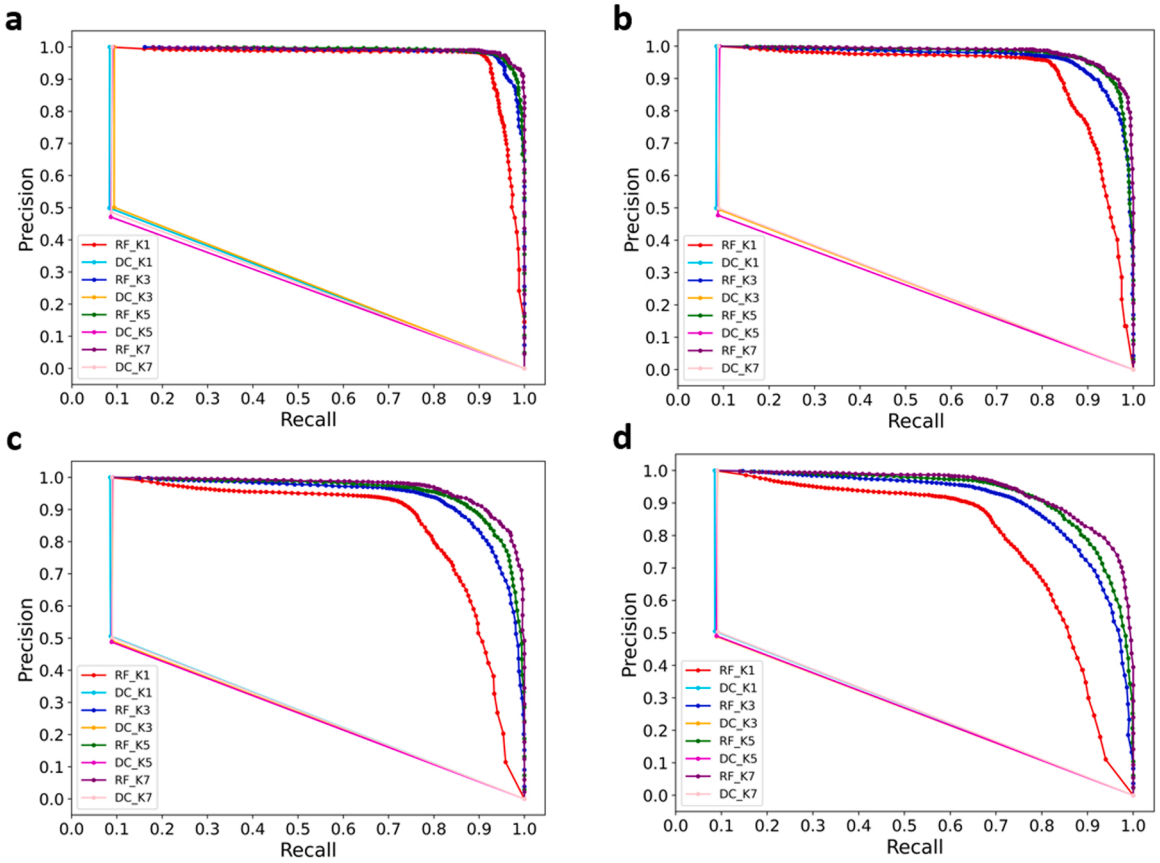


Fig. 7. Precision-Recall curves comparison for the DC as the classification baseline and RF ML model for all kernel sizes **(a)** 10 % hold-out test data **(b)** 20 % hold-out test data **(c)** 30 % hold-out test data and **(d)** 40 % hold-out test data.

More visual comparisons are shown in Section S6 in the SI.

