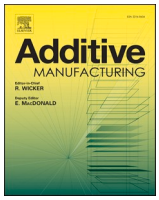
Additive Manufacturing 71 (2023) 103585

Contents lists available at ScienceDirect

Additive Manufacturing

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



Research Paper

Machine learning-based identification of interpretable process-structure linkages in metal additive manufacturing

Marc Ackermann *, Christian Haase

Steel Institute, RWTH Aachen University, Intzestr. 1, 52072 Aachen, Germany

ARTICLE INFO

Keywords:
Metal additive manufacturing
Machine learning
Explainable AI
P-S linkages
Microstructure

ABSTRACT

The use of data-driven methods for metal additive manufacturing (AM) is currently gaining importance as indicated by the increasing number of scientific literature in this field. Incorporation of data-driven methods has the potential to eliminate current bottlenecks in microstructure design given the diverse and complex nature of microstructures in additively manufactured metals. So far, coupling of existing simulation methods, e.g. physics-based process and microstructure models, to simulate AM microstructures with desired morphological characteristics requires extensive computational resources, high computation times and therefore, allows no scalable output. The extension of experimental- and simulation-based approaches by machine learning (ML) algorithms enables fast and computationally efficient predictions. However, the underlying architecture of ML algorithms often does not allow domain experts to interpret how predictions of the model were made and which features are responsible to what extent. This is why ML models are often referred to as black- box models. In this study, we present a data-driven framework based on physics-based simulation data to reveal explainable process-(micro) structure (P-S) linkages for metal AM. We provide an open-source dataset of 960 unique 3D microstructures created by simulation of powder bed fusion in metal AM. We employed the stochastic parallel particle kinetic simulator (SPPARKS) that is based on the kinetic Monte Carlo (kMC) method as an exemplary AM microstructure generator. Selected ML regression algorithms aim to predict 3D chord length distributions (CLDs), as a morphology descriptor, depending on the associated process parameter combinations. Various dimension reduction algorithms are applied for computationally efficient use of the data space. The proposed methodology allows (i) microstructure predictions under given processing conditions and (ii) to navigate experts in the process parameter space to achieve target microstructures. In this context, SHAP (SHapley Additive exPlanations) values are used to decipher the contribution of individual process parameters to the microstructure evolution. In particular, SHAP values calculated in this study unfold the width of the melt pool and the heat-affected zone as dominant features on the model output. We provide open-access to the used dataset and methods for the scientific community to gain experience with the proposed approach.

1. Introduction

Metal additive manufacturing (AM) has strongly increased in importance during the past years and is expected to become one of the key technologies for future sustainable production [1]. However, materials-related obstacles hinder to unravel the full potential of additively manufactured components, such as the lack of AM-adapted alloys, limited reproducibility and highly heterogeneous microstructures [2]. In principle, the microstructural features in additively manufactured metals are similar to those formed during casting or welding (e.g. solidification structures, heat-affected zones (HAZ)) [3]. Although these features are well known in metallurgy, their arrangement and character are highly heterogeneous and the possibility to control those over various length scales is unique to AM, and may provide superior properties [4]. Varying the microstructure, e.g. to provoke a columnar-to equiaxed transition, can be achieved by changing the processing and thus thermal conditions [5]. Nevertheless, the highly variable processing conditions, i.e. potential process parameter combinations, state a major challenge in metal AM, as these provide a vast design space that provides an almost infinite number of possible microstructures. Therefore, optimal microstructures for desired resulting properties can hardly be identified by experimental trial-and-error approaches, but demand

* Corresponding author.
E-mail address: marc.ackermann@iehk.rwth-aachen.de (M. Ackermann).

https://doi.org/10.1016/j.addma.2023.103585
Received 13 November 2022; Received in revised form 24 April 2023; Accepted 30 April 2023
Available online 3 May 2023

efficient and predictive computational tools [6].

Physics-based simulations, especially incorporated in integrated computational materials engineering (ICME) approaches [6,7] are capable to mimic the layer-wise material build-up and appear as the most constructive solution for computational linking of process, (micro) structure, properties and performance (P-S-P-P linkages) in AM metals [8]. Much effort has already been invested in developing and employing physics-based macro-, meso- and micro-scale models to describe AM processes. For instance, powder bed, heat source, molten pool, solidification and residual stress models are now available. The approaches have also been combined with micro-/meso-structure modeling (i.e. phase field, cellular automata, kinetic Monte Carlo [9]) to capture the multi-scale, multi-physics character of metal AM and to derive P-S linkages. Despite its advantages compared to time- and cost-intensive experimental approaches, ICME often requires powerful hardware while creating outputs, e.g., on the microstructure level for a single process-parameter combination. Therefore, such approaches lack generalizability and scalability to screen larger process parameter windows. In this context, emerging data-driven methods are promising tools for learning global linkages among the process chain of AM towards the final microstructure of the material.

Numerous studies focus on revealing process-structure (P-S) [10–12] linkages or structure-property (S-P) [13,14] linkages in metallic materials. All approaches have in common to use a low-dimensional representation of the microstructure for creating microstructure descriptors or features to capture information of e.g. texture, phases or morphological characteristics. These features are often extracted from 2D analysis (based on slices of 3D microstructures [15,16] or 2D experimental data [11,17]) to build meaningful correlations. The problem manifests in the loss of information while transforming data from 3D to 2D. In the recent years, approaches were developed to directly extract features from 3D data [18,19], but often these approaches lack a dataset with physics-based foundation [20,21], or have not yet been applied to the interdisciplinary field of AM and ML.

In this study, we propose a data-driven framework suitable for microstructure design in metal AM consisting of (i) a dataset creation of 3D AM microstructures by physics-based material simulations, (ii) extraction of 3D features to describe the grain morphology encapsulated in a 3D chord length distribution (CLD), (iii) dimensionality reduction of the dataset, and (iv) building and validating an accurate regression model by a comparison of several ML algorithms. For the microstructure simulations, the kinetic Monte Carlo (kMC) method was used as an exemplary AM microstructure generator by employing the open source stochastic parallel particle kinetic simulator (SPPARKS). In addition, unsupervised learning approaches reveal further insights on the high-dimensional design space encoding the complex interplay among extracted features and the process parameter space. Such insights allow an efficient design of microstructures through controlled process parameter selection. For practical use, this knowledge permits controlling the final microstructure by process adjustments, e.g. for a given laser velocity, how should other process parameters be set to create a certain target microstructure? We solve the problem of black-box models in explaining their output with support of SHAP (SHapley Additive exPlanations) values and thus make the predictions understandable for experts. SHAP values reveal further information on individual feature contributions on the P-S linkages on the local and global feature level. Ultimately, SHAP values, as part of concepts of explainable artificial intelligence (XAI), increased the interpretability of the derived ML output and can support experts in explaining model predictions.

2. Methods

2.1. Generating the AM dataset using kMC simulations

The open-source library SPPARKS contains tools for physics-based AM simulation of 3D representative volume elements (RVE). The selected model called potts/additive is a modified version of the Potts kinetic Monte Carlo model applicable for powder bed fusion (PBF) in AM [22]. This model allows the simulation of the evolving melt pool and the accompanying HAZ, while scanning a pre-defined 3D domain according to the laser scan strategy. The scan strategy can be specified for each individual layer. The melt pool geometry, a grain mobility decay factor and the scanning strategy are required parameters for the simulation. The melt pool is defined as double ellipsoid after [23], where each ellipsoid shares a width and depths for the melt pool and HAZ, respectively. On the third axis (laser scan direction), the tail length or the cap height define the ellipsoids of the melt and HAZ. The grain mobility decay factor $f_{mobility}$ represents the coefficient in the grain mobility equation $M = \exp(-f_{mobility} * x)$, with x as shortest distance between lattice site (a unit area of the microstructure) and melt pool boundary.

For this study, previous results on a fully austenitic high manganese steel [5] served as basis to collect input data for the simulation. In total, the variation of 20 input parameter combinations resulted in 960 unique 3D microstructures as model output (Table 1). The selected parameter range reflects an experimentally relevant range, while the values are not intended to be exhaustive. Furthermore, it must be noted that some of the parameters in Table 1 are not independent of each other, leading to variations of parameter combinations that cannot be directly realized in experiments. It is the aim of the parameter combinations considered to identify the individual contribution of each of the parameters and to provide a ML-based framework using kMC simulations as one of multiple tools for AM microstructure generation.

Each cubic RVE with a grid resolution of 100 contains 10^6 elements from which each lattice site, a specific integer spin (described in [22]) is assigned. The aggregation of lattice sites with the same spin value forms a grain, therefore the spin values are later used as an identifier (ID) assigned to each grain. The mobility decay factor was set to 0.15 after conducting a parameter study. For all simulations, the number of layers was set to four and 25 sites represent the hatch distance.

A Linux-computing cluster was used on the node CLAIX-2018-MPI node with two Intel Xeon Platinum 8160 processors, 24 cores per processor and a clock speed of 2.1 GHz. SPPARKS provides functions for parallel computing, therefore we used four processes in parallel to speed up the simulation. Each RVE consumed 4 GB in memory and took around two core hours. Kats et al. [24] recently coupled CA model with finite volume method to predict grain characteristics on a limited amount of process parameter combinations for directed energy deposition.

In contrast to cellular automaton [25,26], we chose the kMC approach given its computational efficiency to cover larger varieties of the process windows.

We use the directional chord length distribution (CLD) as morphology descriptor to capture grain size information in 3D. The chord length is defined as distance between intersections of a grain boundary and a pre-defined scan line (Fig. 1). The orientation of scan lines was varied to analyze the whole RVE and to capture CLDs under different angles. In contrast to previous approaches as the 2D CLD analysis proposed in [27], direct analysis of three dimensions is a main advantage of the implementation used in this work (based on a

Table 1
Parameters as input for kMC simulations. In total 960 unique combinations of parameters serve as input for the microstructure simulations as well as for the ML approach. Further information on the melt pool parameters can be seen in Appendix Fig. A.1.

