International Journal of Material Forming (2026) 19:31
https://doi.org/10.1007/s12289-026-01995-y

RESEARCH

ESAFORM Benchmark 2025: predicting stainless steel PBF-LB part density using statistical, data-driven, and physics-informed machine learning models derived from process parameters and *in-situ* monitoring data

Medad Chiedozie C. Monu[1,2] · Eanna McCarthy[1,2] · Abhilash Puthanveettil Madathil[3] · Josiah C. Chekotu[1,2] · Irina Ilic[4] · Merve Nur Doğu[1,2] · Cian Hughes[1,2] · Rongfei Juan[5] · Ehsan Amini[5] · Junhe Lian[5,6] · Constantinos Vassiades[3] · Olga Bylya[3] · Kim Darosa[7] · Robin Kromer[7] · Abdul Herrim Seidou[8] · Sankhya Mohanty[9] · Anne Marie Habraken[10] · Anne Mertens[8] · Otto Laitinen[4] · Michael R. Tucker[4] · Dermot Brabazon[1,2]

Received: 20 January 2026 / Accepted: 25 February 2026

Abstract

This study benchmarks multiple data-driven methodologies for predicting relative density (RD) of 316 L stainless steel fabricated via Powder Bed Fusion–Laser Beam (PBF-LB), as part of the ESAFORM Benchmark 2025 AMDmodel initiative. Two datasets (DS-01 and DS-02), each with 256 specimens from a 4-factor, 4-level design of experiments, were produced on different PBF-LB systems equipped with equivalent *in-situ* infrared (IR) melt-pool pyrometry. Failed builds (RD = 60%) were retained to allow models to learn from both nominal and catastrophic processing conditions, a scenario rarely addressed in PBF-LB machine learning (ML). Statistical analysis of variance (ANOVA) confirmed that conventional process parameters alone are weak predictors ($R^2 \approx 0.49$). In contrast, sensor-driven supervised ML models using melt-pool thermal descriptors performed substantially better. Recursive feature elimination highlighted the interquartile range and mode of thermal signatures as dominant predictors; an XGBoost model using only these achieved $R^2 = 0.93$ on DS-01. Hybrid models combining parameters and IR descriptors performed slightly worse ($R^2 = 0.92$), indicating mild redundancy. Cross-system transferability was limited: ML models trained on DS-01 underperformed on DS-02 due to IR input-domain divergence despite RD distributions between both domain sources showing high inter-laboratory consistency. To address this, a physics-informed ML framework (PIML) using symbolic regression (QLattice) embedded dimensionless physical priors. Resulting compact expressions dominated by normalized laser power and volumetric energy density achieved $R^2 = 0.83$–0.93 under cross-system validation. Overall, sensor-driven ML models are effective for machine-specific monitoring and layer-wise closed-loop control, whereas PIML provide system-agnostic process parameter-window estimation for design-stage optimization.

Keywords Benchmark, Machine Learning · *In-situ* monitoring · Additive manufacturing · Physics-informed machine learning · Density prediction

Introduction

Machine learning (ML) is transforming metal additive manufacturing (AM) by providing powerful tools for process monitoring, quality assurance, and ultimately, autonomous control [1, 2]. This is especially critical in laser-based Powder Bed Fusion (PBF-LB), where final part quality is governed by complex, transient melt pool dynamics. Data-driven models offer a pathway to predictive control; however, their reliable deployment in industrial settings requires rigorous benchmarking (i.e., a systematic evaluation of algorithms, features, and datasets). Benchmarking is, therefore, not merely an academic exercise but a practical necessity. It enables reproducible comparisons, quantifies algorithm sensitivity to

Extended author information available on the last page of the article

Published online: 24 March 2026

data quality and distribution, and evaluates model generalization across machines and process conditions. These functions are essential for developing trustworthy digital twins capable of emulating process–structure–property relationships and enabling real-time defect detection, optimisation, and closed-loop control. By reducing experimental waste and enforcing process consistency, ML benchmarking also directly contributes to sustainability in AM.

Although deep learning has gained prominence in materials informatics, its "black-box" nature, high data requirements, and limited explainability constrain its practical usefulness in AM. Generalisability across machines, scanning strategies, or operating windows remains particularly challenging [3]. Conventional supervised ML methods, by contrast, offer a more favourable balance between predictive capability, computational efficiency, and interpretability. As Chen et al. [4] emphasise, AM applications often demand a pragmatic compromise between accuracy and transparency, one that supervised learning algorithms are well suited to deliver. Rigorous benchmarking of such algorithms is therefore critical for assessing robustness and establishing actionable insights for process engineers.

Within PBF-LB research, a substantial body of work has examined supervised ML for predicting Archimedes-derived relative density (RD) from process parameters: laser power P [W], scan speed V [mm/s], hatch distance H [mm], spot size f [mm], and layer thickness L [mm]. The most comprehensive comparative study to date for stainless steel is by Barrionuevo et al. [5], who compiled 112 datasets from the literature to train seven supervised ML models. Their dataset spanned RD range (91.20–99.90%) and included the key process parameters: P, V, L, H, and powder particle size. Among the evaluated models (SVM, decision tree, random forest, gradient boosting, Gaussian process, KNN, and MLP), the gradient boosting regressor achieved the best performance (R^2 = 0.63). This moderate accuracy indicates that a substantial portion of the variance in RD cannot be explained by process parameters alone, highlighting both process stochasticity and the limitations of parameter-only based models. Predictions were most reliable at higher RD (> 97%), where defect mechanisms tend to be more uniform. More recently, Barrionuevo et al. [6] published an expanded RD dataset comprising 1,579 observations across multiple alloys, including 316 L stainless steel, AlSi10Mg, 18Ni300, Inconel 718, Ti-6Al-4 V, and CuCrZr, compiled exclusively from published sources. The corresponding input variables include the average particle size (D50), P, f, H, V, L, build-atmosphere gas, printed geometry, and a proposed geometric factor (Eq. 1). This geometric factor accounts for the machine build-volume envelope (V_M), the part volume (V_{part}), and the number of samples produced (n).

$$GF = \left(1 - \frac{V_{part}}{V_M}\right) n \tag{1}$$

Complementary studies have reported higher predictive performance, though typically using smaller or less representative datasets. Gor et al. [7] found that ANN (R^2 = 0.95) and SVM (R^2 = 0.92) outperformed KNN and linear regression for 316 L SS, but their relatively small ~ 40 sample dataset restricts generalisability. Using a broader aggregated dataset of 201 datapoints, Abdulla et al. [8] showed that kernel ridge regression (R^2 = 0.85) surpassed ridge regression and SVR across a wide process window (P = 30–400 W, V = 50–2400 mm/s, H = 0.04–0.3 mm, L = 0.02–0.25 mm, and RD = 75–100%). Additional experimental validation by Eshkabilov et al. [9] demonstrated consistently high performance across six classification models and five SVR models for 316 L components. Collectively, these studies show that ANN, SVM, kernel-based regressors, and gradient-boosting methods can deliver strong predictive accuracy for Archimedes-based RD, though performance remains strongly constrained by small dataset size, a lack of representativeness of a broader range of processing conditions, and a lack of coverage of the defect landscape.

A fundamentally different approach to ground-truth definition was adopted by Wang et al. [10], who evaluated 18 machine-learning models using 2,167 process conditions spanning 50 metal powders. Their dataset combined intrinsic material properties [reflectivity (R, %), powder thermal conductivity (k, $W \cdot m^{-1} \cdot K^{-1}$), specific heat capacity (Cp, $J \cdot g^{-1} \cdot K^{-1}$), and melting point ($T_m$, °C)] with process parameters (P, V, L, H, and VED). In this study, RD was quantified via optical microscopy rather than the Archimedes method and subsequently transformed using a sigmoid function, thereby introducing a distinct measurement basis with its own sources of uncertainty. Nevertheless, the XGBoost model achieved high predictive accuracy for RD values above 98% and demonstrated strong generalisation during experimental validation. Because the RD metric in this work is intrinsically different from gravimetric definitions, the resulting models are methodologically distinct from Archimedes-based studies, underscoring the critical importance of explicitly defining "ground truth" in accordance with the underlying data and measurement techniques.

Collectively, these studies highlight a broader limitation of traditional ML approaches in PBF-LB: achieving robust and transferable predictive performance remains challenging when models rely primarily on static process parameters or empirically derived features. Such inputs often serve only as indirect proxies for the underlying thermo-physical phenomena governing defect formation and densification. To address this limitation, a limited but growing body of work has turned toward *in-situ* sensing to capture process

dynamics more directly. Mahato et al. [11], for example, analysed coaxial infrared (IR) melt-pool emissions during PBF-LB of NiTi to predict porosity. Two modelling strategies were explored: binary classification of layer-level porosity (≥ 1% vs. <1%) and regression-based prediction of RD. The IR-driven models achieved strong performance, with RMSE values of 0.0367 for low-porosity layers and 0.108 for high-porosity layers. Although this study was not conducted on 316 L stainless steel, it demonstrates that consistently high predictive accuracy may require descriptors such as *in-situ* IR signals, that more directly encode process dynamics than static process parameters.

This evolution from parameter-based models to sensor-driven ML naturally motivates the adoption of physics-informed machine-learning (PIML) frameworks. PIML approaches have attracted increasing attention as a means of overcoming the limitations of purely data-driven models in PBF-LB. By embedding physical priors derived from governing equations, process maps, mechanistic constraints, or expert knowledge, PIML seeks to simultaneously improve predictive accuracy, physical interpretability, and generalisation capability [3, 12, 13]. To the best of the authors' knowledge, the application of PIML specifically to RD prediction remains limited, a gap that has been highlighted in several recent review articles [1, 3]. Existing PIML studies in additive manufacturing have instead focused primarily on fatigue life prediction [19, 20], temperature field reconstruction, and melt pool thermal–fluid dynamics [12, 14–17]. In most of these works, physics-informed neural networks (PINNs) constitute the preferred modeling paradigm. While PINNs often achieve high predictive performance, their reliance on deep neural architectures can obscure physical interpretability, thereby limiting their practical utility for real-time monitoring and feedback control in industrial AM systems. One promising strategy to address this limitation is the integration of symbolic regression within the PIML pipeline, replacing neural networks with explicitly interpretable analytical models. Symbolic regression provides a transparent bridge between empirical data fitting and first-principles reasoning, enabling not only accurate RD prediction but also direct insight into dominant physical mechanisms and feasible operating regimes.

Building on a sequence of prior ESAFORM benchmark studies on conventional forming processes with extensive characterisation frameworks [18–20], this ESAFORM 2025 Benchmark initiative (AM-D-model) represents the first benchmark specifically dedicated to PBF-LB AM processes. In this context, the present study advances a rigorous, multi-method evaluation framework for predicting PBF-LB part quality. The benchmark implements a comprehensive comparison across three complementary methodological pillars: (i) statistical analysis of process parameters and their influence on RD, (ii) supervised machine-learning models trained on *in-situ* melt pool thermal signatures, and (iii) physics-informed symbolic regression incorporating dimensionless physical priors. By evaluating forty algorithms across parameter-only, sensor-driven, and hybrid feature sets, this study identifies modeling strategies that optimally balance predictive accuracy, physical consistency, and system-agnostic transferability. Crucially, this broad methodological scope establishes a transparent and reproducible framework for model selection that extends beyond conventional performance-based benchmarking. Rather than merely ranking algorithms by aggregate accuracy, the benchmark enables systematic analysis of model behaviour by resolving feature dominance and sensitivity across the full RD spectrum, including low-density/defect-dominated regimes that are typically excluded in prior work. This capability is particularly novel in the context of PBF-LB quality prediction, as it facilitates direct, like-for-like comparison of parameter-based, sensor-driven, and physics-informed models under identical conditions, thereby revealing fundamental trade-offs between predictive performance, physical interpretability, and generalization. As a result, the benchmark serves not only as a methodological reference for future machine-learning studies in additive manufacturing, but also as a practical decision-support framework for industrial deployment, guiding the selection of robust and interpretable models that remain reliable under process variability and extreme defect conditions, with direct implications for scalable quality assurance and more sustainable production workflows.

Experimental procedure and methods

Experimental design, sample fabrication, and density measurement

The samples for this benchmark study were fabricated using two PBF-LB systems. The first, an Aconity*MINI* system located at Dublin City University, Ireland, is equipped with a 200 W, 1068 nm Yb-fibre laser and a build volume of Ø 140 × H 190 mm. The second, an Aconity*MIDI*+ system at ETH Zurich, Switzerland, has a 500 W, 1068 nm Yb-fibre laser and a build volume of Ø 250 × H 250 mm. In both systems, argon was used as shielding gas, the same powder feedstock was used. The feedstock used was PowderRange® 316 L stainless steel (15–45 µm particle size,) all from the same gas-atomization batch, supplied by the powder manufacturer, Carpenter Additive®. Each sample was cubic (7 × 7 × 7.2 mm) with a tapered base support (2.85 mm high) for stability and ease of detachment (Fig. 1a–b). Samples were printed with a 300 µm-high numerical label, giving

a total height of 10.35 mm. A bi-directional scanning strategy was employed, beginning with a 45° scan vector and rotating each subsequent layer by 90°. In addition, the "Skywriting" feature was enabled to enhance contour accuracy, reduce transient effects, and improve scan quality.

A full-factorial experimental design with four factors at four levels ($4^4 = 256$ parameter combinations) was implemented to comprehensively explore the PBF-LB process space (Table 1). Each combination was fabricated twice (one set on the Aconity*MINI* and another set on the Aconity*MIDI*+) resulting in 512 samples produced. On the Aconity*MINI*, samples were fabricated in batches of 100, with the final batch containing 112 samples (two sets of the last 56 samples) to complete the set. This batching ensured consistent interlayer times and uniform thermal histories. Consequently, data from multiple print sessions captured both systematic and random variations typical of day-to-day industrial/research PBF-LB operations. On the larger Aconity*MIDI*+, all 256 samples were fabricated in a single batch. Inter-sample spacing was maintained at 2.33 mm on both systems. Representative images of 200 samples are shown in Fig. 1c-d.

The derived PBF-LB metrics represent the Volumetric Energy Density (VED) calculated using either the spot size, ***f*** variant (VED_f) or the hatch spacing, ***H*** variant (VED_H), according to the following expressions:

$$VED_f = \frac{P}{V \times L \times f}\left(\frac{J}{mm^3}\right) \tag{2}$$

$$VED_H = \frac{P}{V \times L \times H}\left(\frac{J}{mm^3}\right) \tag{3}$$

Relative density (RD), the target property, was calculated as the ratio of measured sample density to the theoretical density of fully dense 316 L (assumed 8000 kg/m³). Measurements followed Archimedes' principle according to ASTM B962-23 [21], using ethanol as the immersion medium.

In-situ monitoring by infrared (IR) pyrometry and IR data characteristics

In-situ temporal infrared (IR) monitoring was conducted during fabrication using Kleiber® KG 740-LO pyrometers integrated into the Aconity*MINI* and Aconity*MIDI*+ systems. The pyrometers are calibrated against a black-body standard, can measure between 500 and 2500 °C, has a spectral range of 1.58–1.80 µm, a 10 µs response time (T90), and a 10-bit resolution. Both pyrometers were coaxially aligned with the laser beam path, enabling precise capture of melt pool thermal emissions (X, Y coordinates and the sensor signal) for each layer at 100 kHz. This generated ~ 850 GB of raw data (~ 20 million points per layer in PCD (Point Cloud Data) format), far exceeding the capacity of existing commercial analysis tools. An in-house software solution, PyroVista, developed for *in-situ* PBF-LB process quality control, was used to post-process the raw IR data from both Aconity*MINI* and Aconity*MIDI*+, hereafter referred to as data or domain sources DS-01 and DS-02, respectively.

The post-processing workflow comprised thresholding, segmentation, and statistical feature extraction. Signal thresholds of 825 °C for DS-01 and 3 mV for DS-02 were applied to suppress non-physical or process-irrelevant signals, most notably those associated with skywriting during laser repositioning between scan vectors. During skywriting, the laser operates in an off or low-power state that is insufficient to melt powder particles; however, the pyrometers continue to record the laser trajectory. This results in systematic low-temperature traces that do not reflect melt pool behavior and must be excluded prior to analysis. The effect of this thresholding is illustrated in the reconstructed 2D thermograms in Fig. 2a-d. Subsequently, segmentation was performed to isolate melt pool data corresponding to individual samples from the full build-plate dataset.

Following segmentation, statistical feature extraction was conducted on the melt pool thermal distribution for each individual layer using PyroVista. A total of 16 statistical descriptors were extracted, capturing complementary aspects of the thermal distribution, including measures of central tendency (e.g., mean, median, mode), dispersion (e.g., variance, interquartile range), and distribution shape (e.g., skewness, kurtosis), as summarized in Table 2. The selection of this feature set was guided by physical considerations of melt pool thermal behavior in PBF-LB/M. Given that the a *priori* relevance of individual melt pool thermal descriptors to bulk RD is not known, an initially broader feature space was deliberately adopted to avoid prematurely excluding potentially informative characteristics.

Layer-resolved features were subsequently aggregated via averaging across all layers to obtain a single set of global thermal descriptors for each sample. This aggregation was chosen intentionally, as RD is a bulk, post-build property that reflects the cumulative thermal history of the process rather than isolated layer-level events. Moreover, pyrometer measurement signals are subject to significant layer-to-layer melt pool temperature variability arising from scan-strategy–induced transient heating phenomena, including moving point heating phenomena during early scan vectors when laser interacts with relatively cold, unprocessed powder; the transition to moving band heating phenomena as hatch vectors accumulate and thermal overlap increases; and edge-related heat accumulation toward the end of scan vectors as scan vectors shorten [22, 23]. Other variability sources includes, spatter-induced signal disturbances, and

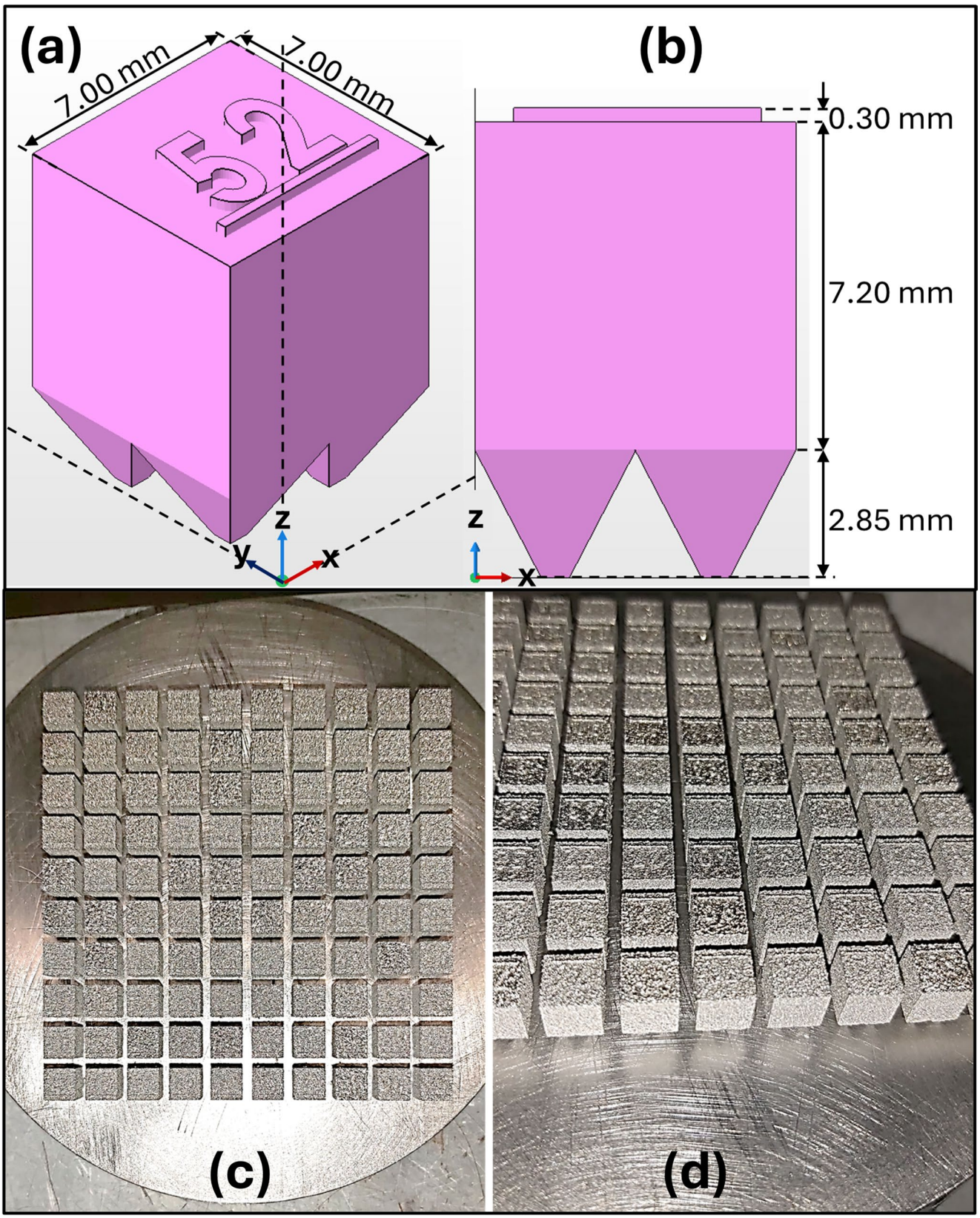


Fig. 1 CAD representations of the sample geometry: (**a**) Isometric view and (**b**) Front view; (**c**) and (**d**) photographs of two plates of printed samples still attached to the build plate for DS-01 batch 1

Table 1 Tabular summary of the 4^4 full factorial experimental design

	Factors	Levels	Units
1.	Laser Power (P)	[80, 120, 160, 200]	W
2.	Scan Speed (*V*)	[600, 800, 1000, 1200]	mm/s
3.	Hatch Spacing (*H*)	[60, 90, 120, 150]	µm
4.	Spot Size (*f*)	[80, 90, 100, 110]	µm
	Derived PBF-LB Metrics	**Range**	
5.	VED_f	[12.12 to 83.33]	J/mm^3
6.	VED_H	[8.89 to 111.11]	J/mm^3

