# A deep learning framework for defect prediction based on thermographic in-situ monitoring in laser powder bed fusion

Simon Oster[1] · Philipp P. Breese[1] · Alexander Ulbricht[1] · Gunther Mohr[1] · Simon J. Altenburg[1]

## Abstract

The prediction of porosity is a crucial task for metal based additive manufacturing techniques such as laser powder bed fusion. Short wave infrared thermography as an in-situ monitoring tool enables the measurement of the surface radiosity during the laser exposure. Based on the thermogram data, the thermal history of the component can be reconstructed which is closely related to the resulting mechanical properties and to the formation of porosity in the part. In this study, we present a novel framework for the local prediction of porosity based on extracted features from thermogram data. The framework consists of a data pre-processing workflow and a supervised deep learning classifier architecture. The data pre-processing workflow generates samples from thermogram feature data by including feature information from multiple subsequent layers. Thereby, the prediction of the occurrence of complex process phenomena such as keyhole pores is enabled. A custom convolutional neural network model is used for classification. The model is trained and tested on a dataset from thermographic in-situ monitoring of the manufacturing of an AISI 316L stainless steel test component. The impact of the pre-processing parameters and the local void distribution on the classification performance is studied in detail. The presented model achieves an *accuracy* of 0.96 and an *f1-Score* of 0.86 for predicting keyhole porosity in small sub-volumes with a dimension of $(700 \times 700 \times 50)$ µm$^3$. Furthermore, we show that pre-processing parameters such as the porosity threshold for sample labeling and the number of included subsequent layers are influential for the model performance. Moreover, the model prediction is shown to be sensitive to local porosity changes although it is trained on binary labeled data that disregards the actual sample porosity.

**Keywords** Laser Powder Bed Fusion (PBF-LB/M, L-PBF) · Selective Laser Melting (SLM) · SWIR thermography · Online monitoring · Flaw detection · Machine learning · Convolutional neural networks (CNN)

## Abbreviations

| | |
|---|---|
| BO | Bayesian optimization |
| CNN | Convolutional neural network |
| EI | Expected improvement |
| FN | False negative |
| FP | False positive |
| FPGA | Field programmable gate array |
| GPU | Graphics processing unit |
| ML | Machine learning |
| MPA | Melt pool area |
| MPE | Melt pool eccentricity |
| MPL | Melt pool length |
| MPP | Melt pool perimeter |
| MPT | Melt pool temperature |
| MPW | Melt pool width |
| NDT | Non-destructive testing |
| NIR | Near infrared |
| PBF-LB/M (also L-PBF) | Laser powder bed fusion |
| PI | Probability of improvement |
| ReLu | Rectified Linear Unit |
| RUS | Random under-sampling |
| SGD | Stochastic gradient descent |
| SLM | Selective laser melting |
| SMOTE | Synthetic minority oversampling technique |

✉ Simon Oster
   simon.oster@bam.de

1  Bundesanstalt für Materialforschung und-prüfung (BAM), Federal Institute for Materials Research and Testing), Unter den Eichen 87, 12205 Berlin, Germany

| SWIR | Short-wave infrared |
| TN | True negative |
| TOT | Time over threshold |
| TP | True positive |
| UCB | Upper confidence band |
| VED | Volumetric energy density |
| XCT | X-ray micro computed tomography |

## Introduction

In-situ monitoring of metal-based additive manufacturing technologies such as laser powder bed fusion (PBF-LB/M, also L-PBF or selective laser melting) has drawn the interest of the scientific community and industry in recent years (Grasso et al., 2021). In-situ sensors offer the possibility to detect defects or irregularities during manufacturing, *e.g.*, surface roughness, porosity, cracking, and delamination. The thereby provided quality control is especially important in safety critical applications as for example in the aerospace industry (McCann et al., 2021). One of the main types of irregularities inherent to PBF-LB/M components is porosity which was found to greatly affect the mechanical performance of a product (Yadollahi et al., 2015). For the detection of porosity, multiple different non-destructive testing (NDT) techniques were applied in recent years, such as acoustic emission (Zhang et al., 2020), X-ray Micro Computed Tomography (XCT) (Fritsch et al., 2021) and eddy current testing (Ehlers et al., 2020). However, the thermal history of the manufactured component provides valuable information which are not covered by these methods. The thermal history can be correlated with the quality of the resulting component, since areas of localized thermal variations had been reported to result in a high probability for the formation of porosity (Lough et al., 2022). Thermography as a radiometric NDT technique captures the local surface radiosity with high temporal and spatial resolution. By monitoring the layer-wise manufacturing, thermography can be used to reconstruct the thermal history of the built component. It had been reported that thermogram features (*e.g.*, melt pool geometry and cooling behavior) could be correlated to unstable process conditions which result in the formation of porosity (Oster et al., 2021). Extracting valid information from thermograms concerning the local build quality requires the analysis of large amounts of complex 2D datasets. In recent studies, Machine Learning (ML) algorithms were introduced to fulfil the task of porosity or void prediction from thermograms (Gaikwad et al., 2022; Krabusch et al., 2020; Lough et al., 2020; Mohr et al., 2020; Smoqi et al., 2022).

In terms of void formation mechanisms in PBF-LB/M, Snow et al. (2020) identified three different characteristic void categories: Lack-of-fusion, gas porosity, and melt pool instabilities. Lack-of-fusion voids generally have an irregular shape and can result from incomplete melting of the feedstock powder. Incomplete melting can be caused by insufficient laser power or increased scan velocity (Bayat et al., 2019) or poorly set hatching parameters such as hatch spacing or layer thickness (Aboulkhair et al., 2014). The formation mechanism of gas porosity voids, which are characterized by small, spherical shapes, is still in debate. Cunningham et al. (2017) reported that inert gas trapped within the feedstock powder is one possible formation factor. If high laser powers and low scan velocities are used for exposure, the PBF-LB/M process transitions from conduction to keyhole mode (Guo et al., 2019), where melt pool instabilities can appear (Snow et al., 2020). In keyhole mode, the vaporization of metal during excessive energy input exerts a recoil pressure onto the melt pool surface. As a result, a slim cavity denoted as keyhole is formed (Rai et al., 2007) in which the laser absorption is increased (Trapp et al., 2017). In unstable keyhole conditions, the oscillating keyhole can collapse temporarily, and its lower end is pinched off. Depending on the oscillation mode, the emerged gas bubbles can escape the keyhole and form a keyhole void (Ren et al., 2023). The resulting voids have a spherical shape and are mostly located at the lower end of the melt pool (Hojjatzadeh et al., 2020).

Using ML to predict porosity or single voids from thermogram feature data is challenging due to a number of aspects: Firstly, features that are related to the thermal history (*e.g.*, the melt pool geometry or cooling rates) need to be evaluated in terms of their benefit to increase the performance of prediction models. In addition, the thermogram data (from which features are extracted) is not sufficiently studied in terms of necessary quality of resolution and temperature calibration. In recent studies, different monitoring setups were discussed utilizing various thermal spectral ranges (*e.g.*, visible (Gobert et al., 2018; Snow et al., 2021)), near infrared (NIR) (Gaikwad et al., 2022; Hooper, 2018), short-wave infrared (SWIR) (Lough et al., 2020; Oster et al., 2021) or mid-wave infrared spectrum (Mohr et al., 2020; Raplee et al., 2020)). Furthermore, different positional camera settings were applied (on-axis and off-axis). These settings differ strongly in acquisition frequency, spatial resolution, and detectable temperature range. Secondly, the prediction of porosity or single voids requires the registration of the thermogram feature data and the associated reference data which is often provided by XCT (Ulbricht et al., 2021). Post-process part deformation and imaging errors are challenging aspects and can result in inferior registration quality. Poorly registered datasets might lead to low performance of prediction models (Oster et al., 2022). Thirdly, due to the layer-wise nature of PBF-LB/M processes, the complexity of the prediction task is increased by effects of re-melting of material ("healing") of previously emerged voids by the laser exposure of subsequent powder layers (Ulbricht et al., 2021).

Despite these challenges, multiple studies were recently published studying model-based prediction of irregularities from in-situ monitoring data. The use of high-resolution digital cameras is a common method to monitor the distribution of powder layers and specimen surfaces pre- and post-exposure (Aminzadeh & Kurfess, 2018; Gobert et al., 2018; Snow et al., 2021; Xiao et al., 2020). This approach is beneficial to detect irregularities in the powder distribution and in the solidified component surface after the printing process. However, the presented methods are not able to acquire information of the thermal history of the manufacturing process since the laser exposure is not monitored.

Other studies report the integration of NIR and SWIR cameras into 3D printers to monitor the build process. In a study by Krabusch et al. (2020), an optical tomography camera was mounted on top of the 3D printer off-axis to the laser path to acquire images during the build job. The images were analyzed by a convolutional neural network (CNN) architecture which included long short-term memory layers. The CNN was used to predict porosity cluster-wise in AlSi10Mg components. They achieved an *f1-Score* of 0.753 for a binary classification using subsequent layers as input data. Mohr et al. (2020) studied the use of a mid-wave infrared camera and optical tomography to detect voids in a PBF-LB/M/316L component. They achieved a *maximum overlap* of 71.4% with voids in the reference XCT data by analyzing the time over threshold (TOT) feature extracted from thermograms. Lough et al. (2022) extracted maximum thermogram temperature and TOT features from SWIR thermograms, and generated probability maps for porosity using Bayesian statistics. Binarized and registered XCT data was used as ground truth. The following *area under curve* values were achieved for porosity prediction in the bulk material of a PBF-LB/M/304L part: 0.88 for maximum temperature features and 0.94 for TOT above 1500 K. Smoqi et al. (2022) used an off-axis dual wavelength imaging pyrometer to acquire melt pool images from the manufacturing of a PBF-LB/M component. The authors extracted four different thermogram features (melt pool length, mean ejecta spread, mean ejecta temperature, and the distribution of the melt pool temperature). They used these features as input for standard machine learning models (*e.g.*, K-nearest neighbor) and a CNN architecture. The extracted features were labelled per component section based on the void type and porosity level. This information was determined by XCT, optical microscopy and scanning electron microscopy. They achieved a maximum *f1-Score* of 0.97. Gaikwad et al. (2022) performed two tasks using standard machine learning algorithms (*e.g.*, support vector machines) and deep neural networks: The identification of systematic laser focus drifts, and the classification into different levels of porosity. They used two high speed NIR cameras operating at wavelengths of 700 μm and 950 μm, respectively (see also Hooper (2018)). The data of cylindric PBF-LB/M components was labeled per layer with regard to the porosity level and pore type which was identified by optical microscopy. They achieved an *f1-Score* of 0.97 for the second classification task with a CNN architecture operating on raw melt pool images. Ren et al. (2023) used thermal cameras (NIR, SWIR) to monitor the melt pool of Ti-6Al-4 V single tracks at high frequencies (50 to 200 kHz). Full-field x-ray images of the single-track melting were acquired during laser exposure. The authors extracted the average keyhole emission intensity and applied wavelet analysis on the 1D time series data. The resulting scalograms were fed into a CNN for the binary classification of "Pore" or "Non-pore". They achieved the highest possible prediction results (*accuracy*, *recall,* and *precision* of 1). Furthermore, they conducted proof-of-concept single-track bare plate experiments using a commercial 3D printer.

The achieved results of the studies mentioned above demonstrated the potential of thermographic in-situ monitoring for porosity prediction. The reviewed studies could be classified into feature-based approaches or raw data approaches which use entire images. We identified three studies where the datasets were labeled per layer or per large component section (Aminzadeh & Kurfess, 2018; Gaikwad et al., 2022; Smoqi et al., 2022). Thereby, the prediction task was strongly simplified since stochastic variances in the local void distribution (Gaikwad et al., 2022) are not addressed. However, the knowledge of the local void distribution is of great importance to predict the service life of a component. Furthermore, many current models insufficiently incorporate the physical background of complex PBF-LB/M phenomena such as keyhole pore formation and closing of voids by re-melting of solidified layers. Additionally, the impact of data pre-processing (*e.g.*, how the ML samples are defined from thermograms and labeled from reference data) on the model performance needs to be examined more precisely. In terms of the evaluation of model performance, there is a lack of information about the spatial accuracy in which voids can be predicted.