Parameter	Values
Rotation angle per layer (°)	***0, 30, 60, 90***
Melt pool width (lattice sites)	***30, 40, 50, 60***
Velocity (lattice sites/Monte Carlo step)	***5, 8, 11, 13, 16***
Melt pool depth (lattice sites)	***25, 35***
Melt pool tail length (lattice sites)	***40, 50, 60***
Heat-affected-zone width (lattice sites)	***5, 10***

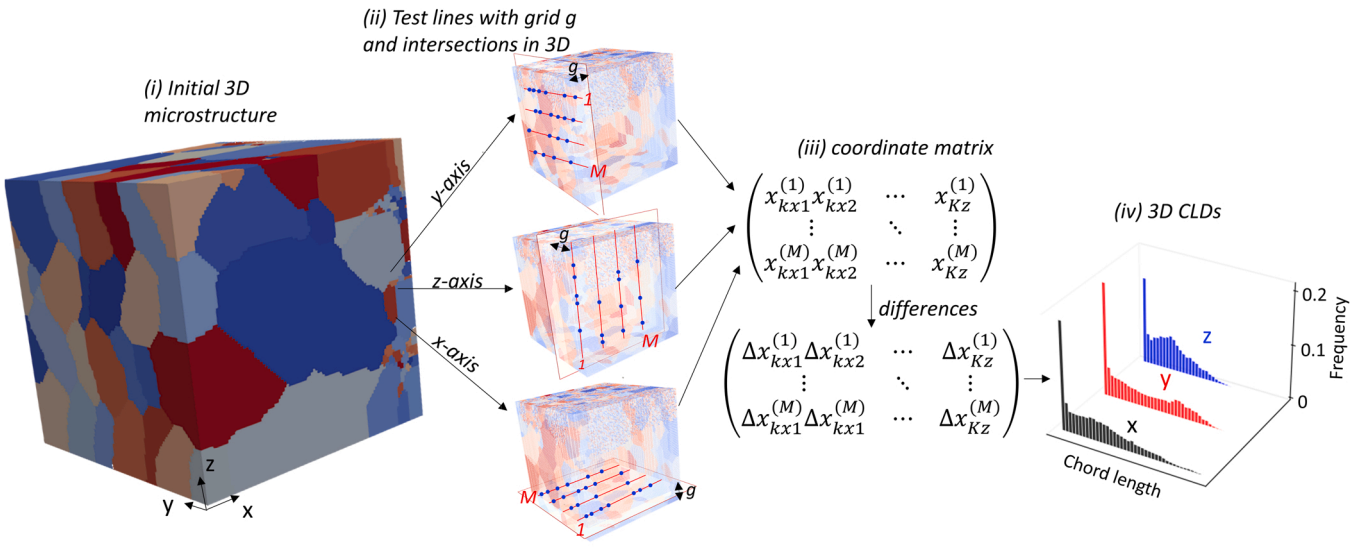


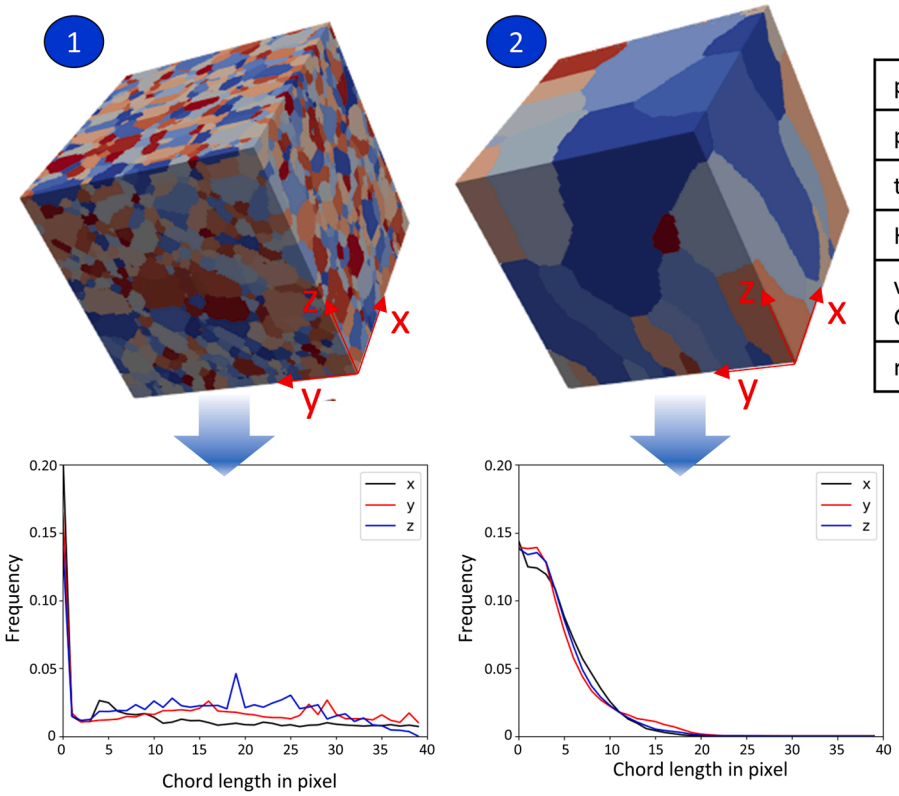
Fig. 1. Illustration of the scanning process along the line variants. (i) Synthetically generated 3D microstructure with elongated grains, (ii) superimposed test lines with grid step size (g) as well as intersections between test lines and boundaries, (iii) sorted coordinate matrix and its row-wise differences, (iv) chord length distribution by binning chord length values, after [27]. It has to be noted that the color scheme of the grains is arbitrary and holds therefore no information except the grain ID.

MATLAB® script, as described in [28]). From the measured and binned chord lengths, the probability P_l of finding a chord of certain length l_i is given as:

$$P_l = \frac{N_i l_i}{\sum_{i=1}^{n} N_i l_i} \tag{1}$$

where N_i defines the number of chords identified from the analyzed microstructure in the interval of the ith (to nth) bin.

Besides capturing 3D information directly (without the necessity of 2D slicing) from the microstructure, the code considers periodicity at the grain boundaries of the RVE, as grains with the same ID might occur on parallel boundaries, e.g. when using DREAM3D [29]. In the implementation, grain boundaries are identified by next neighboring voxels with different grain IDs. Spherical coordinates (transformed by θ as polar angle, and φ as azimuthal angle) were used to discretize scan lines



	1	2
pool width (sites)	30	60
pool depth (sites)	25	35
tail length (sites)	40	60
HAZ width (sites)	5	10
velocity (sites/Monte Carlo steps)	16	5
rotation angle (°)	90	90

Fig. 2. Visualized output of two (out of 960) kMC simulated 3D microstructures with min-max melt pool dimensions and laser velocity with corresponding chord length distributions in three directions. It has to be noted that the color schema of the grains is arbitrary and holds therefore no information except the grain ID.

in the 3D space. For this study, we limited sampling angles to the orthogonal directions x, y, z to calculate CLDs to allow a fast feature extraction. For each scan direction, the grain ID is used to compare line segments. All line segments with the same ID will be summed up until the scan line intersects with the boundary of the RVE. All segments are stored in a matrix ordered by scan direction (rows) and line segments of the same direction (columns). Therefore, plotting one row yields a CLD under a specific direction in 3D space capturing the anisotropy of grain morphologies (Fig. 2). The process parameter matrix in combination with assigned CLDs represent the input dataset for the data-driven framework.

2.2. Framework for building process-structure linkages

The proposed framework (Fig. 3) aims at revealing how parameters of the kMC simulation influence the microstructure evolving in AM in terms of CLDs as morphology descriptor. This knowledge allows building a surrogate model with low computational cost for predictions of unseen process parameter combinations. Additionally, unsupervised learning algorithms provide access to otherwise hidden contributions of features on the model output. Such comprehensive information permits tailored microstructures through controlled adjustments in the process parameter space. At first, the CLD matrix, with directional CLD entries in each row is normalized to obtain zero mean and unit variance of the data, and assigned with corresponding process parameters by adding six additional columns. This results in a high-dimensional data space of *115.200 data points (960 × 126)*. Within a study for parameter optimization, two criteria were defined per individual RVE for data preprocessing: (i) removal of small artefacts within the first two shortest chords (1–2 px in size) with a frequency sum of > 0.5, (ii) removal of outliers with an overall frequency sum of < 2 to filter out outliers with only a few grains per microstructure (shown in the Appendix Fig. A.2). The excluded RVEs declared as outliers show a high number of small melt pool width, all under 0° rotation (first filter) and a larger melt pool width (second filter). After this cleaning step, 917 eligible microstructures were considered for further processing. Further criteria were set for the ML model training. Training and optimization of the model continued until a pre-defined mean average error (MAE) and a minimum R^2 of 0.8 were reached (set as arbitrary values).

The cleaned dataset is further reduced in dimensionality by unsupervised learning algorithms. Principal component analysis (PCA) truncates the dimensions after linear separation of the high-dimensional space by finding orthogonal and linear combinations and ordering these principal components (PCs) in conjunction with their explained variance. The explained variance (Fig. 4) provides information on how much variance can be captured by the most contributing PCs. The CLD data with 120 columns expressed by ten PCs maintains a minimal information loss. Thus, the dimensionality reduction results in a decrease from 120 to 10 dimensions that capture still almost 99 % of the variance in the data.

For non-linearly separable data, other unsupervised learning algorithms are necessary. In this study also kernel PCA and t-distributed neighbor embedding (t-SNE) are tested. In contrast to PCA, kernel PCA [30] utilizes kernel functions for projecting the data into a higher dimensional space where it can be linearly separated, also known as kernel trick [31]. t-SNE [32] searches for similarities of neighboring data points and converts them into Gaussian joint probabilities. A lower dimensionality is the result of minimizing (by gradient descent) the Kullback-Leibler divergence between these joint probabilities in high- and low-dimensional space. This allows identification of local structures or clusters on different scales. The divergence is non-convex in nature, therefore a re-initialization results in a local minimum and subsequently a different low-dimensional embedding. PCA can be also used beforehand to reduce the noise of the dataset and speed up t-SNE for calculating distances of sample pairs. With combination of PCA and t-SNE, the global structure of the dataset can be preserved.

For training selected ML models, we set the PCs of the truncated CLD dataset as target variables. The data matrix associated with process parameters is split into a train (70 %) and test (30 %) set. The test set is used for an overall evaluation of the model performance. Cross-validation is a common practice to evaluate ML models with a limited

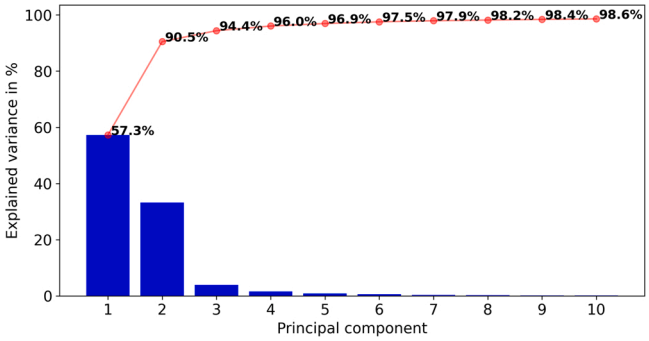


Fig. 4. Explained variance of the PCA ordered by degree of contribution from the first to the tenth PC. Ten PCs capture almost 99 % of the CLD variance. This allows a dimensionality reduction from 120 to 6 components.