3.4. Feature importance analysis

Lastly, we study the feature importance from the RF model with the K5 kernel to access the impact of the thermal feature of the neighboring voxels on the state of the central voxel. Fig. 9 shows the mean Gini feature importance score plot for the case with 30 % hold-out test data. The training data is used to obtain the scores of the features based on the Gini impurity. For computing this, the feature score values corresponding to 100 random splits of 70 %−30 % train-test data for all 250 features in the K5 kernel are calculated and then the average of these scores are taken and shown. We note that the first and second 125 feature scores are corresponding to the time-above threshold (τ) and the peak radiance (T_{max}), respectively. The T_{max} feature score values are much higher than the τ feature score values, suggesting that T_{max} is a more important feature, which is consistent with a recent study [33]. We have further colored the features from voxels belonging to different layers in Fig. 9 as L+2, L+1, L0, L-1 and L-2, where L0 is the present layer, and "+" indicates layers above the present while "-" indicates those below it. The average feature scores of each layer are also shown in Fig. 9. It is understandable that L0 has the highest average feature score of 0.032, L+2 has the second highest of 0.024, while L+1 has the third highest of 0.021. The features in L-1 and L-2, which are layers below the present layer, are not as important as those above it. These findings suggest that the RF model correctly captures the heat transfer nature involved in the LPBF process, where thermal energy in the top layer transferring downwards can impact the micropore of layers beneath it. This is because when pores are formed due to lack of fusion in a region, its thermal conductivity is lower than normal regions, which will in turn raise the temperature of layers above it when the laser scans. For keyhole mode, the higher temperature in a layer generates pores below the current surface. Both scenarios suggest that one would see a much greater effect of the thermal process in layers above the present layer than in those below. In addition, it has been found that pore healing, migration or coalescence can happen as subsequent layers are deposited [40]. Including thermal features of layers above the current layer to predict its voxel states can also take these phenomena into consideration, which may also have contributed to the improved prediction accuracy of our models.

3.5. Perspective on the established ML model

While the pores in the present study were created by varying laser power and exposure time, we believe the model should be generally applicable to detecting defects created by other process conditions (*e.g.*, laser speed) since the model's input is the SWIR signal, which tracks the thermal history of different voxels. It is our belief that it is the thermal history that determines the state of the voxel. Additionally, since we found that the influence of the thermal history is up to the 2nd nearest neighboring voxels, approximately within a range of ~338 μm, the effect is relatively localized. As a result, we believe the model is also applicable to different bulk geometries, although the regions near the surface of the built may need special treatment.

Finally, we note that processing the SWIR data for each layer takes a few seconds, which will allow us to extract the two thermal features (*i.e.*, maximum radiance and time above threshold). To predict the state of all the voxels in a layer, the RF model takes about three seconds using the K5 kernel. If this process can be shortened to within one second *via* means like parallel programming, feedback loop control for *in-situ* correction may be possible since we have a chance to correct the defects in the present layer by changing the process conditions in the next few