Layer thickness (*L*) was kept constant at 50 µm

Table 2 Summary of statistical descriptors derived from the melt pool's thermal emission features based on the *in-situ* collated IR data

Melt pool thermal features		Melt pool thermal features
Max_Temp	9.	Coefficient_of_Variation_Temp
Min_Temp	10.	Range_Temp
Avg_Temp	11.	Variance_Temp
Median_Temp	12.	Q1_Temp
Mode_Temp	13.	Q3_Temp
Stdev_Temp	14.	Interquartile_Range_Temp
Skewness_Temp	15.	Percentile_90_Temp
Kurtosis_Temp	16.	Percentile_10_Temp

For DS-02, the extracted features have '_mV' suffixes instead of '_Temp' (DS-01)

local emissivity variations. Averaging across layers therefore acts as a physically motivated filtering step, suppressing stochastic and geometry-dependent fluctuations while preserving the persistent thermal signatures most relevant to densification. The resulting globally averaged feature vectors provide a compact, stable, and reproducible representation of the melt pool thermal regime for each sample and form the input to the subsequent analysis. *Note*; *The terms "melt pool thermal emission" or "melt pool thermal descriptors", or similar phrasing used throughout the article, is used in a generalized sense, as the two black-body-calibrated pyrometer signals are recorded in different units (°C and mV). Owing to machine-level constraints of the PBF-LB systems, these units cannot be modified/changed, and the data are therefore analysed in their native formats*

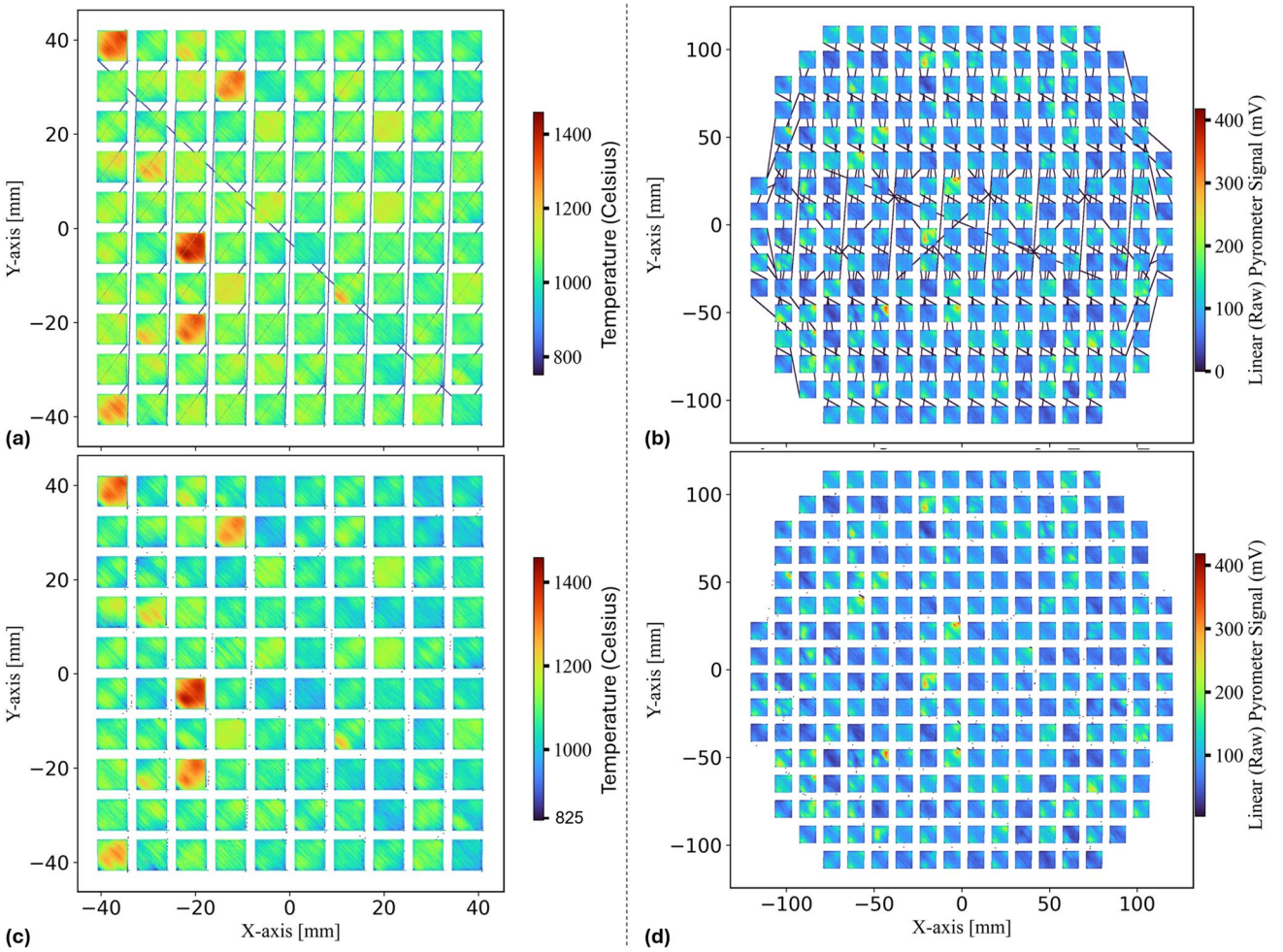


Fig. 2 Thermograms of a build layer of samples reconstructed from collated IR point cloud data illustrating (**a** and **b**) noisy data with false temperature tracks due to skywriting, with degrees Celsius and mV for the DS-01and DS-02 respectively, and (**c** and **d**) denoised thermogram of the same layer after thresholding, with degrees Celsius and mV for the DS-01 and DS-02 respectively

(further discussion is provided in Section "Physical interpretation: aligning feature importance with solidification fundamentals").

Sensor data-driven model benchmarking framework

A structured benchmarking framework was implemented to address common limitations in conventional ML workflows, specifically the risk of optimistically biased performance estimates. The framework enforces methodological rigor through a strict separation of exploratory model screening and confirmatory validation. This process is organized into three sequential phases: Phase I for data preparation and feature selection, Phase II for preliminary model screening and selection using the DS-01 training dataset, and Phase III for comprehensive validation using the independent, held-out DS-02 dataset. A schematic of the entire workflow is shown in Fig. 3.

Phase I: Data preparation and feature selection by recursive feature elimination with cross-validation (RFECV)

DS-01 data was used for model development and validation, while the full set of DS-02 was used for model transferability evaluation. The DS-01 dataset was partitioned via random sampling into a training subset (80%) for model development and a validation subset (20%) for independent evaluation. To ensure a robust validation of model performance across the density range, a constraint was enforced to include a minimum of five low-density samples in the DS-01 validation subset. This procedure resulted in a training set of 205 samples (including 10 low-density samples) and a validation set of 51 samples (including 5 low-density samples). The two subsets were stored separately to prevent data leakage. A hash-based duplication detection algorithm was employed to confirm the complete separation of samples between the training and validation sets. A subsequent distributional analysis was conducted to compare the target variable and feature distributions across the partitioned sets. The analysis revealed that the DS-01 validation set presented a more challenging prediction task, characterized by a lower mean RD (90.78% versus 92.80%) and higher variability (standard deviation of 10.46 versus 7.82) compared to the training set. The value ranges for RD were comparable, spanning from 60.00 to 99.00 in the training set and from 60.00 to 98.62 in the validation set.

Three groups of input features were curated for feature selection stage: (a) the PBF-LB process parameters listed in Table 1; (b) fifteen (15) statistical descriptors of melt pool thermal signature derived from infrared (IR) pyrometer measurements (refer to Table 2), with Min_Temp excluded due to post-processing thresholding; and (c) combination of the PBF-LB and melt pool temperature descriptors, yeilding 21 predictor features per sample. Feature selection was then performed using Recursive Feature Elimination with Cross-Validation (RFECV) [11, 24] exclusively on the DS-01 training dataset. The objective was to identify the most parsimonious feature subset that maximizes predictive performance for RD, thereby enhancing model interpretability. This approach was preferred over consensus-based methods, which often retain marginally beneficial features and lead to unnecessary complexity. This RFECV process, mathematically summarized in Table 3, was implemented with a diverse set of four regressors representing different algorithmic families: Decision Tree for interpretable single-tree modelling, Random Forest for robust ensemble learning, Lasso for L1-regularized linear modelling with inherent feature selection, and Ridge for L2-regularized linear modelling handling multicollinearity. This diversity ensured that feature selection was not biased toward a single algorithmic paradigm.

To ensure robustness, the RFECV procedure was executed through five independent runs per original feature set, each with a distinct random seed (42–46). Each run began with an 80:20 train-test split of the DS-01 training dataset. The RFECV algorithm was configured with shuffled 5-fold cross-validation (CV) using *KFold()*, a step size of one feature per iteration, and R^2 as the optimization metric. All features were normalized using Z-score normalization (*StandardScaler()*) to ensure consistent scaling for all algorithms, a necessary step for the stable performance of distance-based and linear models. The target variable, RD, was retained in its original units (%). For each regressor in each run, the algorithm recursively eliminated the least important features based on model-specific importance metrics (Gini importance for tree-based models, coefficient magnitudes for linear models) while monitoring cross-validation performance. The optimal feature set for a given regressor-run combination was identified as the point that maximized the cross-validation R^2 score. A performance-driven strategy was employed for the final feature selection. For each run, the feature set from the regressor achieving the highest test R^2 was selected, provided it met a minimum performance threshold of $R^2 \geq 0.50$. This approach prioritized empirical predictive power. The overall best feature set for subsequent model screening was then identified as the one yielding the absolute highest test R^2 across all regressors and all RFECV runs.

Note that under specific conditions, the relationship between Pearson's correlation coefficient (r) and the coefficient of determination (R^2) described in Table 3 reduces to $R^2 = r^2$. This equivalence holds in simple linear regression only when the intercept is constrained to zero, and the

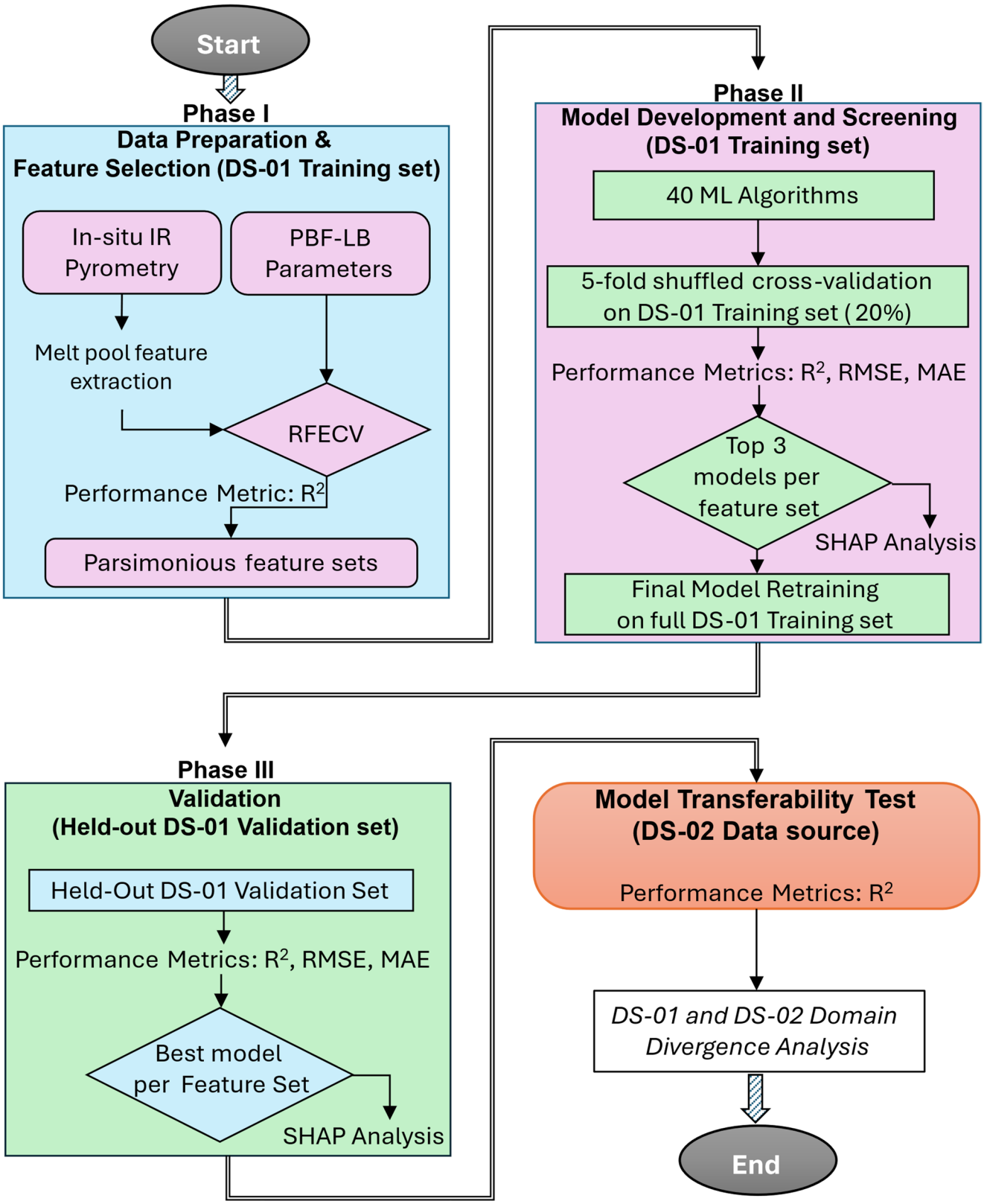


Fig. 3 Sensor data-driven model benchmarking framework

Table 3 Equations of feature selection and model performance metrics

Metric	Formula
Pearson's correlation, r	$\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2\sum_{i=1}^{N}\left(y_i-\bar{y}\right)^2}}$
RFECV score, (S)	$\frac{1}{k}\sum_{i=1}^{k} score\,(S_i)$
Coefficient of determination, R^2	$1-\frac{\sum_{i=1}^{N}(X_i-Y_i)^2}{\sum_{i-1}^{N}\left(\bar{Y}-Y_i\right)^2}$
Root mean squared error, $RMSE$	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-Y_i)^2}$
Mean absolute error, MAE	$\frac{1}{N}\sum_{i=1}^{N}\lvert Y_i-X_i\rvert$
Residuals	$e_i=Y_i-X_i$

where N represents the number of data points, X_i is the predicted i^{th} value, the Y_i is the actual (i.e., true or measured) i^{th} value, and $\bar{Y}$ represents the mean of the true values calculated as: $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$. S is the set of selected features, k is the number of folds in the cross-validation, S_i is the subset of features in fold , and $score\,(S_i)$ is the model's performance score on fold with the subset S_i

predictor and response share identical scaling. Under these conditions, both metrics quantify the proportion of variance in the response explained by a perfectly scaled linear relationship. However, when the regression includes a non-zero intercept or when the predictor and response differ in scale, R^2 will generally be less than r^2. In these situations, Pearson's r measures only the strength of linear association, whereas R^2 additionally penalizes bias (intercept misalignment) and scale mismatch (slope differing from unity). For well-behaved physical relationships, r^2 therefore represents the maximum achievable R^2 after perfect linear calibration.

Phase II: Preliminary model screening and development

This phase consisted of preliminary screening across 40 diverse ML algorithms from scikit-learn [25, 26], representing eleven distinct algorithmic families (Table 4). The screening was conducted using the DS-01 training with feature sets previously optimized through RFECV. Before model training, a distributional analysis confirmed the absence of significant covariate shift between the splits used for screening. Like the RFECV pipeline, model screening and evaluation were performed using an 80:20 split of the DS-01 training subset subjected to a shuffled 5-fold cross-validation (*KFold()*), with all predictor features normalized using Z-score normalization (*StandardScaler()*). RD was retained in its original units (%). Performance was assessed using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE), as defined in Table 3 [27, 28]. Cross-validation consistency was evaluated by calculating the mean and standard deviation of these metrics across the five folds. The top three performing models for each feature set were identified based on their cross-validation performance for progression to the next phase.

Phase III: Comprehensive validation of the top-performing models

The third phase focused on the comprehensive validation of the top-performing models identified during the screening stage at Phase II. Each selected model was retrained on the full DS-01 training dataset to leverage all available data and maximise learning. The final evaluation was then conducted on the completely held-out DS-01 validation set, which was excluded from both the RFECV feature selection and the model screening phases. This provided an unbiased assessment of generalization performance on the specific PBF-LB system from which the data originated. The same evaluation metrics defined in Table 3 were reported to quantify model performance. This external validation stage provided a realistic benchmark of predictive capability, confirming that the selected models were both statistically robust and physically meaningful for PBF-LB process–property prediction.

Methodology for feature importance analysis

Both Phase II and Phase III included detailed error analysis, residual examination, and overfitting assessment to evaluate model reliability and generalization. To enhance interpretability, a feature importance analysis was conducted to identify the relative contribution of process parameters and thermal descriptors to the predictions. SHapley Additive exPlanations (SHAP), a model-agnostic interpretability framework, was applied to the best-performing regressor within each feature set [29]. Positive SHAP values indicate features that increase predicted RD, whereas negative values denote a decreasing influence. Mean absolute SHAP values, computed across all and low-density samples, quantified overall and conditional feature importance. SHAP analyses were performed using the fitted preprocessing pipelines to ensure consistency with cross-validation training. All models, training histories, and hyperparameters were serialized for full reproducibility and potential deployment.

Physics-informed machine learning (PIML) model development

The PIML framework

A physics-informed machine learning (PIML) framework was developed to integrate structured domain knowledge

Table 4 Algorithm groups and the associated models that were examined

Algorithm group	s/*n*	Associated models	Core concept
Single Tree-Based	1.	DecisionTreeRegressor	Uses a single tree structure to make predictions by splitting data on feature values.
	2.	ExtraTreeRegressor (ETR single tree)	
Ensemble (Tree-Based)	3.	ExtraTreesRegressor (ExTRs: extremely randomized trees)	Combines multiple weak tree models (either via bagging or boosting) to create a stronger, more robust model.
	4.	RandomForestRegressor	
	5.	GradientBoostingRegressor	
	6.	HistGradientBoostingRegressor	
	7.	XGBRegressor	
	8.	LGBMRegressor	
	9.	AdaBoostRegressor (often with trees)	
Linear Models	10.	LinearRegression	Models that assume a linear relationship between features and the target. They differ in the loss function and regularization used (L1, L2, both, or none).
	11, 12	Ridge, RidgeCV	
	13, 14	Lasso, LassoCV	
	15, 16	ElasticNet, ElasticNetCV	
	17–20	Lars, LarsCV, LassoLars, LassoLarsIC	
	21.	BayesianRidge	
	22.	SGDRegressor	
	23.	PassiveAggressiveRegressor	
	24.	HuberRegressor	
	25.	QuantileRegressor	
	26.	PoissonRegressor	
	27.	TweedieRegressor	
Support Vector Machines (SVM)	28–30	LinearSVR, SVR, NuSVR	Finds a hyperplane (linear or in a transformed space) to fit the data, often focusing on a margin of error.
Nearest Neighbours (Distance-Based)	31.	KNeighborsRegressor	Predicts based on the average of the k most similar data points in the training set.
Neural Networks	32.	MLPRegressor	A multi-layer perceptron that uses interconnected layers of nodes (neurons) to learn complex, non-linear relationships.
Gaussian Processes	33.	GaussianProcessRegressor	A non-parametric, probabilistic model that defines a distribution over functions and infers the best one from the data.
Kernel Methods	34.	KernelRidge	Combines Ridge regression with the "kernel trick" to model non-linear relationships.
Meta & Utility	35.	BaggingRegressor	Wrappers or utility models that modify the training process or target variable or provide a simple baseline.
	36.	TransformedTargetRegressor	
	37.	DummyRegressor	
Sparse Signal Recovery	38.	OrthogonalMatchingPursuit	Algorithms designed for efficiently recovering sparse signals (where most coefficients are zero).
	39.	OrthogonalMatchingPursuitCV	
Robust Regression	40.	Random Sample Consensus Regressor (RANSAC)	A meta-estimator that is robust to outliers by fitting a model from random subsets of inliers.

with data-driven modeling, with the aim of producing interpretable and physically consistent predictive expressions for RD in PBF-LB. Its key components include: (i) a knowledge base that embeds physical constraints into the learning process, (ii) a data layer that organizes and manages input information, and (iii) a symbolic regression engine that discovers analytical relationships between variables. Model development on the DS-01 and DS-02 domain sources proceeded in three sequential stages: (a) training models using only process parameters (PP), (b) incorporating Physics-Informed Process Descriptors (PIPDs) without outlier filtering, and (c) extending the workflow with outlier removal to improve robustness and cross-dataset transferability. The overall architecture of the framework is shown in Fig. 4.

Data layer and preparation

In this study, the data layer comprises the process parameters listed in Table 1 and their corresponding measured RD values for both DS-01 and DS-02. Data was split 80:20 into training and test sets for each domain source. Outlier filtering was performed independently for each dataset using an interquartile range (IQR) method, applied to both raw process parameters and derived PIPDs. Specifically, an IQR scaling factor of $iqr\_k = 0.75$ and $p = 0.05$ was used for DS-01, while $iqr\_k = 1.25$ and $p = 0.03$ were applied for DS-02. These domain source-specific criteria reduced the influence of extreme or non-physical observations, supporting more robust and generalizable model development.

Knowledge base (KB)

The Knowledge Base (KB) serves as a curated and extensible repository of PBF-LB-specific domain knowledge, which can be updated with newer trends, information as and when it is available via a feedback loop to refine the learnt knowledge. It stores information on process variables and their physical units, material properties, and dimensionless descriptors that capture key aspects of the underlying physics. These descriptors form the foundational priors and constraints guiding the learning process. In this study, the KB is defined by a feature space of eleven physics-informed process descriptors (PIPDs), listed in Table 5, whose base inputs originate from the data layer (Section "Data layer and preparation", Table 1). These include normalized energy densities (NLED, NAED, NVED), hatch-to-layer ratio (H/T), and dimensionless quantities related to power, velocity, beam size, and thermal behaviour [30–33].