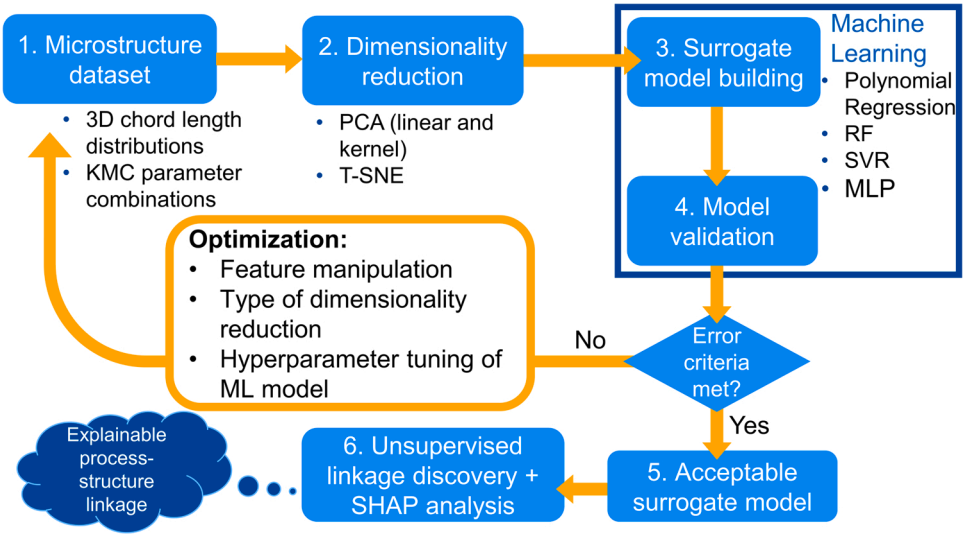


Fig. 3. Framework for building and discovering data-driven process-structure linkages. After creating a microstructure dataset, principal component analysis (PCA, linear or kernel) and t-distributed neighbor embedding (t-SNE) allow dimensionality reduction for linear and non-linear separable data. As error criteria while model training, optimization of each ML model (RF: random forest, SVR: support vector regression, MLP: multilayer perceptron) continued until the mean average error (MAE) falls below 0.25 and the correlation of simulation and model output in terms of R^2 surpasses 0.8. The SHAP analysis on the trained model's output allows explainable process-structure linkages.

amount of data. For the training set, a five-fold cross-validation is used to split the dataset into k smaller sets (called folds). For each new set, the model is trained on the collective of the remaining sets while the model is evaluated on the hold-out set. Thus, each sample will be assigned to the hold-out set once and will be used k-1 times for training. After a screening of potentially relevant ML models, polynomial regression, support vector regression and random forest regression are selected. In addition, those models are compared with a multi-layer perceptron as a representative artificial neural network (ANN) in order to identify if a rather shallow or deep network architecture is necessary to build accurate P-S linkages.

2.2.1. Polynomial regression

In contrast to linear regression, polynomial regression maps a polynomial function of higher degree on the relationship between explanatory variable X and dependent variable Y. The polynomial degree as only hyperparameter indicates its frequent utilization among other ML models [33].

2.2.2. Support vector regression

Support vector machines (SVM) are commonly used classification algorithms. SVMs find decision boundaries referred to as hyperplane based on a linear or kernel function to separate classes in high-dimensional space. These hyperplanes are created with support of extreme data points (support vectors) and a maximum margin by setting the hyperplane at a maximum distance between these data points [34]. SVMs can be applied to regression problems. Then instead of class labels, the model expects floating point values as target variable [35].

2.2.3. Decision tree regression

Decision trees cover another popular branch of ML algorithms. Data is continuously split into nodes by a splitting rule (based on locally optimal decisions of the mean squared error) as a quality measure until further splitting adds no additional value to the model prediction or until a pre-defined depth of the tree is reached. This allows rather simple and interpretable predictions often referred to as white box [36,37] and requires less data preparation (no feature normalization needed), as compared to most ML models with scale variance. A problem arises if small deviations in the data result in unstable and non-reproducible tree constructions. Thus, high variance often occurs after training individual trees, and in presence of unbalanced data single trees tend to overfit the input data. To mitigate such problems, ensemble methods, e.g. a random forest classifier or regressor [38], can induce some randomness by incorporating diverse tree structures. Averaging the predictions of each tree reduces the variance and provides an overall more accurate model.

2.2.4. Multilayer perceptron

ANNs draw increasing attention when it comes to solving machine-learning problems. A multilayer perceptron (MLP) is a forward-predicting neural network trained iteratively on a dataset while minimizing a loss function [39]. For each iteration, trainable parameters (referred to as weights) get updated. Another model functionality, the regularization term, helps to avoid overfitting the model on the training data while computing the loss. Besides, MLPs are probably the most complex types of ML algorithms for optimization given the relatively high number of hyperparameters.

We imported all models from Scikit-learn [40] and conducted model training on Google Colab with Nvidia Tesla P100/K80/T4 GPU.

2.2.5. ML models with explainable features

Among the numerous approaches on the explanation how ML models yield predictions, we use the SHAP library, introduced by Lundberg and Lee [41]. The SHAP library provides model-agnostic explainer functions, and therefore allows applicability on the different ML model types introduced in Sections 2.2.1–2.2.4. Moreover, SHAP values (originally introduced in game theory) can provide information on both, local and global feature effects on the model output to identify individual feature contributions of single data points or global trends of the whole dataset. The approach is based on the calculation of Shapley values as the averaged contribution of a given feature value across all possible feature combinations. In other words, each feature is attributed with a relative importance value depending on its contribution. Therefore, Shapley values of feature values are capable of explaining how the actual prediction deviates from an average prediction.

3. Results

3.1. Exploring the dimensionality-reduced process parameter space

Before applying ML regression for learning linkages between process parameters and CLD characteristics, the reduced CLD space by PCA allows a first glimpse into the distribution of process parameter values among the data population. The first two principal components are used for 2D data visualization (Fig. 5). For instance, a rather random distribution of PC1–PC2 combinations is observed for the indexed datapoints of the rotation angle and the melt pool width. In contrast, the laser velocity shows a clear separation of low to high velocities along the PC2 axis, whereas only low velocities can be found on the entire range of data points along PC1. For the melt pool depth, especially shallow depth values build a cluster at low PC1 values. The tail length of the melt pool shows a broad distribution along PC1 and PC2 for high values, whereas decreasing values can be found at lower PC1 values independent of PC2. Larger values of HAZ width tend to cluster around lower PC1 values and low to medium ranges in PC2. Further analysis on different PCs might result in further trends, but such analysis is out of scope for this study. We encourage interested readers to use the shared code for further data analysis.

The required number of principal components is commonly based on the minimum acceptable loss in variance after truncation or the desired amount of captured variance. The truncation level in this work is determined by the quality of reconstructed CLDs. For a reconstruction based on six PCs (Fig. 6), the reconstruction embodies the principal characteristics of the CLDs in x, y and z direction with smoother curvature compared to the original ones. The effect of additional PCs from only using PC1, PC1 + PC2, up to using five PCs can be seen in Appendix Fig. A.3.

Furthermore, PCA's eigen vectors carry information on how each microstructure differs in terms of its CLDs from the overall average of the CLD characteristics (Fig. 7). Small chords below 5 px are magnified in PC1 direction, while chords with a size in the range of 5–30 px are slightly reduced in PC1 direction. Chords larger than 10 px increase in value for PC2, and decrease for sizes between 1 and 10 px. Largest chords seem to be increased in PC3 direction. PC4 shows a high sensitivity to chords in y direction. Moreover, it becomes clear that PC2 for medium sized chords (10–30 px), and especially the last three PCs, PC4–PC6 capture information on morphological anisotropy associated with pronounced differences between CLDs in all three principal directions, with emphasis in PC5 for CLDs in x direction and in PC6 for CLDs in building direction (z direction).

3.2. Data-driven process-structure linkages

After training of four ML models on a dataset holding six process parameters as features and the reduced CLD space as target variables, polynomial regression with four degrees shows the best performance after applying PCA (Fig. 8). The averaged R^2 score over five times folding of a new validation set from PCA reduction reaches 90.6 %. Random forest, support vector regression and multilayer perceptron show lower R^2 values in contrast to polynomial regression, but still meet the pre-defined criteria (> 80.0 %). The shown results are based on prior hyperparameter tuning for increasing the model accuracy.

The best performing model (for PCA reduced CLD space) was further

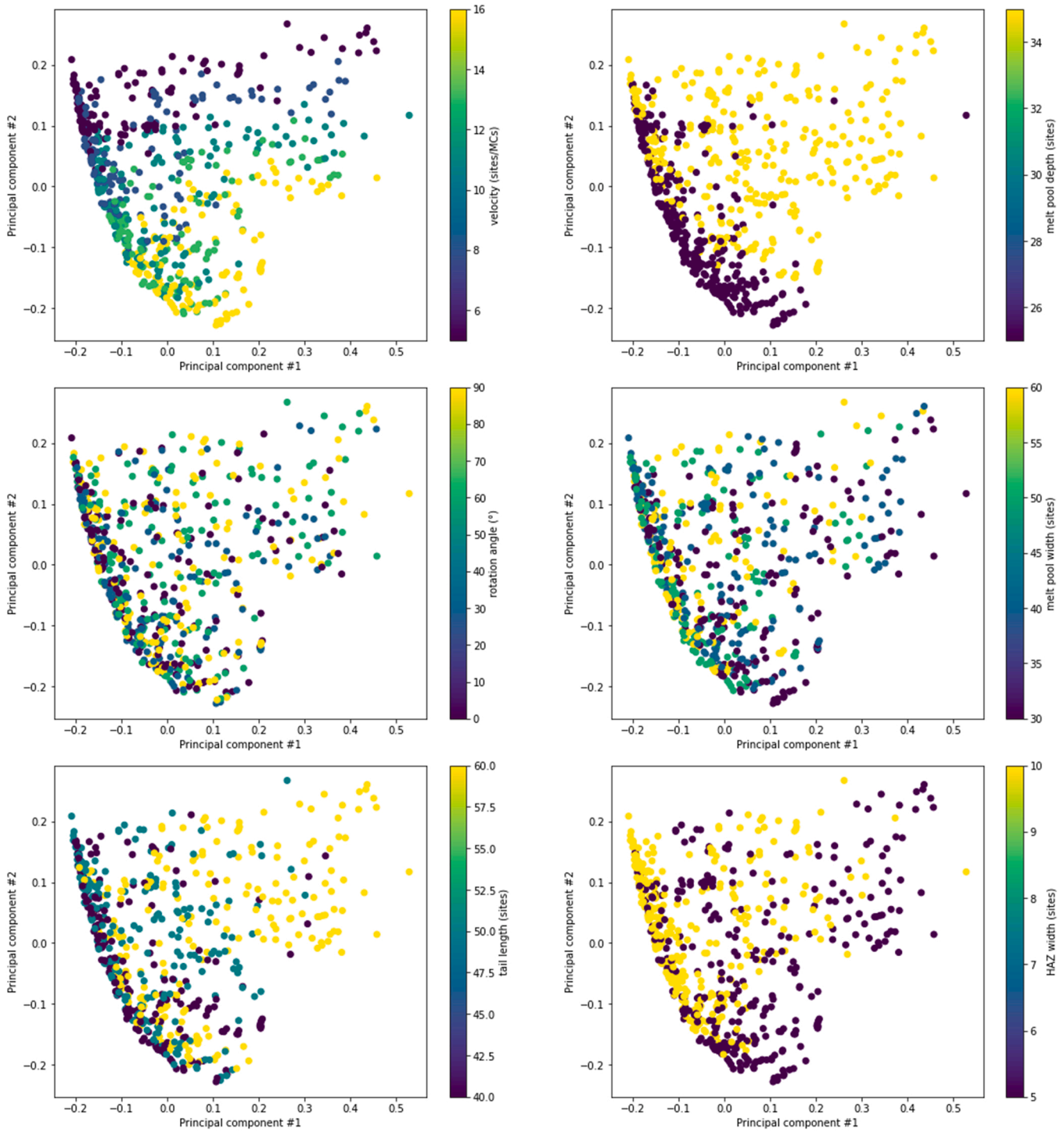


Fig. 5. The shown plots serve as an initial overview of the dataset reduced to two dimensions: PC1 and PC2 indexed with a kMC model parameter. The dataset reduced by PCA shows clusters for certain parameters in the PC1–PC2 space, e.g. the laser speed (velocity) and melt pool depth, while the rotation angle or melt pool width are rather randomly distributed. It hast to be noted that further dimensions might reveal further clustering, but instead we will use SHAP values to interpret the PC-space, as shown in the discussion part.