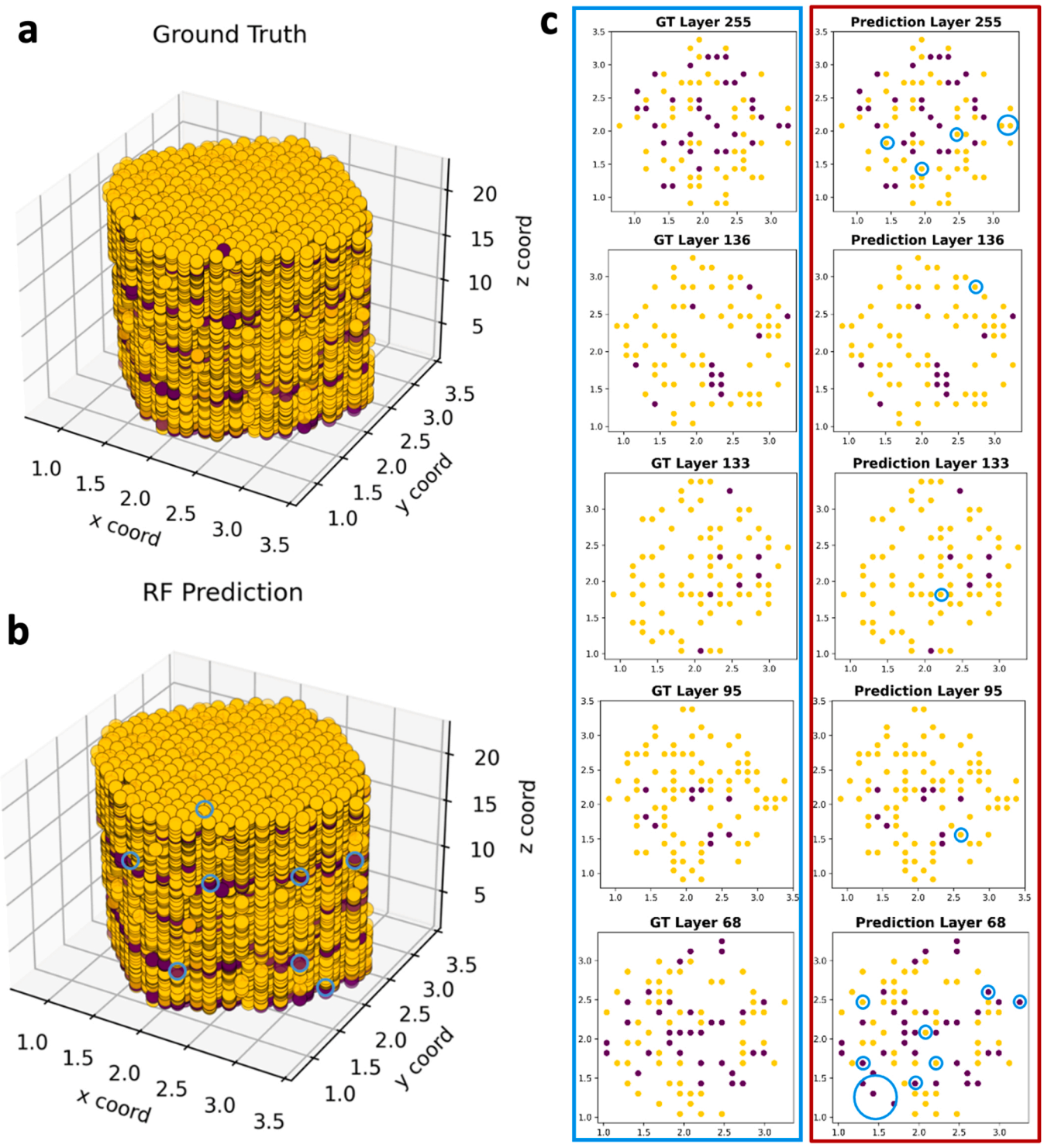


Fig. 8. Visual comparison between the ground truth and RF prediction on the state of the test data set: the 3D view (coordinates in the unit of mm) of **(a)** the ground truth, and **(b)** the RF prediction, and **(c)** the side-by-side comparison of randomly selected layers. GT is representative of the ground truth. The yellow and purple spheres correspond to normal and defective, respectively. The blue circles indicate false predictions.

layers.

Additionally, our results showed the maximum radiance was more important than time above threshold for determining the defects. There is a significant potential benefit to being able to take only the maximum radiance rather than time above threshold experimentally. Specifically, the data used in this paper was collected with the thermal camera in a staring configuration. The field of view is limited in order to maintain a high frame rate. If the camera is placed in a co-axial configuration, following the beam spot, similar spatial resolution can be achieved at the expense of capturing the full temperature history of the voxels. The ability to make high quality classifications with only the maximum radiance data for the pixel, including the nearest neighbors, will allow these techniques to be scaled to much larger parts and builds without sacrificing temporal resolution (maximum temperature accuracy).

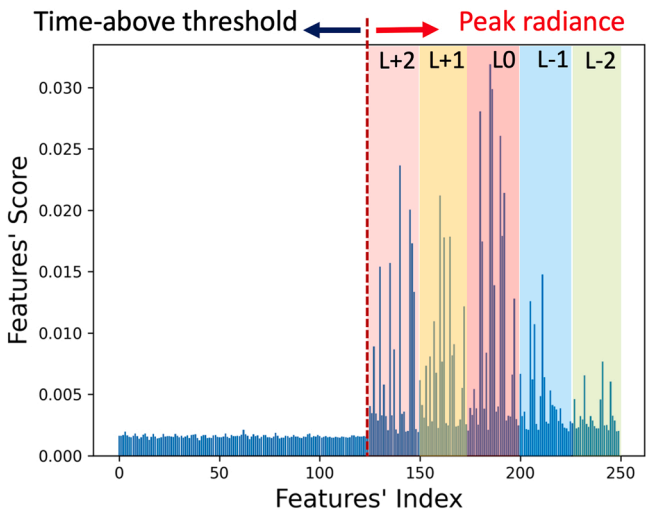


Fig. 9. Mean of the features' score *vs.* features' index corresponding to the $5 \times 5 \times 5$ voxel.