In this study, we present a novel framework for the prediction of keyhole porosity in small sub-volumes ("clusters") based on thermogram features. The framework (Fig. 1) involves a data pre-processing workflow and a custom CNN model architecture. During data pre-processing, feature information from subsequent manufacturing layers is included into the samples. Thereby, we aim to enable the prediction of keyhole porosity. To test this hypothesis, the developed framework is applied on a dataset from the manufacturing of an AISI 316L component. During manufacturing, keyhole porosity formation was forced by increasing the Volumetric Energy Density (VED) in specific component sections. The dataset contains registered thermogram feature and XCT data and was reported in a previously published work (Oster et al., 2022). The prediction task is

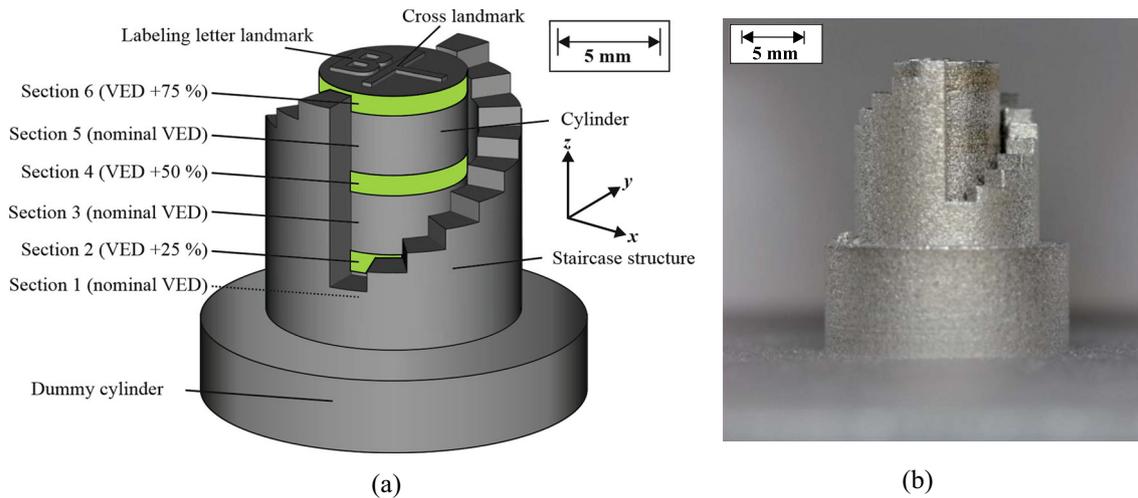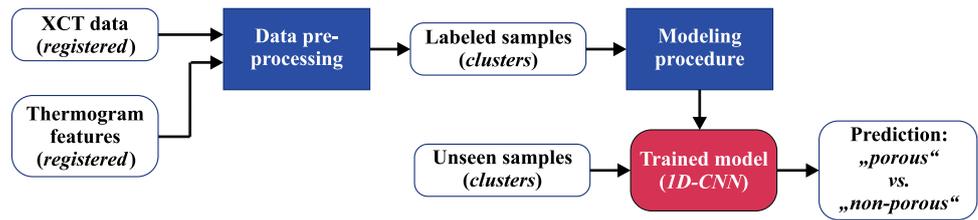**Fig. 1** Illustration of framework for the cluster-wise prediction of local porosity





(a)                                                                        (b)

**Fig. 2** Component design inspired by a study by Gobert et al. (2018) as computer-aided design (**a**) and after manufacturing (**b**). The specimen was manufactured from AISI 316L stainless steel. The outer staircase and the landmarks positioned on the specimen top were utilized for registration purposes (Oster et al., 2022). In the green component section increased VED was utilized to force the formation of keyhole pores (nominal VED of 65.45 J/mm.³). The part was built upon a dummy cylinder to prevent cutting losses. Adapted from Oster et al. (2022) and Oster et al. (2021)

based on binary labels ("porous" and "non-porous"). We show that the model is able to predict porosity above 0.1% with high probability (*accuracy* of 0.95 and *f1-Score* of 0.86) for a cluster dimension of (700 × 700 x 50) $\mu m^3$. Furthermore, we study the impact of pre-processing parameters (porosity threshold for labeling of samples, cluster dimension and number of considered subsequent layers) on the model's performance. The results show that these pre-processing parameters can affect the model's prediction ability detrimentally if poorly set. Furthermore, we analyze the performance of the model with regard to the local void distribution. We show that the model is sensitive to local changes in the porosity level although it is trained only on binary samples ("porous" and "non-porous" classes).

## Materials and methods

### Experimental setup and specimen design

A commercial PBF-LB/M machine (SLM 280 HL, SLM Solutions Group AG, Lübeck, Germany) was used to produce a cylindric specimen (Fig. 2) from AISI 316L stainless steel feedstock powder. In this work, the inner cylinder of the component was observed. The cylinder consisted of 240 layers (layer thickness of 50 μm). The cylinder was built using the optimal machine parameters for this material provided by the machine manufacturer. The resulting nominal VED was 65.45 J/mm³. At three distinct build heights (see green markings in Fig. 2a) the VED was increased in steps of 25% by reducing the scan speed to force the formation of keyhole pores. Further information concerning the powder specifications, the hatch strategy and the build process parametrization can be found in a previously published work (Oster et al., 2021).

A SWIR camera (Goldeye CL-033 TEC1, Allied Vision Technologies GmbH, Stadtroda, Germany) was used to monitor the manufacturing process. The camera was mounted outside of the build chamber and monitored the build plate off-axis the laser path. An image size of (90 × 90) $pix^2$ was chosen for data acquisition and the pixel size was determined to approx. 105 μm. The camera acquired thermograms at an acquisition frequency of 3600 Hz. For further information, refer as well to Oster et al. (2021).

### Feature extraction

A single-point calibration strategy was used for the thermal calibration of the thermograms (Scheuschner et al., 2019). Physically interpretable features were extracted from the

**Table 1** List of extracted melt pool-dependent and time-dependent features from calibrated thermograms

| Feature name | Unit | Abbreviation | Description | Usage for monitoring |
|---|---|---|---|---|
| Melt pool area | pix$^2$ | MPA | Count of pixels associated to melt pool (liquid phase) | Changes in melt pool dimension and shape |
| Melt pool length | pix | MPL | Major axis of elliptical fit of melt pool shape | |
| Melt pool width | pix | MPW | Minor axis of elliptical fit of melt pool shape | |
| Melt pool eccentricity | – | MPE | Ratio of MPW to MPL | |
| Melt pool perimeter | pix | MPP | Perimeter of melt pool pixels via 8-connectivity | |
| Melt pool mean temperature | K | MPT$_{mean}$ | Averaged temperatures of melt pool pixels | Changes in melt pool temperature distribution |
| Melt pool maximum temperature | K | MPT$_{max}$ | Average of three hottest melt pool pixels | |
| Time over threshold of 1200 K | ms | TOT1200K | Time period in which the temperature of a single pixel exceeds threshold temperature (calculated once per layer) | Cooling behavior above lower end of camera sensitivity |
| Time over threshold of 1680 K | ms | TOT1680K | | Cooling behavior above solidification temperature (approx. 1660 K) |
| Time over threshold of 2400 K | ms | TOT2400K | | Cooling behavior above upper end of camera sensitivity |

thermograms to reduce the size of the input data of the ML model. Two classes of features were extracted: Melt pool-dependent features and time-dependent temperature features. The extracted features and their purpose in terms of in-situ monitoring are specified in Table 1. A detailed description of the feature extraction is given in Oster et al. (2021).

## Image registration

The registration procedure of thermogram feature data and XCT data was presented in a previously published study (Oster et al., 2022). The purpose of registration is the accurate spatial alignment of thermogram feature data and XCT reference data. An erroneous registration may result in decreased void prediction performance since feature information and associated voids are spatially disjoined. The registration methodology is presented in Fig. 3. Furthermore, the approximated spatial uncertainty produced by the registration procedure is given in Table 2. For additional information, *e.g.*, concerning XCT machine setup and data reconstruction, the interested reader is referred to Oster et al. (2022).

## Data pre-processing

The data pre-processing workflow was used to generate data samples from the registered feature and XCT datasets for the training of the classifier. The pre-processing workflow contained two major steps: The resampling of the fine-resolved

**Table 2** Lateral geometric registration uncertainty measured in registered datasets

| Error type | Minimum error value in $\mu$m | Maximum error value in $\mu$m |
|---|---|---|
| Lateral translation error | $27 \pm 13$ | $64 \pm 27$ |
| Lateral scaling error | $19 \pm 8$ | $55 \pm 23$ |

The error was determined per component section (compare Fig. 2). Here, the minimum and maximum values with the respective standard deviations are given for a translation transformation function (Oster et al., 2022)

datasets to coarser resolved sub-volumes ("clusters"), and the generation of sample matrices $M$ as model input by using thermogram feature information of subsequent layers.

## Resampling to clusters

During registration (see Fig. 3), the thermogram feature datasets were up-sampled to the XCT voxel size of $(10 \times 10 \times 10)$ $\mu$m$^3$ using linear interpolation. Despite this fine voxel resolution, the feature data incorporated multiple sources of spatial uncertainty: Firstly, the original pixel resolution of the SWIR camera was determined to 105 $\mu$m/pix. Secondly, the melt pool travelled a distance in the range of 110 to 200 $\mu$m between two frames (this value depends on the static hatch distance and the chosen scan velocity). Thirdly,
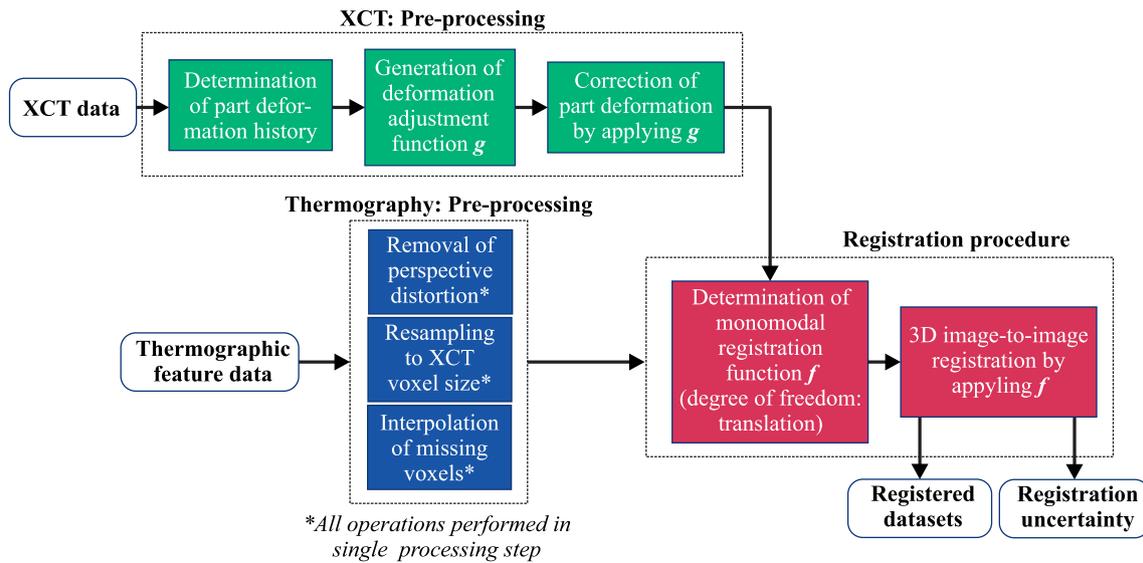
**Fig. 3** Flow chart of the registration process of XCT data and thermographic feature data
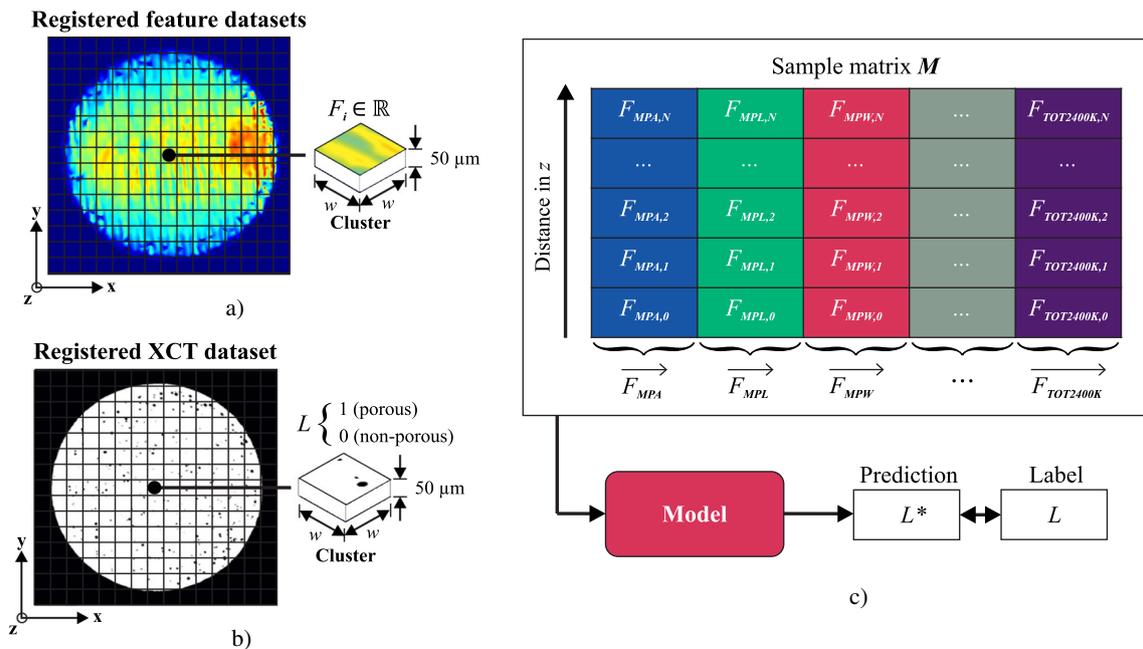