tested on its ability to be generalized on unseen data. Therefore, the model prediction of each point of the test dataset was plotted against the kMC simulation output declared as ground truth (Fig. 9). The comparison between training and test dataset shows for PC1 to PC3 lower R^2 values as for the training dataset with a maximum difference of 0.095 (0.896–0.801) for PC1, while the smaller differences for PC2 and PC3 indicate a satisfactory ability of the model to generalize predictions within PC2 and PC3 space. Since PC1 captures most of the variance in the dataset, it has to be noted that at the same time for PC1 some points occur as outliers in contrast to PC2 and PC3 in both, training and test set. These outliers seem to be an underestimation by the model compared to the ground truth values.

3.3. Case study: process parameter selection for maximum laser velocity

For a practical application of the data-driven framework, a scenario

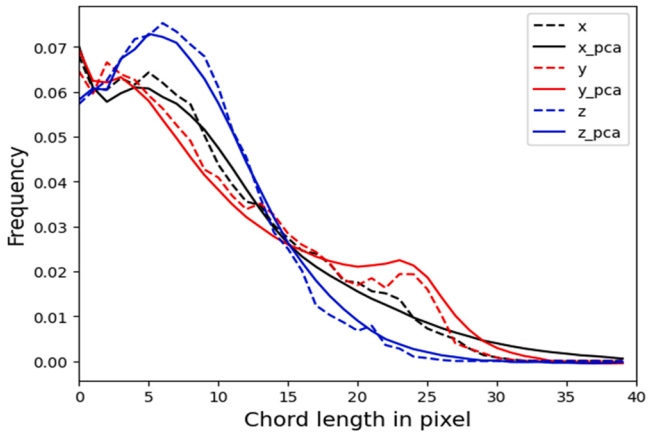


Fig. 6. Reconstruction of CLDs based on six PCs (solid lines) compared to the original CLDs (dashed lines) of a randomly selected microstructure.

can be assumed where processing a previously additively manufactured material with known microstructure requires a change of laser speed (e.g. due to a replacement of the laser source). In this scenario, it is unknown how to set the remaining process parameters to reach the same microstructure characteristics. With the proposed framework, the dataspace can be filtered to extract only points of interest in the PC space, e.g. with maximum laser speed (Fig. 10). Specifically aiming at selected target microstructures created under maximum laser velocity,

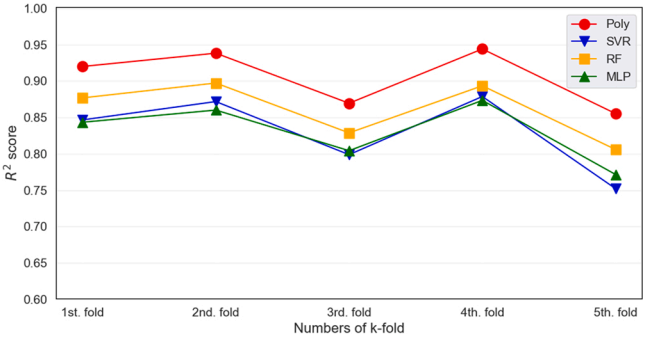


Fig. 8. a) R^2 scores for PCA-reduced data of the top three models tested in this study (Poly: Polynomial Regression, SVR: Support Vector Regression, RF: Random Forest, MLP: Multilayer Perceptron) for 5-fold cross-validation. A polynomial degree of four reaches the highest accuracy of 90.6 % followed by 86.1 % for RF, 83.3 % for MLP, and 83.0 % for SVR.

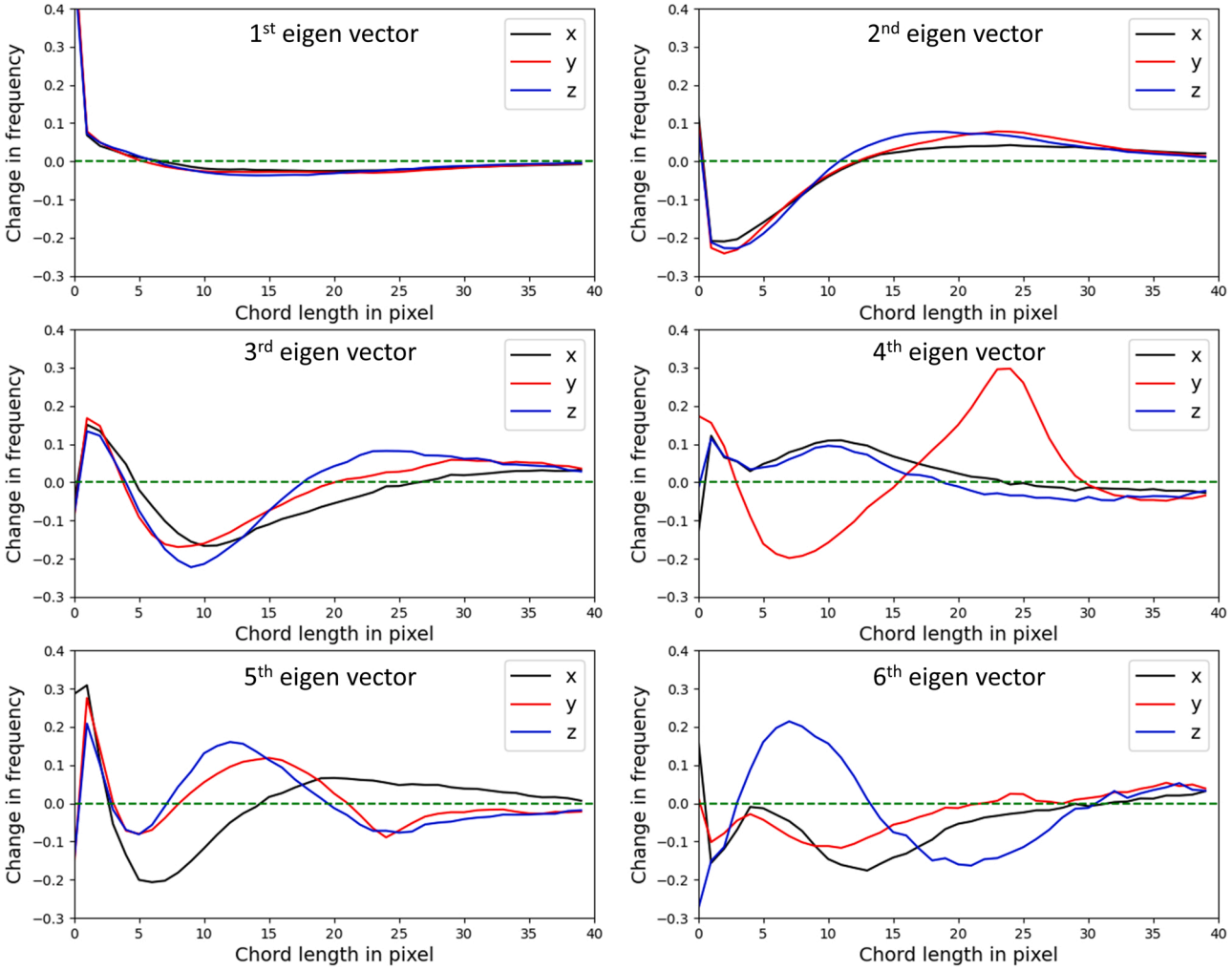


Fig. 7. First six basis vectors of PCA principal components (eigen vectors) for chord length distributions. Eigen vectors store information on the chords in x, y and z direction, e.g. the first vector increases densities for chords < 5 pixels, while the second vector decreases densities for chords < 10 pixels. From the differences between x, y and z, PC scores 3–6 carry additional information on anisotropy. The green horizontal line shows the main differences between the individual data points and the ensemble mean in their original dimensionality.

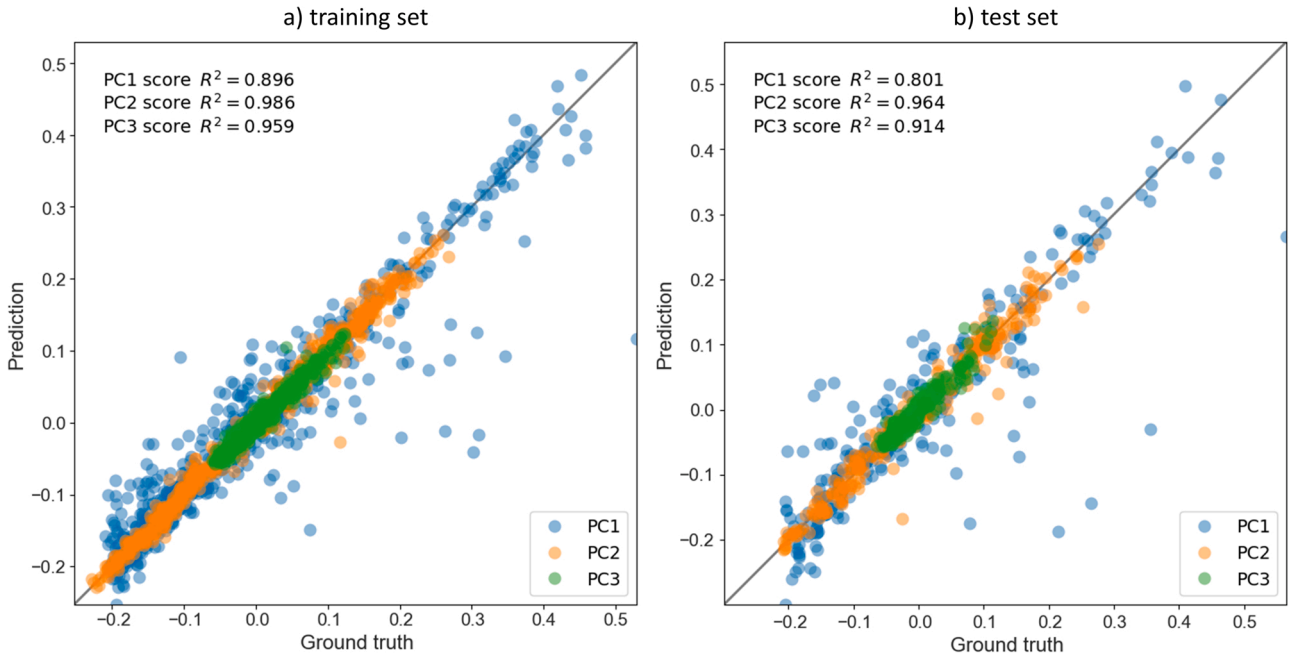


Fig. 9. The parity plot for polynomial regression (PCA-reduced dataset) with degree four on a) the training dataset and b) the test dataset for the first three PCs.