4. Conclusions

In this work, we have developed machine learning models that can use *in-situ* thermographic data to predict the voxel state of LPBF stainless steel materials. Two features from the thermograph, including the time above the apparent melting threshold (τ) and the maximum radiance (T_{max}) of each voxel, are used as inputs, and the binary state of each voxel, either defective or normal, is the output. Six different ML models are trained and tested for the binary classification task, with RF showing the highest predictive performance with the highest F1 score reaching 0.966. Besides using the thermal features of each voxel to predict its own state, those of neighboring voxels are also included, which is shown to improve prediction accuracy. Such a finding suggests that the thermal history around a voxel can also impact its state due to heat transfer. Results also show that useful information can be collected from up to the 2nd nearest neighbor to improve model accuracy, but including scope beyond the 2nd nearest neighbor shows limited effect. Feature importance analysis based on the RF model shows that T_{max} is more important to the voxel state than τ. The analysis also finds that the thermal history of the voxels above the present voxel is more influential than those beneath it, which agrees with the heat transfer nature in the LPBF process. Our study demonstrates the viability to use *in-situ* thermographic data to predict porosity in LPBF materials. Since ML models are fast, they may play integral roles in the optimization and control of such AM technologies.

CRediT authorship contribution statement

Sina Malakpour Estalaki: Conceptualization, Methodology, Software, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing, **Cody S. Lough:** Data curation, Investigation, Visualization, Writing – review & editing, **Robert G. Landers:** Writing – review & editing, **Edward C. Kinzel:** Investigation, Writing – review & editing, **Tengfei Luo:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

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.

Acknowledgements

The authors are grateful for the data provided by Honeywell Federal Manufacturing and Technologies and obtained during work funded by Contract No. DE-NA0002839 with the U.S. Department of Energy. The computations are supported by the University of Notre Dame, Center for Research Computing (CRC).

Code availability

The Python code required to reproduce these findings are available to download from https://github.com/sinaDFT/ML-LPBF-AM upon publication.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.addma.2022.103008.

References

[1] Z. Jin, Z. Zhang, K. Demir, G.X. Gu, Machine learning for advanced additive manufacturing, Matter 3 (2020) 1541–1556, https://doi.org/10.1016/j.matt.2020.08.023.
[2] S.L. Sing, C.N. Kuo, C.T. Shih, C.C. Ho, C.K. Chua, Perspectives of using machine learning in laser powder bed fusion for metal additive manufacturing, Virtual Phys. Prototyp. 16 (2021) 372–386, https://doi.org/10.1080/17452759.2021.1944229.
[3] S.K. Everton, M. Hirsch, P.I. Stavroulakis, R.K. Leach, A.T. Clare, Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing, Mater. Des. 95 (2016) 431–445, https://doi.org/10.1016/j.matdes.2016.01.099.
