Adv. Manuf.
https://doi.org/10.1007/s40436-026-00596-x

Multi-objective process parameter optimization using a hybrid reinforcement/machine learning model in metal additive manufacturing

Arash Seifoddini[1] · Arefeh Azad[1] · Mohammad Hossein Mosallanejad[2,3,4] · Abdollah Saboori[3,4]

Received: 23 February 2025 / Revised: 29 April 2025 / Accepted: 26 January 2026

Abstract Additive manufacturing enables the production of complex geometries and lightweight structures. However, producing defect-free components fundamentally depends on the careful optimization of process parameters that influence critical factors, such as melt pool depth, in laser powder bed fusion (L-PBF), an important class of metal additive manufacturing methods. Hence, for the first time, this paper presents a novel data-driven framework that integrates machine learning (ML) with reinforcement learning (RL) to optimize process parameters, ensuring the desired melt pool characteristics while meeting multi-objective design criteria. Experimental data from the literature were extracted to train several ML models, with Gaussian process regression (GPR) featuring a Matérn kernel proved to be the most accurate predictor. This model served as the digital twin of the L-PBF process and was integrated with a Q-learning-based RL agent to derive the optimal process parameters. The GPR model achieved a high prediction accuracy (root mean squared error (RMSE): 34 μm, the coefficient of determination R^2: 96.05%) for L-PBF Ti6Al4V alloy using minimal training data. Integrating a Q-learning framework with ML demonstrated superior performance in deriving process parameters that achieved the target melt pool depth while minimizing the normalized enthalpy, directly impacting the formation of the keyhole defect.

Keywords Metal additive manufacturing · Laser powder bed fusion (L-PBF) · Machine learning (ML) · Q-learning · Normalized enthalpy · Keyhole

1 Introduction

Additive manufacturing (AM) is a method of building objects layer by layer based on digital models. The technique offers several advantages compared to the traditional methods. It allows for producing lightweight components with complex shapes, supports rapid prototyping, reduces environmental impact by minimizing material waste, and offers greater design flexibility [1–4]. There are various metal AM methods, such as laser powder bed fusion (L-PBF), electron beam powder bed fusion (EB-PBF), and directed energy deposition (DED), which are commonly used in various sectors, including automotive, aerospace, energy, and medical industries [5–8]. In L-PBF technology, a focused laser beam selectively melts metal powder particles to create a three-dimensional (3D) object. Following the melting of each layer, the build platform is lowered, and a fresh layer of powder is uniformly distributed over the previously solidified layer. This sequence of melting and layering continues iteratively until the final component is fully realized [9]. L-PBF has several advantages over EB-PBF and DED, such as offering high precision and a wide range of composition choices, including various alloys and composites [10, 11]. It also reduces the need for post-processing by producing better surface roughness and increases the reliability of consistency and repeatability in production [12, 13].

✉ Abdollah Saboori
abdollah.saboori@polito.it

1 Department of Mechanical and Aerospace Engineering, Polytechnic University of Turin, 10129 Torino, Italy

2 Department of Materials Engineering, Isfahan University of Technology, Isfahan 84156–83111, Iran

3 Department of Management and Production Engineering, Polytechnic University of Turin, 10129 Torino, Italy

4 Integrated Additive Manufacturing Center (IAM@PoliTo), Polytechnic University of Turin, 10129 Torino, Italy

Published online: 20 March 2026

Although there are numerous advantages to such manufacturing methods, fabrication of defect-free parts using AM techniques requires obtaining the optimal process parameters that lead to the required target characteristics. The depth of the melt pool is a critical factor in the L-PBF process, as improper control over this factor can negatively impact the mechanical and physical properties of the final product. Defects such as lack-of-fusion, porosity, keyholing, residual stresses, and surface irregularities are directly linked to the melt pool formation condition during the L-PBF process [14–20]. If the laser power is not sufficiently high, it leads to inadequate bonding between the powders, causing porosity. Furthermore, keyholing occurs when the energy density is high enough to boil the metal locally, and the accompanying recoil pressure pushes the molten pool surface to create a cavity or depression [15]. Additionally, the melt pool depth is crucial for determining the cooling rate and solidification phases, influencing residual stresses in the product. For instance, inconsistent melt pool depth can result in surface roughness and anisotropy [16]. To mitigate these issues, each target property may require adjusting the process parameters to values that could conflict with other properties. This results in a multi-objective optimization problem where weight and importance are assigned to each property, and the optimal parameters are determined by minimizing the objective function [21, 22].