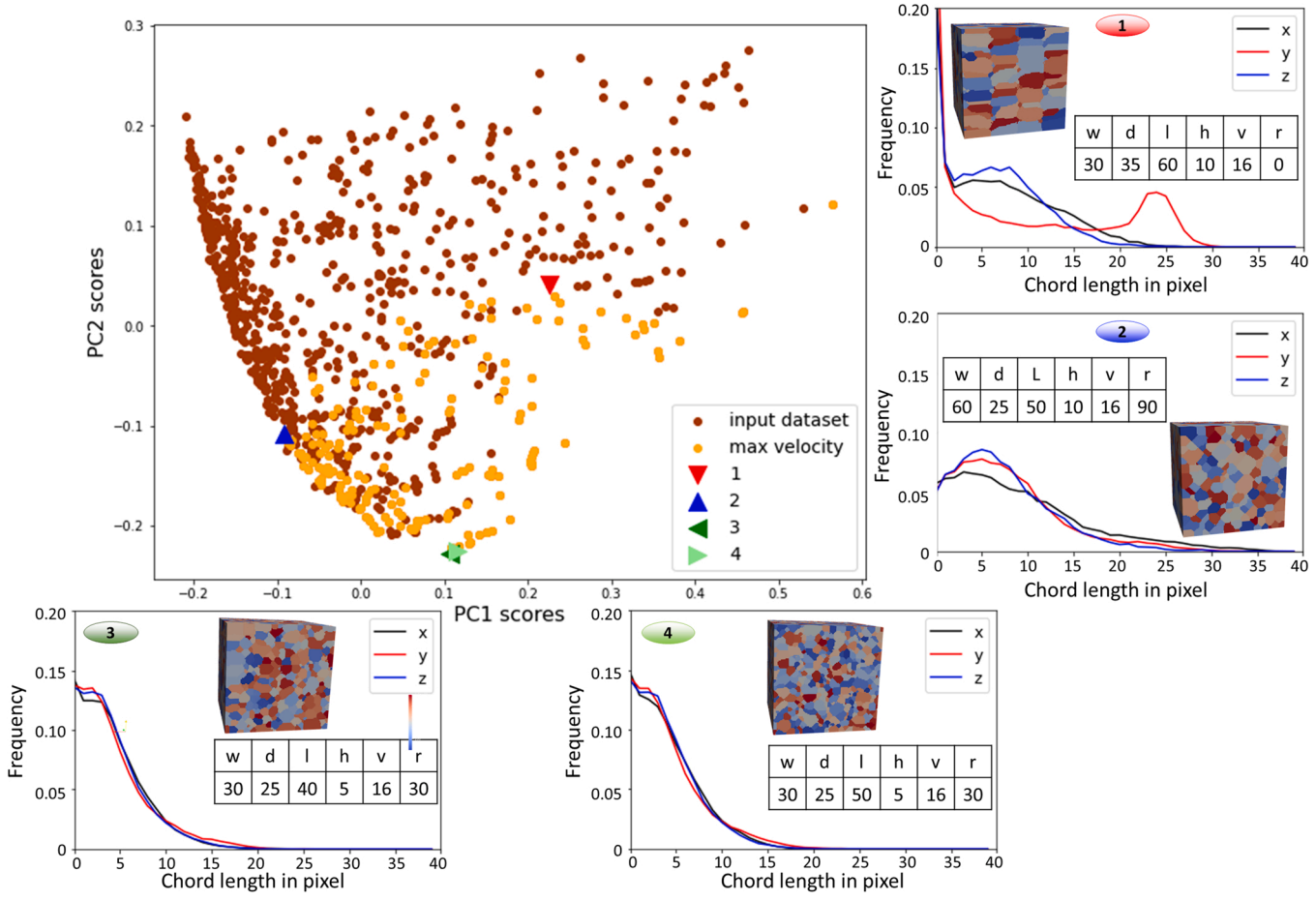


Fig. 10. Process-structure linkages of selected microstructures (no. 1–4) in PC1–PC2 space of the input dataset (917 microstructures) with corresponding 3D CLD distribution and process parameter combination: melt pool width (w), depth (d), tail length (l), HAZ width (h), laser velocity (v), rotation angle (r). Orange points emphasize maximum laser velocity values in the input dataset, where using a parameter combination of microstructure no. 1 yields anisotropic CLD characteristics with high PC2 score, and for no. 2 low PC1 score. Besides microstructure no. 3, no. 4 mimics similar PC1–PC2 characteristics with a variation in tail length.

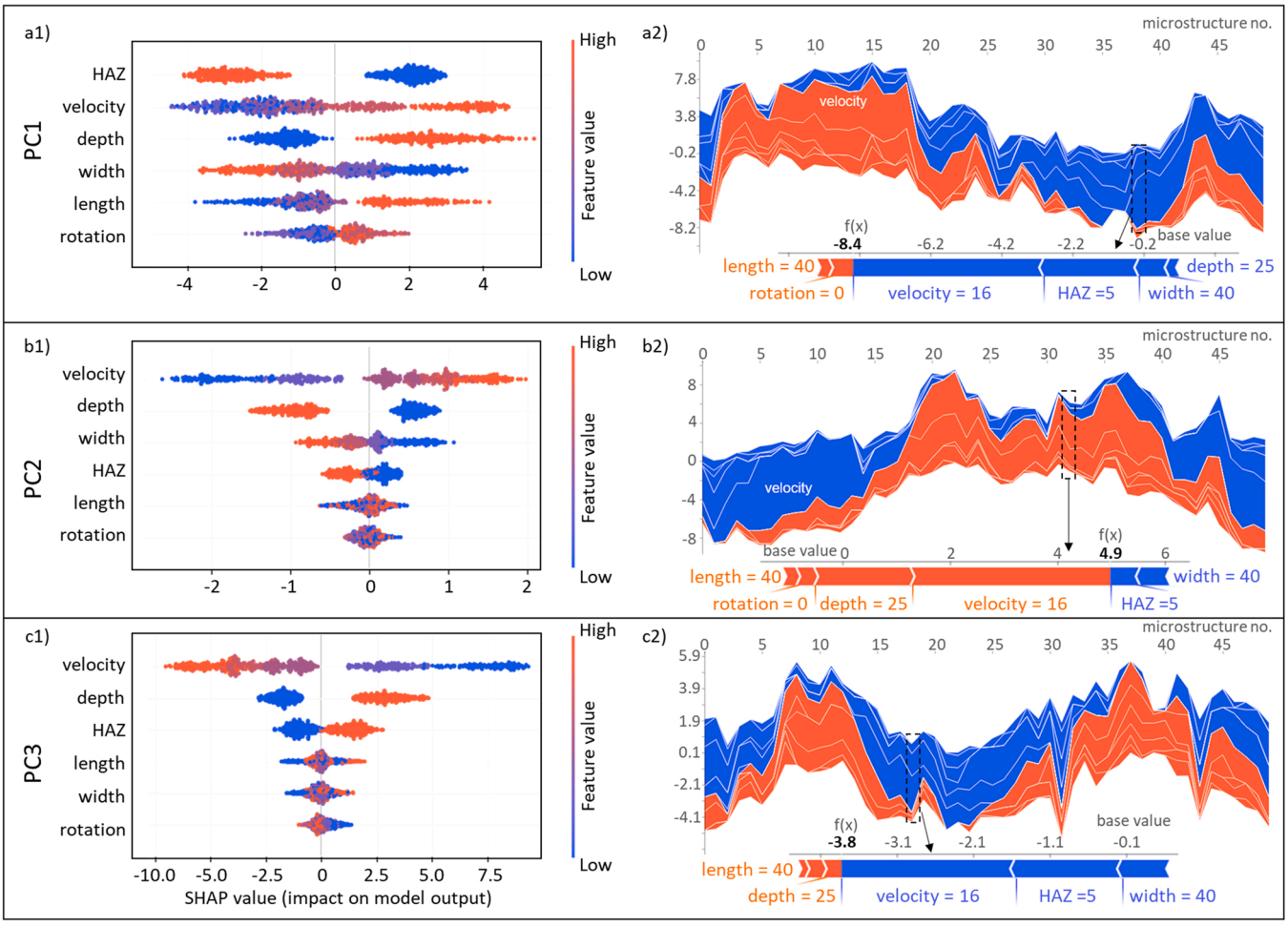


Fig. 11. Global (left column) and local (right column) feature contributions on the output of neural network-type ML based on the t-SNE-reduced data space to three PCs. The left column shows a hierarchical order of most contributing features (e.g. PC1: 1. Laser velocity, 2. melt pool depth, etc.) and for each feature a horizontal distribution of SHAP values with positive (> 0) or negative (< 0) contribution indexed as high feature value (marked in orange) and blue as low values (marked in blue). For explaining the model output of a single microstructure (right column) in PC1 space, a high laser velocity value together with HAZ and depth pushed the SHAP value to the right. Individual data points originate from randomly selected 50 data points of the training set for identification of local contributions. The base value indicates the averaged value of all contributions.

and indexing their corresponding process parameters, reveals information on the range of CLD characteristics present in the sub-dataset. For example, a scan strategy without rotation (selected data point no. 1) yields CLDs with high PC2 score and pronounced CLDs with larger size in y direction (captured by a high PC4 score). In case of maximum laser speed, the CLD space of selected microstructure no. 2 with largest values