The formulations of these descriptors adopt standard thermal properties for 316 L stainless steel at STP (standard room temperature and pressure), including thermal conductivity k = 26 $W \cdot m^{-1} \cdot K^{-1}$ and a thermal diffusivity $\alpha \approx 6.5 \times 10^{-6}$ $m^2 \cdot s^{-1}$, derived from $\alpha = k/(\rho c_p)$; taking the theoretical density as $\rho \approx 8000 kg \cdot m^{-3}$ and a specific heat capacity of $c_p \approx 500 J \cdot kg^{-1} \cdot K^{-1}$ [34]. An absorptivity of $\eta = 0.6$, consistent with literature values for 316 L powder (typically between 0.3 and 0.7) under a 1068 nm laser [35], and an assumed temperature difference of ΔT = 1200 °C were also used. Beyond defining features, the KB encodes physical rules such as dimensional consistency, monotonicity expectations, and bounded RD ($\rho_{rel} \in [0, 1]$), which constrain the

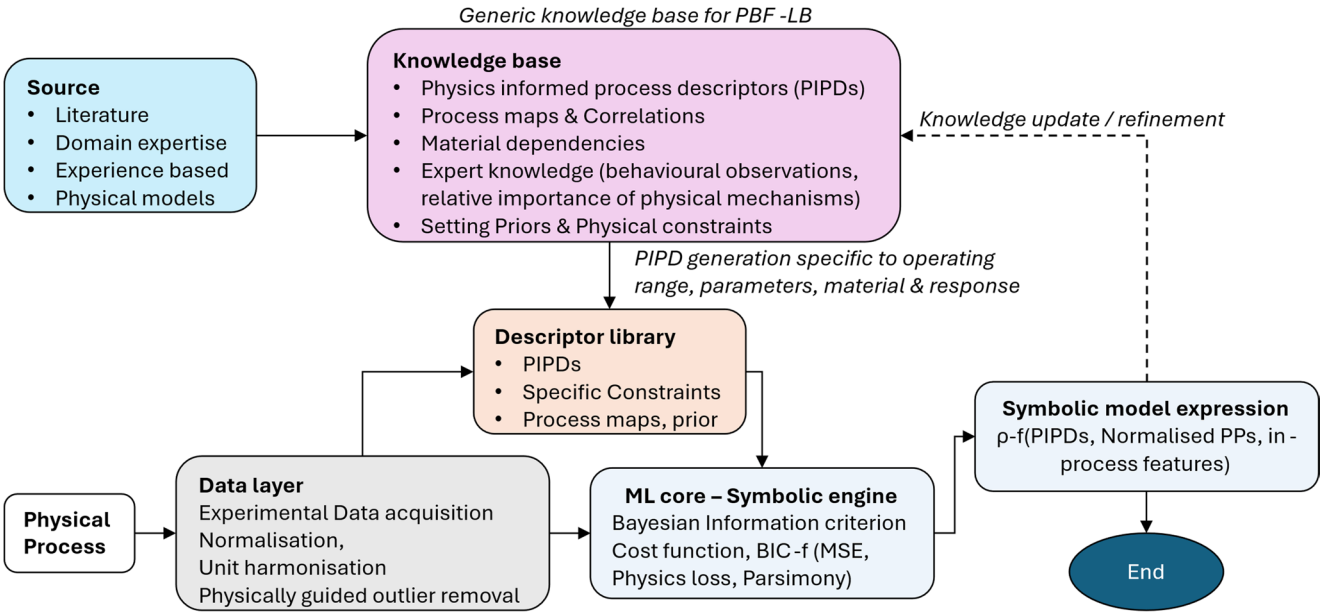


Fig. 4 Physics-informed machine-learning (PIML) framework architecture

Table 5 The feature space of Physics-Informed Process Descriptors (PIPDs)

Descriptor	Meaning	Formula	Feature sign
NLED	Normalized Linear Energy Density	$Norm\left(\frac{P}{V}\right)$	x_1
NAED	Normalized Areal Energy Density	$Norm\left(\frac{P}{(V\times(\frac{H}{1000}))}\right)$	x_2
NVED	Normalized Volumetric Energy Density	$Norm\left(\frac{P}{(V\times(\frac{H}{1000}))\times t}\right)$	x_3
NLP	Normalized Laser Power	$Norm\,(P)$	x_4
HT	Hatch-to-Layer Ratio	$\frac{H/1000}{t}$	x_5
Overlap	Track Overlap Fraction	$1-\frac{H}{f}$	x_6
NHSR	Normalized Hatch-to-Spot Size Ratio	$Norm\left(\frac{H}{f}\right)$	x_7
DLP	Dimensionless Laser Power	$\frac{\eta P}{k\Delta T f}$	x_8
DBV	Dimensionless Beam Velocity	$\frac{V\times f}{\alpha}$	x_9
HSR	Hatch-to-Spot Size Ratio	$\frac{H}{f}$	x_{10}
NLT	Normalized Layer Thickness	$\frac{t}{f}$	x_{11}

'Norm' refers to Min-Max Normalization between 0 and 1, based on process parameter range presented in Table 1

learning process toward mechanistically plausible, interpretable, and transferable solutions.

Symbolic regression engine

Symbolic regression was performed using a Qlattice ML library, which searches for analytical expressions linking PIPDs to RD [36, 37]. The search process is guided by the KB, encouraging physically admissible forms. Candidate models are represented as computation graphs constructed from primitive operations (addition, multiplication) and nonlinear functions (exponential, logarithmic, hyperbolic tangent) [37]. During training, thousands of candidate graph structures are sampled from a probabilistic distribution reflecting the Qlattice's evolving assessment of promising model architectures. Each candidate is fitted to the training data by optimizing free numerical parameters via gradient-based backpropagation, minimizing mean-squared error. Models are then ranked using the Bayesian Information Criterion (BIC), which penalizes unnecessary complexity [38]. BIC is computed as:

$$BIC = k_{BIC}\,ln\,(n) - 2\,ln\,(L_{BIC}) \tag{4}$$

where k_{BIC} is the number of free parameters, n the number of data points, and L_{BIC} is the maximised likelihood; lower values indicate a better fit–parsimony balance.

Models with poor predictive performance or redundant structure are progressively pruned based on both their age in training epochs and their relative performance, though the best-performing candidate is always retained [38]. Following pruning, the Qlattice updates its internal probability distribution to reinforce the structural motifs of the surviving models. This update biases subsequent sampling toward more promising regions of the search space while maintaining stochastic exploration. This sample–fit–prune–update procedure continues until convergence, or a predefined complexity threshold is reached, yielding an ensemble of high-performing symbolic models.

Feature selection is intrinsic to this process. Each sampled model contains only a subset of available descriptors, and a maximum complexity cap of 10 graph edges limits the number of features and interactions. As BIC penalizes complexity, descriptors compete for inclusion, leading to the exclusion of redundant or weakly informative PIPDs through the selection of a single representative among highly correlated features. To ensure robustness and avoid premature convergence, structurally diverse top-ranking candidates are retained for interpretability and evaluation [37]. This integrated framework outputs concise, analytical expressions for RD.

Model Validation and Computational Considerations

Although k-fold cross-validation can be applied, it is not strictly required within a Qlattice workflow. The model-selection process already incorporates iterative resampling, pruning, and BIC penalization, which suppress overfitting and eliminate unnecessarily complex structures during training. For the physics-informed models, robustness

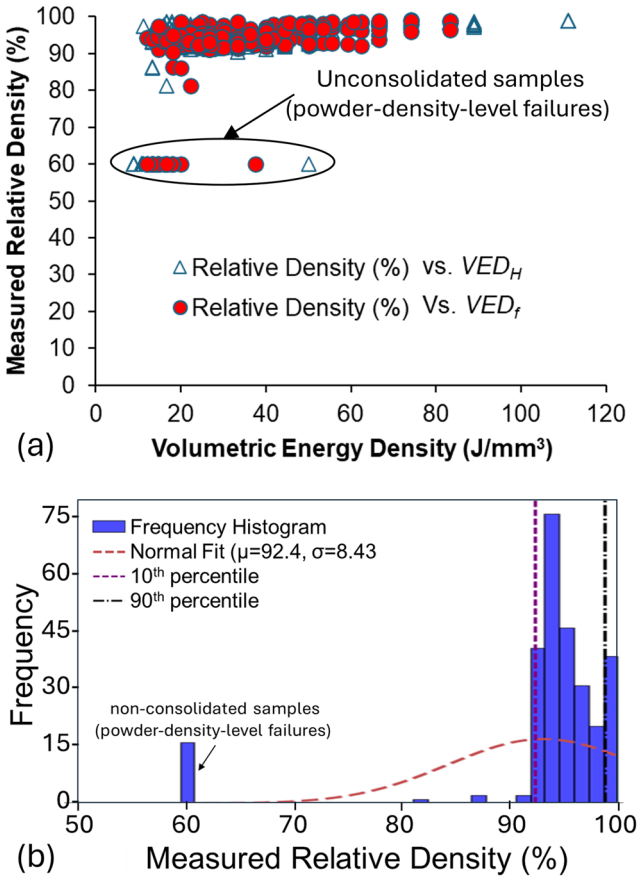


Fig. 5 Relative density, RD (%) for the sample set (n=256): (**a**) variation with respect to volumetric energy density (VED), and (**b**) frequency histogram illustrating the distribution of RD values. The dashed red curve represents a fitted normal distribution (μ=92.4%, σ=8.43%) shown for reference. VED_f represents VED derived using spot size, and VED_H VED derived using hatch spacing

Table 6 Descriptive statistical summary of the observed Relative Density (RD) data for the set of 256 samples

Statistics	value
count	256
mean	92.40
std	8.43
min	60.00
10th Percentile	91.65
median	93.77
90th Percentile	97.89
max	99.00
skewness	-3.31
kurtosis	10.16
mode	60.00

was instead assessed through independent validation across the DS-01 and DS-02 domain sources using the respectively trained Qlattice models, providing a more stringent and domain-relevant generalization test with emphasis on model transferability, using the evaluation metrics defined in Table 3.

Moreover, the Qlattice workflow incurs substantially higher computational cost than conventional supervised ML pipelines, as it must repeatedly sample, train, prune, and update large populations of candidate symbolic models (over 8,000 in this study, with training durations on the order of minutes), rather than fitting a single parametric estimator, which typically trains within milliseconds.

Results and discussion

Understanding the measured relative density (RD)

The DS-01 RD (%) of 256 processed samples was analysed to evaluate material consolidation and process consistency, see Fig. 5; Table 6. Figure 5a shows little to no information on the variation of RD with both definition variants of volumetric energy density. In contrast, from Table 6; Fig. 5b, the mean (μ) RD was 92.4%, with a standard deviation of 8.43%, and values ranged from 60% to 99%, indicating substantial variability. Most samples achieved high RD (median = 93.77%, 90th percentile = 97.89%), while a small subset of low-density samples (1st percentile = 60%, mode = 60%) highlights incomplete densification. Consequently, the distribution exhibited strong left skewness (–3.31) and pronounced kurtosis (10.16) as shown in Table 6, reflecting clustering of samples near high RD with a tail of low-density anomalies at the powder packing density level. The Shapiro–Wilk test [39] confirmed significant deviation from normality (W = 0.495, $p < 0.001$), consistent with the frequency histogram plot (Fig. 5b).

The pronounced skewness in the RD distribution (–3.31) is largely attributable to 15 samples that crumbled post-fabrication (non-consolidated samples). These included samples 17, 24, 37, 54, 64, 85, 116, 133, 152, 166, 173, 206, and 236, all manufactured under a constant laser power of 80 W with scan speeds ranging from 800 to 1200 mm/s. These samples were classified as powder-density-level failures (PDLF) that exhibit RD no lower than the powder bed's packing density (PBPD) during PBF-LB processing. The PBPD of 316 L SS generally ranges between 50% and 60% of theoretical alloy density, depending on powder characteristics and spreading techniques [40–42]. Consequently, these failed samples were assigned a nominal RD of 60% to more accurately reflect their physical state and to curb undue bias in subsequent analysis. Note: *Although these samples were not physically subjected to Archimedes' test, assigning a value of 0% implies a perfect vacuum, which is impossible given the inherent density of the powder particles by itself and the packing density of these powders during a typical PBF-LB process.*

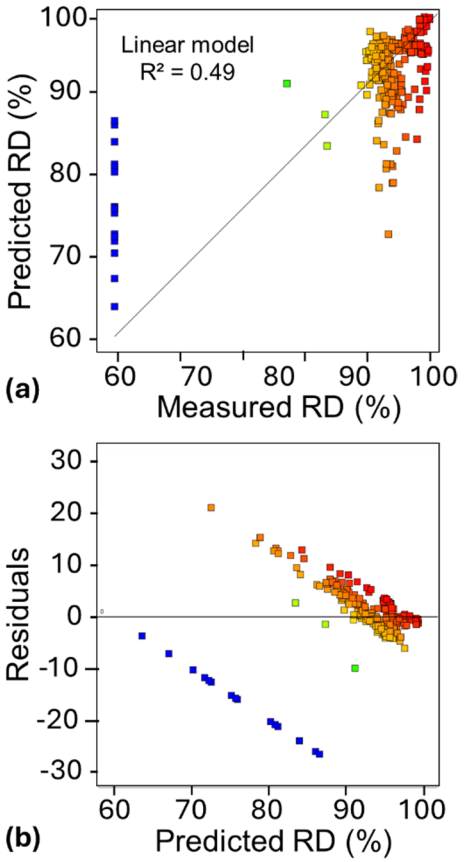


Fig. 6 Statistical linear model performance based on the DS-01 dataset showing (**a**) predicted vs. measured RD (%) for all PBF-LB processed 316 L SS samples ($n=256$); and (**b**) residuals plot

Interestingly, at least one such failed sample was observed across every level of hatch spacing and spot size in the four-factor, four-level full-factorial design (refer to Table 1). This observation suggests that the occurrence of PDLF is governed less by hatch spacing or spot size alone, and more by the combined effect of insufficient volumetric energy input associated with low laser power and elevated scan speed. Importantly, the inclusion of these failed samples represents an important unique contribution of this dataset, as prior ML studies in PBF-LB have largely excluded such extreme failures, focusing solely on higher density samples (typically > 80%). By retaining these physically meaningful powder-density observations, the dataset captures the full spectrum of process outcomes, enabling predictive models to learn from critical failure modes and improving the ability to forecast both extreme and nominal part densities.

By contrast, the highest-density specimen (sample 58) was produced with a laser power of 200 W, a scan speed of 600 mm/s, and, most notably, a matched hatch spacing and laser spot size of 90 µm (Appendix Table 12) This parameter set highlights a key finding: matching the hatch spacing to the spot size appears to create a favourable condition

Table 7 ANOVA results for linear and extra terms model (Aliased) based on PBF-LB process parameters

Source	F-value	*p*-value	significant
ANOVA Linear Model	4.62	< 0.0001	True
Laser Power (W)	0.15	0.7241	*False*
Scan Speed (mm/s)	0.14	0.7059	*False*
Hatch Spacing (µm)	23.35	< 0.0001	True
Spot Size (µm)	2.87	0.0918	*False*
VEDh (J/mm^3)	10.30	0.0015	True
VEDf (J/mm^3)	1.77	0.0057	True
Residual			
Lack of Fit	10.08	0.0174	True
Fit Statistics			
R^2	**0.49**		
Adjusted R^2	0.38		
Predicted R^2	0.15		

for achieving high RD. Among the 17 samples fabricated with matched hatch and spot size, RD ranged from 91% to the maximum of 99% across all levels of laser power and scanning speed used, indicating a plausibly good processing window. Within this matched condition, the hatch-based (VED_H) and spot-size-based (VED_f) volumetric energy densities of 74.1 J/mm³ served as a successful energy input value for densification, though the matching of geometrical parameters was the foundational factor.

To move beyond descriptive observations, it is necessary to statistically model and quantify how individual process parameters and derived energy density metrics contribute to variations in RD. For this purpose, correlation analyses were conducted using Pearson's correlation (for linear dependencies), Spearman's rank correlation (for monotonic but potentially nonlinear trends), and Kendall's tau coefficient (for rank-order concordance).

Statistical modelling of relative density (RD): evidence of nonlinear thermal effects and the need for machine learning

Weak predictive power of standard process parameters

Correlation analysis (Fig. 7; Table 8) and linear modelling (Fig. 6; Table 7) showed that standard PBF-LB process parameters were weak predictors of RD. Pearson correlations were low (–0.28 to 0.42), with VED_H, VED_f, and laser power explaining only ~ 18, 17, and 12% of variance in the RD, respectively (i.e., $R^2 \approx 0.18, 0.17$, and 0.12). ANOVA with interaction terms confirmed this limited explanatory power as summarised in Table 7. While the overall model was significant ($F=4.62$, $p<0.0001$), only hatch spacing, VED_H, and VED_f were strongly significant ($p<0.001$). Laser power (W) and scan speed (mm/s) were non-significant ($p>0.70$), spot size (µm) was marginal ($p \approx 0.09$), and

layer thickness was constant and contributed no variance. Model fit remained poor ($R^2 = 0.49$), and as shown in Fig. 6a and b, predictions were very poor for both PDLF cases (blue square markers) and high-density samples (red square markers). The large gap between Predicted R^2 (0.15) and Adjusted R^2 (0.39) exceeded the acceptable 0.2 threshold, indicating poor generalization. The highly significant Lack of Fit ($F = 10.08$, $p = 0.0174$) (Table 7) further demonstrates that linear ANOVA, even with interactions, does not capture the nonlinear, high-order relationships underlying RD in PBF-LB processed 316 L SS.

In line with existing literature [43, 44], VED remains a valuable and statistically significant aggregate indicator of energy input per unit volume. However, it is fundamentally a simplified descriptor that cannot fully capture the underlying thermophysical phenomena such as melt pool dynamics, laser–material interaction variability, and spatiotemporal heat accumulation, that critically influence densification. This has been to topic of discussion in PBF-LB processing of 316 L SS [45, 46], and other alloys like NiTi (nitinol) [23]. Thus, while VED offers useful first-order insights, its limited descriptive power, even when combined with laser power, is incapable of modelling nonlinear and coupled process–property relationships in PBF-LB.

Strong predictive power of melt pool thermal stability

In contrast to the process parameters, thermal distribution metrics of the melt pool measured by pyrometers were strongly correlated with RD (Fig. 7; Table 8). Pearson coefficients for the mean, median, and variance of melt pool temperature were strongly negative, ranging from −0.60 to -0.85. The variance in temperature (*Variance_Temp*) was the strongest single predictor ($r = -0.85$, $R^2 \approx 0.72$), followed by *Interquartile_Range_Temp* (a measure of thermal consistency/variability) with $r = -0.83$, $R^2 \approx 0.69$. Both features suggests that melt pool thermal stability (minimization of melt pool temperature variance), rather than absolute energy input, controls densification.

Evidence of non-linear relationships from divergent correlation metrics

The comparison of linear (Pearson's r) and rank-based (Spearman's ρ, Kendall's τ) correlation metrics revealed significant divergences, providing strong evidence of non-linear process dynamics (Table 8). This is particularly evident for the melt pool thermal descriptor features. For instance, metrics like *Avg_Temp* and *Median_Temp* showed strong negative linear correlations (Pearson's $r \approx$ -0.6 to -0.7) but weak-to-moderate positive monotonic correlations (Spearman's ρ, Kendall's $\tau \approx +0.3$). This sign reversal indicates that the underlying relationship is fundamentally non-linear and non-monotonic in a complex way, which a simple linear model fails to capture.

By contrast, a subset of PBF-LB parameters, VED_H, VED_f and Laser Power (W), demonstrated moderate positive correlations with RD that were consistent across all three correlation methods. This consistency suggests their relationship with density is more reliably monotonic, making them stable predictors despite the overall non-linear landscape. Furthermore, higher-order statistical descriptors of the melt pool thermal emission features (e.g., *Kurtosis_Temp* and *Percentile_10_Temp*) also showed stable, moderate correlations, suggesting that the shape and uniformity of the thermal distribution are also predictive, not just its central tendency.

The observed non-linearity aligns with the established densification behaviour in PBF- LB [47, 48]:

1. Insufficient energy leads to lack-of-fusion porosity.
2. Excessive energy causes keyholing and vaporization, reducing density.
3. An optimal intermediate range produces stable melt pools and optimal consolidation.

The sign reversal observed in the thermal feature correlations likely reflects this phenomenon: at very high energy inputs, increasing average temperature becomes associated with keyholing and defect generation, violating the linear assumption and causing Pearson's r to turn negative.

Synthesis and implications for predictive modelling of relative density (RD)

The afore-discussed statistical modelling and analysis (Section "Weak predictive power of standard process parameters" to "Evidence of non-linear relationships from divergent correlation metrics") confirms that RD is governed primarily by melt pool thermal stability, a phenomenon not fully captured by standard process parameters such as VED. The stark contrast between linear (Pearson's r and ANOVA linear model) and rank-based (Spearman's ρ, Kendall's τ) correlations underscores this conclusion: while linear models primarily highlight the destabilizing influence of extreme thermal variance, the rank-based approaches reveal the underlying monotonicity within a constrained, non-linear process window.