**Fig. 4** Data pre-processing for the generation of samples. **a** Resampling of thermogram feature data to larger clusters with an edge length $w$ and a height of 50 µm. Shown here is a 2D feature slice extracted from the MPA dataset. **b** Resampling of XCT dataset to larger clusters. The label $L$ is calculated based on the porosity threshold $P_{thresh}$ (see "Sample construction"). **c** Construction of feature matrix $M$ using mean values of feature cluster. The row position in $M$ corresponds to the vertical distance to the cluster whose label shall be predicted. For each label $L$, a peculiar matrix $M$ is constructed and utilized as input to produce the prediction $L*$

after the registration process, we measured a lateral displacement of up to $(64 \pm 27)$ µm and scaling difference of up to $(55 \pm 23)$ µm between the datasets (Table 2). To study the influence of spatial uncertainty on the porosity prediction, we down-sampled feature and XCT datasets to coarser resolved cuboid-shaped clusters (Fig. 4a and b). We varied the cluster edge length $w$ from 100 to 1000 µm with steps of 100 µm. The cluster height was set to the layer thickness of 50 µm

to simplify interpretation. During down-sampling, all voxel values inside of a newly created feature cluster were averaged. Thereby, discrete 1D sample values were acquired and spatially assigned to the cluster center point. In the following, the mean value of an arbitrary feature cluster of a feature $i$ is denoted as $\overline{F_i}$. Only clusters that were entirely located inside of the component bulk were regarded as valid.
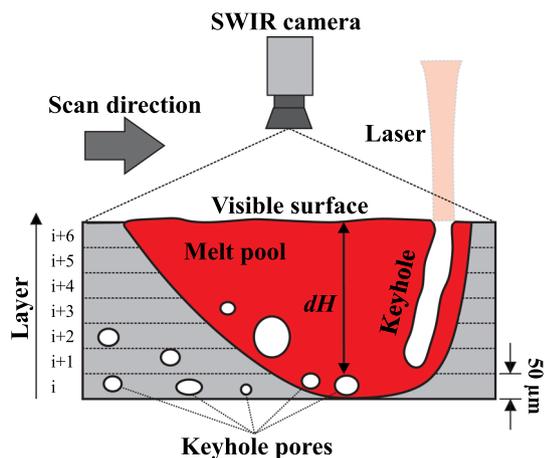
**Fig. 5** Schematical keyhole pore formation with height difference *dH* between the part surface and forming keyhole pores at the bottom of the melt pool

## Sample construction

In this study, the cluster-wise porosity prediction was performed as binary classification, so clusters were labeled either as "non-porous" or "porous". In the following, we denote the cluster whose label shall be predicted as "target cluster". A cluster label *L* was calculated based on the binary XCT voxels within a cluster:

$$L = \begin{cases} 1 \text{("porous"); } if \frac{N_{vox,0}}{N_{vox}} > P_{thresh} \\ 0 \text{("non-porous"); } if \frac{N_{vox,0}}{N_{vox}} \leq P_{thresh} \end{cases} \quad (1)$$

In the given equation, $P_{thresh}$ corresponds to the porosity threshold. Furthermore, $N_{vox,0}$ corresponds to the number of cluster voxels with a digital value of 0 and $N_{vox}$ is equivalent to the total number of cluster voxels. The ratio of $N_{vox,0}$ and $N_{vox}$ is defined as the mean porosity of a single cluster. We note that Eq. 1 considers also single cluster voxels with a digital value of 0, since no strict definition of voids was implemented.

Since the majority of voids in the observed component were keyhole pores, we included the physics of keyhole pore formation (Fig. 5) into the sample construction (Fig. 4c). Keyhole pores form predominantly at the bottom of the melt pool (Hojjatzadeh et al., 2020). Since the melt pool can have a depth of multiple layer thicknesses for standard PBF-LB/M parameters (Mohr et al., 2020), a height difference *dH* can exist between the production surface and the emerged pore. However, the SWIR camera only captures the production surface radiosity. As a result, the emerged pore and its associated thermogram features can be equally disjoined by the vertical distance *dH* in their respective dataset location. Therefore, we postulate that the prediction of keyhole pores demands

the inclusion of feature information from subsequent layers with similar lateral position as the pore.

Hence, we constructed a thermogram feature matrix *M* (Fig. 4c) using feature clusters of *N* subsequent layers as model input. *M* represents the local thermal history of a small vertical region of interest above the target cluster. The idea of using *M* is to enable the model to recognize the complex keyhole pore formation mechanism and further effects such as remelting of pores. Each entry in *M* is a floating number and can be interpreted to correspond to a single layer above the target cluster due to the fixed height of 50 μm (see "Feature Extraction").

For each target cluster in the component, a peculiar matrix *M* was constructed, and a label *L* was defined based on Eq. (1). The last *N* component layers were disregarded for sample construction since the sampling method included *N* + 1 specimen layers to generate a sample. In the case of the last *N* layers, data from outside the component bulk would be used.

## Proposed prediction model

For the porosity prediction based on thermogram features, a custom 1D-CNN classifier was designed. In contrast to standard ML algorithms, CNNs combine feature engineering (or feature extraction) and classification (Baumgartl et al., 2020; Westphal & Seitz, 2021). Even though extensive feature engineering was already performed during pre-processing, the use of 1D filter kernels offers the possibility to detect patterns inside the feature vectors of *M*. Thereby, the classifier is enabled to learn thermal patterns from the exposure of subsequent layers that are assumably connected to keyhole pore formation or healing effects (Ulbricht et al., 2021).

Bayesian Optimization (BO) was used to design the model architecture and to optimize important hyperparameters. BO is an iterative optimization technique which approximates an unknown function (*e.g.*, a ML model) by a surrogate model and makes decisions about the choice of parameters (*e.g.*, model hyperparameters) to maximize a function output (Snoek et al., 2012). For each iteration, the choice of the next hyperparameter set is made by an acquisition function. As acquisition function, Expected Improvement (EI), the Probability of Improvement (PI) or Upper Confidence Band (UCB) can be used. In this study, we chose gaussian processes as surrogate model due to their reported descriptive power and analytic traceability (Klein et al., 2017; Williams & Rasmussen, 2006). As acquisition function, we chose EI since it was found to show improved results in comparison to PI (Snoek et al., 2012) and, in contrast to UCB, did not require its own hyperparameter tuning. A basic model architecture and a limited range of hyperparameters (Table 3) were chosen based on preliminary tests to limit the computational

**Table 3** Hyperparameter space for Bayesian Optimization of the CNN architecture

| Hyperparameter | Type | Parameter space |
|---|---|---|
| Number of convolutional layers | Discrete | [1, 2, 3, 4] |
| Number of filters in convolutional layers | Discrete | [16, 32, 64, 128, 256, 512, 1024, 2056] |
| Kernel size | Discrete | [2, 3, 4, 5] |
| Pooling size | Discrete | [2, 3, 4, 5] |
| Number of dense layers | Discrete | [1, 2, 3, 4, 5] |
| Number of neurons in fully connected layers | Discrete | [16, 32, 64, 128, 256, 512, 1024, 2056] |
| Activation function | Discrete | [Sigmoid, ReLu, tanh] |
| Optimizer | Discrete | [Adam, SGD] |
| Learning rate | Continuous | $[10^{-6}, 10^{-2}]$ |

cost of the model optimization. The basic model architecture consisted of the following elements:

1. Block of one or more convolutional layers
2. Maximum pooling layer
3. Flatten layer
4. Block of one or more fully connected layers
5. Fully connected single output neuron with a sigmoid activation function

We chose the *f1-Score* as target to be maximized during BO. The *f1-Score* is defined in the following chapter.

Two optimization runs were performed to obtain the final model architecture and hyperparameter setup. After the first run (40 random guesses, 40 optimization runs), significant trends could be derived for certain hyperparameters. Therefore, for the second run, these hyperparameters were fixated to decrease the complexity of the optimization problem: Kernel size and pooling size were set to 2, while Rectified Linear Unit (ReLu) was chosen as activation function and Adam (Kingma & Ba, 2015) as optimizer. From the second run (5

random guesses, 15 optimization runs), an optimized hyperparameter setup was derived which is shown in Fig. 6. A learning rate of $1.26 \cdot 10^{-4}$ was used. The model had approximately $3.83 \cdot 10^6$ trainable weights and biases and was used for the experiments in chapter 3.

## Training procedure and model evaluation

75% of samples from of the entire dataset were used for training and 25% for testing. The size of the dataset varied depending on the chosen cluster edge length $w$ from 4972 samples ($w = 1000\,\mu\text{m}$) to 836,426 samples ($w = 100\,\mu\text{m}$). The dataset was split randomly. A subset of $0.2 \cdot 75\% = 15\%$ of the training data was dedicated for the model validation during training. We used early stopping as a callback method and restored the best weights. This was performed since partial overfitting was observed during training, resulting in diverging binary cross-entropy losses. *Accuracy*, *recall*, *precision*, and *f1-Score* were calculated to test the model's performance. These scores are defined by the following equations:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$f1 - Score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \tag{5}$$

Here, *TP* corresponds to the true positives (correctly identified "porous" clusters), *FP* to the false positives ("non-porous" clusters that are falsely identified as "porous"), *FN* to the false negatives ("porous" clusters that are falsely identified as "non-porous") and *TN* to the true negatives (correctly identified "non-porous" clusters). While the *accuracy* monitors the total number of correct predictions, the remaining scores concentrate on the prediction of "porous" samples. In



**Fig. 6** Model architecture derived from hyperparameter tuning by BO. The number of filters/neurons is given by the number in the block center. The layer dimension represents the number of incorporated thermogram features and is given by the number below the block
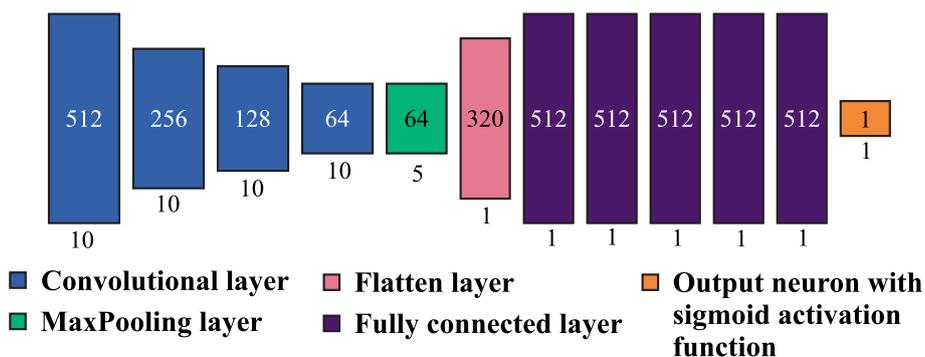
**Fig. 7** Modeling procedure. The application of sampling approaches such as Random Oversampling (RUS) or the Synthetic Minority Oversampling Technique (SMOTE) is optional and is described in the appendix
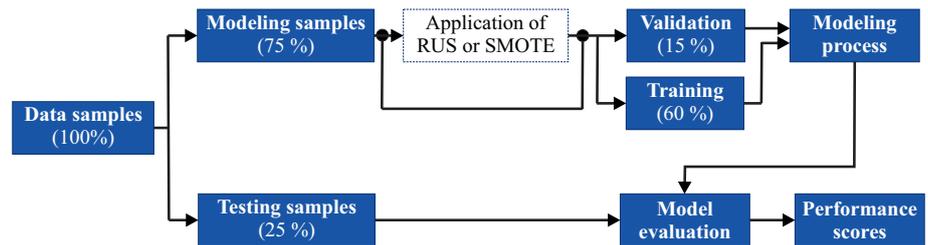


**Table 4** Statistical analysis of average melt pool and time dependent features per component section including the corresponding standard deviation

| Component sections | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| VED in J/mm$^3$ | 65.45 | 81.84 | 65.45 | 98.21 | 65.45 | 114.58 |
| Mean porosity in % | 0.015 | 0.024 | 0.042 | 0.268 | 0.306 | 1.59 |
| MPA in pix | 61.9 ± 17.6 | 68.7 ± 19.8 | 60.8 ± 17.3 | 77.6 ± 23.2 | 61.6 ± 17.3 | 88.6 ± 26.9 |
| MPL in pix | 16.3 ± 5.2 | 17.9 ± 5.8 | 16 ± 5.2 | 19.2 ± 6.1 | 16.3 ± 5.2 | 20.7 ± 6.5 |
| MPW in pix | 5.5 ± 0.9 | 5.7 ± 0.9 | 5.5 ± 1 | 5.9 ± 1.1 | 5.5 ± 1 | 6.1 ± 1.1 |
| MPE [−] | 0.86 ± 0.17 | 0.87 ± 0.16 | 0.86 ± 0.17 | 0.88 ± 0.16 | 0.86 ± 0.17 | 0.88 ± 0.16 |
| MPP in pix | 35.5 ± 10.3 | 38.6 ± 11.2 | 35.1 ± 10.2 | 41.3 ± 11.8 | 35.6 ± 10.2 | 44.4 ± 12.4 |
| MPT$_{mean}$ in K | 1817 ± 281 | 1805 ± 280 | 1816 ± 283 | 1813 ± 281 | 1817 ± 283 | 1820 ± 277 |
| MPT$_{max}$ in K | 2199 ± 357 | 2188 ± 355 | 2192 ± 358 | 2239 ± 363 | 2202 ± 360 | 2281 ± 366 |
| TOT1200K in ms | 62.2 ± 9.8 | 93.6 ± 15.7 | 59.9 ± 8.4 | 130.5 ± 23.6 | 59.1 ± 8.7 | 164.2 ± 34.2 |
| TOT1680K in ms | 8.6 ± 1.5 | 11.8 ± 2.7 | 8.4 ± 1.5 | 16.2 ± 4.6 | 8.5 ± 1.5 | 21.5 ± 7.1 |
| TOT2400K in ms | 0.13 ± 0.11 | 0.13 ± 0.13 | 0.11 ± 0.1 | 0.33 ± 0.26 | 0.13 ± 0.11 | 0.59 ± 0.39 |

the context of porosity prediction for PBF-LB/M parts, undiscovered "porous" clusters could represent a serious risk to component safety. Therefore, the *recall* is especially important to monitor since it includes *FN* samples. In comparison, the *precision* monitors "false alarms" by including *FP*. The *f1-Score* depicts the harmonic mean of *precision* and *recall*.