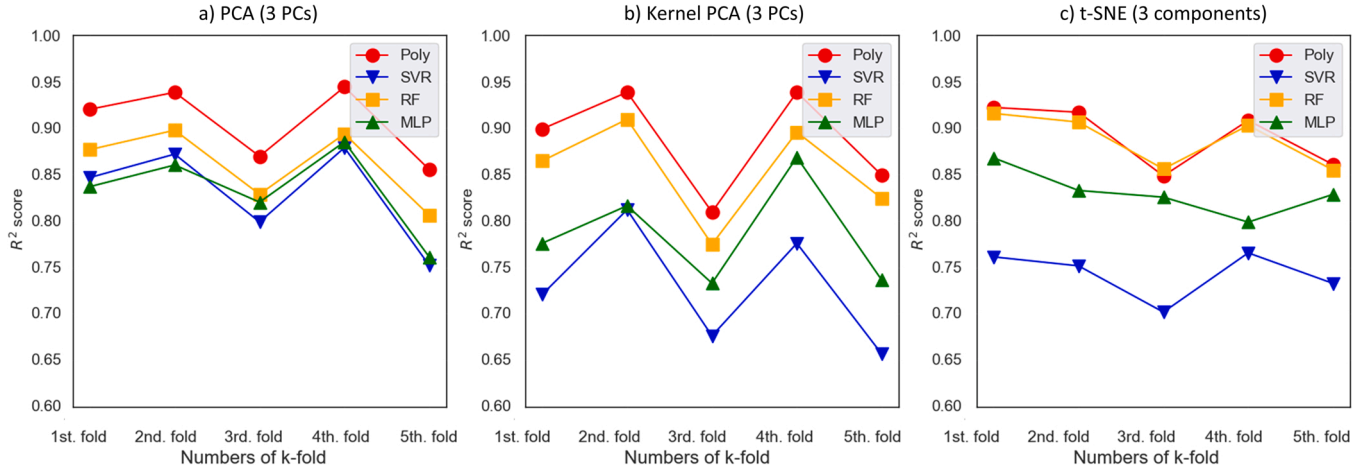


Fig. 12. Effect of dimensionality reduction method on ML model accuracy. For each method, a hyperparameter search was carried out for individual model tuning. Polynomial regression yields the best performance among all dimensionality reduction methods with 90.6 % (PCA), 88.7 % (kernel PCA) and 89.1 % (t-SNE). Model parameters can be seen in Appendix Table B.1.

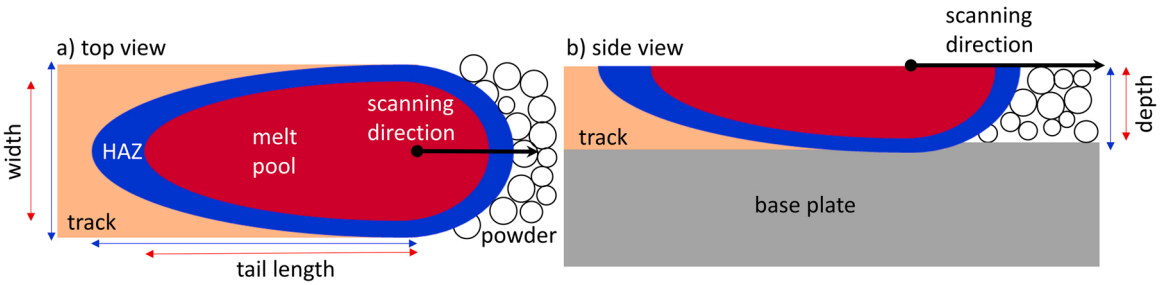


Fig. A.1. Parameters of melt pool and heat affected zone (HAZ) as input variables for the kMC simulation, after [11].

for melt pool and HAZ width causes a higher anisotropy, and results in low PC1 values. In contrast, a rather small melt pool geometry at a maximum laser speed combined with 30° rotation between subsequent layers (no. 3) creates CLDs with a high frequency of smaller sizes with low anisotropy in x, y and z direction. Extracting process parameter combinations of close datapoints in the PC space can be used to mimic target microstructures. Such information of datapoints in the proximity of the microstructure of interest allows experts to choose from further alternatives for setting the process parameters. For example, a change in the tail length still produces similar microstructure characteristics given data point no. 4 in the proximity of no. 3 in PC1–PC2 space (their proximity is also true for PC3–PC4 which is not shown here).

4. Discussion

4.1. Interpretation of the process-structure linkages

So far, the trained model represents a black-box with hidden information of the actual process-structure linkages. In this context, neural networks lack further information on how the output was created by transforming input data layer-wise until a model output is presented. To solve this issue, SHAP values, originally introduced in game theory [42], are capable to explain feature contributions on the local and global feature level. Calculation of such values is carried out by averaging the contribution of all features under every possible feature combination. The python library SHAP [41] was used to calculate and plot the SHAP values to allow interpretability of the MLP predictions (Fig. 11). SHAP values provide global (Fig. 11 a1–c1) and local (Fig. 11 a2–c2) feature interpretations on model predictions. Low HAZ and high values of laser velocity increase the PC1 score, i.e. result in low chord lengths/grain sizes, which is in agreement with the expected correlation between higher cooling rate due to the high laser velocity and small, solidified grains [5]. On the contrary, high velocity, low melt pool depth and width are the top three contributing factors for increasing the model output in PC2 direction. Hence, such melt pool characteristics result in larger chord lengths/grain sizes, independent from the orientation of the grains' major axis. For PC3, low velocity, high melt pool depth and high HAZ dominate the increase of the model output, while length, width and rotation show a minor contribution.

On the local level, applying SHAP values to single data points from the 3D t-SNE space, gives a more detailed view on the origin of the final predictions. SHAP values of individual microstructures (PC1: − 8.4, PC2: − 4.9, PC3: − 3.8) corresponding to model outputs can be subdivided into negative contributions of the HAZ width and laser velocity and a counteracting influence of a rotation of 0° for PC1. For PC2, a higher weight on melt pool depth and velocity increases the SHAP value, while for PC3, laser velocity, depth, and melt pool width and HAZ width yield a negative value.

Indexing selected points in the region of interest within the PCA-reduced data space or alternatively, the use of unsupervised learning models allows exploration of the process parameter space associated with tendencies towards the corresponding CLD characteristics. With support of the trained and tested ML model, PC values can be predicted from new process parameter combinations, which allows reconstruction of CLDs in 3D. With support of the SHAP analysis, visualized data points in PC1-PC2 space can be further interpreted in Fig. 10 (Section 3.3). Other data points in the vicinity of data point no. 3 in Fig. 10 reflect e.g. a change in tail length (data point no. 4) or a change in rotation (not shown in Fig. 10), under otherwise constant parameters. From the PC1–PC2-plots, it becomes clear that these two features are the least

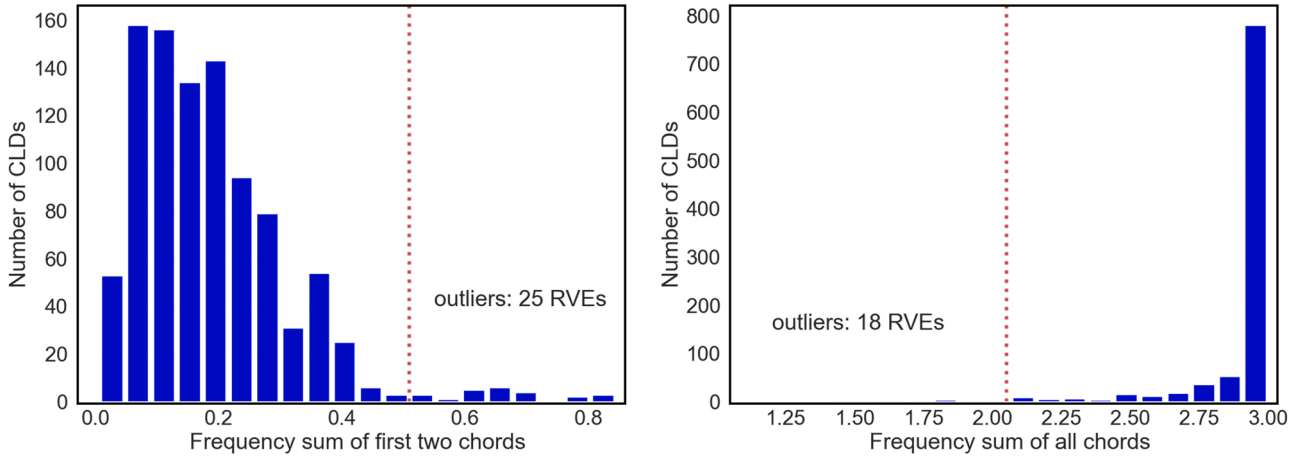


Fig. A.2. Outlier removal based on frequency sum of the first two chords (left) and the frequency sum of all chords (right).

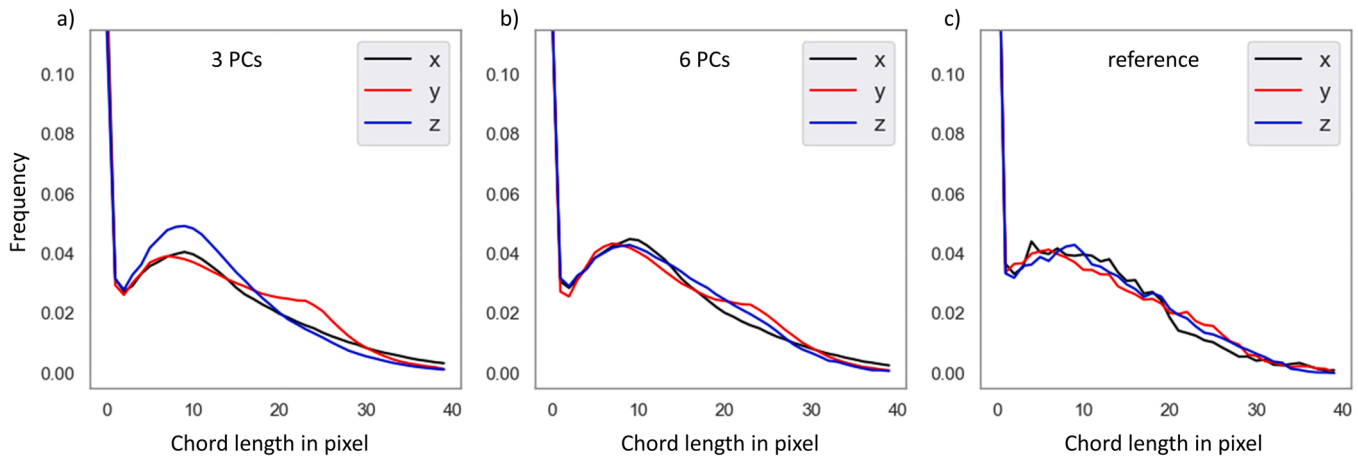


Fig. A.3. Reconstruction of CLDs based on up to a) three PCs, b) six PCs compared to c) original CLDs.

important to contribute to the model output. Together with the HAZ (only for PC2) and width (only for PC3), these features offer some degree of freedom for selection for reaching similar microstructure characteristics. Under constant velocity, PC1–PC2 values are influenced most by adjustments of the HAZ value. This can be seen from the same HAZ value for no. 3 and no. 4 with low PC2, whereas both, no. 1 and no. 2 with higher PC2 values are associated with higher HAZ values. Such insights allow controlled navigation in the process parameter space to guide the microstructure design for metal AM.