These findings have two key implications. First, they emphasize the critical value of *in-situ* AM monitoring capable of capturing these complex thermal signatures [49, 50], such as the coaxial infra-red pyrometry used in this study. Second, they expose the limitations of classical statistical models, which can identify dominant factors but struggle to

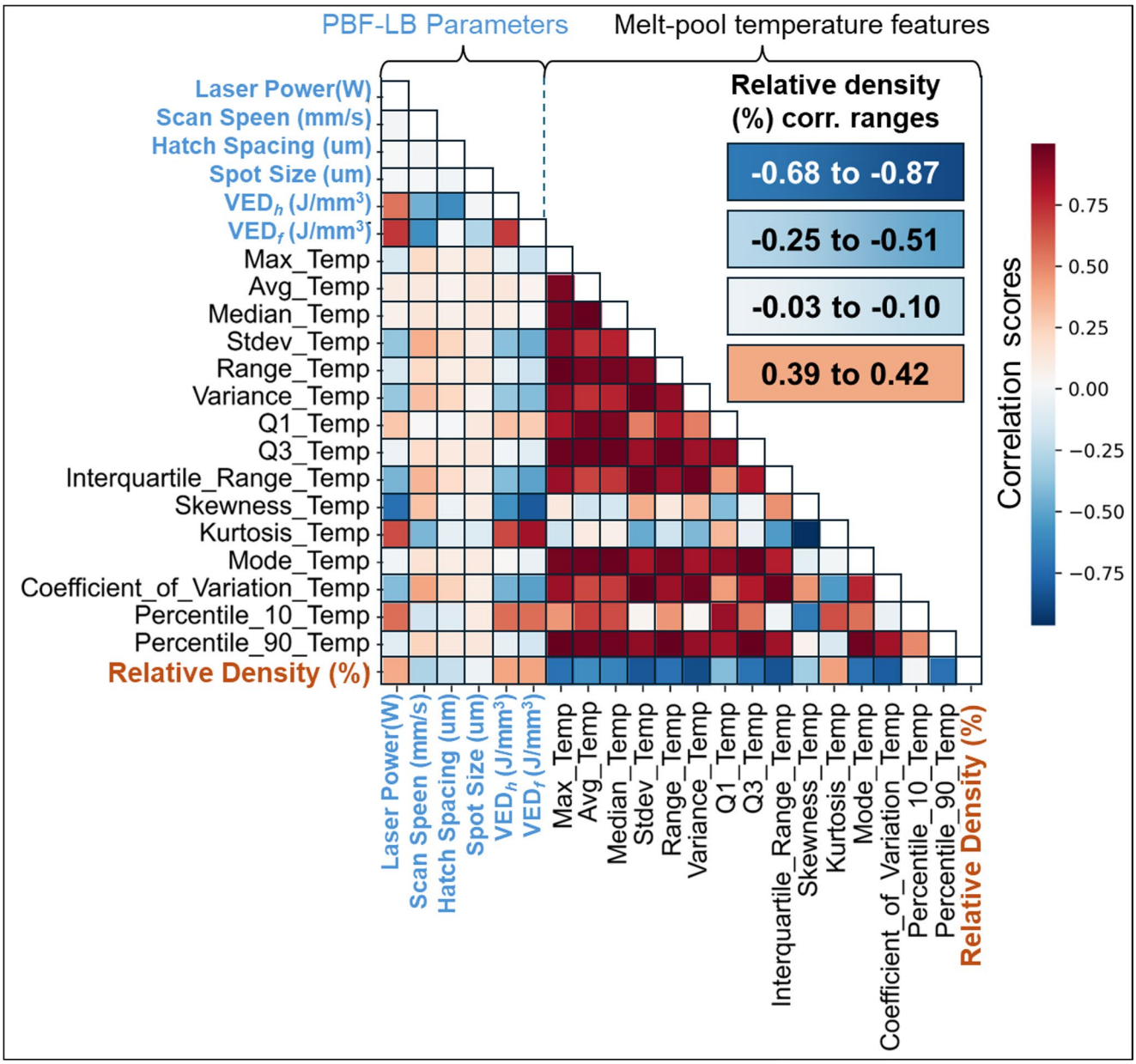


Fig. 7 Pearson correlation coefficients (linear correlations) for the combined feature set of process parameters (*PBF-LB Feature set*) and melt-pool temperature features (*IR Feature set* extracted from *in-situ* infrared pyrometer measurements during PBF-LB processing. The target property is the global RD of each of the 256 samples. Melt-pool temperature characteristics represent layer-wise averages aggregated across the build

represent the subtle, multi-parameter dependencies inherent to the PBF-LB process. This limitation directly motivates the application of data-driven approaches. Supervised ML methods are uniquely suited to interpretably capture the non-linear trends and high-dimensional interactions identified here, promising more accurate prediction of RD and more actionable insights for process optimization and defect mitigation in PBF-LB processing of 316 L SS.

Sensor data-driven machine learning

Consideration of class imbalance and threshold sensitivity

The DS-01 dataset exhibited a natural class imbalance, with only 15 samples (≈6%) showing RD < 60% (Section "Weak predictive power of standard process parameters", Fig. 5b). This sparsity of low-density samples focuses the sensor

Table 8 Correlation scores between process parameters, melt-pool temperature features, and relative density for 256 PBF-LB processed 316 L stainless steel samples

Pearson's correlation	*r* score	R^2	Spearman's correlation	ρ rank	Kendall's tau correlation	τ rank
Relative Density (%)	1	1	Relative Density (%)	1	Relative Density (%)	1
Kurtosis_Temp	0.42	0.18	Kurtosis_Temp	0.46	Percentile_10_Temp	0.34
VEDf	*0.42*	*0.17*	***VED$_f$***	***0.46***	***VED$_f$***	***0.33***
VEDh	*0.40*	*0.16*	Percentile_10_Temp	0.46	Kurtosis_Temp	0.33
Laser Power (W)	*0.39*	*0.15*	***VED$_h$***	***0.43***	***VED$_h$***	***0.3***
Percentile_10_Temp	0.00	0.00	***Laser Power (W)***	***0.36***	***Laser Power (W)***	***0.28***
Spot Size (um)	*-0.03*	*0.00*	Q1_Temp	0.33	Avg_Temp	0.27
Hatch Spacing (um)	*-0.21*	*0.04*	Avg_Temp	0.33	Mode_Temp	0.26
Scan Speed (mm/s)	*-0.28*	*0.08*	Mode_Temp	0.32	Median_Temp	0.26
Skewness_Temp	-0.32	0.10	Median_Temp	0.31	Q1_Temp	0.26
Q1_Temp	-0.41	0.17	Q3_Temp	0.31	Q3_Temp	0.26
Avg_Temp	-0.60	0.36	Percentile_90_Temp	0.29	Percentile_90_Temp	0.24
Median_Temp	-0.64	0.41	Max_Temp	0.27	Max_Temp	0.22
Q3_Temp	-0.70	0.49	Range_Temp	0.27	Range_Temp	0.22
Max_Temp	-0.71	0.51	*Spot Size (µm)*	*0.03*	*Spot Size (µm)*	*0.02*
Range_Temp	-0.71	0.51	Variance_Temp	-0.11	Variance_Temp	-0.07
Mode_Temp	-0.71	0.51	Stdev_Temp	-0.12	Stdev_Temp	-0.07
Percentile_90_Temp	-0.72	0.51	*Hatch Spacing (µm)*	*-0.14*	*Hatch Spacing (µm)*	*-0.1*
Coefficient_of_Variation_Temp	-0.80	0.63	Interquartile_Range_Temp	-0.19	Interquartile_Range_Temp	-0.13
Stdev_Temp	-0.82	0.67	Coeff_of_Var_Temp	-0.2	Coeff_of_Var_Temp	-0.13
Interquartile_Range_Temp	-0.83	0.69	*Scan Speed (mm/s)*	*-0.37*	Skewness_Temp	-0.27
Variance_Temp	-0.85	0.72	Skewness_Temp	-0.38	*Scan Speed (mm/s)*	*-0.28*

a. *Italicised features* (also in blue font colour) represent PBF-LB process parameters.

b. Temp is an abbreviation for Temperature

c. Coeffiecient_of_Variance_Temp is shortened to Coeff_of_Var_Temp

Table 9 Features from RFECV used for subsequent model development and validation

Feature sets	Selected parsimonious feature
PBF-LB feature set	VED_H
IR feature set	Interquartile_Range_Temp, Mode_Temp
Hybrid feature set	Hatch Spacing (um), Spot Size (um), VEDh, VEDf, Stdev_Temp, Variance_Temp, Interquartile_Range_Temp, Mode_Temp
Hybrid enforced	Laser Power (W), Scan Speed (mm/s), VEDh, Interquartile_Range_Temp, Mode_Temp

data-driven ML framework on distinguishing between stable process conditions (RD > 90%) and catastrophic lack-of-fusion near the powder bed packing density (PBPD) level.

The presence of a more continuous distribution of samples, particularly within the 70–85% RD range, would enable learning subtler, non-linear relationships between thermal history and gradual density loss. However, the fundamental predictive relationship of the predictor features, would remain valid. While generating such a dataset would require a parameter investigation broader than the 4 × 4 full factorial DOE used here, its absence clarifies the scope of the present models: they are robustly trained to identify severe process failure.

Given the low cardinality and high-dimensional nature of the sensor-derived feature space (melt pool thermal signature), synthetic oversampling techniques (e.g., SMOTE) were deemed unsuitable, as they risk generating non-physical meltpool feature artifacts and model bias. Therefore, the analysis relied on the inherent robustness of the selected regression algorithms. To further ensure model robustness and efficiency, recursive feature elimination with cross-validation (RFECV) was employed to identify a parsimonious subset of the most predictive features.

Feature subsets from RFECV

The RFECV-derived feature subsets are summarized in Table 9. The analysis revealed that within the PBF-LB parameter set, the volumetric energy density (VED_H) was consistently selected by all regressors and represented the most parsimonious predictive feature, in agreement with the ANOVA results as shown in Table 7 (p-value = 0.0015). For the *IR feature set*, RFECV identified a compact subset comprising just two (2) complementary melt pool temperature descriptors ($Interquartile\_Range\_Temp, Mode\_Temp$). The *Hybrid feature set* included nine (9) features, combining elements from both domains: VED_H, VED_f, *Hatch Spacing (µm)*, and *Spot Size (µm)* from the PBF-LB feature

group, together with $Interquartile\_Range\_Temp$ and *Mode_Temp*.

Notably, many of these hybrid features are either derived quantities or hard-to-change machine parameters. For real-time process control and digital-twin applications, however, physically meaningful parameters such as laser power and scan speed are preferred, as they can be dynamically adjusted to achieve the target RD and minimize defect formation during fabrication. In contrast, most commercial PBF-LB systems do not allow on-the-fly modification of parameters such as laser spot size or hatch spacing once a build has begun. Accordingly, a *hybrid-enforced feature set* was developed by manually introducing laser power and scan speed to the stable melt-pool descriptors (*Interquartile_Range_Temp*, *Mode_*Temp) identified in both the IR and hybrid RFECV subsets (Table 9).

Feature selection and engineering are essential steps to remove redundant and irrelevant predictors that can hinder learning efficiency, increase model complexity, and reduce interpretability [51]. Beyond methodological benefits, these steps have clear implications for computational and energy sustainability. By reducing the original 23 descriptors to six (or ≤ 6) stable predictors, training time decreased by approximately 60–70%, directly lowering the computational cost and energy demand of the model benchmarking process. Such reductions are particularly impactful when scaling to larger datasets or embedding models in real-time control systems, where leaner architectures enhance responsiveness and reduce the carbon footprint of machine-learning workflows in additive manufacturing.

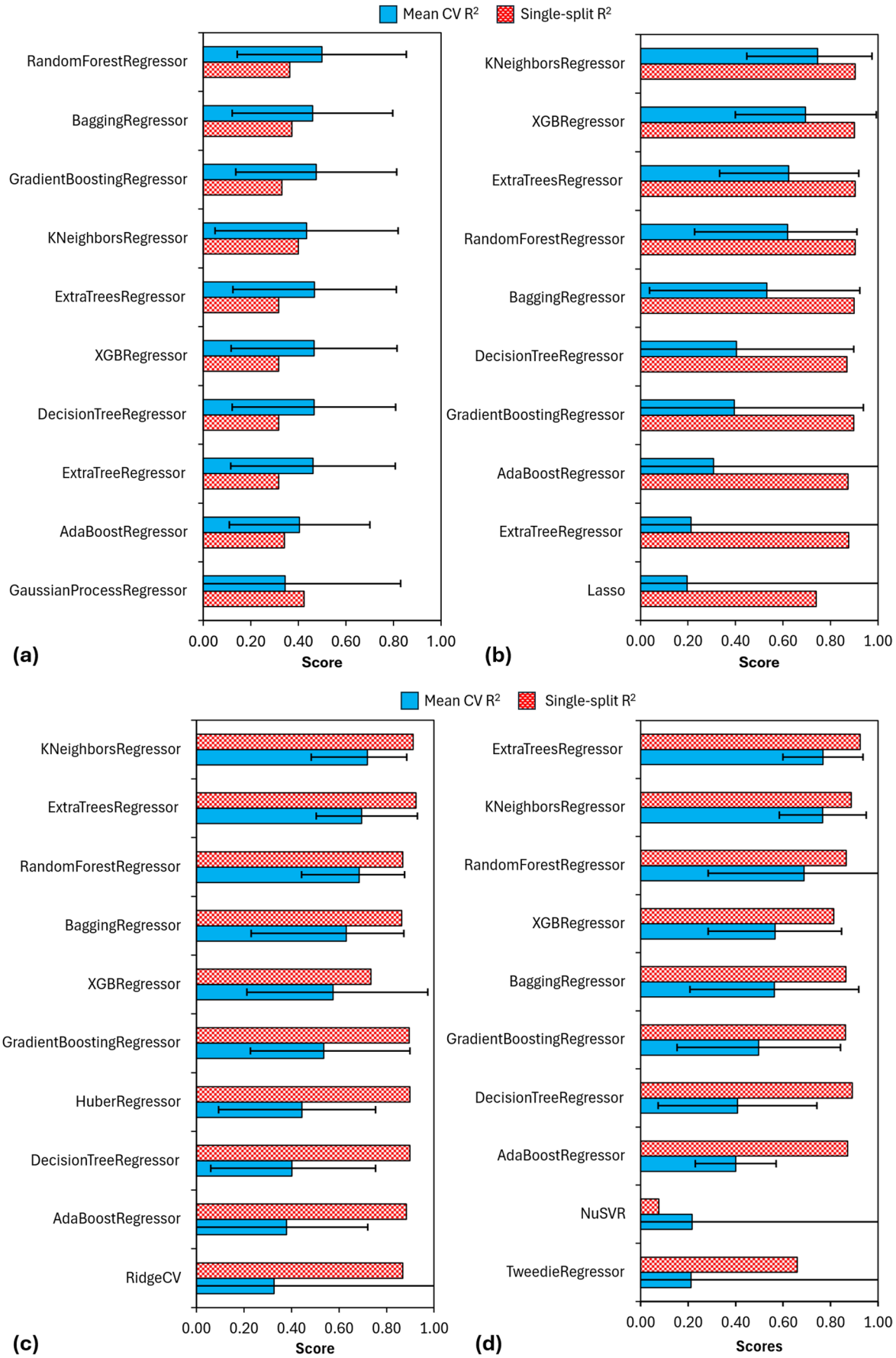
Comprehensive algorithm screening using the training dataset

The comprehensive screening of 40 regression algorithms across five (5) shuffled cross-validation (CV) folds was conducted using the RFECV-derived feature subsets (Table 9). For clarity, only the training R^2 for the ten best-performing models shown in Fig. 8a – d. The results were consistently dominated by ensemble-based and distance-based algorithms, indicating their suitability for modelling the RD of PBF-LB processed 316 L stainless steel. A key finding is that models incorporating *in-situ* melt pool temperature descriptors consistently achieved predictive performance approximately double that of models using only process parameters, with single-split (from the 20% test set of the DS-01 training subset) R^2 values approaching 0.9 for the highest-performing algorithms. This demonstrates that infrared pyrometry captures critical physical variance inaccessible through preset parameters alone. For subsequent validation, the top three models from each feature set, selected based on their cross-validation performance and stability as illustrated in Fig. 9, were selected for further analysis.

Fig. 8 Ranking of the top 10 (out of 44) regression models based on cross-validation performance over five shuffled folds of the DS-01 training dataset, showing the dominance of ensemble-based and distance-based algorithms. Performance is shown for (**a**) the *PBF-LB feature set*, (**b**) the *IR feature set*, (**c**) the *Hybrid feature set*, and (**d**) the *Hybrid-Enforced feature set*, for predicting the RD of PBF-LB–processed 316 L stainless steel samples

Among these three best-performing models, distinct performance patterns were observed depending on the feature domain, highlighting the varying predictive relevance of the underlying inputs. For the *PBF-LB feature set*, which was reduced to a single parsimonious predictor VED_H, the RandomForestRegressor (RFR) achieved the highest mean CV $R^2$of 0.50 ± 0.36 and a corresponding single-split R^2 of 0.36. Despite this, the relatively high variance across folds and modest explanatory power suggest a weak and unstable predictive relationship between VED_H and the RD of 316 L-SS. The GradientBoostingRegressor (GBR) and ExtraTreesRegressor (ETR) produced nearly equivalent outcomes, with mean CV R^2values of 0.48 ± 0.34 and 0.47 ± 0.34, respectively. This convergence suggests that different ensemble tree architectures arrive at a similarly limited predictive ceiling when restricted to this single process-derived descriptor. Notably, expanding the feature set to include the full suite of PBF-LB process parameters (Table 1) did not improve predictive performance. Instead, model training generalization slightly deteriorated. The RFR achieved a single-split test R^2 of 0.51 and mean CV $R^2$of only 0.49 ± 0.31, and the ETR performed poorly (CV $R^2 = 0.34 \pm 0.48$), despite similar model complexity. This degradation suggests that the additional parameters introduced noise or multicollinearity without contributing meaningful explanatory variance. The performance of these ML models is comparable to the statistical ANOVA model fit ($R^2 = 0.49$; Table 7), collectively indicating that PBF-LB process parameters alone, including the RFECV-identified VED_H and the broader parameter sets (Table 1), possess only marginal predictive utility. These parameter-based models capture a constrained portion of the physical variability underlying RD, a limitation that becomes starkly apparent when compared to models incorporating *in-situ* sensor data, as discussed in the subsequent section.

In contrast, the models trained on the *IR feature set* (comprising only *Interquartile_Range_Temp* and *Mode_Temp*) exhibited a substantial improvement in predictive performance compared with those based solely on PBF-LB parameters (Fig. 9). The mean CV $R^2$of the top models increased from approximately 0.48 (*PBF-LB feature set* average) to 0.69–0.75, representing an average enhancement of around 50–60% in explained variance. Concurrently, the mean CV RMSE decreased from 4.37 to 3.05 ($\approx$30% reduction


Mean CV R²
Single-split R²
RandomForestRegressor
BaggingRegressor
GradientBoostingRegressor
KNeighborsRegressor
ExtraTreesRegressor
XGBRegressor
DecisionTreeRegressor
ExtraTreeRegressor
AdaBoostRegressor
GaussianProcessRegressor
0.00 0.20 0.40 0.60 0.80 1.00
Score
(a)
KNeighborsRegressor
XGBRegressor
ExtraTreesRegressor
RandomForestRegressor
BaggingRegressor
DecisionTreeRegressor
GradientBoostingRegressor
AdaBoostRegressor
ExtraTreeRegressor
Lasso
0.00 0.20 0.40 0.60 0.80 1.00
Score
(b)
Mean CV R²
Single-split R²
KNeighborsRegressor
ExtraTreesRegressor
RandomForestRegressor
BaggingRegressor
XGBRegressor
GradientBoostingRegressor
HuberRegressor
DecisionTreeRegressor
AdaBoostRegressor
RidgeCV
0.00 0.20 0.40 0.60 0.80 1.00
Score
(c)
ExtraTreesRegressor
KNeighborsRegressor
RandomForestRegressor
XGBRegressor
BaggingRegressor
GradientBoostingRegressor
DecisionTreeRegressor
AdaBoostRegressor
NuSVR
TweedieRegressor
0.00 0.20 0.40 0.60 0.80 1.00
Scores
(d)

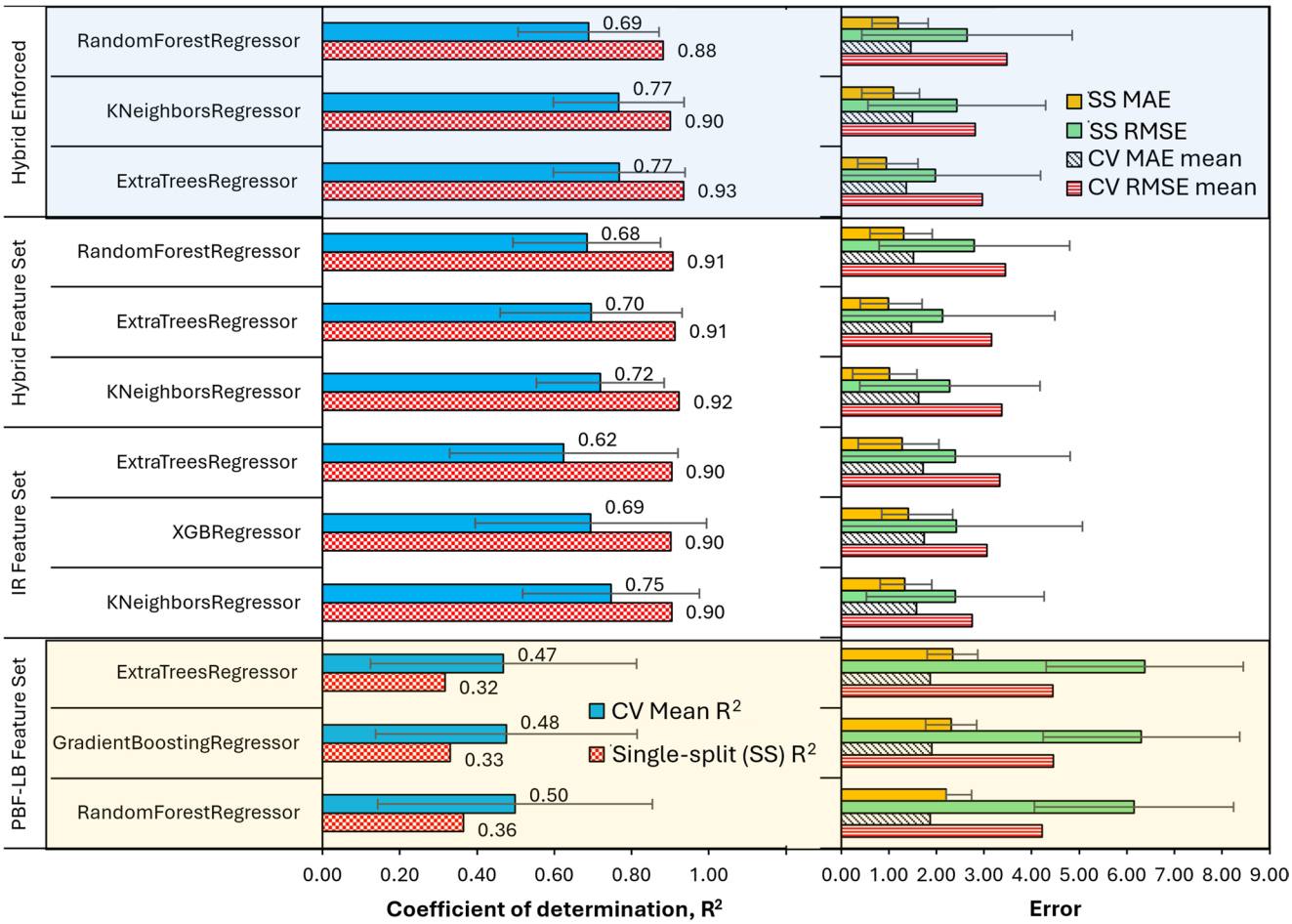


Fig. 9 Predictive performance of the top three models during development, comparing single-split (SS) metrics (from the 20% test subset of DS-01 training data) against 5-fold cross-validation results. Both R^2 and error metrics (RMSE, MAE) are shown to illustrate accuracy and stability across feature sets

in prediction error), accompanied by a notable decrease in mean absolute error (from 1.88 to 1.68, ≈11% improvement). The KNN model achieved the highest mean CV performance (R^2 = 0.75 ± 0.23) with comparatively low validation errors (CV RMSE = 2.75 ± 1.87; CV MAE = 1.58 ± 0.56), followed closely by XGBRegressor, XGB (R^2 = 0.69 ± 0.30) and ETR (R^2 = 0.62 ± 0.30). Overall, the CV std across folds is still slightly high, partly attributable to the limited number of PDLF (failed) samples (n = 10), which can cause imbalances in the cross-validation splits. Nevertheless, all three models displayed strong consistency between training single-split and CV scores, with no indication of overfitting. These results clearly demonstrate that the *in-situ* melt pool temperature descriptors obtained via infrared pyrometry encode substantially more predictive information than the nominal process parameters. Both distance-based (KNN) and ensemble learning approaches (ETR and XGB) were able to leverage these descriptors to achieve consistently higher accuracy, effectively capturing a greater proportion of the variance underlying the RD of the 316 L SS samples by directly reflecting the local heat flow and solidification dynamics governing part densification.