Each model was trained five-fold using the *ShuffleSplit* method of the python package scikit-learn (Buitinck et al., 2013) and the performance results were averaged. Thereby, possible model inconsistencies were eliminated that may result from the random splitting into test and training subsets. TensorFlow (Abadi et al., 2016) was used for the construction of the model architecture and further modeling procedure. The training was performed on a Nvidia Geforce RTX3090 GPU with computational times ranging from 40 to 150 s. This computational time was dependent on the number of samples and the chosen batch size. The entire data training procedure is visualized in Fig. 7.

## Results and discussion

### Statistical evaluation of thermogram feature and porosity interaction

A first statistical evaluation of the extracted thermogram features was performed to fundamentally assess the complexity

of porosity prediction, and to examine the necessity of ML in this matter.

In Table 4, the mean thermogram feature values and the corresponding standard deviations are given regarding the different component sections. Furthermore, the VED and the averaged section porosity are listed. The results show that most melt pool-based features and all time-dependent features are sensitive to changes in VED. Additionally, they seem to be indicators for global changes in porosity. MPE and MPT$_{mean}$ are exceptions that remain nearly constant for changing VED.

Furthermore, we calculated the correlation coefficients for the feature cluster values and the porosity labels that were constructed in "Data pre-processing" section. This analysis was carried out to evaluate if a linear correlation is present between porosity and features on a local component scale. The correlation coefficient is given by the following equation (Kamath, 2016):

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{6}$$

*Cov* describes the covariance of inputs *X* and *Y,* and $\sigma_X, \sigma_Y$ correspond to the respective standard deviation. *Corr(X,Y)* is a measure of the linear relationship between two populations. For the calculation of *Corr(X,Y)*, we chose the following pre-processing parameters based on the results of preliminary

**Voxel size = (100 x 100 x 50) μm³**

| Subsequent layer | MPA | MPL | MPW | MPE | MPP | MPT$_{mean}$ | MPT$_{max}$ | TOT1200K | TOT1680K | TOT2400K |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | -0.17 | -0.21 | -0.24 | -0.36 | -0.22 | -0.33 | -0.33 | 0.15 | 0.13 | 0.11 |
| 8 | -0.17 | -0.21 | -0.24 | -0.36 | -0.23 | -0.33 | -0.33 | 0.14 | 0.13 | 0.12 |
| 7 | -0.18 | -0.22 | -0.24 | -0.36 | -0.23 | -0.33 | -0.33 | 0.13 | 0.13 | 0.11 |
| 6 | -0.19 | -0.22 | -0.25 | -0.36 | -0.23 | -0.33 | -0.33 | 0.12 | 0.11 | 0.11 |
| 5 | -0.19 | -0.22 | -0.25 | -0.36 | -0.24 | -0.33 | -0.33 | 0.11 | 0.10 | 0.10 |
| 4 | -0.20 | -0.23 | -0.25 | -0.36 | -0.24 | -0.33 | -0.33 | 0.09 | 0.09 | 0.08 |
| 3 | -0.22 | -0.24 | -0.26 | -0.37 | -0.25 | -0.33 | -0.33 | 0.07 | 0.07 | 0.07 |
| 2 | -0.23 | -0.24 | -0.27 | -0.37 | -0.26 | -0.33 | -0.33 | 0.05 | 0.05 | 0.06 |
| 1 | -0.24 | -0.25 | -0.27 | -0.37 | -0.26 | -0.33 | -0.34 | 0.03 | 0.03 | 0.04 |
| 0 | -0.25 | -0.26 | -0.27 | -0.37 | -0.27 | -0.33 | -0.34 | 0.00 | 0.00 | 0.02 |

a)

**Voxel size = (500 x 500 x 50) μm³**

| Subsequent layer | MPA | MPL | MPW | MPE | MPP | MPT$_{mean}$ | MPT$_{max}$ | TOT1200K | TOT1680K | TOT2400K |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 0.41 | 0.21 | 0.44 | -0.04 | 0.23 | 0.02 | 0.30 | 0.70 | 0.68 | 0.55 |
| 8 | 0.40 | 0.21 | 0.43 | -0.04 | 0.22 | 0.02 | 0.30 | 0.70 | 0.68 | 0.56 |
| 7 | 0.38 | 0.20 | 0.41 | -0.05 | 0.21 | 0.03 | 0.29 | 0.68 | 0.66 | 0.55 |
| 6 | 0.36 | 0.18 | 0.40 | -0.05 | 0.20 | 0.04 | 0.29 | 0.65 | 0.64 | 0.54 |
| 5 | 0.34 | 0.17 | 0.36 | -0.05 | 0.18 | 0.04 | 0.27 | 0.61 | 0.61 | 0.52 |
| 4 | 0.30 | 0.14 | 0.34 | -0.05 | 0.16 | 0.04 | 0.25 | 0.57 | 0.57 | 0.48 |
| 3 | 0.27 | 0.12 | 0.32 | -0.06 | 0.14 | 0.05 | 0.23 | 0.53 | 0.53 | 0.46 |
| 2 | 0.24 | 0.11 | 0.28 | -0.06 | 0.12 | 0.05 | 0.21 | 0.48 | 0.49 | 0.43 |
| 1 | 0.20 | 0.08 | 0.25 | -0.06 | 0.09 | 0.04 | 0.17 | 0.43 | 0.44 | 0.38 |
| 0 | 0.17 | 0.06 | 0.22 | -0.07 | 0.07 | 0.04 | 0.15 | 0.37 | 0.40 | 0.34 |

b)

**Voxel size = (1000 x 1000 x 50) μm³**

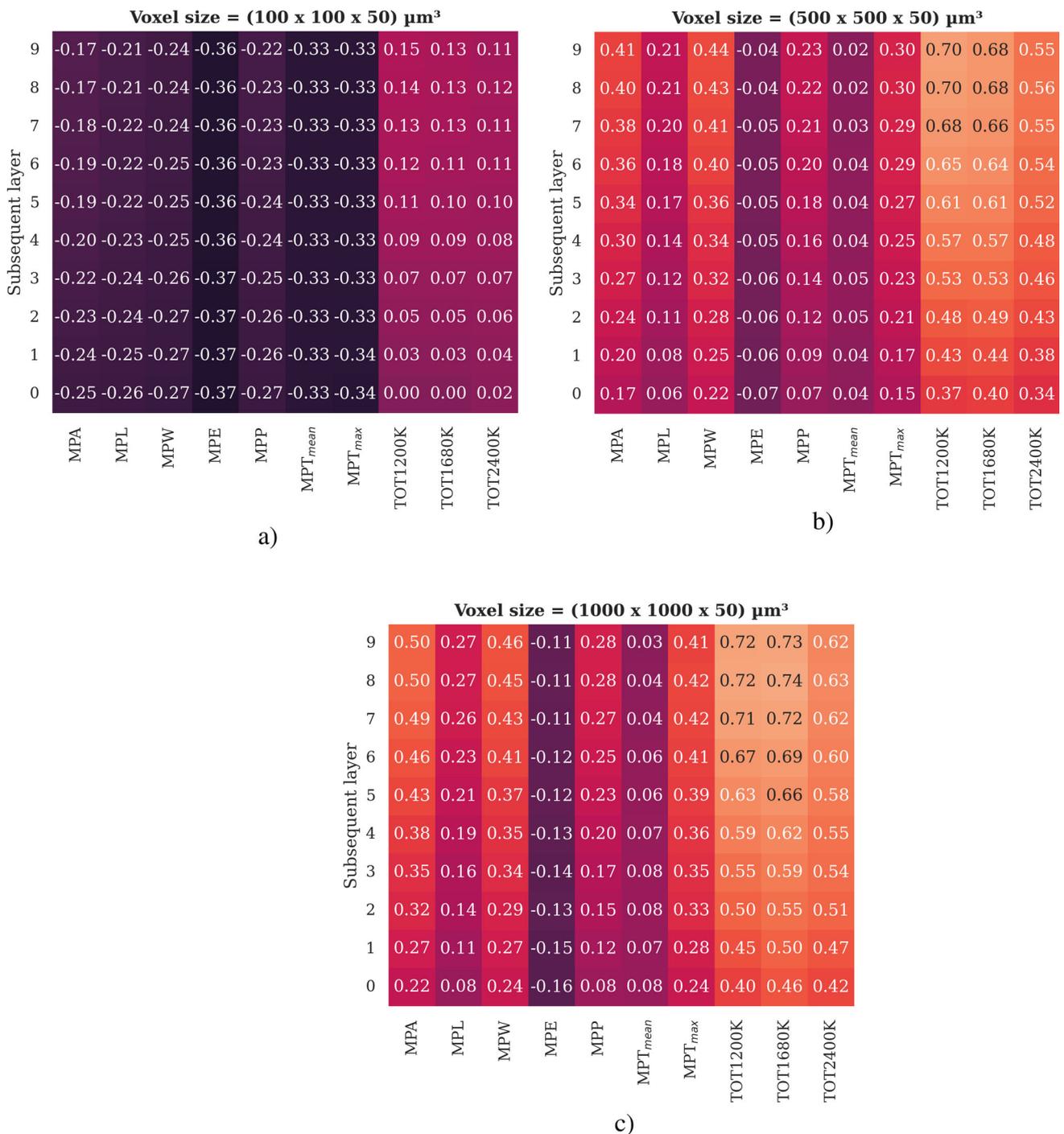| Subsequent layer | MPA | MPL | MPW | MPE | MPP | MPT$_{mean}$ | MPT$_{max}$ | TOT1200K | TOT1680K | TOT2400K |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 0.50 | 0.27 | 0.46 | -0.11 | 0.28 | 0.03 | 0.41 | 0.72 | 0.73 | 0.62 |
| 8 | 0.50 | 0.27 | 0.45 | -0.11 | 0.28 | 0.04 | 0.42 | 0.72 | 0.74 | 0.63 |
| 7 | 0.49 | 0.26 | 0.43 | -0.11 | 0.27 | 0.04 | 0.42 | 0.71 | 0.72 | 0.62 |
| 6 | 0.46 | 0.23 | 0.41 | -0.12 | 0.25 | 0.06 | 0.41 | 0.67 | 0.69 | 0.60 |
| 5 | 0.43 | 0.21 | 0.37 | -0.12 | 0.23 | 0.06 | 0.39 | 0.63 | 0.66 | 0.58 |
| 4 | 0.38 | 0.19 | 0.35 | -0.13 | 0.20 | 0.07 | 0.36 | 0.59 | 0.62 | 0.55 |
| 3 | 0.35 | 0.16 | 0.34 | -0.14 | 0.17 | 0.08 | 0.35 | 0.55 | 0.59 | 0.54 |
| 2 | 0.32 | 0.14 | 0.29 | -0.13 | 0.15 | 0.08 | 0.33 | 0.50 | 0.55 | 0.51 |
| 1 | 0.27 | 0.11 | 0.27 | -0.15 | 0.12 | 0.07 | 0.28 | 0.45 | 0.50 | 0.47 |
| 0 | 0.22 | 0.08 | 0.24 | -0.16 | 0.08 | 0.08 | 0.24 | 0.40 | 0.46 | 0.42 |

c)

**Fig. 8** Correlation coefficients between labels and features determined for three different sets of cluster dimensions. The extracted thermographic features are given on the abscissa and the voxel position above the label is shown on the ordinate

tests: The porosity threshold $P_{thresh}$ (see Eq. 1) was set to 0.1% and we included 9 subsequent layers for the sample construction. Since the influence of the cluster dimension on the prediction performance was unknown, we calculated *Corr(X,Y)* for cluster edge lengths $w$ of 100 μm, 500 μm and 1000 μm for a first assessment. The results are shown in Fig. 8.