This multi-objective optimization problem is traditionally solved using the design of experiment (DOE) sampling method. A finite sample space is generated, and the most significant parameters that influence the target properties are identified using analysis tools such as S/N ratio and analysis of variance (ANOVA) [23]. These methods are time-consuming, costly, and impractical for processes with numerous parameters, which result in a high dimensional sample space. Therefore, new approaches have been emerging to simplify the process, including artificial intelligence (AI) based methods such as machine learning (ML). These black-box models, which have gained prominence in recent years, can accurately model various stages of the design and manufacturing processes and consequently predict the desired targets based on the trained algorithms [24]. This can be attributed to the increasing volume of experimental data available in the literature, the emergence of advanced open-source ML models, and improvements in computational power, all of which have facilitated their use. Consequently, these models provide highly accurate predictions, improving efficiency and product quality while offering a cheaper and faster design process [25, 26].

Significant progress has been made in applying surrogate models for predicting target outcomes in metal AM processes [27–32]. A variety of supervised regression algorithms have been leveraged to predict critical aspects of AM technology, including mechanical properties (such as fatigue resistance, hardness, and ultimate tensile strength), as well as physical characteristics like resultant microstructure, relative density and product quality metrics (such as surface roughness, wear rate, and porosity) [33–40]. These algorithms have been applied both individually and in comparative analyses to evaluate their predictive capabilities. For instance, some studies have utilized various ML models to assess the accuracy of predicting the melt pool geometries [41–43]. While existing studies have examined the effectiveness of ML models in predicting target outputs, they do not provide insights on how to select the process parameters needed to achieve desired characteristics and properties. Therefore, it is important to utilize additional methods, such as reinforcement learning (RL), to effectively tackle the optimization challenges mentioned.

RL utilizes computational agents that learn to perform tasks by interacting with their environment through a trial-and-error approach [44], which can be applied to various aspects of the process. A recent systematic review by Sousa et al. [45] examined AI applications for real-time control in laser-based AM, identifying RL as one of the most promising strategies for managing the complex, multiscale, and highly dynamic behaviors inherent to these processes. Furthermore, researchers have been utilizing RL in AM processes to optimize and control process parameters. For instance, a deep deterministic policy gradient agent was developed to control a wire arc additive manufacturing (WAAM) process in a simulator utilizing a reduced-order model of the system [46]. Knaak et al. [47] used high dynamic range optical imaging with convolutional neural networks (CNNs) to predict surface roughness in real time during an L-PBF process. Meta-RL agents utilized the predictions by the CNN model and a dynamic model to extract the optimal process parameters under varying and uncertain conditions. In Ref. [48], a double dueling deep *Q*-network controller was employed to address process instabilities in laser remelting for surface topography modification, aiming to control melt pool dynamics and enhance surface quality. RL agents were adopted

Table 1 Material properties of Ti6Al4V [63]

Parameter	Value
Density/(g·mm^{-3})	0.004 43
Thermal diffusivity /(mm^2·s^{-1})	8.6
Heat capacity /(J· (g · K)$^{-1}$)	0.83
Melting temperature /K	1 923
Ambient temperature /K	293
Absorption coefficient	0.3

Table 2 Laser features and their values

Parameter	Value
Laser power/W	50 – 400
Laser speed /(mm·s^{-1})	200 – 2 250
Laser radius/mm	0.05 – 0.13
Layer thickness/mm	0.02– 0.05

in another work to derive optimal deposition paths for producing parts with complex geometries using WAAM methods [49, 50]. In L-PBF process, Vagenas et al. [51] introduced an RL-based multi-layer process control framework capable of mitigating thermal accumulation across successive layers. The proposed control strategy demonstrated stable convergence during training and outperformed conventional proportional-integral-derivative (PID)-based controllers in maintaining thermal stability throughout 3D builds.