When combining both process parameters and melt pool temperature descriptors, the RFECV-derived *hybrid feature set* further improved the balance between model complexity and generalization. The KNN, ETR, and RFR achieved mean CV R^2scores of 0.72±0.17, 0.70±0.24, and 0.68±0.19, respectively, demonstrating enhanced consistency and reduced predictive error compared with models trained exclusively on process parameters. Notably, the *hybrid enforced feature set* (constructed through the deliberate inclusion of *Laser Power (W)*, *Scan Speed (mm/s)*, and the intentional combination of the *PBF-LB feature set (VEDh)* with the *IR-feature set*) yielded the strongest overall predictive performance. Here, ETR achieved a single split $R^2$of 0.93 and a CV $R^2$of 0.77±0.17, with correspondingly low training and validation errors (Train RMSE = 1.98; CV RMSE = 2.96±2.21), indicating a high

degree of generalization without evidence of overfitting. The KNN model in this configuration achieved a similarly high CV $R^2$of 0.77±0.17, while the RFR followed closely with 0.69±0.18.

Collectively, these findings demonstrate that multi-source feature integration, especially when guided by domain knowledge, substantially enhances model robustness, predictive accuracy, and the physical interpretability of feature importance. This approach, which captures both preset process conditions and real-time melt pool thermal dynamics, provides a superior foundation for developing reliable digital twins and feedback control systems for PBF-LB.

SHAP feature importance analysis for relative density prediction

SHAP analysis reveals a clear hierarchy of predictive features, confirming the dominant influence of melt pool thermal descriptors across most modeling configurations. Analysis of mean absolute SHAP values shows that *Mode_Temp* and *Interquartile_Range_Temp* are consistently the most critical predictors across different algorithms and feature set compositions, as illustrated in Fig. 10a–d. The only exception occurs in the *PBF-LB Feature Set* (Fig. 10a), where VED_H is the sole parsimonious predictive feature, leading to poor model performance, with the best mean CV $R^2 = 0.50 \pm 0.36$and a single-split $R^2 = 0.36$ achieved by RFR model. In the IR feature set, these two thermal descriptors are solely responsible for the significant predictive improvement. As shown in the stacked bar chart in Fig. 10b, *Interquartile_Range_Temp* attains a mean absolute SHAP value of 2.19 in the XGB model, followed closely by *Mode_Temp*. An identical ranking is observed for the KNN model, and a reversed ranking for the ETR model, underscoring that melt pool thermal stability and central tendency are the primary drivers of RD prediction.

For the *Hybrid Feature Set*, the relative importance of melt pool descriptors is slightly reduced but remains clearly dominant (Fig. 10c). In the KNN model, *Variance_Temp* is ranked highest (SHAP = 1.02), followed by *Mode_Temp* and a shared contribution from *Interquartile_Range_Temp* and *Stdev_Temp* (0.67 each), indicating that statistical descriptors of the temperature distribution play a prominent role. Across all three models (RFR, KNN, and ETR), *Mode_Temp* consistently emerges as an important predictor, with mean SHAP values of 1.08, 0.70, and 0.94, respectively. *Interquartile_Range_Temp* likewise exhibits consistently high contributions, with mean absolute SHAP values ranging from 0.63 to 0.79, together with other percentile-based melt pool thermal descriptors. In contrast, the introduced process parameters display substantially lower SHAP values, supporting the conclusion that intrinsic melt pool thermal characteristics captured *in-situ* provide a more direct predictive signal than preset manufacturing parameters.

This trend persists in the *Hybrid enforced feature set*, where models are constrained to include key laser power and

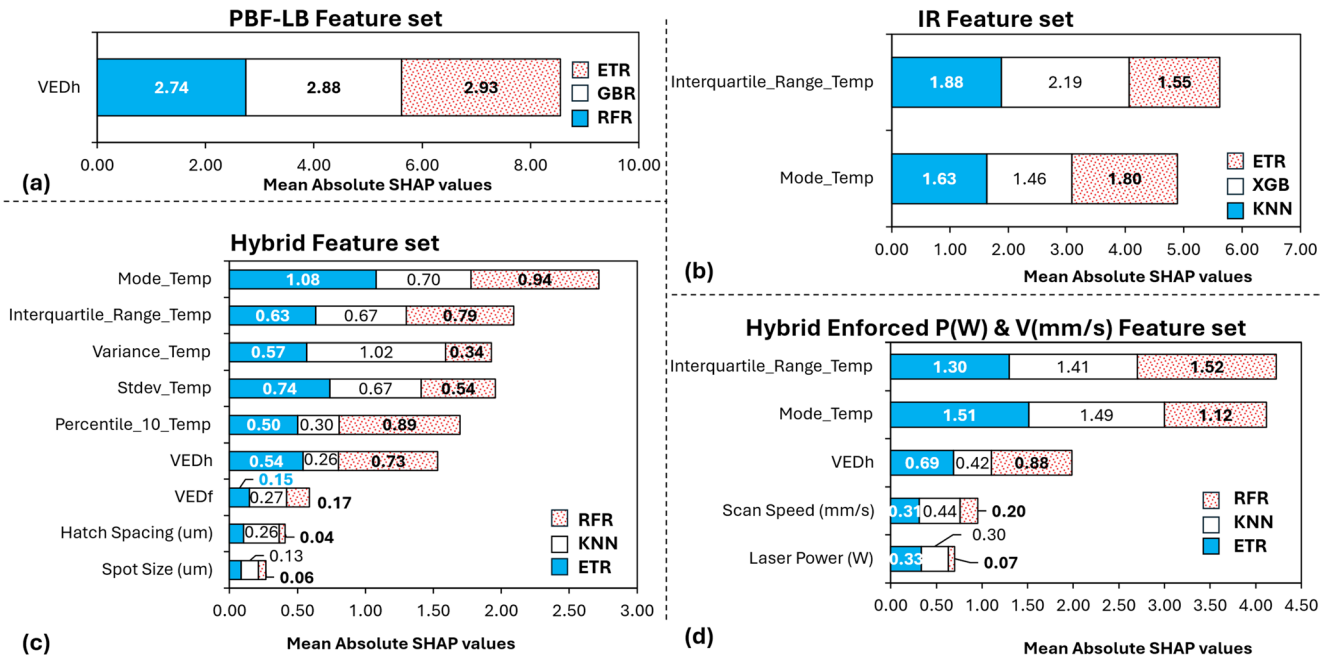


Fig. 10 SHAP feature importance summary for the top three models in predicting the relative density of PBF-LB processed 316 L stainless steel using (**a**) the PBF-LB feature set, (**b**) the IR feature set, (**c**) the Hybrid feature set, and (**d**) the Hybrid enforced feature set. All analyses were performed using the training dataset

scan speed variables (Fig. 10d). Even under this enforced configuration, *Mode_Temp* and *Interquartile_Range_Temp* remain the most influential features across all algorithms (KNN, ETR, and RFR), with mean absolute SHAP ranges of 1.12–1.51 and 1.30–1.52, respectively. Meanwhile, Laser Power and Scan Speed exhibit lower SHAP value ranges of 0.07–0.33 and 0.20–0.44, respectively, and the derived VED_H shows intermediate importance with SHAP values between 0.42 and 0.88. Although their contribution is secondary to the dynamic melt pool characteristics, these process parameters still play a physically meaningful role in modulating the melt pool's thermal response.

The consistent dominance of *Mode_Temp* and *Interquartile_Range_Temp* across all model configurations demonstrates that these descriptors capture the essential thermal dynamics governing process stability and material consolidation in PBF-LB of 316 L SS. This underscores that thermal behaviour, rather than nominal process parameters, is the principal driver of build quality, with parameters serving primarily as boundary conditions that shape the melt pool response. The strong predictive performance of models based on thermal features, particularly within the IR and hybrid feature sets, confirms that straightforward statistical representations of melt pool temperature encapsulate the underlying process physics while maintaining both accuracy and interpretability.

Model validation and generalization assessment using validation dataset

Independent validation using the DS-01 held-out validation dataset provided a rigorous evaluation of each model's ability to generalize beyond the training data. All top three models from each feature set were subjected to this validation process; however, only the best-performing model per feature set, determined by the validation R^2, is summarized in Table 10. The corresponding residual and predicted versus true RD plots are presented in Fig. 11, illustrating each model's predictive accuracy and error distribution. Notably, the target variable distributions show that the validation set has slightly lower mean RD (90.78% versus 92.80%) and higher variability (10.46 versus 7.82 standard deviation), indicating the validation set presents a more challenging prediction task.

The overall best model from this final validation stage, as detailed in Table 10, was the XGBRegressor (XGB) model utilizing the *IR feature set*. This model demonstrated exceptional predictive accuracy on the DS-01 held-out validation data, achieving a $R^2$of 0.93, alongside an RMSE of 2.83 and an MAE of 1.67. Its performance on the validation data was notably superior to its own cross-validation estimate, with the validation R^2 exceeding the mean CV $R^2$of 0.69. Two RFR models, one employing the *Hybrid feature set* and the other a *Hybrid Enforced set*, also exhibited strong and nearly identical validation performance, both attaining a validation R^2 of 0.92. The RFR model with the *Hybrid feature set* yielded a validation RMSE of 2.98 and MAE of 1.62, while the model with *Hybrid Enforced features* produced a validation RMSE of 3.01 and MAE of 1.59.

A comparative analysis of training and validation metrics reveals insights into model generalizability. All three top-performing models showed a high degree of consistency between their performance on the training and validation datasets, indicating robust generalization. This is evidenced by the minimal differences between single-split R^2 value during model development (Phase II) and validation R^2 values (Phase III), with the IR-XGB model showing a difference of only − 0.023, the Hybrid-RFR a difference of -0.049, and the Hybrid Enforced-RFR a difference of -0.033. While the RFR models, particularly the one with *Hybrid Enforced features*, achieved slightly lower MAE values on the validation set, suggesting a marginal advantage in median-centric prediction accuracy, the IR-XGB model's superior R^2 and competitive RMSE on the independent validation set solidify its position as the most accurate model overall. The results suggest that the *IR feature set*, when paired with the XGB algorithm, provided the most effective combination for predictive modelling on this dataset, though the hybrid feature sets also facilitated the development of highly robust and accurate models using the RFR algorithm.

Furthermore, an analysis of the residual standard deviation (residual STD in Fig. 11) provides additional insight into the precision and error distribution of the models. The standard deviation of the residuals on the independent validation set was consistently higher than that observed on the initial single-split training data for all top models, which is indicative of the expected performance degradation when applying a model to unseen data. The XGB model with the IR feature set, which

Table 10 Summary of the best-performing models for each feature set based on DS-01 validation R^2 (Val R^2)

Phase I (Feature selection)	Model	Phase II (DS-01 Training dataset)				Phase III (DS-01 Validation data set)			R^2 Diff
Feature Set		Single-split R^2	Single-split RMSE	Single-split MAE	Train CV R^2 mean	Val R^2	Val RMSE	Val MAE	
IR	XGB	0.90	2.42	1.41	0.69	0.93	2.83	1.67	-0.023
Hybrid	RFR	0.87	2.80	1.32	0.68	0.92	2.98	1.62	-0.049
Hybrid Enforced	RFR	0.88	2.65	1.20	0.69	0.92	3.01	1.59	-0.033

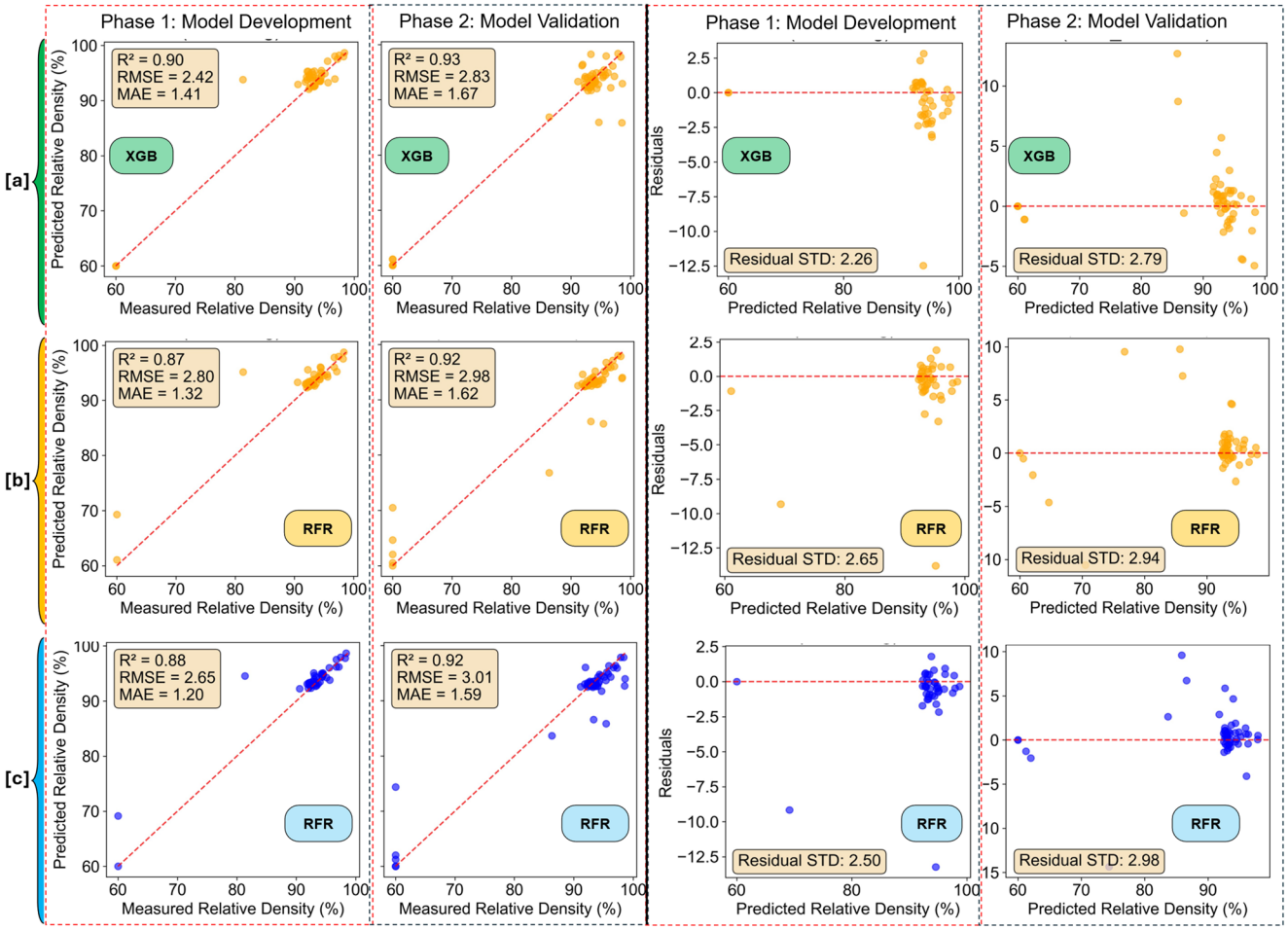


Fig. 11 Predicted versus measured RD and residual plots illustrating the comparative performance of the overall best model for each feature set on the training and validation datasets. (**a**) IR feature set, (**b**) Hybrid feature set, and (**c**) hybrid enforced feature set. The overall best model in each feature set is identified based on its validation R^2 obtained from the held-out validation dataset. XGB and RFR refer to XGBoost and Random Forrest Regressors, respectively

was identified as the overall best performer, demonstrated the most favourable residual characteristics. Its residual standard deviation increased from 2.26 on the training data to 2.79 on the validation set, representing the smallest absolute value and the most constrained error spread among the evaluated models on the held-out data. The predicted versus measured RD plots for the IR-XGB model (Fig. 11) further demonstrate its robustness, showing accurate predictions for nearly all low-density samples, with only a single validation data point deviating slightly above the regression line (actual RD = 60%, predicted RD 62). In contrast, the two RFR models displayed greater dispersion in their prediction errors, particularly for the lower-density samples, which tended to fall slightly above the regression line. Consequently, the Hybrid-RFR model exhibited an increase in residual standard deviation from 2.65 to 2.94, while the Hybrid Enforced-RFR model rose from 2.50 to 2.98. Notably, all mis-predicted low-density samples had RD values of ≤75%, which in practice would still indicate potentially defective parts within a real-time feedback control or digital twin monitoring system. These patterns are consistent with the corresponding RMSE and R^2results, collectively confirming that the IR-XGB model not only achieved the highest predictive accuracy but also maintained the most stable and least variable error distribution across the validation dataset, demonstrating superior robustness and generalization capability.

Overall, these validation results confirm the robustness and physical consistency of the IR-based and hybrid feature sets. Models trained on these descriptors not only generalized accurately to unseen data but also maintained the same feature importance hierarchy observed during cross-validation. This consistency demonstrates that melt pool thermal-response features (specifically, *Mode_Temp* and *Interquartile_Range_Temp*) provide a stable, high-fidelity representation of the underlying process physics, enabling reliable prediction of both dense and defective (PDLF) build regimes within the context of data (DS-01) from the same PBF-LB machine.

Physical interpretation: aligning feature importance with solidification fundamentals

The SHAP analysis (Section "Comprehensive algorithm screening using the training dataset"), model validation and generalisation (Section "SHAP feature importance analysis for relative density prediction", Table 10, and Fig. 11) shows that *Mode_Temp* and *Interquartile_Range_Temp*, are strongly aligned with fundamental metallurgical principles governing solidification and densification in PBF-LB. As noted earlier, these features reflect the central role of thermal history and melts pool stability in controlling porosity, rather than being a purely statistical outcome, for the following reasons:

Firstly, *Mode_Temp* functions as a robust proxy for the Peak Temperature (T_p) of the melt pool. In physical terms, while T_p represents the instantaneous maximum thermal intensity at the laser's focal point, *Mode_Temp* represents the most frequently realized thermal state in the monitored field. In a stable, Gaussian-like thermal distribution typically observed in PBF-LB, the mode of the temperature distribution directly correlates with the peak intensity. Consequently, *Mode_Temp* dictates the steady-state melt pool size, penetration depth, and layer-to-layer remelting behavior essential for suppressing lack-of-fusion (LoF) porosity and ensuring strong intertrack bonding. Experimental and numerical studies on 316 L SS confirm that process conditions which stabilize the melt pool and maintain a sufficient Peak Temperature, produce components with high RD (≥ 99%). Conversely, conditions that depress the peak thermal state reduce melt pool depth, promoting LoF defects and irregular track boundaries [52–54]. These observations are consistent with broader PBF-LB studies with other metal alloy systems demonstrating that heat buildup and local peak temperatures (reflected here by *Mode_Temp*) strongly correlate with keyhole porosity, warpage, and microstructure coarsening [55–58]. The strong SHAP weight assigned to *Mode_Temp* therefore provides quantitative evidence that the central thermal peak is a first-order control parameter for the densification of PBF-LB 316 L SS.

Secondly, *Interquartile_Range_Temp* extends this interpretation by capturing the spatial and temporal variability of the thermal field. In metallurgical terms, this variance serves as a data-driven surrogate for the local temperature gradient G and solidification rate R, which are the fundamental coordinates of solidification microstructure maps. In the present work, a narrow *Interquartile_Range_Temp* is interpreted as a signature of a uniform, steady-state melt pool characterized by consistent G and R conditions. Conversely, a broader interquartile range indicates increased fluctuations in the melt-pool thermal profile. Previous *in-situ* monitoring and numerical modeling studies have linked such instability to erratic keyhole behavior, plume dynamics, and spatter, all of which facilitate the nucleation of gas pores and irregular solidification fronts. For example, solidification microstructure selection maps constructed in G–V (or G–R) space for rapidly solidified alloy systems demonstrate that maintaining a high thermal gradient and high growth rate favour ultrafine cellular or micro-dendritic structures. Departures from this regime, characterized by high thermal variance, lead to coarser morphologies and increased defect content [56–61]. Such phenomena are been observed in Ti-6Al-4 V [62, 63] and 316 L [64, 65], where thermal instability and excessive heat accumulation promote heterogeneous microstructures and keyhole porosity. By identifying *Interquartile_Range_Temp* as a strong co-predictor for 316 L RD, the SHAP analysis and model performance provide a quantitative analogue to these solidification-map concepts: thermal stability and constrained variance in local G and R are necessary conditions for consistent melt pool consolidation and pore suppression.