[4] L.J. Ladani, Applications of artificial intelligence and machine learning in metal additive manufacturing, J. Phys Mater. 4 (2021), 042009, https://doi.org/10.1088/2515-7639/ac2791.
[5] C. Tian, T. Li, J. Bustillos, S. Bhattacharya, T. Turnham, J. Yeo, A. Moridi, Data-driven approaches toward smarter additive manufacturing, Adv. Intell. Syst. 3 (2021) 2100014, https://doi.org/10.1002/aisy.202100014.
[6] J.F. Arinez, Q. Chang, R.X. Gao, C. Xu, J. Zhang, Artificial intelligence in advanced manufacturing: current status and future outlook, 110804-1- 110804-16, Manuf. Sci. Eng. Trans. ASME 142 (2020), https://doi.org/10.1115/1.4047855.
[7] C.S. Lough, L.I. Escano, M. Qu, C.C. Smith, R.G. Landers, D.A. Bristow, L. Chen, E. C. Kinzel, In-situ optical emission spectroscopy of selective laser melting, Manuf. Process. 53 (2020) 336–341, https://doi.org/10.1016/j.jmapro.2020.02.016.
[8] F.H. Kim, H. Yeung, E.J. Garboczi, Characterizing the effects of laser control in laser powder bed fusion on near-surface pore formation via combined analysis of in-situ melt pool monitoring and X-ray computed tomography, Addit. Manuf. 48 (2021), 102372, https://doi.org/10.1016/j.addma.2021.102372.
[9] M.O. Alabi, K. Nixon, I. Botef, A survey on recent applications of machine learning with big data in additive manufacturing industry, Am. J. Eng. Appl. Sci. 11 (2018) 1114–1124, https://doi.org/10.3844/ajeassp.2018.1114.1124.
[10] Z. Snow, B. Diehl, E.W. Reutzel, A. Nassar, Toward in-situ flaw detection in laser powder bed fusion additive manufacturing through layerwise imagery and machine learning, Manuf. Syst. 59 (2021) 12–26, https://doi.org/10.1016/j.jmsy.2021.01.008.
[11] B. Yuan, G.M. Guss, A.C. Wilson, S.P. Hau-Riege, P.J. DePond, S. McMains, M. J. Matthews, B. Giera, Machine-learning-based monitoring of laser powder bed fusion, Adv. Mater. Technol. 3 (2018) 1800136, https://doi.org/10.1002/admt.201800136.
[12] C. Liu, L. le Roux, Z. Ji, P. Kerfriden, F. Lacan, S. Bigot, Machine learning-enabled feedback loops for metal powder bed fusion additive manufacturing, Procedia Comput. Sci. 176 (2020) 2586–2595, https://doi.org/10.1016/j.procs.2020.09.314.
[13] Q. Liu, H. Wu, M.J. Paul, P. He, Z. Peng, B. Gludovatz, J.J. Kruzic, C.H. Wang, X. Li, Machine-learning assisted laser powder bed fusion process optimization for AlSi10Mg: new microstructure description indices and fracture mechanisms, Acta Mater. 201 (2020) 316–328, https://doi.org/10.1016/j.actamat.2020.10.010.
[14] R. Liu, S. Liu, X. Zhang, A physics-informed machine learning model for porosity analysis in laser powder bed fusion additive manufacturing, Int. J. Adv. Manuf. Technol. 113 (2021) 1943–1958. ⟨http://arxiv.org/abs/2101.05605⟩.
[15] N.S. Johnson, P.S. Vulimiri, A.C. To, X. Zhang, C.A. Brice, B.B. Kappes, A. P. Stebner, Invited review: machine learning for materials developments in metals additive manufacturing, Addit. Manuf. 36 (2020), 101641, https://doi.org/10.1016/j.addma.2020.101641.
[16] D. Mahmoud, M. Magolon, J. Boer, M.A. Elbestawi, M.G. Mohammadi, Applications of machine learning in process monitoring and controls of LPBF additive manufacturing: a review, Appl. Sci. 11 (2021) 11910, https://doi.org/10.3390/app112411910.
[17] T. Li, J. Yeo, Strengthening the sustainability of additive manufacturing through data-driven approaches and workforce development, Adv. Intell. Syst. 3 (2021) 2100069, https://doi.org/10.1002/aisy.202100069.