It has to be noted that SHAP values show indeed individual feature correlations to the model output, but these values provide no information on causality. If the feature space incorporates biases or pair-wise correlations, SHAP values can be misleading. Therefore, an actual causality of features and model output requires experimental design for validation of the interpretations based on the SHAP analysis.

4.2. Accuracy of the ML models

The best performing regression model in this study reached 87.3 % (PCA-reduced data). Changes along the processing steps of the framework (Fig. 3) allow further improvements of the ML models to close the gap towards higher accuracies. For instance, increasing the directions of scanned CLDs will capture more information, which might be hidden between the principal axes. In [43], directional CLDs were sampled with a step size of 1° providing more information in 2D, but sampling 3D microstructures requires more care to sample uniformly through the 3D space. Otherwise, a simple sampling rule such as sampling by steps of 30° rotations in polar angle θ and azimuthal angle φ ends up in a skewed dataset. Additionally, shorter line spacing distance of test lines might record more grain-related information but increase computation time, while large distances allow fast calculations coupled with information loss. During pre-processing, the number of CLDs was cut-off at 40 chords in each direction yielding an array of 120 in length per microstructure. An equal number of elements per array (to provide a $n \times m$ matrix with n rows and m columns) is a pre-requisite to perform PCA, but possibly removes grain-related information after the cut-off. Furthermore, data-cleaning criteria were introduced to remove outliers. Here, the threshold for identification of outliers represents a hyperparameter, which has ultimately an effect on the model accuracy. Or the RVE size or simulated layers can be seen as hyperparameter, where larger sizes than the dimensions used in this study (four layers within a 100^3 voxel cube) might capture more information comparable to characteristics of real microstructures.

Dimensionality reduction compresses the data and in case of PCA, orders PCs as data descriptors in a hierarchical manner. The difficult part of PCA lies in finding the right truncation level. The elbow method finds a cut-off point as threshold while plotting the explained variance over PCs. Comparable to the elbow method [44], eigenvalues plotted against PCs (referred to as scree plot [45]) identifies the threshold when additional PCs cause a transition from sharp decrease to a flat curvature. Such hard thresholds might cause to maintain noise in the data. Or the expert sets a fixed variance to be explained as threshold for finding the

Table B.1
Optimized parameters for ML models (POLY: polynomial regression, SVR: support vector regression, RF: random forest, MLP: multilayer perceptron) after using GridSearchCV from sklearn library. The number of components were varied as parameters of the dimensionality reduction method (PCA: principal component analysis, kPCA: kernel PCA), while additionally different kernel functions for kPCA was tested. POLY, SVR and RF share the same model parameters for different dimensionality reduction techniques.

Model	Dimensionality reduction	Hyperparameters
POLY	PCA, kPCA, t-SNE	degree: 4, interaction_only: False
SVR	PCA, kPCA, t-SNE	degree: 3, C: 500
RF	PCA	max_depth: 8, max_features: 'sqrt', min_samples_split: 2, n_estimators: 300
RF	kPCA, t-SNE	max_depth: 8, max_features: 'sqrt', min_samples_split: 2, n_estimators: 500
MLP	PCA	activation: 'logistic', alpha: 0.1, hidden_layer_sizes: (100, 100, 50, 25), learning_rate: 'constant', max_iter: 50, solver: 'adam'
MLP	kPCA	activation: 'tanh', alpha: 0.1, hidden_layer_sizes: (100, 100, 50, 25), learning_rate: 'adaptive', 'max_iter': 600, 'solver': 'adam'
MLP	t-SNE	activation: 'tanh', alpha: 0.1, hidden_layer_sizes: (100, 100, 50, 25), learning_rate: 'constant', max_iter: 200, solver: 'adam'
-	PCA	n_components: 3
-	kPCA	n_components: 3, kernel: "poly", gamma: 10, alpha: 0.1
-	t-SNE	n_components: 3, learning_rate: 'auto', init: 'random'

right number of PCs. Commonly these approaches suffer from their subjectivity. With the Gavish-Donoho method [46], a threshold where additional ranks yield more noise to the data is seen as optimal. This might be the most objective method among the statistical approaches for threshold identification. Furthermore, a closer look on the feature space of the data is necessary to find the most suitable data reduction technique, whether the data space is linearly or non-linearly seperable. For the latter, the kernel trick in terms of a kernel function transforms the data space to a higher dimension and thus, simplifies to find a decision surface. But finding the most suitable kernel function (e.g. a polynomial or radial kernel) and choosing a regularization term are essential for avoiding overfitting the model.

Training of ML models commonly requires setting of hyperparameters specific to the model. Tuning of simple ML model architectures (with one or two hyperparameters) is a rather straightforward task. In case of polynomial regression this gave the best results among the tested models. In contrast, ANNs that inherit a higher model complexity might increase the prediction accuracy, but deep networks are prone to overfitting, in particular if the available data is limited. Python libraries as Optuna [47] or Scikit-optimize [48] allow a systematic parameter search for individual models. Such libraries become powerful for models with numerous hyperparameters, e.g. for ANNs.

Tuning of the models after dimensionality reduction (Fig. 12) with kernel PCA and t-SNE yields a maximum performance of 88.7 % (kernel PCA-reduced data) and for the latter 89.1 % using polynomial regression (degree 4). Random forest reaches 85.4 % (kernel-PCA) and after using t-SNE surpasses with 88.3 % the accuracy compared to PCA-reduction. The lowest performance occurs for SVR, independent of the reduction technique. The MLP shows the best performance of 83.3 % after PCA reduction, while applying kernel-PCA and t-SNE cause an accuracy of 78.6 % and 83.2 %, respectively. Further information on the model parameters can be found in Appendix Table B.1.

In particular for process-structure-property (P-S-P) tasks, data is commonly sparse. Therefore incorporating physics into the learning process can guide ML models to learn linkages with higher accuracy. These physics can be represented by tailored and physically-meaningful features in the input data [24,49], or directly to the model architecture of physics-informed neural networks by adjusting e.g. the loss function [50,51]. Such data-driven approaches already show their capabilities in building linkages along composition-process property [49] or P-S-P [52] combinations.

5. Conclusions

In this study, we introduce a data-driven framework for scalable and physics-based exploration of the vast design space in metal additive manufacturing (AM). The developed open-source framework serves as a basis to enable predictive quantification of process-structure linkages by employing machine-learning (ML) models. In the presented study, the models were trained on physics-based, low-dimensional microstructure data extracted from kinetic Monte Carlo microstructure simulations as an exemplary microstructure generation tool. The 3D chord length distribution (CLD) is used as morphology descriptor of the as-built grain structure. Besides established principal component analysis (PCA) as dimensionality reduction technique, kernel-PCA and t-SNE are identified as powerful alternatives, in particular for non-linearly separable data. A selection of ML models is then trained for regression of the process parameter space as input towards the reduced CLD space as target variables. The ML model output is further interpreted in terms of explainable artificial intelligence (XAI) using SHapley Additive exPlanations (SHAP) values. The proposed approach is rather intended to provide a methodology on finding explainable process-microstructure linkages instead of obtaining highly accurate physical models. The following conclusions can be drawn:

- The applied methodology enables computational-efficient identification of process-structure linkages in metal AM. This includes dimensionality reduction of high-dimensional microstructure data to allow fast identification of process parameter combinations that yield tailored AM microstructures.
- Under constant laser velocity, the HAZ width showed significant contributions on the correlation between process parameters and corresponding microstructure, while the tail length of the melt pool and a change of the rotation angle appear less important. No rotation (0°) of the laser scan vector between subsequent layers results in microstructures with a higher degree of anisotropy.
- With the rise of black-box-type ML approaches, SHAP values are a promising tool to reveal global and local influences of input features on the model predictions. In metal AM, a SHAP analysis could support experts to interpret ML model outputs with respect to contributions of individual process parameters to their effect on the evolving AM microstructure.
- The highest ML model accuracy on learning investigated P-S linkages reached 90.6 %. Tuning the selected ML models to achieve more accurate P-S linkages is a tedious task and depends on multiple aspects, beginning with underlying characteristics of the CLDs in the dataset, the applied dimensionality reduction technique combined with its level of truncation, and the hyperparameters of the ML model. For the latter, open-source libraries (e.g. GridSearchCV in sklearn) allow systematic optimization of the model.

CRediT authorship contribution statement

Christian Haase: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Marc Ackermann:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

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

Data Availability

The collected dataset and code are publicly available at ⟨https://github.com/mrc989crm/InterpretablePSLinkagesInMetalAdditiveManufacturing⟩.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanýs Excellence Strategy – EXC-2023 Internet of Production – 390621612.

We would also like to acknowledge the support of the German Federal Ministry of Education and Research within the NanoMatFutur project “MatAM - Design of additively manufactured high-performance alloys for automotive applications” (Project ID 03XP0264).

Appendix A

See Figs. A.1–A.3.

Appendix B

See Table B.1.

References

[1] S.H. Huang, P. Liu, A. Mokasdar, L. Hou, Additive manufacturing and its societal impact: a literature review, Int. J. Adv. Manuf. Technol. 67 (2013) 1191–1203, https://doi.org/10.1007/s00170-012-4558-5.

[2] T. DebRoy, H.L. Wei, J.S. Zuback, T. Mukherjee, J.W. Elmer, J.O. Milewski, A. M. Beese, A. Wilson-Heid, A. De, W. Zhang, Additive manufacturing of metallic components – process, structure and properties, Prog. Mater. Sci. 92 (2018) 112–224, https://doi.org/10.1016/j.pmatsci.2017.10.001.

[3] W.J. Sames, F.A. List, S. Pannala, R.R. Dehoff, S.S. Babu, The metallurgy and processing science of metal additive manufacturing, Int. Mater. Rev. 61 (2016) 315–360, https://doi.org/10.1080/09506608.2015.1116649.

[4] Y.M. Wang, T. Voisin, J.T. McKeown, J. Ye, N.P. Calta, Z. Li, Z. Zeng, Y. Zhang, W. Chen, T.T. Roehling, R.T. Ott, M.K. Santala, P.J. Depond, M.J. Matthews, A. V. Hamza, T. Zhu, Additively manufactured hierarchical stainless steels with high strength and ductility, Nat. Mater. 17 (2018) 63–71, https://doi.org/10.1038/nmat5021.

[5] P. Köhnen, M. Létang, M. Voshage, J.H. Schleifenbaum, C. Haase, Understanding the process-microstructure correlations for tailoring the mechanical properties of L-PBF produced austenitic advanced high strength steel, Addit. Manuf. 30 (2019), 100914, https://doi.org/10.1016/j.addma.2019.100914.

[6] S.A.H. Motaman, F. Kies, P. Köhnen, M. Létang, M. Lin, A. Molotnikov, C. Haase, Optimal design for metal additive manufacturing: an integrated computational materials engineering (ICME) approach, JOM 72 (2020) 1092–1104, https://doi.org/10.1007/s11837-020-04028-4.

[7] M.F. Horstemeyer (Ed.), Integrated Computational Materials Engineering (ICME) for Metals, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2012.