Low correlation coefficients in the range of −0.37 to 0.15 were determined for small cluster dimensions of $(100 \times 100 \times 50)$ μm³. The melt pool-dependent features were correlated more strongly with the porosity labels than

the time-dependent features, but with entirely negative coefficients. This is surprising since it is expected that melt pool size and temperature increase in areas of increased porosity, as was shown in the results listed in Table 4. Furthermore, we observed little differences in correlation between subsequent layers.

For larger cluster dimensions of (500 × 500 x 50) $\mu m^3$ and (1000 × 1000 x 50) $\mu m^3$, we observed larger and mostly positive correlation coefficients in the range of -0.16 to 0.73. Especially the TOT features showed increased coefficients with values up to 0.73 (Fig. 8c). In comparison, the coefficients for MPE and $MPT_{mean}$ were small and ranged from -0.16 to 0.08 (Fig. 8b and c).

These observed variations in correlation indicate that specific features (such as TOT) are more significant indicators then others (MPE, $MPT_{mean}$) for the void formation. The results reveal a further important trend: The coefficients from higher subsequent layers tend to correlate stronger with the target cluster label ("porous" or "non-porous") than those of lower layers. This indicates that features from subsequent layers contain crucial information related to void formation. However, only small deviations are observed between coefficients of two subsequent layers. Furthermore, for both larger cluster dimensions, we observed a high degree of similarity in the correlation coefficient values and their distribution.

In the following, we summarize the main findings of chapter 3.1:

1. A large variety of features from different subsequent layers have weak to moderate (−0.17 to 0.73) linear correlation with the porosity label. This meets the set expectations that the complex stochastic nature of void formation is hard to capture with a simple, linear model due to highly dynamic melt pool properties (Gaikwad et al., 2022). Therefore, the use of a multivariate nonlinear ML model is necessary for the prediction of porosity labels. The correlation coefficients indicate that specific features might have a larger impact than others. Nonetheless, all extracted thermogram features are used as input for the model. Thereby, no potentially important feature information is disregarded. A detailed investigation of the feature influence on the prediction ability is out of the scope of this study and will be tackled in future work.

2. The results show that the correlation coefficients are sensitive to changes of pre-processing parameters such as the cluster dimension. Furthermore, we found increasing correlation coefficients for increasing numbers of included subsequent layers. From these observations, we conclude that pre-processing parameters have a serious impact on the prediction performance. Additional results concerning the impact of pre-processing parameters on the prediction performance are shown in the upcoming chapters.
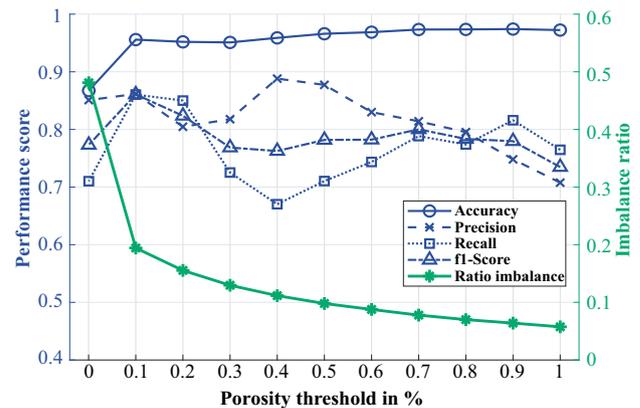


**Fig. 9** Model performance depending on the set porosity threshold and imbalance ratio between both classes for a cluster dimension of (700 × 700 x 50) $\mu m^3$ and nine included subsequent layers

## Influence of porosity threshold on prediction performance

In the following, we present our study on the influence of the porosity threshold on the prediction performance. The porosity threshold was used to label the data samples into "porous" or "non-porous" and, therefore, affected the intrinsic composition of the dataset. We trained models for eleven thresholds $P_{thresh}$ ranging from values of 0% to 1%. We kept the cluster dimensions of (700 × 700 × 50) $\mu m^3$ constant and chose the number of considered voxels of subsequent layers to $N = 9$. The constant pre-processing parameters were set based on preliminary tests to keep the computational effort within reasonable limits. Under these conditions, the dataset consisted of 11,752 samples. Furthermore, we monitored the ratio of "porous" to "non-porous" clusters to describe the class imbalance in the dataset since its influence on the prediction performance was unknown. The results are visualized in Fig. 9.

The results showed that the model performances fluctuated for changing $P_{thresh}$ values. A threshold of 0.1% produced the highest *f1-Score* (~ 0.86) and high *accuracy* (~ 0.95). For 0% and from 0.2% to 1%, the *f1-Score* decreased below 0.8. We observed that the *precision* score was mostly higher (at maximum 0.89) than the f1-Score, while equal values were achieved for $P_{thresh}$ values of 0.1% and 0.8%. For 0.2%, 0.9% and 1%, the *precision* values were decreased in comparison to the *f1-Score*. For the *recall* score, we found reverse behavior with approximately equal differences to *f1-Scores* when compared to the *precision* score. Generally, the *accuracy* score remained at a high level above 0.86 regardless of the threshold. Furthermore, the imbalance ratio decreased from 0.48 to 0.06 for increasing $P_{thresh}$ values. A large drop is present for threshold values between 0 and 0.1%.

For varying $P_{thresh}$ values, the *accuracy* remained high above 0.86. This observation was made for all experiments in

this study (see "Influence of cluster dimension on prediction performance" and "Influence of inclusion of subsequent layers on prediction performance"). It can be concluded that the *accuracy* as a performance score has very limited capability to give reasonable information about the model performance. The low significance of *accuracy* for imbalanced datasets is a known issue in the machine learning community (Chawla, 2005). Therefore, further interpretations are mainly based on *recall*, *precision* and *f1-Score*.

The increased performance scores for $P_{thresh} = 0.1\%$ are especially outstanding in the scope of the studied porosity thresholds. A possible explanation is that $P_{thresh} = 0.1\%$ separates the induced porosity (keyhole voids) in component sections of increased VED from the inherent porosity in component sections of nominal VED. Thereby, the data samples are split into classes whose void formation mechanisms are physically different. This might not be the case for a randomly chosen threshold. The drop in imbalance ratio between $P_{thresh} = 0\%$ and $0.1\%$ supports this explanation approach: It shows that the number of "porous" clusters is significantly increased for $P_{thresh} = 0\%$ in comparison to $P_{thresh} = 0.1\%$. This outstanding rise indicates that clusters are labelled as "porous" that contain inherent porosity from component sections of nominal VED. If $P_{thresh} = 0\%$ is set, all clusters are labeled as "porous" that contain already one or more XCT voxel with a binarized grey level of 0. This would include small voids or potential errors from the XCT data binarization which might be difficult to be predicted by the model.

The results indicate that the porosity threshold as a pre-processing parameter should be set carefully since it can influence model performance detrimentally. Since we aim to predict small porosity values with high probability, a porosity threshold of 0.1% is identified as optimum for this study.

The intrinsic dataset imbalance of majority ("non-porous") and minority ("porous") classes (Fig. 9) can be detrimental for the model performance. Class imbalance is a known challenge in ML classification tasks (Han et al., 2005). Therefore, we performed additional experiments for datasets that were adjusted by two different sampling methods, Random Undersampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE). The results showed that the sampling methods provided no performance improvement in comparison to original dataset (see appendix "Study on influence of class imbalance"). Therefore, the application of sampling methods was not regarded for the following experiments.

## Influence of cluster dimension on prediction performance

As already pointed out in "Statistical evaluation of thermogram feature and porosity interaction", the impact of
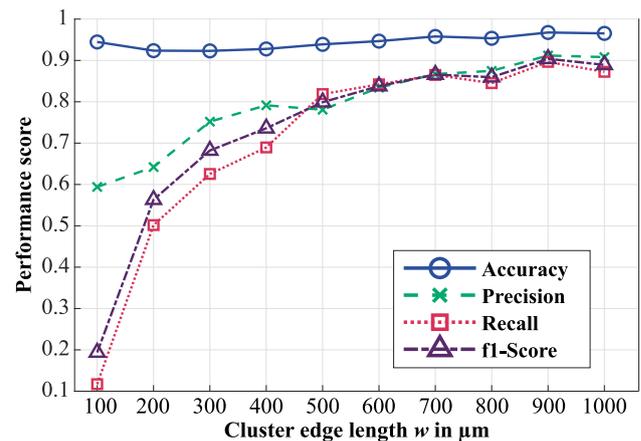


**Fig. 10** Influence of cluster dimension on model performance for a porosity threshold of 0.1% and considered voxels from $N = 9$ subsequent layers

the cluster dimension on the prediction performance of the classifier is of interest. In order to study this impact, we trained models using ten datasets with different cluster edge lengths $w$. The edge lengths ranged from of 100 μm (836426 samples) to 1000 μm (4972 samples). The cluster height remained constant at 50 μm (corresponding to the layer thickness). Thereby, each cluster corresponded to a single subsequent layer. For the experiment, $P_{thresh}$ was set to 0.1% and $N = 9$ subsequent layers were included to build the sample matrix $\boldsymbol{M}$.

The results (Fig. 10) showed that the *precision*, *recall* and *f1-Scores* increased with increasing $w$ for $w \geq 200$ μm. A plateau was reached for approx. $w \approx 700$ μm. The highest performance scores were observed at a cluster dimension of $(1000 \times 1000 \times 50)$ μm³. For $w$ equals 100 μm and 200 μm, the model showed decreased performance scores of 0.65 or below.

Especially the low *recall* scores (< 0.51) emphasize the limited capability of predicting porosity for small cluster dimension with $w \leq 200$ μm. Possible reasons for the low prediction performance at small cluster dimensions are the different causes of spatial uncertainty within the thermogram feature data (see chapter 3.1). Firstly, the camera parametrization and the resulting spatial resolution may lead to a decrease of the porosity prediction scores. The available data points for the construction of 3D feature datasets are limited due to the spatial resolution of approx. 105 μm/pix and the melt pool travelling distance between frames of 110 μm to 200 μm (see "Resampling to clusters"). This data "sparsity" may affect the performance of the model especially at lower cluster dimensions.

Secondly, the registration uncertainties may be a detrimental factor. We calculated maximum uncertainties of $64 \pm 27$ μm for lateral dislocation and $55 \pm 23$ μm for
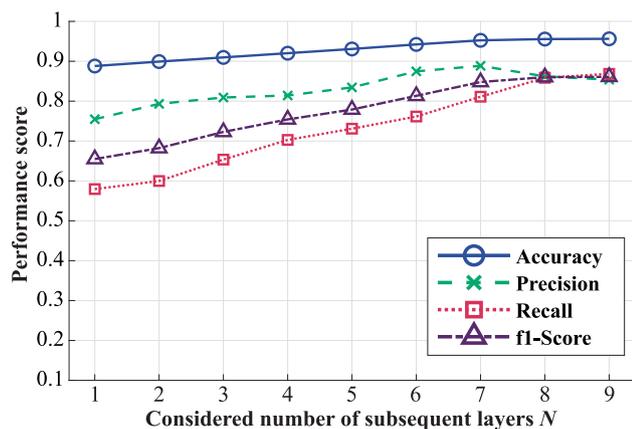
**Fig. 11** Influence of considered number of subsequent layers on model performance for a porosity threshold of 0.1% and a target voxel size of $(700 \times 700 \times 50)\,\mu m^3$

lateral scaling (Table 1). Especially for small cluster dimensions, the registration uncertainties could lead to insufficient alignment between thermogram feature and XCT clusters. Thereby, the model would be trained on spatially disjoint samples which could result in a low ability to predict porosity.

The results indicate that the cluster dimension needs to be increased until the influence of the spatial data uncertainties on model performance becomes neglectable.

## Influence of inclusion of subsequent layers on prediction performance

The inclusion of subsequent layers for the construction of $M$ is of interest since it might enable the model to recognize complex effects such as re-melting of material and keyhole pore formation. Therefore, the number of included subsequent layers $N$ is expected to be significant for the porosity prediction (see "Statistical evaluation of thermogram feature and porosity interaction"). We trained models for varying $N$ in the range of 1 to 9 which resulted in input vector dimensions ranging from [2,1] to [10,1]. A minimum vector dimension of [2,1] was mandatory to perform the convolution since the kernel size was set to 2 (see chapter 2.5). $P_{thresh}$ was kept constant at 0.1% and the cluster dimension was chosen to $(700 \times 700 \times 50)\,\mu m^3$ (11,752 samples).