The comparatively lower SHAP importance (Fig. 10c-d) attributed to nominal process parameters (e.g. laser power, scan speed, hatch spacing) within the *Hybrid enforced* models aligns with physics-based interpretations. These parameters function as boundary conditions that shape the energy input; however, they influence porosity and microstructure only through the resulting thermal fields. These fields are further modulated by complex variables including build geometry, scan strategy, material properties, and powder-bed characteristics. Numerous studies report that microstructure, residual stress, and density correlate more robustly with simulated or measured thermal histories (specifically peak temperature (T_p), cooling rate, and melt pool depth), than with any single nominal parameter [55]. The SHAP hierarchy in this work captures this precise causal chain by assigning greater importance to *in-situ* thermal descriptors (*Mode_Temp* and *Interquartile_Range_Temp*) than to preset machine parameters. This effectively identifies these thermal metrics as more proximal predictors of pore formation and densification.

Taken together, these features provide a quantitative and interpretable confirmation of key PBF-LB metallurgical principles. The central tendency of the melt pool temperature field (represented by *Mode_Temp*) and its stability (captured by *Interquartile_Range_Temp*, which serves as a proxy for controlled G–R conditions) emerge as the principal drivers of dense, defect-lean builds in PBF-LB 316 L SS. These findings are fully consistent with empirical solidification maps and thermal-history–based PBF-LB design frameworks reported in the literature. Therefore, the trained models (and its performance discussed in Section "SHAP feature importance analysis for relative density prediction") does not simply identify statistical correlations; they decode

the *in-situ* thermal signal into features that directly map onto the physical variables (specificallly, peak melt pool temperatures and thermal gradient) known to dictate part quality. This alignment establishes the trained models as a robust, physically grounded tool for optimization within the DS-01 framework.

Transferability limitations: domain divergence analysis between DS-01 and DS-02 domain sources

The DS-01 driven ML models, as identified in Section "SHAP feature importance analysis for relative density prediction", Table 10, were applied directly to the DS-02 dataset to assess their transferability. This transfer resulted in a complete degradation of model performance, yielding negative R^2 values. This indicates that the statistical relationships learned from DS-01 were not transferable to the data distribution of DS-02, demonstrating a critical limitation of cross-machine model deployment without domain adaptation. The performance loss is attributable to two fundamental categories of domain divergence: (1) a physical covariate shift in the process state (e.g., inter-layer time altering thermal statistics) and (2) a technical measurement shift in sensor data acquisition (e.g., calibration, signal conditioning).

A primary factor is the longer inter-layer times inherent to the larger build volume of the AconityMIDI+ system (DS-02) compared to the AconityMINI (DS-01). Inter-layer time (ILT) (also called inter-layer dwell time, ILDT), defined as the interval between successive laser exposures at a given point, is governed by the total duration of scanning, repositioning, and recoating across the entire build field. Increasing the effective build diameter from Ø 140 mm to Ø 250 mm systematically lengthens this thermal dwell, as the laser must traverse longer paths per layer (or process more parts), thereby increasing ILT [66]. Numerical and experimental work shows that shorter ILT leads to pronounced layer wise heat accumulation and higher surface temperatures (sometimes by hundreds of degrees) near the end of the build, while longer ILT provides greater opportunity for heat dissipation between layers, which reduces layer-to-layer heat accumulation and lowers the residual part temperature [67, 68]. This cooling effect alters melt-pool thermal statistics by compressing temperature variability and attenuated peak intensities, as the melt pool inherits its starting temperature from the reduced residual heat of the prior layer [69] (statistical evidence provided later). Studies indicate that ILT below approximately 30–40 s promotes significant *in-situ* heat accumulation, while longer ILT can mitigate global temperature buildup, albeit at the cost of increased build times and potential intra-layer thermal gradients [69–71].

Another contributing factor is the lack of standardized protocols for pyrometer integration and commissioning across different PBF-LB machine platforms. Although both sensors are nominally coaxial with the laser beam and coupled through reflective mirrors and beam-splitter optics, small installation-dependent variations can modify the effective focal position, incidence angle, and overall optical throughput of the pyrometric system. Even minor geometric deviations may shift the measurement spot relative to the true melt-pool centre or alter the fraction of emitted radiation reaching the detector, introducing systematic differences in signal magnitude and temporal response. While both pyrometers are black-body calibrated in degrees Celsius by KLIEBER®, the post-installation signal conditioning and output configuration differ between the two systems. Specifically, the DS-01 pyrometer produces signal values in the range of approximately 800–2000, whereas the DS-02 signal spans 0–400 mV. In addition, differences in emissivity settings cannot be ruled out. Collectively, these factors reflect the absence of harmonized standards for pyrometer installation, alignment, and signal normalization across PBF-LB platforms, which contributes to inter-machine discrepancies in recorded thermal data despite nominally identical sensor hardwar.

Despite identical post-processing, the mismatch in the output signals means its magnitudes do not correspond directly to absolute/physical melt pool temperatures. Consequently, the derived descriptors encode different underlying physics, rendering the feature-target relationships learned from DS-01 inapplicable to DS-02. It is critical to note that the model development pipeline for the tree-based estimators (Hybrid-RFR, Hybrid Enforced-RFR, and IR-XGB) inherently includes z-score normalization. While this step is technically redundant for the scale-invariance of tree-based models, it was a fixed component of the pre-processing sequence. Its failure to enable transferability conclusively demonstrates that standardization, while adjusting for mean and variance, cannot reconcile fundamental mismatches in sample thermal history, sensor alignment, response curves, or higher-order distributional characteristics. The combined effect of these physical and technical disparities is a profound domain divergence, which will be quantitatively analysed hereafter. This shift manifests not merely as a difference in scale but as a fundamental alteration in the joint probability distribution, $P(x, y)$ of the features between the two domain sources, DS-01 and DS-02.

To quantify distributional divergence between datasets, we employed the Kolmogorov–Smirnov (*KS*) statistic and the first-order Wasserstein distance (*WD*), complemented by probability density function (PDF) and cumulative distribution function (CDF) visualisations. The two-sample *KS* statistic measures the maximum absolute difference between the empirical cumulative distribution functions (ECDFs) of two samples, $F_1(x)$and $F_2(x)$, and is defined as:

$$D_{\mathrm{KS}} = \sup_{x} \mid F_1(x) - F_2(x) \mid \tag{5}$$

where $\sup_{x}$ is the supremum function. The KS test is non-parametric and sensitive to differences in both location and shape of the underlying distributions, with associated p-values used to assess statistical significance [72–74].

WD (also referred to as the Earth Mover's Distance (EMD) [75]) quantifies the minimum "cost" required to transform one distribution into another and is particularly sensitive to differences in distributional scale and spread. For one-dimensional distributions, the first-order WD is defined as:

$$W_1(F_1, F_2) = \int_0^1 \mid F_1^{-1}(u) - F_2^{-1}(u) \mid \mathrm{du} \tag{6}$$

where $F^{-1}(u)$denotes the quantile function. Unlike KS, WD provides a metric distance with the same physical units as the underlying variable, enabling direct interpretation of magnitude differences [76]. The *KS* and WD metrics revealed substantial statistical discrepancies across nearly all melt pool thermal descriptors (Appendix Tables 13 and 14). For most features, *KS* values approached unity (KS ≈ 1.0, $p < 10^{-152}$) with concomitantly large WD values (> 900) Fig. 12a-1. Notably, *Interquartile_Range_Temp* and *Mode_Temp*, which are core components of the IR, Hybrid, and Hybrid Enforced feature sets, exhibited significant, though comparatively lower, distributional shifts (*KS* = 0.72 and 1.00; *WD* = 29.18 and 998.12, respectively; Fig. 12b-1, c-11). The CDF plots (Fig. 12b-3, c-3) corroborate these findings. The mean value of *Interquartile_Range_Temp* halved from 62.8 in DS-01 to 33.6 in DS-02, indicating a substantial reduction in measured thermal variability. The mean *Mode_Temp* dropped dramatically from 1081.6 to 83.5, underscoring the profound scale inconsistency arising from the differing calibration units.

Following z-score normalization, *KS* and *WD* metrics were recalculated. While *KS* values were markedly reduced from their pre-normalized values, several variability-related descriptors, including *Variance_Temp* (*KS* = 0.316, $p < 10^{-11}$, WD = 0.391), *Stdev_Temp* (*KS* = 0.206, $p < 10^{-5}$, *WD* = 0.380), *Mode_Temp* (KS = 0.190, $p < 10^{-4}$, WD = 0.410), and *Interquartile_Range_Temp* (*KS* = 0.195, $p < 10^{-4}$, *WD* = 0.311), remained statistically distinct (Fig. 12a-2). The corresponding PDFs and CDF plots for these features clearly illustrate this persistent divergence, indicating inherent differences in melt pool fluctuation dynamics and steady-state temperature regimes between the two datasets, driven largely by differences in inter-layer cooling and global heat accumulation (Fig. 12b-2, b-4, c-2, c-4).

Notably, the shapes of these PDFs provide direct visual evidence of the influence of the ILT on melt pool temperature. In Fig. 12b-2, b-4, the normalised DS-01 *Interquartile_Range_Temp* and *Mode_Temp* features exhibit taller and narrower PDFs, consistent with the shorter inter-layer times and higher heat accumulation that preserve sharper thermal gradients. Conversely, the normalised DS-02 features shown in Fig. 12c-2, c-4 display shorter and broader PDFs, reflecting the longer ILT that promote greater heat dissipation, smoothing thermal gradients and attenuating extreme temperature values. Therefore, Fig. 12 visually corroborates the direct relationship between the ILT and the observed statistical distribution of melt pool descriptors.

In contrast, the target variable, RD, exhibited only minor divergence (*KS* = 0.157, *p* = 0.003, *WD* = 1.75) and nearly identical mean values (92.40% for DS-01 vs. 93.75% for DS-02). The strong agreement between the RD distributions, evidenced by the small Wasserstein distance and the close alignment of the CDF curves (Fig. 12d-1, d-2), confirms consistent densification behaviour across both PBF-LB systems. Given that the datasets were generated independently on different machines and evaluated by separate teams using the Archimedes method per ASTM B962-23 [21], this level of agreement represents a realistic measure of inter-laboratory consistency. This confirms that the source of transferability failure lies not in the target domain but in the machine-dependent input domain, specifically the IR pyrometer signal response and its translation into thermal descriptors.

To further investigate this, we performed an experiment where 20% of the DS-02 dataset (51 samples) was combined with the DS-01 training data to form an augmented set. Retraining the complete model development pipeline on this data yielded only a marginal improvement in performance on the remaining, unseen 80% of DS-02 ($R^2 \sim 0.093$). This negligible gain indicates that a small subset of DS-02 data is insufficient to bridge the cross-machine domain gap, and a larger inclusion would undermine the practical goal of a transferable model.

To address this challenge, physics-informed machine learning offers a promising path forward. By embedding physical relationships such as energy balance, heat conduction, and melt pool solidification dynamics into the learning process, PIML models can achieve better interpretability and cross-system adaptability. For instance, normalizing thermal features using physically derived quantities like effective laser energy input could help decouple models from machine-specific thermal histories. Incorporating dimensionless groups or energy-based scaling factors would enable learned relationships to reflect underlying process physics rather than empirical correlations alone. Ultimately, developing such transferable, physics-informed ML frameworks is essential for realizing machine-independent digital twins and advancing robust, certifiable control strategies for PBF-LB processes.

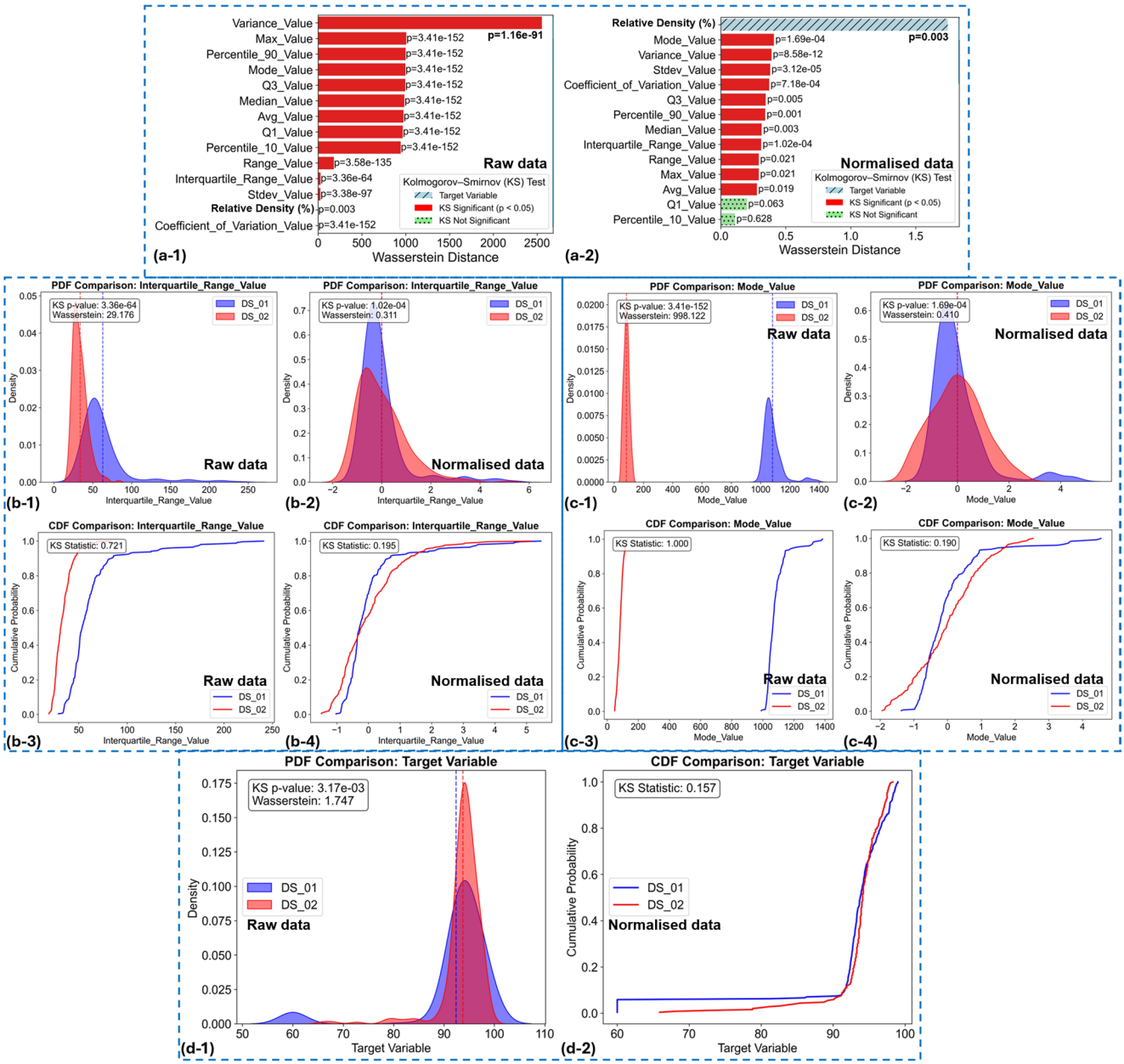


Fig. 12 Domain divergence analysis of *in-situ* melt-pool thermal descriptor datasets collected using coaxial IR pyrometers on the Aconity*MINI* (DS-01) and Aconity*MIDI*+ (DS-02) systems. Ranked Wasserstein distances with annotated Kolmogorov–Smirnov (KS) statistics are shown for (**a-1**) raw and (**a-2**) normalized (z-score standardized between 0 and 1) data. Probability Density Function (PDF) and Cumulative Distribution Function (CDF) plots are presented for *Interquartile_Range_Value* (**b-1** to **b-4**) and *Mode_Value* (**c-1** to **c-4**) for both raw and normalized data. (**d-1**) and (**d-2**) show PDF and CDF plots of RD using raw data (%). *Mode_Value* represents the most frequently observed pyrometer signals of melt pool thermal emissions in degrees Celsius (*Mode_Temp* for DS-01) or millivolt (*Mode_mV* for DS-02), capturing the central tendency of the thermal distribution. *Interquartile_Range_Value* represents the difference between the 75th and 25th percentiles, capturing the variability or spread of the recorded pyrometer temperature (*Interquartile_Range_Temp* for DS-01) or millivolt signal (*Interquartile_Range_mV* for DS-02)

Insights from physics-informed machine learning (PIML)

Baseline performance using PBF-LB feature set

The baseline QLattice model, which used only raw process parameters and assigned a powder-bed packing density (PBPB) of 60% to failed components in accordance with the PDLF criteria described in Section "Understanding the measured relative density (RD)", revealed a clear performance gap between the two domain sources (Table 11 and Fig. 13). For DS-01, the test R^2 reached 0.55, with hatch spacing, laser power, and laser scan speed identified as the most influential parameters. This baseline performance is slightly better than that of the best-performing Random Forest Regressor (RFR), which achieved a single-split test R^2 of 0.51 and mean CV $R^2$of 0.49±0.31 (Section "Feature subsets from RFECV"), and it also surpassed the statistical ANOVA model (R^2= 0.49, Section "Weak Predictive Power of Standard Process Parameters", Table 8). However, DS-01 exhibited consistently larger prediction errors across all metrics, including a test RMSE of 5.27 compared to 2.13 in DS-02, indicating a broader absolute error scale.

In contrast, the baseline model performance for DS-02 was lower (test R^2 = 0.44), but with low performance errors. The model achieved training and testing Pearson's r of 0.71 and 0.67, corresponding to R^2 values of 0.50 and 0.44, respectively (Table 11). Hatch spacing, laser power, scan speed, and spot size were identified as the most influential parameters. Taken together, these results demonstrate that raw process parameters exhibit only limited predictive power for RD in PBF-LB processed 316 L SS and highlights the need for physics-informed descriptors to capture higher-order melt-pool behaviour (Fig. 13).

Impact of physics-informed feature engineering

Transforming raw parameters into the *PIPD Feature set* constrained by KB as listed in Section "Knowledge base (KB)", Table 5, improved generalizability for both DS-01 and DS-02, as evidenced by increased test R^2 values (DS-01: 0.55 to 0.69; DS-02: 0.44 to 0.50). This represents approximately 13% gain in predictive performance (Table 11). This confirmed that incorporating domain knowledge through dimensionless descriptors enhances model robustness beyond what is achievable with purely data-driven models/correlations. The identified influential features shifted from basic parameters to composite descriptors like Dimensionless Laser Power (DLP) and Normalised Volumetric Energy Density (NVED) for DS-01. For DS-02, the Normalized Volumetric Energy Density (NVED), Normalized Areal Energy Density (NAED), and Normalized Hatch Spacing

Table 11 Comparison of performance across DS-01 and DS-02 domain sources

Domain source		PBF-LB parameter set				PIPD Feature set (PIML without outlier removal)				PIPD Feature set (PIML with outlier removal)			
		r	R^2	RMSE	MAE	r	R^2	RMSE	MAE	r	R^2	RMSE	MAE
DS-02	**train**	0.71	0.50	2.75	1.57	0.80	0.64	2.24	1.53	0.95	0.90	0.83	0.64
	test	0.66	0.44	2.13	1.68	0.71	0.50	2.01	1.36	0.95	0.90	0.77	0.61
DS-01	**train**	0.82	0.67	5.24	3.34	0.84	0.71	5.01	3.09	0.91	0.83	0.92	0.64
	test	0.74	0.55	5.27	3.33	0.83	0.69	4.41	3.31	0.93	0.86	0.76	0.64

r refers to Pearson's correlation coefficient

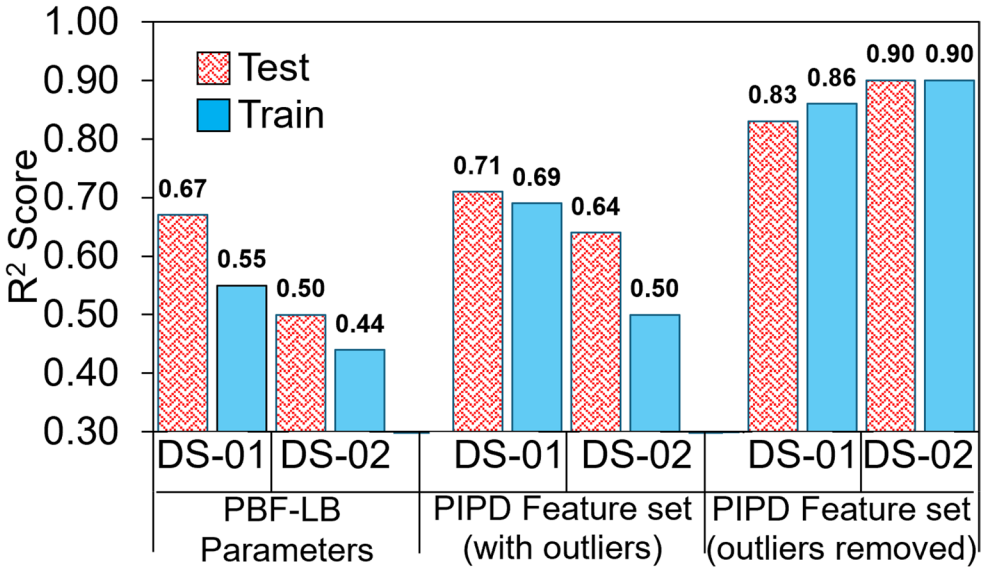


Fig. 13 Performance comparison across the DS-01 and DS-02 domain sources

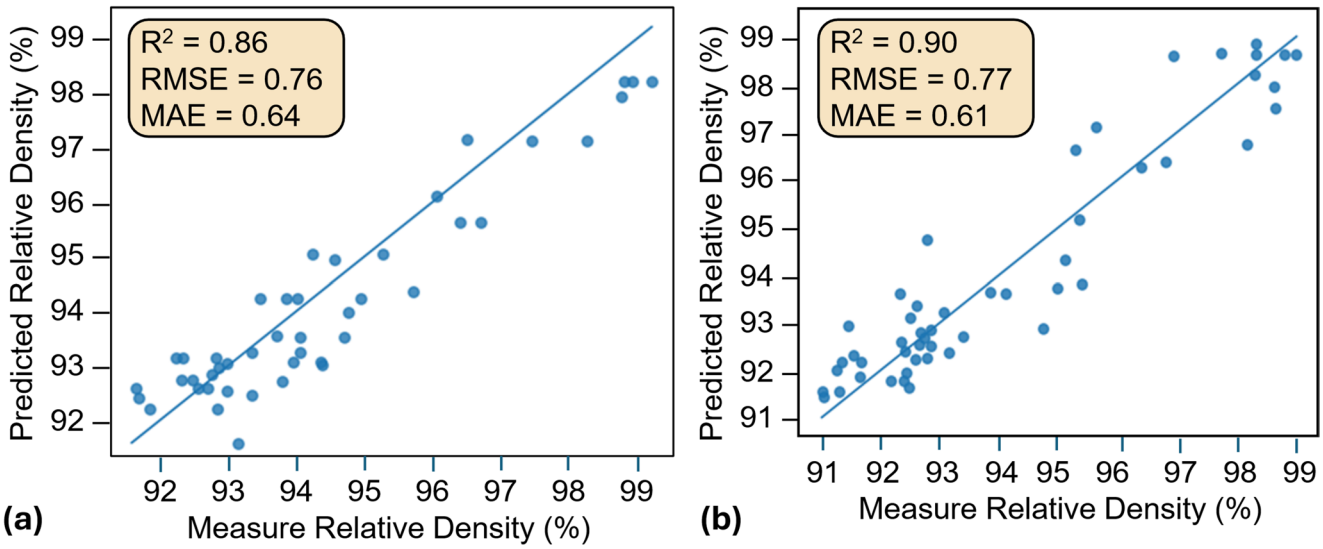


Fig. 14 Predicted vs. actuals plot of RD based on (**a**) DS-01, (**b**) DS-02 test domain source

emerged as the dominant contributors. These shifts illustrate how physically meaningful features allow the model to better represent energy-balance conditions, melt-pool geometry, and broader process–structure relationships.