Furthermore, this methodology can be used for multi-objective optimization to determine optimal parameters that meet various complex design criteria. For example, Dharmadhikari et al. [52] employed an RL agent that interacted with a digital twin of the environment to find the optimal process parameters for generating a desired melt pool depth. The digital twin was derived by calibrating the parameters of the Eagar-Tsai equations using experimental data. The off-policy RL framework was based on Q-learning [53], with the reward function reflecting the difference between desired and evaluated depth. Also, Vaghefi et al. [54] employed a soft actor-critic based RL agent [55] to interact with a finite element simulation of the L-PBF process, estimating the optimal process parameters for various component geometries. Additionally, model-based RL, specifically Dyna-Q+, has been applied to improve temperature control in DED. This method outperforms traditional Q-learning by requiring fewer samples and enhancing control and efficiency in evaluating melt pool geometry and temporal characteristics [56]. In another study, Shi et al. [57] developed a deep RL framework using proximal policy optimization to dynamically adjust process parameters, such as laser power, scan speed, and path sequence, based on a temperature simulation model. The approach significantly improved temperature uniformity and reduced mechanical property variability in Inconel 718 components. These contributions collectively underscore the growing role of RL in achieving both offline and in-situ, data-driven process control and optimization in metal AM.

While many studies adopt RL models to optimize the AM processes, a key innovation of the current study lies in the integration of an ML surrogate model with an RL framework, where the ML model serves as the environment for the RL agent. This unified, fully data-driven approach advances beyond prior works, which typically treat ML-based prediction and RL-based optimization in isolation. This eliminates dependency on physics-based simulations and enables dynamic exploration of parameter combinations in a digital twin of the L-PBF process. The RL agent utilizes a customizable reward function to balance conflicting design objectives, overcoming the rigidity of traditional multi-objective methods like weighted-sum optimization. In addition, the proposed framework is designed to be data-efficient, achieving high accuracy in training, testing, and estimation tasks using a minimal dataset curated from existing literature. This is particularly significant in the context of AM, where high production costs often limit the feasibility of large-scale data collection. Consequently, a framework has been developed in this study to streamline the process of deriving optimal parameters that satisfy complex design criteria. Firstly, a small sample space of experimental data was extracted and preprocessed from available literatures. Afterwards, various ML models were analyzed and trained on the extracted datasets, and the model with the best prediction performance was chosen to represent the process on the L-PBF machine as its digital twin. Lastly, a model-free Q-learning based RL agent was developed to utilize the surrogate model for simulating the process and derive the value of all possible state-action pairs in a discrete space based on a reward function tailored to achieving complex design criteria.

2 Methodology

2.1 Data extraction

The experimental data were extracted from Refs. [59–62], considering only the L-PBF process. The figures and tables in the mentioned papers were used to extract the data by employing plot digitizing tools. The composition of interest in this study was Ti6Al4V [58], with the characteristics listed in Table 1 [63]. The extracted dataset contains 176 data points with the process parameters consisting of laser power (P), laser speed (V), laser radius (R), and layer thickness (T) (see Table 2). The material properties were not considered as process parameters, as they were assumed to remain constant and independent of temperature variations. The target property was the melt pool depth (d), which was the distance between the deposited layer surface and the lowest point of the melt pool.

2.2 Feature selection

Feature selection is a critical step in the ML process that determines the most relevant parameters and features of the dataset for developing the ML model. It helps reduce the risk of overfitting, enhances model generalization, and increases computational efficiency.

One tool used in this process is SHapley Additive exPlanations (SHAP), a game-theoretic approach that assigns an importance value to each feature representing its contribution to the model-predicted output. Features with positive values positively influence the prediction, whereas those with negative values do the opposite. The magnitude of these values indicates their importance. A feature selection process was conducted for each of the designed ML models using the training dataset.

2.3 Dataset splitting

Cross-validation (CV) is a popular method for estimating model prediction errors. One simple approach is the hold-out method, which splits the dataset into training and validation parts. Despite its simplicity, the hold-out method has some drawbacks. Firstly, it provides only one estimate of the model's accuracy. Secondly, the test set, which contains useful information for training, is ignored.

In this work, k-fold cross-validation is used due to its effectiveness and robustness. The dataset is randomly divided into k parts. One of the k parts is used for the testing, while the remaining k–1 parts are used for training the model. The benefit of this approach is that the process can be repeated several times with new random splits of the data, allowing for the creation of a histogram of the errors. The value of k typically ranges from 5 to 10 in practical machine-learning scenarios.

2.4 ML models

Several criteria can be used to evaluate model performance, including root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

RMSE quantifies the average magnitude of the prediction errors by measuring the differences between the observed target y_i and the prediction $\widehat{y}_i$ of the trained model for each sample (x_i, y_i)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}, \tag{1}$$

where n is the number of data points.