The results (Fig. 11) showed that the prediction performance increased with an increasing $N$. For $N = 7$, a plateau was reached. Although the *recall* score was constantly lower than the *f1-Score*, both were surpassed by the *precision*.

The resulting performance scores indicate that the inclusion of subsequent layers improves the prediction performance. This assertion is supported by the physics of the keyhole void formation: As stated beforehand, the melt pool penetrates multiple layers of powder and solid material causing a height difference $dH$ between pore location
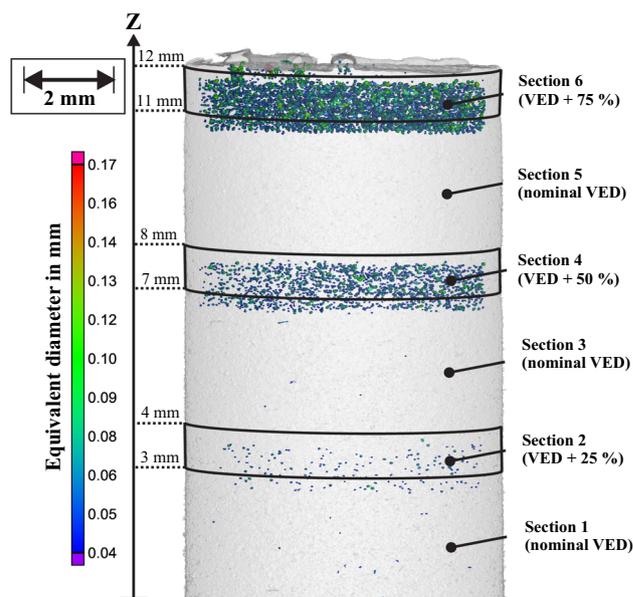


**Fig. 12** Visualization of the void distribution of the XCT data. Component sections of applied increased VED are highlighted. The equivalent void diameter is indicated by the color and corresponds to the diameter of a sphere with similar volume. The component sections are indicated by the dashed lines. Pores belonging to the contour scan are blanked out for increased visibility

and the location of the thermal information. Depending on the applied VED, the depth of the melt pool varies (Mohr et al., 2020), and therefore, also the location of the corresponding thermal information may vary in the sample matrix $M$. In Oster et al. (2022), we approximated the melt pool depth of the PBF-LB/M process to vary in a range of $213 \pm 19\,\mu m$ and $471 \pm 54$ which corresponds to 4 to 9 layers. The results shown in Fig. 11 indicate that a mean melt pool depth of approximately 350 to 450 $\mu m$ was present when voids emerged. This assumption is based on the observation that the largest number of "porous" clusters were identified correctly if $N$ exceeded seven layers.

The results indicate that the best model performance is achieved if the added heights of the included subsequent layers exceed the maximum melt pool depth during manufacturing.

## Evaluation of local model performance

In this chapter, we evaluate the model performance by component section since the void formation was controlled by section-wise changes of the VED. Thereby, the model's capability of porosity prediction for varying void distributions and sizes is studied. Therefore, we visualized the void distribution in the component bulk of the XCT data (Fig. 12). Furthermore, we trained models with the optimal

**Table 5** Performance scores per component section averaged from five independent training and testing cycles

| Component section | Number of "non-porous" clusters | Number of "porous" clusters | Mean porosity of "porous" clusters in % | *Accuracy* | *Recall* | *Precision* | *f1-Score* |
|---|---|---|---|---|---|---|---|
| 1 | 3794 | 166 | 0.22 ± 0.13 | 0.956 | 0.084 | 0.368 | 0.137 |
| 2 | 1239 | 89 | 0.23 ± 0.13 | 0.923 | 0.056 | 0.227 | 0.09 |
| 3 | 3379 | 365 | 0.41 ± 0.29 | 0.959 | 0.923 | 0.731 | 0.816 |
| 4 | 589 | 710 | 0.49 ± 0.38 | 0.854 | 0.977 | 0.799 | 0.879 |
| 5 | 3340 | 613 | 1.42 ± 0.81 | 0.989 | 0.951 | 0.983 | 0.967 |
| 6 | 0 | 397 | 1.51 ± 0.85 | 1.0 | 1.0 | 1.0 | 1.0 |
| Entire sample | 11,341 | 2340 | 0.84 ± 0.78 | 0.955 | 0.867 | 0.853 | 0.86 |

pre-processing parameters identified in the previous chapters. $P_{thresh}$ was set to 0.1%, the cluster dimension was set to $(700 \times 700 \times 50) \, \mu m^3$ and $N = 9$ subsequent layers were considered (resulting in 11,752 samples). Five different models were independently trained using the *ShuffleSplit* method based on the entire component dataset. Each model was tested on six subsets, each corresponding to a specific component section (see Fig. 2). We calculated the mean *accuracy*, *recall*, *precision,* and *f1-Scores* from all five independent models (Table 5).

The void distribution in the component bulk (Fig. 12) showed that voids occurred not only in the component sections of increased VED but also in the upper parts of the subjacent component sections. This distribution results from the explained formation mechanism of keyhole pores (see Fig. 5). Furthermore, only few voids were located in the upper 40% of layers in component sections 2, 4 and 6. Therefore, especially component sections 3 to 6 contained both "porous" and "non-porous" clusters. In component sections 1 and 2, a similar observation was made. However, the numbers of voids were significantly lower in these sections. Furthermore, the equivalent void diameter showed that the void size increased with increasing VED.

The results showed that small performance scores (*precision*, *recall* and *f1-Score* below 0.368) were achieved in component sections 1 and 2 (Table 5). Possible reasons for this might be the low number of "porous" clusters with low mean porosity of (0.22 ± 0.13) % and (0.23 ± 0.13) %, respectively. This indicates that the model was trained on few samples with porosity values in the vicinity to $P_{thresh} = 0.1\%$. In comparison, "porous" clusters from component sections 3 to 6 show higher mean porosity values in this regard. This indicates that the porosity prediction is challenging if "porous" target clusters appear sparsely and have a mean porosity close to $P_{thresh}$.

The results in component sections 3 and 4 showed high *f1-Scores* (> 0.816) with increased *recall* (> 0.923) and decreased *precision* (> 0.731). An increased number of "porous" target clusters with increased mean porosity could

be a possible reason for this result. Remarkably, the increased *recall* scores indicate that the model predicts the present keyhole pores with high probability despite the described height difference *dH* between thermal information and emerged void (see "Resampling to clusters").

Especially in component sections 5 and 6, "porous" clusters were predicted correctly with a very high probability (> 0.951). The overwhelming number of "porous" clusters with very high mean porosity compared to the porosity of "non-porous" clusters might help the classifier. Similar to component sections 3 and 4, the present keyhole pores are predicted almost entirely correct.

Regarding the entire sample, the results showed a mean *f1-Score* of 0.86 and a mean *recall* score of 0.867. Furthermore, a mean porosity of 0.84 ± 0.78% for "porous" clusters was found. Regarding the cluster dimension of $(700 \times 700 \times 50) \, \mu m^3$, this mean porosity corresponds to an average void volume of approximately $5.6 \cdot 10^3 \, \mu m^3$.

To better illustrate this result, the following example can be used: In the hypothetical case where the entire volume is referred to a single circular void, the corresponding void diameter would be 73.3 μm. If we conservatively consider the upper limit of the standard deviation (mean porosity of 1.62%), the resulting void volume would be $8.7 \cdot 10^3 \, \mu m^3$ and the corresponding void diameter would be 93.2 μm, respectively. This hypothetical example leads to the interpretation that the model would detect the presence of single circular void with a diameter of 73.3 μm (or 93.2 μm, respectively) with a probability of 86.7% (*recall*). However, we note that usually more than one void was present per cluster.

Furthermore, we closely studied the impact of differences in local porosity on the model. Therefore, we observed a component subsection comprised of component sections 3 and 4 since a high variance in cluster porosity was found here (see Fig. 13). This subsection corresponded to a component height from Z = 6.5 mm to 7.5 mm and contained a major part of "porous" clusters of the named component sections. Here, we recorded the raw output of the model's final sigmoid neuron (see Fig. 6). The output of a sigmoid
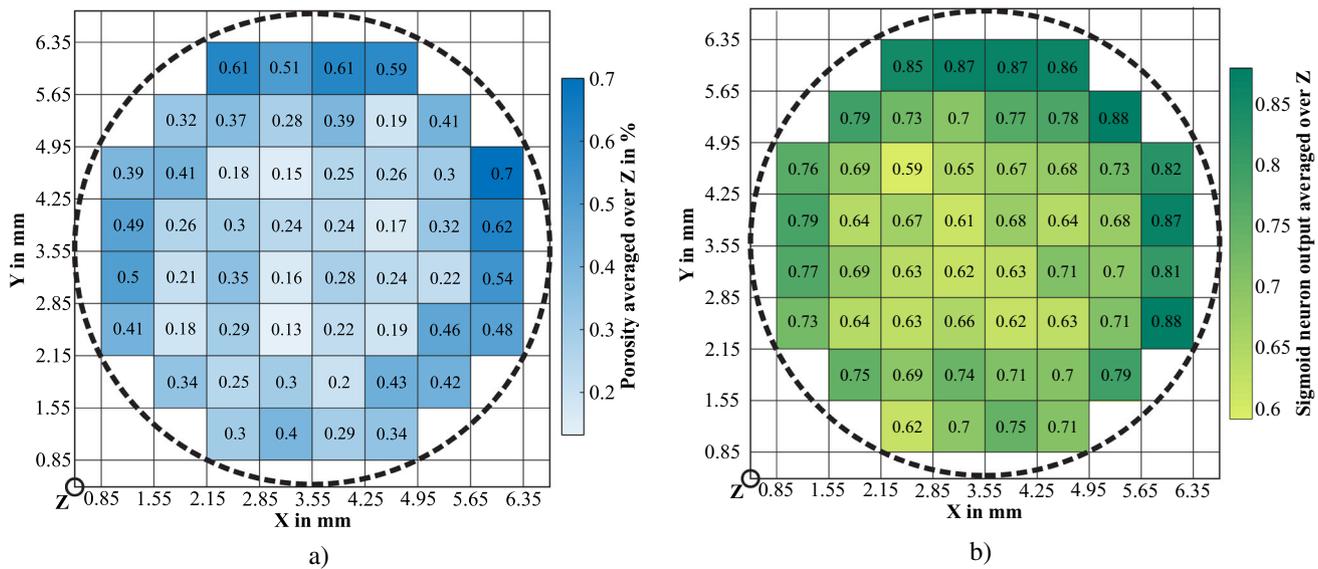
**Fig. 13** Porosity vs. raw model prediction averaged along the Z axis from 6.5 to 7.5 mm. In this area, a major part of "porous" clusters was located resulting from the VED increase in component sections 3, 4 and 6. The dashed line indicates the outer rim of the cylindric specimen. **a** Average porosity per target voxel. **b** Average output of sigmoid neuron before the classification threshold is applied. It is shown that the model produced increased outputs for higher porosity values especially for the component rim

neuron produces a floating-point number between 0 and 1. For binary classification, a threshold (from here denoted as "classification threshold") is applied to distinguish between both classes. In this study, we used the standard classification threshold of 0.5. We averaged the cluster porosity and the sigmoid output vertically over the component height Z to visualize a possible correlation (see Fig. 13).

The results (Fig. 13a) showed that the rim clusters had increased porosity compared to the clusters close to the component center. As shown in Fig. 13b), the model mostly produced increased sigmoid outputs at these rim clusters. In other words, the model predicted an increased probability for the presence of voids for clusters with increased porosity. This is remarkable since the model was trained using samples that did not contain information about the actual cluster porosity. For each sample, only the binary cluster label ("porous" or "non-porous") was available. Nonetheless, the results indicate that the model is able to distinguish between clusters of increased and decreased porosity based on the thermogram features. This can be also seen in the specimen center, where the model produces lower outputs near to the classification threshold of 0.5 for porosity values closer to $P_{thresh} = 0.1\%$.

This is, to our knowledge, the first time that comparable observations are made for in-situ monitoring of PBF-LB/M. Based on these observations, we conclude that crucial information about the probability of local void formation is encoded in the thermogram feature data. Furthermore, we postulate that a sufficiently powerful model can predict changes in local porosity based on the thermogram feature input.