The hybrid outlier-removal procedure (combining physics-based constraints with statistical interquartile filtering) before model training produced the most substantial improvement across both domain sources. Approximately ~ 10% of DS-01 and ~ 3% from DS-02 were removed, including failed builds assigned artificially low powder bed packing density values of 60% and additional non-physical observations. For DS-01, hybrid filtering yielded the highest overall performance, with a test R^2 of 0.86. Although the relative improvement was smaller than in DS-02, the final RMSE (0.76) and MAE (0.64) became closely aligned with the corresponding metrics for DS-02. This shows that the PIML framework effectively harmonised the predictive accuracy of two datasets with inherently different characteristics. The most influential descriptors for DS-01 were the Normalised Areal Energy Density, the Normalised Hatch Spacing, and the Normalised Volumetric Energy Density.

For DS-02, the benefits of hybrid outlier removal were even more substantial. The test R^2 increased to 0.90, and the RMSE dropped from 2.01 to 0.77 (Table 11), indicating that a significant portion of the initial error originated from points that conflicted with underlying physics or lay at extreme statistical distances. The dominant descriptors for DS-02 were the Volumetric Energy Density, Areal Energy Density,

Overlap, and Deposited Layer Power. The corresponding predicted-versus-actual plot is shown in Fig. 14a-b.

Two insights emerge from this analysis. First, the higher baseline errors in DS-01 are likely linked to its batching strategy discussed in Section "Experimental design, sample fabrication, and density measurement", which is physically realistic of typical PBF-LB operations. Each DS-01 batch required independent operator setup, introducing realistic but more complex nonlinear interactions that increased data variability. The PIML pipeline, and particularly the hybrid outlier-removal step, proved effective in identifying and filtering such points, reducing both systematic and random errors. Second, the convergence of both datasets to high R^2 values (>0.86) and to similarly low error magnitudes (RMSE$\approx$0.76–0.77, MAE$\approx$0.61–0.64) after completing the full PIML workflow demonstrates the framework's ability to enhance model transferability. By constructing a physics-constrained feature space, the approach enables models trained on fundamentally different domain sources to achieve consistent, reliable, and comparable predictive accuracy.

When dimensional constraints were applied to restrict the models to only two predictors, both DS-01 (Eq. 7) and DS-02 (Eq. 8) converged to simplified expressions based solely on Normalised Laser Power (NLP) and Normalised Volumetric Energy Density (NVED). Although this reduction in dimensionality produced the expected decrease in predictive performance, the resulting models retained strong explanatory power, achieving test R^2 values of 0.72 for DS-01 and 0.89 for DS-02. These compact, physics-meaningful models demonstrate that a minimal descriptor set composed of NLP and NVED can still capture the majority of variance in RD, offering a practical trade-off between model interpretability and accuracy. The resulting two-variable symbolic regression models are given by:

$$\rho_{rel} = 94 - 3.98log(-0.479x4 + 0.795 + e - 65.3(x3 - 0.179)2) \quad (7)$$

$$\rho_{rel} = 54.5\tanh\left(0.023x_8 - e^{-27.4(0.155-x_3)^2 - 2.47(0.016x_8-1)^2} + 1.08\right) + 46.1 \quad (8)$$

where x_4 is normalised laser power (NLP), x_8 is dimensionless laser power (DLP) and x_3 is normalized volumetric energy density (NVED) (refer to Table 5).

Model generalisability and transferability

The concise two-descriptor expressions for RD were next evaluated for their ability to generalise across independent domain sources. Because DS-01 was produced through batch printing, each consisting of 100 samples with repeated machine setup and operator interventions, it inherently captures a wider range of complex, nonlinear process interactions. This makes DS-01 the more challenging dataset, but also the most valuable benchmark for assessing whether models trained under controlled conditions can remain robust in the presence of realistic process variation.

The DS-02 model (Eq. 7) was first applied to DS-01, achieving an $R^2 = 0.62$ (Pearson's $r = 0.79$). This level of performance is notable given the additional complexity embedded in DS-01 and aligns with the statistical analysis showing that the RD target variable exhibits only minor divergence between the datasets (refer to Section "Physical interpretation: aligning feature importance with solidification fundamentals", Fig. 12d). The small Wasserstein distance and nearly identical RD distributions confirm highly consistent densification behaviour across both PBF-LB systems, indicating that the underlying physics of melt-pool consolidation is stable across machines. Thus, the residual 38% unexplained variance in the DS-02 to DS-01 transfer task is not due to differences in densification outcomes, but rather due to the richer process variability introduced by the batch-based printing workflow, including subtle machine differences, environmental fluctuations, and operator-specific influences not captured by the two selected dimensionless descriptors.

Conversely, when the DS-01 model (Eq. 6) was applied to DS-02, it achieved a substantially higher R^2 of 0.86 (Pearson's $r = 0.93$). This reflects the lower machine-specific variability in DS-02, which was manufactured in a single, stable 256-sample build, and further supports the conclusion drawn from the RD-domain analysis: both systems produce statistically consistent densities, and divergence arises primarily from differences in the input process signatures rather than the target variable itself. In this context, DS-01 serves as a rigorous stress-test dataset that exposes the model to realistic physical and practical manufacturing variability, while DS-02 functions as a clean validation set that confirms the stability of the learned physical relationships.

Physical insights from PIML

Both the derived symbolic expressions consistently selected variants of laser power and normalised volumetric energy density (NVED) as the dominant predictors of RD across DS-01 and DS-02, reinforcing the earlier observation that the model generation procedure discovers descriptors that remain transferable between machines and process settings. The ability of these expressions to generalise across datasets indicates that the governing mechanisms of defect formation are not dataset-specific artefacts but instead arise from the fundamental way energy is delivered and distributed in the melt pool. This explains why these dimensionless

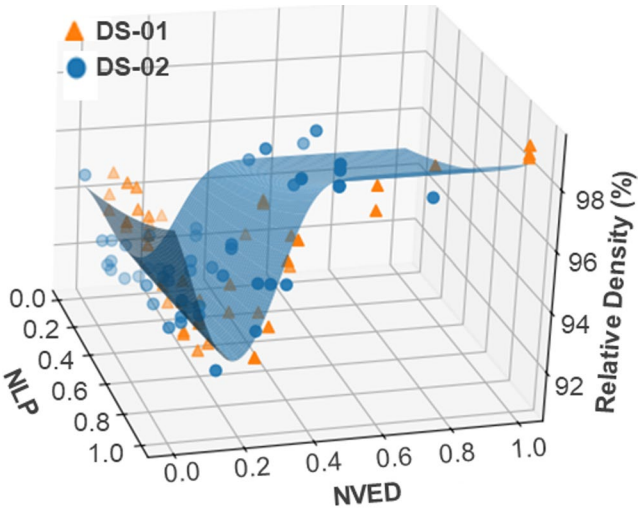


Fig. 15 Final two-variable model surface as a function of NLP and NVED, overlaid with RD datapoints from both domain sources

parameters emerged repeatedly during the constrained symbolic regression stage and why they preserved predictive performance even when evaluated on data originating from different machines and scan strategies.

The final two-variable model surface in Fig. 15 illustrates this transferability concretely. Both datasets collapse onto the same physically meaningful structure, revealing a defect-prone band at an NVED value of roughly 0.18, corresponding to a transition region where melt-pool instability such as keyholing or balling becomes likely. At lower NLP values, the model captures the lack-of-fusion regime, characterised by insufficient energy input and incomplete melting. Across the joint dataset, the resulting surface delineates a shared operating window for 316 L processed under same process parameter window, with boundaries defined by under-melting on one side and excessive energy deposition on the other.

The fact that this physically consistent representation emerges from a unified model trained on heterogeneous datasets highlights the advantage of the PIML framework for cross-platform parameter selection and model transfer. Because the model relies on physics-derived descriptors rather than machine-specific thermal signatures, it remains robust when applied to new systems or modified process envelopes. In this sense, the approach offers a complementary pathway to sensor-driven monitoring: it enables generalisable, offline predictions of defect risk and density targets without requiring in-process data from each machine. As the framework evolves, the hybrid outlier-removal step could be replaced by a dedicated classifier that automatically identifies unstable process regimes, thereby ensuring that subsequent property-prediction models are applied only on mechanically meaningful, transferable datasets.

Conclusion

This study benchmarked 40 supervised machine learning (ML) algorithms for predicting relative density (RD) in PBF-LB–processed 316 L stainless steel using 256 samples characterized by melt-pool temperature features extracted from *in-situ* infrared (IR) pyrometry. Notably, some samples exhibited powder-bed packing density failures (specimens that disintegrated after fabrication and were assigned a nominal RD of 60%). Retaining these extreme outcomes preserved the full physical spectrum of build quality and enabled models to learn from both successful and failed fabrication scenarios, an aspect rarely included in prior PBF-LB ML studies.

The benchmarking revealed the limited predictive capability of conventional process parameters alone. Statistical ANOVA identified only hatch spacing and volumetric energy density as significant contributors to RD variability, with very poor performance (R^2 = 49). In contrast, incorporating *in-situ* pyrometry features substantially improved model accuracy. Recursive feature elimination identified melt-pool thermal stability descriptors, specifically *Interquartile_Range_Temp* and *Mode_Temp*, as the most influential predictors. Models using this sensor-driven feature set achieved nearly double the predictive performance of parameter-only models (representing ~30% improvement in R^2 and ≈50% reduction in error metrics). An XGBoost model demonstrated strong generalization on a held-out validation dataset from the same machine (DS-01), achieving R^2 = 0.93 with accurate prediction of low-density failure cases. Combining sensor features with process parameters (*hybrid* and *hybrid-enforced feature sets*) resulted in a slight performance decrease (validation R^2 = 0.92; RMSE increasing from 2.83 to ~3.01). This suggests that the process parameters introduced mild redundancy, as their predictive signal was already captured by the more dominant thermal descriptors, a finding consistent with the SHAP analysis.

However, a critical limitation emerged in cross-system transferability. When applied to an independent domain source (DS-02) collected on a different PBF-LB machine, sensor-driven models failed. This is due to domain divergence in pyrometer signal characteristics, driven by differences in build area and the corresponding effect on interlayer times. This was quantitatively confirmed using Kolmogorov–Smirnov statistics and Wasserstein distance. In contrast, the target variable

(RD) showed only minor divergence (KS = 0.157, p = 0.003, WD = 1.75) and virtually identical mean values (92.40% vs. 93.75%), indicating strong inter-laboratory consistency and confirming that transferability failure originated from the input domain, not the output domain.

More broadly, these findings highlight that the scalability of sensor-driven ML in PBF-LB is constrained not only by model architecture, but also by the lack of system-level standardization in *in-situ* sensor integration and signal representation. Even when nominally identical pyrometric hardware is employed, variations in optical configuration, alignment, emissivity settings, and post-installation signal conditioning can yield non-comparable thermal signatures across machines. Establishing harmonized practices for sensor integration, calibration transfer, and signal normalization therefore represents a key enabling step toward transferable, data-driven process monitoring and control in industrial PBF-LB systems.

To address the transferability challenge, a physics-informed machine learning (PIML) framework was implemented using the QLattice symbolic regression library, yielding compact and interpretable expressions dominated by normalized laser power and volumetric energy density. These models achieved substantially improved cross-system generalization, with one expression explaining 93% of the variance during validation (R^2 = 0.93), further confirming the similarity of RD outcomes in DS-01 and DS-02 fabricated under nominally identical PBF-LB parameters. By embedding dimensionless physical priors constrained by established process knowledge, the PIML approach mitigates the machine-specific dependencies that limit the portability of raw sensor-based descriptors, thereby bridging empirical prediction and mechanistic understanding.

Importantly, this does not diminish the value of conventional sensor-driven ML pipelines. Models trained on *in-situ* melt-pool thermal signatures remain the most suitable approach for layer-by-layer monitoring, defect detection, and real-time closed-loop feedback control. This is because thermal profiles vary strongly with specimen geometry even when nominal process parameters are unchanged. Such geometry-dependent thermal dynamics, together with machine-specific sensor characteristics, make high-resolution, sensor-based models uniquely effective for capturing local deviations during a build. In contrast, PIML serves a complementary role: it enables global process-window estimation, and build planning before manufacturing, supporting scalable deployment across machines, facilities, and material batches without extensive sensor calibration or retraining. Ultimately, this study demonstrates a dual-pathway framework: sensor-based ML for high-resolution, layer-by-layer monitoring within a single-machine, and physics-informed models for scalable, system-agnostic process planning.

Recommendations for future work

While the dual-pathway framework established in this study provides a robust foundation for quality assurance in PBF-LB, several avenues remain for further exploration to enhance industrial scalability, predictive depth, and critically, cross-machine transferability:

1. **Addressing Domain Divergence for Transferable Models**: The core limitation identified in Section "Physical interpretation: aligning feature importance with solidification fundamentals", the failure of sensor-data-driven models trained on one machine (DS-01) to perform on another (DS-02), must be systematically addressed to enable industrial deployment. Future work should develop and compare specific mitigation strategies, including:
    a Domain Adaptation Techniques: Implementing feature-space alignment methods (e.g., Correlation Alignment, CORAL [77]) or adversarial training [78] to learn machine-invariant representations from pyrometer data.
    b Physics-Based Signal Normalization: Creating transfer functions grounded in thermal physics or sensor physics to map disparate pyrometer signals (e.g., mV vs. ^{o}C units) onto a common, physically interpretable scale, moving beyond statistical standardization.
    c Explicit Covariate Integration: Compute and direct incorporation of layer-by-layer thermal history and inter-layer dwell times into model architectures, to mechanistically account for the physical source of domain shift, as outlined in *Recommendation No. 4* below.

2. **Multi-Material Generalization and Feature Universality**: Although 316 L SS served as the primary vehicle for this study, a critical next step is to evaluate the portability of the IR pyrometer sensor-driven and PIML frameworks to other alloy systems, such as NiTi, Ti-6Al-4 V, and IN718. Future research should specifically investigate whether *Mode_Temp* and *Interquartile_Range_Temp* maintain their high SHAP feature importance across materials with vastly different thermophysical properties. For instance, in alloys with wider solidification ranges (e.g., IN718) or those prone to phase transformations (e.g., NiTi), the "stability" captured by these features may correlate differently with defect formation than in the relatively stable 316 L SS matrix. Validating whether these specific IR-derived features consistently serve as the primary "proximal

predictors" across varying thermal conductivities and solidification kinetics would confirm the framework's utility as a universal, material-agnostic tool for PBF-LB process optimization.

3. **Advanced Temporal and Spatial Feature Engineering**: This study utilized global thermal descriptors averaged across all layers. Future work should incorporate the temporal evolution of the melt pool and spatial "thermal neighborhoods" (accounting for heat accumulation from preceding tracks and layers). Integrating these into thermo-mechanical analytical formulations could enable the prediction of not only RD but also structural integrity defects such as macro/micro-cracks, delamination, and warping, which serve as proxies for residual stress accumulation.
4. **Hybrid Synthetic-Experimental Data Pipelines & Direct Machine-Divergence Analysis**: To reduce the reliance on extensive experimental campaigns (as performed in this study with 512 samples) and directly address the cross-machine challenge, future work should integrate high-fidelity physics-based thermal modeling. Specifically, Computational Fluid Dynamics (CFD) or Finite Element Analysis (FEA) could be employed to simulate key descriptors like *Mode_Temp* and *Interquartile_Range_Temp* across an even broader processing space and across simulated machine conditions (e.g., varying inter-layer times, chamber geometries). This would allow for a mechanistic investigation into the signal divergence observed between DS-01 and DS-02, coupling physics-informed simulations with ML to generate synthetic data that bridges domain gaps.
5. **Closed-Loop Integration with Domain-Aware Control**: The sensor-driven models and the PIML regression formulas identified in this work are ideal candidates for real-time deployment. Future efforts should focus on integrating these model architectures into the machine control unit to enable real-time feed-forward or feedback loops, where laser power or scan speed is adjusted mid-build to maintain thermal descriptors within a "stability zone" for a desired RD target. This control logic must be informed by the domain adaptation strategies in *Recommendation No.1* to function robustly across different machines.

Appendix

Table 12 Relative density results of samples with matching Volumetric energy density based on hatch spacing and spot size. Layer thickness is constant at 50 μm. Table is sorted by descending relative density value.

Sample	*P* (W)	V (mm/s)	H (μm)	f (μm)	VED_H (J/mm^3)	VED_f (J/mm^3)	Relative Density (%)
58	200	600	90	90	74.07	74.07	99
48	160	600	90	90	59.26	59.26	97.91
45	200	800	90	90	55.56	55.56	97.88
40	200	1000	90	90	44.44	44.44	96.57
199	160	800	90	90	44.44	44.44	96.22
52	120	600	90	90	44.44	44.44	94.76
123	80	1200	90	90	14.81	14.81	94.24
174	80	800	90	90	22.22	22.22	94.13
16	120	1200	90	90	22.22	22.22	94.06
18	160	1000	90	90	35.56	35.56	93.26
81	200	1200	90	90	37.04	37.04	93.26
200	80	1000	90	90	17.78	17.78	92.7
182	120	800	90	90	33.33	33.33	92.28
87	80	600	90	90	29.63	29.63	92.07
242	160	1200	90	90	29.63	29.63	91.52
165	120	1000	90	90	26.67	26.67	91.22

Table 13 Summary of raw feature distribution divergence between DS-01 and DS-02 domain sources

10	KS Statistic	KS *p*-value	WD	DS_01 Mean	DS_02 Mean
Variance_Temp	0.84	1.16E-91	2555.55	3448.65	893.10
Max_Temp	1.00	3.41E-152	1007.12	1197.80	190.68
Percentile_90_Temp	1.00	3.41E-152	999.10	1126.01	126.91
Mode_Temp	1.00	3.41E-152	998.12	1081.59	83.46
Q3_Temp	1.00	3.41E-152	996.10	1102.62	106.51
Median_Temp	1.00	3.41E-152	985.35	1073.52	88.17
Avg_Temp	1.00	3.41E-152	976.23	1066.22	89.99
Q1_Temp	1.00	3.41E-152	966.92	1039.81	72.89
Percentile_10_Temp	1.00	3.41E-152	943.05	1001.02	57.97
Range_Temp	0.97	3.58E-135	180.14	366.80	186.65
Interquartile_Range_Temp	0.72	3.36E-64	29.18	62.80	33.63
Stdev_Temp	0.86	3.38E-97	26.08	54.46	28.38
Coefficient_of_Variation_Temp	1.00	3.41E-152	0.27	0.05	0.32
Measured Relative Density (%)	0.16	0.0032	1.75	92.40	93.75

KS refers to Kolmogorov–Smirnov, and WD refers to the Wasserstein Distance

Table 14 Summary of raw feature distribution divergence between DS-01 and DS-02 domain sources

Feature	KS Statistic	KS *p*-value	WD	DS-01 Mean	DS_02 Mean
Mode_Value	0.1898	0.0002	0.4098	-7.08E-16	1.49E-16
Variance_Value	0.3164	0.0000	0.3907	6.94E-17	2.53E-16
Stdev_Value	0.2062	0.0000	0.3802	-9.37E-17	3.93E-16
Coefficient_of_Variation_Value	0.1744	0.0007	0.3727	-2.88E-16	4.49E-16
Q3_Value	0.1522	0.0047	0.3462	-8.57E-16	3.30E-16
Percentile_90_Value	0.1678	0.0013	0.3421	-2.09E-15	-2.81E-17
Median_Value	0.1570	0.0032	0.3152	1.21E-15	2.11E-17
Interquartile_Range_Value	0.1947	0.0001	0.3109	-2.64E-16	-1.69E-16
Range_Value	0.1316	0.0213	0.2926	-8.40E-16	2.81E-16
Max_Value	0.1316	0.0213	0.2926	-6.73E-16	0.00E+00
Avg_Value	0.1334	0.0189	0.2782	4.16E-17	7.09E-16
Q1_Value	0.1147	0.0629	0.2009	3.99E-15	2.43E-16
Percentile_10_Value	0.0647	0.6281	0.1096	-8.79E-16	-1.40E-17

KS refers to Kolmogorov–Smirnov, and WD refers to the Wasserstein Distance

Relative density was not normalised and hence excluded in this table

Acknowledgements The authors would like to thank the ESAFORM European Scientific Association for Material Forming for funding this research through the ESAFORM Benchmark 2025 AMDmodel initiative. The collective effort of all involved institutions has been invaluable to the success of this benchmark study.
Authors from Dublin City University acknowledge research funding support from Research Ireland's I-Form Advanced Manufacturing Research Centre under Grant Number 21/RC/10295_P2, co-funded under the European Regional Development Fund.
Authors from NMIS, University of Strathclyde acknowledge the support from the Data-driven Design and Manufacturing CoLab (D3M_CoLab) grant (project ref: 10054855, 10158298), funded by UKRI.
Authors from ETH Zurich gratefully acknowledge the internal funding, equipment, and broader support provided by the Professorship for Advanced Manufacturing at ETH Zurich.
All authors from RWTH Aachen University and Aalto University acknowledge the funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (HIGMAM, grant agreement No. 101078829).
Authors from Technical University of Denmark acknowledge research funding support from the Villum Fonden (MicroAM) under grant number VIL54495.
As Research Director of F.R.S-FNRS, Anne M. Habraken acknowledges the support of "Fonds de la Recherche Scientifique – FNRS".
Authors from ULiege acknowledge the CAREM of the ULiege for providing SEM/EDS/EBSD facilities. They also thank Sylvie Salieri for the sample preparation.