[8] S.M. Hashemi, S. Parvizi, H. Baghbanijavid, A.T.L. Tan, M. Nematollahi, A. Ramazani, N.X. Fang, M. Elahinia, Computational modelling of process–structure–property–performance relationships in metal additive manufacturing: a review, Int. Mater. Rev. 67 (2022) 1–46, https://doi.org/10.1080/09506608.2020.1868889.

[9] M. Markl, C. Körner, Multiscale modeling of powder bed–based additive manufacturing, Annu. Rev. Mater. Res. 46 (2016) 93–123, https://doi.org/10.1146/annurev-matsci-070115-032158.

[10] N.H. Paulson, M.W. Priddy, D.L. McDowell, S.R. Kalidindi, Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics, Acta Mater. 129 (2017) 428–438, https://doi.org/10.1016/j.actamat.2017.03.009.

[11] E. Popova, T.M. Rodgers, X. Gong, A. Cecen, J.D. Madison, S.R. Kalidindi, Process-structure linkages using a data science approach: application to simulated additive manufacturing data, Integr. Mater. Manuf. Innov. 6 (2017) 54–68, https://doi.org/10.1007/s40192-017-0088-1.

[12] B.L. DeCost, T. Francis, E.A. Holm, Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures, Acta Mater. 133 (2017) 30–40, https://doi.org/10.1016/j.actamat.2017.05.014.

[13] P. Fernandez-Zelaia, Y.C. Yabansu, S.R. Kalidindi, A comparative study of the efficacy of local/global and parametric/nonparametric machine learning methods for establishing structure–property linkages in high-contrast 3D elastic composites, Integr. Mater. Manuf. Innov. 8 (2019) 67–81, https://doi.org/10.1007/s40192-019-00129-4.

[14] D. Montes de Oca Zapiain, E. Popova, F. Abdeljawad, J.W. Foulk, S.R. Kalidindi, H. Lim, Reduced-order microstructure-sensitive models for damage initiation in two-phase composites, Integr. Mater. Manuf. Innov. 7 (2018) 97–115, https://doi.org/10.1007/s40192-018-0112-0.

[15] B. Caglar, G. Broggi, M.A. Ali, L. Orgéas, V. Michaud, Deep learning accelerated prediction of the permeability of fibrous microstructures, Compos. Part A: Appl. Sci. Manuf. 158 (2022), 106973, https://doi.org/10.1016/j.compositesa.2022.106973.

[16] A. Bhutada, S. Kumar, D. Gunasegaram, A. Alankar, Machine learning based methods for obtaining correlations between microstructures and thermal stresses, Metals 11 (2021) 1167, https://doi.org/10.3390/met11081167.

[17] M. Ahmed, O.M. Horst, A. Obaied, I. Steinbach, I. Roslyakova, Automated image analysis for quantification of materials microstructure evolution, Model. Simul. Mater. Sci. Eng. 29 (2021) 55012, https://doi.org/10.1088/1361-651X/abfd1a.

[18] Z.-L. Wang, Y. Adachi, Property prediction and properties-to-microstructure inverse analysis of steels by a machine-learning approach, Mater. Sci. Eng.: A 744 (2019) 661–670, https://doi.org/10.1016/j.msea.2018.12.049.

[19] C. Herriott, A.D. Spear, Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine- and deep-learning methods, Comput. Mater. Sci. 175 (2020), 109599, https://doi.org/10.1016/j.commatsci.2020.109599.

[20] A. Pandey, R. Pokharel, Machine learning based surrogate modeling approach for mapping crystal deformation in three dimensions, Scr. Mater. 193 (2021) 1–5, https://doi.org/10.1016/j.scriptamat.2020.10.028.

[21] J. Jung, J.I. Yoon, H.K. Park, J.Y. Kim, H.S. Kim, An efficient machine learning approach to establish structure-property linkages, Comput. Mater. Sci. 156 (2019) 17–25, https://doi.org/10.1016/j.commatsci.2018.09.034.

[22] T.M. Rodgers, J.D. Madison, V. Tikare, Simulation of metal additive manufacturing microstructures using kinetic Monte Carlo, Comput. Mater. Sci. 135 (2017) 78–89, https://doi.org/10.1016/j.commatsci.2017.03.053.

[23] J. Goldak, A. Chakravarti, M. Bibby, A new finite element model for welding heat sources, MTB 15 (1984) 299–305, https://doi.org/10.1007/BF02667333.

[24] D. Kats, Z. Wang, Z. Gan, W.K. Liu, G.J. Wagner, Y. Lian, A physics-informed machine learning method for predicting grain structure characteristics in directed energy deposition, Comput. Mater. Sci. 202 (2022), 110958, https://doi.org/10.1016/j.commatsci.2021.110958.

[25] Y. Lian, Z. Gan, C. Yu, D. Kats, W.K. Liu, G.J. Wagner, A cellular automaton finite volume method for microstructure evolution during additive manufacturing, Mater. Des. 169 (2019), 107672, https://doi.org/10.1016/j.matdes.2019.107672.

[26] K. Teferra, D.J. Rowenhorst, Optimizing the cellular automata finite element model for additive manufacturing to simulate large microstructures, Acta Mater. 213 (2021), 116930, https://doi.org/10.1016/j.actamat.2021.116930.

[27] M.I. Latypov, M. Kühbach, I.J. Beyerlein, J.-C. Stinville, L.S. Toth, T.M. Pollock, S. R. Kalidindi, Application of chord length distributions and principal component analysis for quantification and representation of diverse polycrystalline microstructures, Mater. Charact. 145 (2018) 671–685, https://doi.org/10.1016/j.matchar.2018.09.020.

[28] S.A.H. Motaman, D. Kibaroglu, The anisotropic grain size effect on the mechanical response of polycrystals: the role of columnar grain morphology in additively manufactured metals, 2022. arXiv preprint arXiv:2211.05879.

[29] M.A. Groeber, M.A. Jackson, DREAM.3D: a digital representation environment for the analysis of microstructure in 3D, Integr. Mater. Manuf. Innov. 3 (2014) 56–72, https://doi.org/10.1186/2193-9772-3-5.

[30] B. Schölkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, in: G. Goos, J. Hartmanis, J. van Leeuwen, W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (Eds.), Artificial Neural Networks — ICANN'97, Springer Berlin Heidelberg, Berlin, Heidelberg, 1997, pp. 583–588.

[31] T. Hofmann, B. Schölkopf, A.J. Smola, Kernel methods in machine learning, Ann. Stat. 36 (2008), https://doi.org/10.1214/009053607000000677.

[32] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 11.

[33] S. Brandt, Data Analysis, Springer International Publishing, Cham, 2014.

[34] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn 20 (1995) 273–297, https://doi.org/10.1007/BF00994018.

[35] M. Awad, R. Khanna, Support vector regression, in: M. Awad, R. Khanna (Eds.), Efficient Learning Machines, Apress, Berkeley, CA, 2015, pp. 67–80.

[36] T. Rieg, J. Frick, H. Baumgartl, R. Buettner, Demonstration of the potential of white-box machine learning approaches to gain insights from cardiovascular disease electrocardiograms, PLoS One 15 (2020), e0243615, https://doi.org/10.1371/journal.pone.0243615.

[37] M. Wu, X. Zhong, Q. Peng, M. Xu, S. Huang, J. Yuan, J. Ma, T. Tan, Prediction of molecular subtypes of breast cancer using BI-RADS features based on a "white box" machine learning approach in a multi-modal imaging setting, Eur. J. Radiol. 114 (2019) 175–184, https://doi.org/10.1016/j.ejrad.2019.03.015.

[38] A. Liaw, Matthew Wiener, Classification and regression by randomForest, R News 2 (3) (2002) 18–22.

[39] M. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmos. Environ. 32 (1998) 2627–2636, https://doi.org/10.1016/S1352-2310(97)00447-0.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[41] Scott M. Lundberg, Su-In Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[42] M. Sundararajan, A. Najmi, The many Shapley values for model explanation, in: Proceedings of the International Conference on Machine Learning, PMLR, 2020.

[43] Y.S. Fan, X.G. Yang, D.Q. Shi, L. Tan, W.Q. Huang, Quantitative mapping of service process-microstructural degradation-property deterioration for a Ni-based superalloy based on chord length distribution imaging process, Mater. Des. 203 (2021), 109561, https://doi.org/10.1016/j.matdes.2021.109561.

[44] L. Ferré, Selection of components in principal component analysis: a comparison of methods, Comput. Stat. Data Anal. 19 (1995) 669–682, https://doi.org/10.1016/0167-9473(94)00020-J.

[45] M. Zhu, A. Ghodsi, Automatic dimensionality selection from the scree plot via the use of profile likelihood, Comput. Stat. Data Anal. 51 (2006) 918–930, https://doi.org/10.1016/j.csda.2005.09.010.

[46] M. Gavish, D.L. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$, IEEE Trans. Inform. Theory 60 (2014) 5040–5053, https://doi.org/10.1109/TIT.2014.2323359.

[47] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna, in: A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, G. Karypis (Eds.), Proceedings of the 25th ACM

SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, New York, NY, USA, 2019, pp. 2623–31.
[48] Tim Head, MechCoder, Gilles Louppe, Iaroslav Shcherbatyi, fcharras, Zé Vinícius, cmmalone, Christopher Schröder, nel215, Nuno Campos, Todd Young, Stefano Cereda, Thomas Fan, rene-rex, Kejia (KJ) Shi, Justus Schwabedal, carlosdanielcsantos, Hvass-Labs, Mikhail Pak, SoManyUsernamesTaken, Fred Callaway, Loïc Estève, Lilian Besson, Mehdi Cherti, Karlson Pfannschmidt, Fabian Linzberger, Christophe Cauet, Anna Gut, Andreas Mueller, Alexander Fabisch, Scikit-Optimize/Scikit-Optimize: V0.5.2, Zenodo, 2018.
[49] S. Liu, B.B. Kappes, B. Amin-ahmadi, O. Benafan, X. Zhang, A.P. Stebner, Physics-informed machine learning for composition – process – property design: shape memory alloy demonstration, Appl. Mater. Today 22 (2021), 100898, https://doi.org/10.1016/j.apmt.2020.100898.
[50] E. McGowan, V. Gawade, W.G. Guo, A physics-informed convolutional neural network with custom loss functions for porosity prediction in laser metal deposition, Sensors 22 (2022), https://doi.org/10.3390/s22020494.
[51] B. Kim, V.C. Azevedo, N. Thuerey, T. Kim, M. Gross, B. Solenthaler, Deep fluids: a generative network for parameterized fluid simulations, Comput. Graph. Forum 38 (2019) 59–70, https://doi.org/10.1111/cgf.13619.
[52] R.N. Saunders, K. Teferra, A. Elwany, J.G. Michopoulos, D. Lagoudas, Metal AM process-structure-property relational linkages using Gaussian process surrogates, Addit. Manuf. 62 (2023), 103398, https://doi.org/10.1016/j.addma.2023.103398.