## General discussion of framework

In comparison to related irregularity prediction frameworks (Aminzadeh & Kurfess, 2018; Gaikwad et al., 2022; Gobert et al., 2018; Krabusch et al., 2020; Lough et al., 2022; Mohr et al., 2020; Ren et al., 2023; Smoqi et al., 2022; Snow et al., 2021; Xiao et al., 2020), the presented approach offers significant advantages: Firstly, the model predicts local porosity clusters-wise instead of predicting data that was labelled on a large component scale such as entire layers (Aminzadeh & Kurfess, 2018; Smoqi et al., 2022) or entire specimen (Gaikwad et al., 2022). A cluster-based prediction is advantageous because it enables the detection of safety–critical porosity in small-scale component regions that are subjected to high mechanical loads during operation. Secondly, the inclusion of subsequent layers into the model input significantly increases the recognition of complex PBF-LB/M phenomena such as keyhole pore formation. The advantages of this method were already recognized by Krabusch et al., 2020. This indicates that models for keyhole porosity prediction which are based for example on single track data (Ren et al., 2023) might have decreased performance if applied to real 3D printing data. Thirdly, due to the enhanced pre-processing and feature engineering, the necessary model architecture

remains simple especially compared to raw data approaches (Krabusch et al., 2020; Westphal & Seitz, 2021). Therefore, the training process is computationally cheap (training time below 3 min on a Nvidia Geforce RTX3090 GPU) and the model's small layer depth (only twelve model layers) and dimension benefit its interpretability (Samek et al., 2021).

A possible drawback in comparison to other irregularity prediction approaches is the limited real-time capability of the framework which is strongly dependent on the number of included subsequent layers $N$. If $N = 9$ is assumed, a first valid prediction result would be achieved after the exposure of 10 layers. However, in recent studies, the real-time analysis of data was identified as a crucial factor for in-situ defect prediction (Grasso et al., 2021; McCann et al., 2021). This demand for real-time analysis should be reconsidered since it ignores PBF-LB/M phenomena such as re-melting of material to close voids and the keyhole pore formation. Due to these phenomena, the manufacturing of a layer can significantly change the void distribution of subjacent layers. Therefore, a void prediction in real-time based on data from a single layer without considering subsequent layers has only little significance in this context. However, as pointed out by Grasso et al. (2021), it is still important to process large amounts of data in adequate time to meet the manufacturing speed of PBF-LB/M processes.

For the application of the framework in an industrial environment, a suitable camera and, *e.g.*, an FPGA device could be used for feature extraction, data pre-processing and model execution at an appropriate computational speed. Comparable implementations were already shown by multiple studies (Kwon et al., 2020; Lane & Yeung, 2020; Modaresialam et al., 2022). Another challenging aspect for the framework application is the hardware limitation of the used infrared camera. To meet the PBF-LB/M dynamics, high acquisition frequencies are necessary which limit the size of the camera's field-of-view. Thereby, the effective component surface is reduced which can be monitored. However, a decrease of the acquisition frequency leads to a detrimental reduction of the available thermal process information. The camera used in this study was restricted to a maximum field-of-view of $(9.5 \times 9.5)$ mm$^2$ using an acquisition frequency of 3600 Hz. A change from off-axis to on-axis monitoring (field-of-view changing with the laser position) could help to overcome this issue since a smaller field-of-view could be used.

Concerning the chosen model architecture, the prediction results shown in Fig. 13 indicate that the model can detect porosity differences from thermogram feature data. This suggests that the underlying keyhole prediction can be solved using a regression model. In this study, we decided to perform a simple binary classification task since the potential of the feature data for porosity prediction was a-priori unclear. Since the achieved results are promising but not optimal (*e.g.*, *f1-Score* of 0.86), we expect that a regression task would lead to reduced model performances. The use of enhanced acquisition hardware or the expansion of the thermogram feature portfolio (*e.g.*, by spatter features) could be enabling factors to use regression models in future works.

During manufacturing, we forced the formation of keyhole porosity by decreasing the scan velocity and, thereby, increasing the VED. However, unstable keyhole conditions can be caused by multiple other manufacturing conditions. Prominent examples are high laser power settings, spot size drifts, or unfavorable scan strategies which result in local overheating (Ren et al., 2023). Hence, the presented framework might show reduced performances when facing unseen data produced under these conditions. One main underlying problem is that the described further manufacturing conditions are not represented in the training data used in this study. Due to the same reason, data from different materials and machines could also pose a challenge to the framework.

However, as was stated by Lough et al. (2022), it is assumed that the void formation events described in this study are closely related to thermal events of the thermal history. Therefore, we speculate that the presented prediction framework might show comparable performance for different keyhole void mechanisms if we expand the training datasets accordingly. In the scope of this study, the used dataset was sufficient to emphasize the potential and the advantages of the presented framework for keyhole porosity prediction. For our future work, we aim to apply the framework on further datasets that may include additional manufacturing conditions, materials, and machines.

## Conclusion

In this study, we presented a framework consisting of a data pre-processing workflow and a supervised CNN architecture for the prediction of porosity in local sub-volumes ("clusters") from thermogram features. We used registered thermogram features and XCT data from the manufacturing of a PBF-LB/M/316L part (Oster et al., 2022) as data basis. The pre-processing workflow included the resampling to coarser resolved clusters and the construction of a sample matrix including thermogram feature information of subsequent layers. We labeled the data based on the mean cluster porosity into "porous" and "non-porous". An 1D-CNN classifier was developed and optimized using Bayesian Optimization. We studied the prediction performance under the influence of the following pre-processing parameters: The porosity threshold $P_{thresh}$, the dimension of the clusters (by adjusting the cluster edge length $w$) and the number of included subsequent layers $N$. Furthermore, we performed an in-depth evaluation of the relationship between the local porosity and the model prediction. In the following, we summarize the main findings of this paper:

1. The presented framework achieved an *accuracy* of 0.96, an *f1-Score* of 0.86, a *recall* of 0.87 and a *precision* of 0.85 for the prediction of porosity above a threshold of 0.1% in clusters with the dimension of $(700 \times 700 \times 50)\,\mu m^3$. These results were achieved using 25% of the entire dataset for testing. Furthermore, the model showed increased *recall* scores of above 0.92 in component sections 4 to 6 that exhibit porosity from an increase of VED (Volumetric Energy Density) of at least 50%. In terms of detectable void size, we approximated a hypothetical detectable void diameter of 73.3 $\mu m$ if a single circular void is assumed. The achieved prediction results are promising, but we expect further improvement in the future, *e.g.,* from using enhanced acquisition hardware or extending the portfolio of monitored features.

2. We showed that pre-processing parameters have high impact on the model performance. A reduction of the cluster dimension below a size of $(700 \times 700 \times 50)\,\mu m^3$ led to decreased performance scores. The spatial uncertainty in the feature data (*e.g.*, from registration errors) might be a limiting factor when minimizing the cluster dimension. Furthermore, the porosity threshold can influence the porosity prediction detrimentally and should be set with care. Moreover, the consideration of feature information from subsequent layers led to increased performance scores. We showed that the model performs best if the number of included layers is equal or larger than the approximated melt pool depth. We postulate that the presented pre-processing workflow enables the model to identify complex PBF-LB/M phenomena such as keyhole pore formation and re-melting of material ("healing").

3. The model was shown to be sensitive to changes in local porosity. The results show that the raw model output was increased for clusters with increased porosity values. This is remarkable since the model was trained on binary data without considering the actual cluster porosity. This illustrates the potential of porosity prediction based on thermographic in-situ monitoring data and is, to the authors' knowledge, the first time that comparable results are presented to the AM community.

From the findings of this study, relevant future tasks can be derived: We aim to study the impact of single features on the prediction performance to identify features of high significance. Thereby, the model architecture can be further simplified, *e.g.*, by reducing the dimension of the convolutional layers. Furthermore, the role of spatial uncertainty inherent in the thermogram feature data needs to be studied in terms of prediction performance. Moreover, we aim to generate a larger data base that includes different machines, materials and different types of irregularities to increase the model's generalizability.

**Data availability** The datasets used, generated, or analyzed during the current study are publicly available in an online repository (https://doi.org/10.17605/OSF.IO/ZMDG4). The code developed and used in this study is available upon request.

**Code availability** The software used is mentioned in the work and is commercially available.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## Appendix

### Study on influence of class imbalance

Dataset imbalance is a known challenge for machine learning classifiers. If one class appears rarely in the dataset in comparison to other classes, the model can behave undesirable especially when the minority class is of interest (Han et al., 2005). Furthermore, the significance of traditional performance measures such as the accuracy score is limited if class imbalance is present (Chawla, 2005). Common approaches to deal with class imbalance are sampling approaches that change the intrinsic data distribution to allow the use of standard learning systems (Torgo et al., 2013). Prominent representatives of sampling approaches are the

**Fig. 14** Flow chart for the creation of synthetic data samples with the help of SMOTE
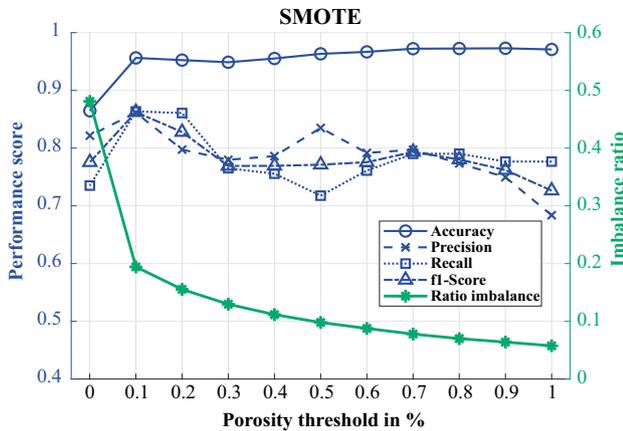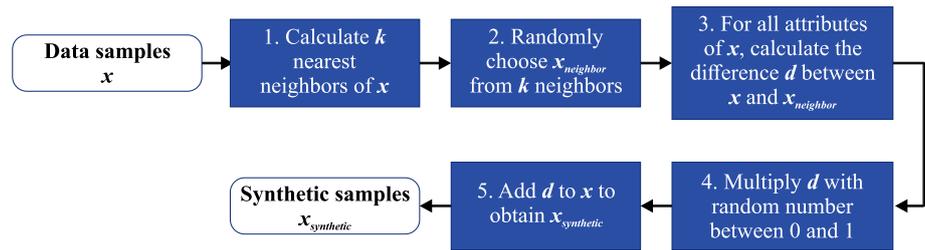




**Fig. 15** Model performance in dependency on the chosen porosity threshold with application of SMOTE sampling method. The performance scores remain comparably similar for the usage of SMOTE in comparison to the original dataset
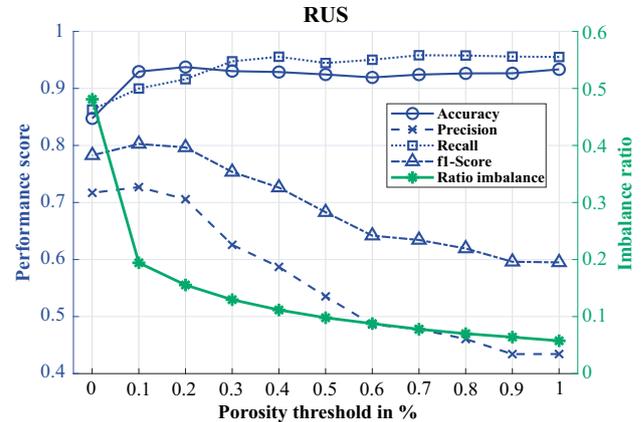
**Fig. 16** Model performance in dependency on the chosen porosity threshold with application of RUS sampling method. In comparison to SMOTE and the original dataset the application of RUS result in very high recall scores, while f1-Score and precision are significantly lower

Random Under-Sampling (RUS) technique or the Synthetic Minority Over-sampling Technique (SMOTE). The RUS technique decreases the number of majority samples to adjust the ratio between the classes. The majority samples that are deleted are chosen randomly (Torgo et al., 2013). SMOTE is an over-sampling method where synthetic minority samples are created based on the random K-nearest-neighbor method (Chawla et al., 2002; Torgo et al., 2013). The ratio between original samples and synthetic samples is specified by the over-sampling rate $R$. The parameter $k$ corresponds to the number of randomly chosen neighbors. $k$ and $R$ are chosen by the user. Common values for $k$ found in literature were ranging between 5 (Chawla et al., 2002) and 20 (Wang et al., 2019). For a chosen sample $x$, the creation of synthetic samples is performed as illustrated in Fig. 14.

In the present study of porosity prediction, the number of "porous" clusters was lower than the number of "non-porous" clusters. The resulting class imbalance was dependent on the chosen cluster dimension and the porosity threshold that differentiates between "porous" and "non-porous" samples. For SMOTE, $k$ was chosen equal to 10 based on preliminary tests and a ratio of 1 between majority and minority classes was achieved by the appropriate adjustment of $R$. The sampling methods were applied on the training data as shown in Fig. 7 in the main paper.

For the study of the influence of the sampling methods we used the same pre-processing conditions as in "Influence of porosity threshold on prediction performance". The cluster dimension was chosen to $(700 \times 700 \times 50)\ \mu m^3$ and 9 subsequent layers were considered for the construction of the sample matrix $M$. The results for the application of SMOTE and RUS are shown in Figs. 15 and 16, respectively.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Víegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, Savannah, GA, USA.