[18] Y. Fu, A.R.J. Downey, L. Yuan, T. Zhang, A. Pratt, Y. Balogun, Machine learning algorithms for defect detection in metal laser-based additive manufacturing: a review, Manuf. Process. 75 (2022) 693–710, https://doi.org/10.1016/j.jmapro.2021.12.061.
[19] C. Wang, X.P. Tan, S.B. Tor, C.S. Lim, Machine learning in additive manufacturing: state-of-the-art and perspectives, Addit. Manuf. 36 (2020), 101538, https://doi.org/10.1016/j.addma.2020.101538.
[20] V. Kadam, S. Kumar, A. Bongale, S. Wazarkar, P. Kamat, S. Patil, Enhancing surface fault detection using machine learning for 3D printed products, Appl. Syst. Innov. 4 (2021) 34, https://doi.org/10.3390/asi4020034.
[21] R. Li, M. Jin, V.C. Paquit, Geometrical defect detection for additive manufacturing with machine learning models, Mater. Des. 206 (2021), 109726, https://doi.org/10.1016/j.matdes.2021.109726.
[22] Z. Jin, Z. Zhang, X. Shao, G.X. Gu, Monitoring anomalies in 3D bioprinting with deep neural networks, ACS Biomater. Sci. Eng. (2021), https://doi.org/10.1021/acsbiomaterials.0c01761.
[23] S. Moon, R. Ma, R. Attardo, C. Tomonto, M. Nordin, P. Wheelock, M. Glavicic, M. Layman, R. Billo, T. Luo, Impact of surface and pore characteristics on fatigue life of laser powder bed fusion Ti–6Al–4V alloy described by neural network models, Sci. Rep. 11 (2021) 1–17, https://doi.org/10.1038/s41598-021-99959-6.
[24] W.E. King, H.D. Barth, V.M. Castillo, G.F. Gallegos, J.W. Gibbs, D.E. Hahn, C. Kamath, A.M. Rubenchik, Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing, Mater. Process. Technol. 214 (2014) 2915–2925, https://doi.org/10.1016/j.jmatprotec.2014.06.005.
[25] Y. Chen, S.J. Clark, C.L.A. Leung, L. Sinclair, S. Marussi, M.P. Olbinado, E. Boller, A. Rack, I. Todd, P.D. Lee, In-situ synchrotron imaging of keyhole mode multi-layer laser powder bed fusion additive manufacturing, Appl. Mater. Today 20 (2020), 100650, https://doi.org/10.1016/j.apmt.2020.100650.
[26] R. Cunningham, C. Zhao, N. Parab, C. Kantzos, J. Pauza, K. Fezzaa, T. Sun, A. D. Rollett, Keyhole threshold and morphology in laser melting revealed by ultrahigh-speed X-ray imaging (https://doi.org/ DOI:), Science 363 (2019) 849–852, https://doi.org/10.1126/science.aav4687.
[27] S.M.H. Hojjatzadeh, N.D. Parab, W. Yan, Q. Guo, L. Xiong, C. Zhao, M. Qu, L. I. Escano, X. Xiao, K. Fezzaa, W. Everhart, T. Sun, L. Chen, Pore elimination mechanisms during 3D printing of metals, Nat. Commun. 10 (2019) 1–8, https://doi.org/10.1038/s41467-019-10973-9.
[28] N.H. Paulson, B. Gould, S.J. Wolff, M. Stan, A.C. Greco, Correlations between thermal history and keyhole porosity in laser powder bed fusion, Addit. Manuf. 34 (2020), 101213, https://doi.org/10.1016/j.addma.2020.101213.
[29] L. Scime, J. Beuth, Using machine learning to identify in-situ melt pool signatures indicative of flaw formation in a laser powder bed fusion additive manufacturing process, Addit. Manuf. 25 (2019) 151–165, https://doi.org/10.1016/j.addma.2018.11.010.
[30] C. Gobert, E.W. Reutzel, J. Petrich, A.R. Nassar, S. Phoha, Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging, Addit. Manuf. 21 (2018) 517–528, https://doi.org/10.1016/j.addma.2018.04.005.
[31] H. Baumgartl, J. Tomas, R. Buettner, M. Merkel, A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring, Prog. Addit. Manuf. 5 (2020) 277–285, https://doi.org/10.1007/s40964-019-00108-3.
[32] C.S. Lough, X. Wang, C.C. Smith, R.G. Landers, D.A. Bristow, J.A. Drallmeier, B. Brown, E.C. Kinzel, Correlation of SWIR Imaging with LPBF 304L stainless steel part properties, Addit. Manuf. 35 (2020), 101359, https://doi.org/10.1016/j.addma.2020.101359.
[33] C.S. Lough, T. Liu, X. Wang, B. Brown, R.G. Landers, D.A. Bristow, J.A. Drallmeier, E.C. Kinzel, Local prediction of laser powder bed fusion porosity by short-wave infrared imaging thermal feature porosity probability maps, Mater. Process. Technol. 302 (2022), 117473, https://doi.org/10.1016/j.jmatprotec.2021.117473.
[34] P.A. Hooper, Melt pool temperature and cooling rates in laser powder bed fusion, Addit. Manuf. 22 (2018) 548–559, https://doi.org/10.1016/j.addma.2018.05.032.
[35] H. Han, W.Y. Wang, B.H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. International conference on intelligent computing, Springer,, Berlin, Heidelberg, 2005.
[36] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[37] S. Nembrini, I.R. König, M.N. Wright, The revival of the gini importance? Bioinformatics 34 (2018) 3711–3718, https://doi.org/10.1093/bioinformatics/bty373.
[38] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury Google, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, Advances in neural information processing systems 32 (2019) 8024–8035.
[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, Mach. Learn. Res. 12 (2011) 2825–2830. ⟨http://scikit-learn.sourceforge.net⟩.
[40] C.L.A. Leung, S. Marussi, M. Towrie, R.C. Atwood, P.J. Withers, P.D. Lee, The effect of powder oxidation on defect formation in laser additive manufacturing, Acta Mater. 166 (2019) 294–305, https://doi.org/10.1016/j.actamat.2018.12.027.