Author contributions This manuscript reports the outcomes of the ESAFORM 2025 Benchmark initiative (AM-D-model), a coordinated, community-based effort involving 9 universities and 5 research institutes across 8 countries. All authors contributed to the conception and design of the benchmark, data generation and/or analysis, interpretation of results, and critical review of the manuscript. All authors read and approved the final version of the manuscript.

Funding Open Access funding provided by the IReL Consortium.

Data availability Density dataset is available at: https://invenio.am-d-model.eu/records/b2wwb-tjf65.

Declarations

Competing interests The authors declare no competing interests.

Conflict of 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.

References

1. Farrag A, Yang Y, Cao N et al (2025) Physics-Informed Machine Learning for metal additive manufacturing. Prog Addit Manuf 10:171–185. https://doi.org/10.1007/s40964-024-00612-1
2. Kapusuzoglu B, Mahadevan S (2020) Physics-Informed and Hybrid Machine Learning in Additive Manufacturing: Application to Fused Filament Fabrication. JOM 72:4695–4705. https://doi.org/10.1007/s11837-020-04438-4
3. Yang S, Peng S, Guo J, Wang F (2024) A Review on Physics-Informed Machine Learning for Monitoring Metal Additive Manufacturing Process. Adv Manuf 1:1–2. https://doi.org/10.55092/am20240008
4. Chen K, Zhang P, Yan H et al (2024) A review of machine learning in additive manufacturing: design and process. Int J Adv Manuf Technol 135:1051–1087. https://doi.org/10.1007/s00170-024-14543-2
5. Barrionuevo GO, Ramos-Grez JA, Walczak M, Betancourt CA (2021) Comparative evaluation of supervised machine learning algorithms in the prediction of the relative density of 316L stainless steel fabricated by selective laser melting. Int J Adv Manuf Technol 113:419–433. https://doi.org/10.1007/s00170-021-06596-4
6. Barrionuevo GO, La Fé-Perdomo I, Ramos-Grez JA (2025) Laser powder bed fusion dataset for relative density prediction of commercial metallic alloys. Sci Data 12:1–11. https://doi.org/10.1038/s41597-025-04576-x
7. Gor M, Dobriyal A, Wankhede V et al (2022) Density Prediction in Powder Bed Fusion Additive Manufacturing: Machine Learning-Based Techniques. Appl Sci 12:7271. https://doi.org/10.3390/app12147271
8. Abdulla H, Maalouf M, Barsoum I, An H (2022) Truncated newton kernel ridge regression for prediction of porosity in additive manufactured SS316L. Appl Sci 12. https://doi.org/10.3390/app12094252
9. Eshkabilov S, Ara I, Azarmi F (2022) A comprehensive investigation on application of machine learning for optimization of process parameters of laser powder bed fusion – processed 316L stainless steel. Int J Adv Manuf Technol 2733–2756. https://doi.org/10.1007/s00170-022-10331-y
10. Wang J, Jeong SG, Kim ES et al (2023) Material-agnostic machine learning approach enables high relative density in powder bed fusion products. Nat Commun 14:6557. https://doi.org/10.1038/s41467-023-42319-x
11. Mahato V, Chatterjee S, Nyabadza A et al (2024) Layer porosity in powder-bed fusion prediction using regression machine learning models and time-series features. Int J AI Mater Des 1:33. https://doi.org/10.36922/ijamd.4812
12. Zhu Q, Liu Z, Yan J (2021) Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks. Comput Mech 67:619–635. https://doi.org/10.1007/s00466-020-01952-9
13. WuY, Sicard B (2024) A review of physics-informed machine learning methods with applications to condition monitoring and anomaly detection. arXivLabs. https://doi.org/10.48550/arXiv.2401.11860
14. Grose J, Liao A, Foong CS, Cullinan M (2023) Data-efficient surrogate model for rapid prediction of temperature evolution in a microscale selective laser sintering system. J Micro- Nano-Manufacturing 11. https://doi.org/10.1115/1.4064106
15. Peng B, Panesar A (2023) Predicting temperature field for mental additive manufacturing using PINN. Int Solid Free Fabr Symp 881–896 https://doi.org/10.26153/tsw/50997
16. Sajadi P, Dehaghani MR, Tang Y, Wang GG (2024) Real-time 2D temperature field prediction in metal additive manufacturing using physics-informed neural networks. arXiv org 1–42 https://doi.org/10.48550/arXiv.2401.02403
17. Jiang F, Xia M, Hu Y (2024) Physics-Informed Machine Learning for Accurate Prediction of Temperature and Melt Pool Dimension in Metal Additive Manufacturing. 3D Print Addit Manuf 11:e1679–e1689. https://doi.org/10.1089/3dp.2022.0363
18. Habraken AM, Aksen TA, Alves JL et al (2022) Analysis of ESAFORM 2021 cup drawing benchmark of an Al alloy, critical factors for accuracy and efficiency of FE simulations. Int J Mater Form 15:61. https://doi.org/10.1007/s12289-022-01672-w
19. Agirre J, Bernal D, Flipon B et al (2025) The ESAFORM benchmark 2023: interlaboratory comparison benchmark for the characterization of microstructural grain growth and dynamic recrystallization kinetics of a single-phase Ni-base superalloy. Int J Mater Form 18:33. https://doi.org/10.1007/s12289-025-01893-9
20. Vanhulst M, Lee Y, Steinfels D et al (2025) ESAFORM benchmark 2024: study on the geometric accuracy of a complex shape with single point incremental forming. Int J Mater Form 18:72. https://doi.org/10.1007/s12289-025-01928-1
21. ASTM B962-23 (2023) Test methods for density of compacted or sintered powder metallurgy (PM) products using archimedes principle. ASTM B962-23 i:1–7
22. Carter LN, Martin C, Withers PJ, Attallah MM (2014) The influence of the laser scan strategy on grain structure and cracking behaviour in SLM powder-bed fabricated nickel superalloy. J Alloys Compd 615:338–347. https://doi.org/10.1016/j.jallcom.2014.06.172
23. Monu MCC, Ekoi EJ, Hughes C et al (2022) Resultant physical properties of as-built nitinol processed at specific volumetric energy densities and correlation with in-situ melt pool temperatures. J Mater Res Technol 21:2757–2777. https://doi.org/10.1016/j.jmrt.2022.10.073
24. Awad M, Fraihat S (2023) Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. J Sens Actuator Networks 12:67. https://doi.org/10.3390/jsan12050067
25. Fabian P, Gael V, Alexandre G et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830
26. Buitinck L, Louppe G, Blondel M et al (2013) API design for machine learning software: experiences from the scikit-learn project. 1–15 https://doi.org/10.48550/arXiv.1309.0238
27. BÜYÜKKEÇECİ M, OKUR MC (2023) A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning. Gazi Univ J Sci 36:1506–1520. https://doi.org/10.35378/gujs.993763
28. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623. https://doi.org/10.7717/peerj-cs.623
29. Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst 2017–Decem:4766–4775
30. Thomas M, Baxter GJ, Todd I (2016) Normalised Model-Based Processing Diagrams for Additive Layer Manufacture of Engineering Alloys. Acta Mater 108:26–35
31. Manikandan P, Venkatesan K (2024) The Influence of Linear Energy Density on Density, Defect Formation, Residual Stress, Microstructure, and Texture in 310 Austenitic Stainless Steel by

Laser Powder Bed Fusion. J Manuf Process 131:2191–2207. https://doi.org/10.1016/j.jmapro.2024.10.022
32. Wang Z, Zhang J, Zhang F, Qi C (2024) Impact of laser energy density on the structure and properties of laser-deposited Fe–Ni–Ti composite coatings. Front Mater 11:1408333. https://doi.org/10.3389/fmats.2024.1408333
33. Buhairi MA, Foudzi FM, Jamhari FI et al (2023) Review on volumetric energy density: influence on morphology and mechanical properties of Ti6Al4V manufactured via laser powder bed fusion. Prog Addit Manuf 8:265–283. https://doi.org/10.1007/s40964-022-00328-0
34. Minervini I, Yang P, Higgins H, Ziade E (2023) Thermal conductivity of am fabricated stainless steel 316L Part. SANDIA-REPORT: 2023-00333. https://doi.org/10.2172/2432179
35. Demir P, Kacar E, Akman E, Demir A (2014) Single Pulse Laser Ablation of AISI 316L Stainless Steel Surface Using Nd : YAG Laser Irradiation. ACTA Phys Pol A 125:439–441. https://doi.org/10.12693/APhysPolA.125.439
36. Aldeia GSI, Zhang H, Bomarito G et al (2025) Call for Action: towards the next generation of symbolic regression benchmark. Proc Genet Evol Comput Conf Companion 1:2529–2538. https://doi.org/10.1145/3712255.3734309
37. Broløs KR, Machado MV, Cave C et al (2021) An approach to symbolic regression using Feyn. https://arXiv.org/abs/210405417
38. Cave C, Kasak J (2025) Fitting models. In: Feyn Doc. https://docs.abzu.ai/docs/guides/primitives/fit_models. Accessed 12 Nov 2025
39. González-Estrada E, Cosmes W (2019) Shapiro–Wilk test for skew normal distributions based on data transformations. J Stat Comput Simul 89:3258–3272. https://doi.org/10.1080/00949655.2019.1658763
40. Haferkamp L, Spierings A, Rusch M et al (2021) Effect of Particle size of monomodal 316L powder on powder layer density in powder bed fusion. Prog Addit Manuf 6:367–374. https://doi.org/10.1007/s40964-020-00152-4
41. Choi J-P, Shin G-H, Lee H-S et al (2017) Evaluation of Powder Layer Density for the Selective Laser Melting (SLM) Process. Mater Trans 58:294–297. https://doi.org/10.2320/matertrans.M2016364
42. Huang G, Chen H, Ma Z et al (2025) Microstructure and mechanical properties of stainless steel addictively manufactured via laser powder bed fusion in high-dense process parameter window. Mater Sci Eng A 928:148033. https://doi.org/10.1016/j.msea.2025.148033
43. Jaskari M, Hamada A, Allam T et al (2025) Effects of volumetric energy density on defect structure and fatigue behaviour of powder bed fusion manufactured 316L stainless steel. Mater Sci Eng A 925:147868. https://doi.org/10.1016/j.msea.2025.147868
44. Tan HT, Lopez E, Selbmann A et al (2024) Parameter analysis to correlate density with surface roughness and productivity in Powder Bed Fusion – Laser Beam (PBF-LB) of AlSi10Mg. Procedia CIRP 124:157–162. https://doi.org/10.1016/j.procir.2024.08.090
45. Scipioni Bertoli U, Wolfer AJ, Matthews MJ et al (2017) On the limitations of Volumetric Energy Density as a design parameter for Selective Laser Melting. Mater Des 113:331–340. https://doi.org/10.1016/j.matdes.2016.10.037
46. Smith C, Hommer G, Keeler M et al (2025) Assessing Volumetric Energy Density as a Predictor of Defects in Laser Powder Bed Fusion 316L Stainless Steel. Jom 77:737–748. https://doi.org/10.1007/s11837-024-06946-z
47. Buhairi MA, Foudzi FM, Jamhari FI et al (2022) Review on volumetric energy density: influence on morphology and mechanical properties of Ti6Al4V manufactured via laser powder bed fusion. Prog Addit Manuf 2022 1–19. https://doi.org/10.1007/S40964-022-00328-0
48. Forien JB, Calta NP, DePond PJ et al (2020) Detecting keyhole pore defects and monitoring process signatures during laser powder bed fusion: A correlation between in situ pyrometry and ex situ X-ray radiography. Addit Manuf 35:101336. https://doi.org/10.1016/j.addma.2020.101336
49. Yadav P, Rigo O, Arvieu C et al (2020) In Situ Monitoring Systems of The SLM Process: On the Need to Develop Machine Learning Models for Data Processing. Crystals 10:524. https://doi.org/10.3390/cryst10060524
50. McCann R, Obeidi MA, Hughes C et al (2021) In-situ sensing, process monitoring and machine control in Laser Powder Bed Fusion: A review. Addit Manuf 45:102058. https://doi.org/10.1016/j.addma.2021.102058
51. Yu L, Liu H (2004) Efficient Feature Selection via Analysis of Relevance and Redundancy. J ofMachine Learn Res 5:1205–1224
52. Manvatkar V, De A, DebRoy T (2015) Spatial variation of melt pool geometry, peak temperature and solidification parameters during laser assisted additive manufacturing process. Mater Sci Technol (United Kingdom) 31:924–930. https://doi.org/10.1179/1743284714Y.0000000701
53. Mohr G, Altenburg SJ, Ulbricht A et al (2020) In-situ defect detection in laser powder bed fusion by using thermography and optical tomography—comparison to computed tomography. Met (Basel) 10. https://doi.org/10.3390/met10010103
54. Yan Z, Liu W, Tang Z et al (2018) Review on thermal analysis in laser-based additive manufacturing. Opt Laser Technol 106:427–441. https://doi.org/10.1016/j.optlastec.2018.04.034
55. Riensche A, Bevans BD, Smoqi Z et al (2022) Feedforward control of thermal history in laser powder bed fusion: Toward physics-based optimization of processing parameters. Mater Des 224:111351. https://doi.org/10.1016/j.matdes.2022.111351
56. Hu L, Li Y, Luo G et al (2024) Constructing solidification microstructure selection map for AlSi10Mg alloy processed by laser powder bed fusion. J Alloys Compd 976:173308. https://doi.org/10.1016/j.jallcom.2023.173308
57. Huang Y, Fleming TG, Clark SJ et al (2022) Keyhole fluctuation and pore formation mechanisms during laser powder bed fusion additive manufacturing. Nat Commun 13:1–11. https://doi.org/10.1038/s41467-022-28694-x
58. Forien J, Calta N, DePond P, Guss G, Roehling T (2019) MM Detecting keyhole porosity and monitoring process signatures in additive manufacturing : an in situ pyrometry and ex situ X-ray radiography correlation. 1–3. https://doi.org/LLNL-JRNL-782115
59. Kotadia HR, Gibbons G, Das A, Howes PD (2021) A review of Laser Powder Bed Fusion Additive Manufacturing of aluminium alloys: Microstructure and properties. Addit Manuf 46:102155. https://doi.org/10.1016/j.addma.2021.102155
60. Kurz W, Fisher DJ, Rappaz M (2023) Fundamentals of Solidification, 5th edn. Trans Tech Publications Ltd https://digital.casalini.it/9783036410159
61. Santos Macías JG, Douillard T, Zhao L et al (2020) Influence on microstructure, strength and ductility of build platform temperature during laser powder bed fusion of AlSi10Mg. Acta Mater 201:231–243. https://doi.org/10.1016/j.actamat.2020.10.001
62. Khorasani M, Ghasemi AH, Leary M et al (2022) A comprehensive study on meltpool depth in laser-based powder bed fusion of Inconel 718. Int J Adv Manuf Technol 120:2345–2362. https://doi.org/10.1007/s00170-021-08618-7
63. Leung CLA, Marussi S, Atwood RC et al (2018) In situ X-ray imaging of defect and molten pool dynamics in laser additive manufacturing. Nat Commun 9:1355. https://doi.org/10.1038/s41467-018-03734-7
64. Scipioni Bertoli U, Guss G, Wu S et al (2017) In-situ characterization of laser-powder interaction and cooling rates through high-speed imaging of powder bed fusion additive manufacturing. Mater Des 135:385–396. https://doi.org/10.1016/j.matdes.2017.09.044

65. Abrami MB, Tocci M, Obeidi MA et al (2022) Prediction of Microstructure for AISI316L Steel from Numerical Simulation of Laser Powder Bed Fusion. Met Mater Int 28:2735–2746. https://doi.org/10.1007/s12540-022-01168-x
66. Mohr G, Altenburg SJ, Hilgenberg K (2020) Effects of inter layer time and build height on resulting properties of 316L stainless steel processed by laser powder bed fusion. Addit Manuf 32:101080. https://doi.org/10.1016/j.addma.2020.101080
67. Vallabh CKP, Hinnebusch S, To AC, Zhao X (2022) Layer-wise melt pool temperature evolution in laser powder bed fusion: an experimental study using a single camera based two-wavelength imaging pyrometry, vol 2. Manufacturing processes; manufacturing systems. American Society of Mechanical Engineers https://doi.org/10.1115/MSEC2022-85387
68. Krishna C, Vallabh P, Zhang H et al (2024) Melt pool width measurement in a multi – track, multi – layer laser powder bed fusion print using single – camera two – wavelength imaging pyrometry. Int J Adv Manuf Technol 2575–2585. https://doi.org/10.1007/s00170-024-13486-y
69. Kusano M, Takata Y, Yumoto A, Watanabe M (2024) Effects of time per layer and part geometry on thermal history and microcracking in the fabrication of nickel superalloy samples by laser powder bed fusion. Addit Manuf 80:103987. https://doi.org/10.1016/j.addma.2024.103987
70. Medvedev AE, Brudler S, Piegert S et al (2025) Interlayer time as a robust, geometry-agnostic predictor of microstructural and mechanical properties evolution in PBF-LB/M Ti6Al4V alloy. J Mater Process Technol 340:118858. https://doi.org/10.1016/j.jmatprotec.2025.118858
71. Zhang W, Tong M, Harrison NM (2023) Multipart Build Effects on Temperature and Residual Stress by Laser Beam Powder Bed Fusion Additive Manufacturing. 3D Print Addit Manuf 10:749–761. https://doi.org/10.1089/3dp.2021.0143
72. Karson M (1968) Handbook of methods of applied statistics. Volume I: techniques of computation descriptive methods, and statistical inference. Volume II: Planning of surveys and experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967, $9.00. J Am Stat Assoc 63:1047–1049. https://doi.org/10.1080/01621459.1968.11009335
73. Calixto E (2016) Lifetime Data Analysis. In: Calixto E (ed) Gas and Oil Reliability Engineering, Second Edi. Elsevier, Boston, pp 1–92
74. Vegas Sánchez-Ferrero G, Alberola-López C (2024) Estimation and inference. In: Frangi AF, Prince JL, Sonka M (eds) Medical Image Analysis. Elsevier, pp 85–111 https://doi.org/10.1016/B978-0-12-813657-7.00016-9
75. Ramdas A, Trillos N, Cuturi M (2017) On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests. Entropy 19:47. https://doi.org/10.3390/e19020047
76. Panaretos VM, Zemel Y (2019) Statistical Aspects of Wasserstein Distances. Annu Rev Stat Its Appl 6:405–431. https://doi.org/10.1146/annurev-statistics-030718-104938
77. Sun B, Feng J, Saenko K (2016) Return of frustratingly easy domain adaptation. Proc AAAI. Conf Artif Intell 30. https://doi.org/10.1609/aaai.v30i1.10306
78. Rahman MM, Fookes C, Baktashmotlagh M, Sridharan S (2020) Correlation-aware adversarial domain adaptation and generalization. Pattern Recognit 100:107124. https://doi.org/10.1016/j.patcog.2019.107124

Authors and Affiliations

Medad Chiedozie C. Monu[1,2] · Eanna McCarthy[1,2] · Abhilash Puthanveettil Madathil[3] · Josiah C. Chekotu[1,2] · Irina Ilic[4] · Merve Nur Doğu[1,2] · Cian Hughes[1,2] · Rongfei Juan[5] · Ehsan Amini[5] · Junhe Lian[5,6] · Constantinos Vassiades[3] · Olga Bylya[3] · Kim Darosa[7] · Robin Kromer[7] · Abdul Herrim Seidou[8] · Sankhya Mohanty[9] · Anne Marie Habraken[10] · Anne Mertens[8] · Otto Laitinen[4] · Michael R. Tucker[4] · Dermot Brabazon[1,2]

✉ Medad Chiedozie C. Monu
monumedadcc@gmail.com

[1] I-Form, Advanced Manufacturing Research Centre, Dublin City University, Glasnevin, Dublin 9, Dublin, Ireland

[2] DCU Institute for Advanced Processing Technology, School of Mechanical and Manufacturing Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland

[3] National Manufacturing Institute Scotland, University of Strathclyde, Paisley, Renfrew PA3 2EF, UK

[4] Advanced Manufacturing Laboratory, Department of Mechanical and Process Engineering, ETH Zurich, Technoparkstrasse 1, Zurich 8005, Switzerland

[5] Materials to Products, Department of Energy and Mechanical Engineering, Aalto University, Puumiehenkuja 3, Espoo, Finland

[6] Institute of Metal Forming, RWTH Aachen University, Intzestrasse 10, Aachen, Germany

[7] University of Bordeaux, Arts et Métiers, CNRS, I2M UMR, Talence, France

[8] Aerospace and Mechanics Department, MMS, University of Liège, 9 Allée de la Découverte, Liège 4000, Belgium

[9] Department of Civil and Mechanical Engineering, Technical University of Denmark, Lyngby, DK, Denmark

[10] ArGEnCo dpt, MSM team, University of Liège, 9 Allée de la Découverte, Liège 4000, Belgium