Aboulkhair, N. T., Everitt, N. M., Ashcroft, I., & Tuck, C. (2014). Reducing porosity in AlSi10Mg parts processed by selective laser melting. *Additive Manufacturing, 1–4*, 77–86. https://doi.org/10.1016/j.addma.2014.08.001

Aminzadeh, M., & Kurfess, T. R. (2018). Online quality inspection using Bayesian classification in powder-bed additive manufacturing from high-resolution visual camera images. *Journal of Intelligent Manufacturing, 30*(6), 2505–2523. https://doi.org/10.1007/s10845-018-1412-0

Baumgartl, H., Tomas, J., Buettner, R., & Merkel, M. (2020). A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. *Progress in Additive Manufacturing, 5*(3), 277–285. https://doi.org/10.1007/s40964-019-00108-3

Bayat, M., Thanki, A., Mohanty, S., Witvrouw, A., Yang, S., Thorborg, J., Tiedje, N. S., & Hattel, J. H. (2019). Keyhole-induced porosities in Laser-based Powder Bed Fusion (L-PBF) of Ti6Al4V: High-fidelity modelling and experimental validation. *Additive Manufacturing*. https://doi.org/10.1016/j.addma.2019.100835

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Müller, A. C., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases (2013)*, Prague, Czech. https://doi.org/10.48550/arXiv.1309.0238

Chawla, N. V. (2005). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 853–867). New York: Springer.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Cunningham, R., Narra, S. P., Montgomery, C., Beuth, J., & Rollet, A. D. (2017). Synchroton-based X-ray microtomography characterization of the effect of processing variables on porosity formation in laser power-bed additive manufacturing of Ti-6Al-4V. *JOM Journal of the Minerals Metals and Materials Society, 26*(3), 479–484. https://doi.org/10.1007/s11837-016-2234-1

Ehlers, H., Pelkner, M., & Thewes, R. (2020). Heterodyne eddy current testing using magnetoresistive sensors for additive manufacturing purposes. *IEEE Sensors Journal, 20*(11), 5793–5800. https://doi.org/10.1109/jsen.2020.2973547

Fritsch, T., Farahbod-Sternahl, L., Serrano-Muñoz, I., Léonard, F., Haberland, C., & Bruno, G. (2021). 3D computed tomography quantifies the dependence of bulk porosity, surface roughness, and re-entrant features on build angle in additively manufactured IN625 lattice struts. *Advanced Engineering Materials*. https://doi.org/10.1002/adem.202100689

Gaikwad, A., Williams, R. J., de Winton, H., Bevans, B. D., Smoqi, Z., Rao, P., & Hooper, P. A. (2022). Multi phenomena melt pool sensor data fusion for enhanced process monitoring of laser powder bed fusion additive manufacturing. *Materials & Design*. https://doi.org/10.1016/j.matdes.2022.110919

Gobert, C., Reutzel, E. W., Petrich, J., Nassar, A. R., & Phoha, S. (2018). Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. *Additive Manufacturing, 21*, 517–528. https://doi.org/10.1016/j.addma.2018.04.005

Grasso, M., Remani, A., Dickins, A., Colosimo, B. M., & Leach, R. K. (2021). In-situ measurement and monitoring methods for metal powder bed fusion: An updated review. *Measurement Science and Technology*. https://doi.org/10.1088/1361-6501/ac0b6b

Guo, Q., Zhao, C., Qu, M., Xiong, L., Escano, L. I., Hojjatzadeh, S. M. H., Parab, N. D., Fezzaa, K., Everhart, W., Sun, T., & Chen, L. (2019). In-situ characterization and quantification of melt pool variation under constant input energy density in laser powder bed fusion additive manufacturing process. *Additive Manufacturing, 28*, 600–609. https://doi.org/10.1016/j.addma.2019.04.021

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. *International Conference on Intelligent Computing (ICIC 2005)*, Hefei, China. *Advances in Intelligent Computing*, pp. 878–887. https://doi.org/10.1007/11538059_91

Hojjatzadeh, S. M. H., Parab, N. D., Guo, Q., Qu, M., Xiong, L., Zhao, C., Escano, L. I., Fezzaa, K., Everhart, W., Sun, T., & Chen, L. (2020). Direct observation of pore formation mechanisms during LPBF additive manufacturing process and high energy density laser welding. *International Journal of Machine Tools & Manufacture*. https://doi.org/10.1016/j.ijmachtools.2020.103555

Hooper, P. A. (2018). Melt pool temperature and cooling rates in laser powder bed fusion. *Additive Manufacturing, 22*, 548–559. https://doi.org/10.1016/j.addma.2018.05.032

Kamath, C. (2016). Data mining and statistical inference in selective laser melting. *The International Journal of Advanced Manufacturing Technology, 86*(5–8), 1659–1677. https://doi.org/10.1007/s00170-015-8289-2

Kingma, D. P., & Ba, L. J. (2015). Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations*, San Diego, USA. https://doi.org/10.48550/arXiv.1412.6980

Klein, A., Falkner, S., Bartels, S., Hennig, P., & Hutter, F. (2017). Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, Florida, USA. *PMLR: W&CP*, 54, pp. 528–536.

Krabusch, J., Meixlsperger, M., Burkert, T., & Schleifenbaum, J. H. (2020). Prediction of the Quality of L-PBF Parts Using Process Monitoring Image Data and Deep Learning Models. *Fraunhofer Direct Digital Manufacturing Conference*, Berlin, Germany.

Kwon, O., Kim, H. G., Ham, M. J., Kim, W., Kim, G.-H., Cho, J.-H., Kim, N. I., & Kim, K. (2020). A deep neural network for classification of melt-pool images in metal additive manufacturing. *Journal of Intelligent Manufacturing, 31*(2), 375–386. https://doi.org/10.1007/s10845-018-1451-6

Lane, B., & Yeung, H. (2020). Process monitoring dataset from the Additive Manufacturing Metrology Testbed (AMMT): Overhang Part X4. *Journal of Research of the National Institute of Standards and Technology*. https://doi.org/10.6028/jres.125.027

Lough, C. S., Liu, T., Wang, X., Brown, B., Landers, R. G., Bristow, D. A., Drallmeier, J. A., & Kinzel, E. C. (2022). Local prediction of laser powder bed fusion porosity by short-wave infrared imaging thermal feature porosity probability maps. *Journal of Materials Processing Technology*. https://doi.org/10.1016/j.jmatprotec.2021.117473

Lough, C. S., Wang, X., Smith, C. C., Landers, R. G., Bristow, D. A., Drallmeier, J. A., Brown, B., & Kinzel, E. C. (2020). Correlation of SWIR imaging with LPBF 304L stainless steel part properties. *Additive Manufacturing*. https://doi.org/10.1016/j.addma.2020.101359

McCann, R., Obeidi, M. A., Hughes, C., McCarthy, É., Egan, D. S., Vijayaraghavan, R. K., Joshi, A. M., Acinas Garzon, V., Dowling, D. P., McNally, P. J., & Brabazon, D. (2021). In-situ sensing, process monitoring and machine control in laser powder bed fusion: A review. *Additive Manufacturing*. https://doi.org/10.1016/j.addma.2021.102058

Modaresialam, M., Roozbahani, H., Alizadeh, M., Salminen, A., & Handroos, H. (2022). In-Situ monitoring and defect detection of selective laser melting process and impact of process parameters on the quality of fabricated SS 316L. *IEEE Access, 10*, 46100–46113. https://doi.org/10.1109/access.2022.3169509

Mohr, G., Altenburg, S. J., Ulbricht, A., Heinrich, P., Baum, D., Maierhofer, C., & Hilgenberg, K. (2020). In-Situ defect detection in laser

powder bed fusion by using thermography and optical tomography—comparison to computed tomography. *Metals*. https://doi.org/10.3390/met10010103

Oster, S., Fritsch, T., Ulbricht, A., Mohr, G., Bruno, G., Maierhofer, C., & Altenburg, S. J. (2022). On the registration of thermographic in situ monitoring data and computed tomography reference data in the scope of defect prediction in laser powder bed fusion. *Metals*. https://doi.org/10.3390/met12060947

Oster, S., Maierhofer, C., Mohr, G., Hilgenberg, K., Ulbricht, A., & Altenburg, S. J. (2021). Investigation of the thermal history of L-PBF metal parts by feature extraction from in-situ SWIR thermography. *Thermosense: Thermal Infrared Applications XLIII*. https://doi.org/10.1117/12.2587913

Rai, R., Elmer, J. W., Palmer, T. A., & DebRoy, T. (2007). Heat transfer and fluid flow during keyhole mode laser welding of tantalum, Ti-6Al-4V, 304L stainless steel and vanadium. *Journal of Physics D: Applied Physics, 40*, 5753–5766. https://doi.org/10.1088/0022-3727/40/18/037

Raplee, J., Gockel, J., List, F., III., Carver, K., Foster, S., McFalls, T., Paquit, V., Rao, R., Gandy, D. W., & Babu, S. S. (2020). Towards process consistency and in-situ evaluation of porosity during laser powder bed additive manufacturing. *Science and Technology of Welding and Joining, 25*(8), 679–689. https://doi.org/10.1080/13621718.2020.1823654

Ren, Z., Gao, L., Clark, S. J., Fezzaa, K., Shevchenko, P., Choi, A., Everhart, W., Rollett, A. D., Chen, L., & Sun, T. (2023). Machine learning-aided real-time detection of keyhole pore generation in laser powder bed fusion. *Science, 379*, 89–94. https://doi.org/10.1126/science.add4667

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE, 109*(3), 247–278. https://doi.org/10.1109/jproc.2021.3060483

Scheuschner, N., Strasse, A., Altenburg, S. J., Gumenyuk, A., & Maierhofer, C. (2019). In-situ thermographic monitoring of the laser metal deposition process. *II International Conference on Simulation for Additive Manufacturing – Sim-AM*, Pavia, Italy.

Smoqi, Z., Gaikwad, A., Bevans, B., Kobir, M. H., Craig, J., Abul-Haj, A., Peralta, A., & Rao, P. (2022). Monitoring and prediction of porosity in laser powder bed fusion using physics-informed meltpool signatures and machine learning. *Journal of Materials Processing Technology*. https://doi.org/10.1016/j.jmatprotec.2022.117550

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, Lake Tahoe, Nevada, USA.

Snow, Z., Diehl, B., Reutzel, E. W., & Nassar, A. (2021). Toward in-situ flaw detection in laser powder bed fusion additive manufacturing through layerwise imagery and machine learning. *Journal of Manufacturing Systems, 59*, 12–26. https://doi.org/10.1016/j.jmsy.2021.01.008

Snow, Z., Nassar, A. R., & Reutzel, E. W. (2020). Review of the formation and impact of flaws in powder bed fusion additive manufacturing. *Additive Manufacturing*. https://doi.org/10.1016/j.addma.2020.101457

Torgo, L., Ribeiro, R. P., Pfahringer, B., & Branco, P. (2013). Smote for regression. *Progress in Artificial Intelligence*. https://doi.org/10.1007/978-3-642-40669-0_33

Trapp, J., Rubenchik, A. M., Guss, G., & Matthews, M. J. (2017). In situ absorptivity measurements of metallic powders during laser powder-bed fusion additive manufacturing. *Applied Materials Today, 9*, 341–349. https://doi.org/10.1016/j.apmt.2017.08.006

Ulbricht, A., Mohr, G., Altenburg, S. J., Oster, S., Maierhofer, C., & Bruno, G. (2021). Can potential defects in LPBF Be healed from the laser exposure of subsequent layers? A quantitative study. *Metals*. https://doi.org/10.3390/met11071012

Wang, Y., Wu, X., Chen, Z., Ren, F., Feng, L., & Du, Q. (2019). Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using smote for Lishui City in Zhejiang Province, China. *International Journal of Environmental Research and Public Health, 16*, 3. https://doi.org/10.3390/ijerph16030368

Westphal, E., & Seitz, H. (2021). A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. *Additive Manufacturing*. https://doi.org/10.1016/j.addma.2021.101965

Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT press.

Xiao, L., Lu, M., & Huang, H. (2020). Detection of powder bed defects in selective laser sintering using convolutional neural network. *The International Journal of Advanced Manufacturing Technology, 107*(5–6), 2485–2496. https://doi.org/10.1007/s00170-020-05205-0

Yadollahi, A., Shamsaei, N., Thompson, S. M., & Seely, D. W. (2015). Effects of process time interval and heat treatment on the mechanical and microstructural properties of direct laser deposited 316L stainless steel. *Materials Science & Engineering A, 644*, 171–183. https://doi.org/10.1016/j.msea.2015.07.056

Zhang, W., Abranovic, B., Hanson-Regalado, J., Koz, C., Duvvuri, B., Shimada, K., Beuth, J., & Kara, L. B. (2020). Flaw Detection in Metal Additive Manufacturing Using Deep Learned Acoustic Features. *Workshop on machine learning for engineering modeling, simulation and design*, online.