The MAE metric treats all errors equally and is less sensitive to outliers.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|. \tag{2}$$

MAPE metric calculates the average absolute percentage error between predicted and actual values and is easier to interpret than MAE.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{3}$$

The R^2 value is a statistical tool that represents the portion of the variance in the dependent variable that is explained by the independent variables and ranges from 0 to 1, where a value closer to 1 indicates a stronger fit. It is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}, \tag{4}$$

where $\overline{y}_i$ is the mean of the observed target. The numerator represents the sum of the squared of residuals of the model, and the denominator represents the total sum of squares, which measures the total variance in the actual data.

In this study, various regression models were implemented to identify the model that achieved the lowest RMSE, MAE, and MAPE values while obtaining the highest R^2 score. The models described below have been analyzed, and their features are summarized.

2.4.1 Support vector regression (SVR)

SVR is a type of support vector machine algorithm that is used for regression problems [64]. The objective of SVR is to find a hyperplane with the minimum distance from the data points in an N-dimensional space, where N is the number of features. The first step is to choose a kernel, which is a function that transforms the data into a higher-dimensional space, making it possible to fit complex non-linear models. The most common types of kernels are linear, polynomial, sigmoid, and radial basis function (RBF). In this study, an SVR model with an RBF kernel was employed due to its ability to effectively manage high-dimensional data and its robustness against outliers.

2.4.2 Gaussian process regression (GPR)

GPR provides a probabilistic prediction of continuous values using the Bayesian approach [65]. Two of the most common covariance functions are the RBF and the Matérn kernel. The RBF kernel excels at capturing smooth, stationary trends due to its infinite differentiability, making it ideal for modeling continuous, noise-free relationships. In contrast, the Matérn kernel generalizes the RBF by introducing a tunable smoothness parameter (ν), enabling greater flexibility for datasets with irregular or moderately noisy variations.

2.4.3 Random forest regression

Random forest is a type of bagging ensemble learning method where several decision tree regressors are trained on a random subset of training data with replacement [66]. These individual trees generate multiple predictions, which are then averaged to achieve the final prediction. This method reduces the risk of overfitting and increases robustness.

2.4.4 Gradient boosting regression (GBR)

GBR is a type of boosting ensemble learning technique that combines multiple weak learners (often shallow decision trees) to form a stronger predictor [67]. Each new weak learner is trained to correct the errors of the combined ensemble of all previous learners. While this method is accurate and flexible, it is vulnerable to overfitting.

2.4.5 XGBoost regression

XGBoost regression is an optimized version of the gradient boosting algorithm, offering a good balance between bias and variance by introducing additional features [68]. The advantages of using this method include built-in cross-validation capability, efficient handling of missing data, regularization to avoid overfitting, cache awareness, tree pruning, and parallelized tree building.

2.5 RL-based optimization framework

RL is a computational method that relies on an agent learning through direct interaction with its environment to maximize some notion of cumulative reward. The problem is often modeled as a Markov decision process that defines the interactions between the learning agent and its environment via states (s), actions (a), and rewards (r). The agent takes actions a in states s, either by following the current policy $\pi(a|s)$ (on-policy), or by exploring to find the optimal policy (off-policy). The resulting state is obtained from the environment and used in the reward function to compute a scalar reward associated with the state action pair. The resulting reward can be used to optimize the policy either directly (policy-based) or indirectly (value-based).

Policy-based methods focus on directly optimizing the policy, which is a probability distribution over actions given a state that can be stochastic or deterministic.

$$\pi(a|s) = P\left(A_t = a, S_t = s\right), \tag{5}$$

where S_t and A_t are the state and the action taken at time t, respectively.

On the other hand, value-based methods focus on estimating the value functions, which are used to derive the optimal policy, by determining the value of being in a particular state ($V^\pi(s)$) or taking a specific action in that state ($Q^\pi(s,a)$).

$$V^\pi(s) = \mathrm{E}^\pi\left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_t = s\right], \tag{6}$$

$$Q^\pi(s,a) = \mathrm{E}^\pi\left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_t = s, A_t = a\right], \tag{7}$$

where γ is the discount factor, and R_{t+1} is the reward received at time $t+1$.

In this study, a Q-learning-based agent is used as an off-policy, model-free, and value-based RL algorithm in which the agent learns the optimal actions in Markovian domains through direct experience without needing to construct a detailed representation of the environment.

The agent is represented by the laser of the L-PBF machine, while the environment consists of the powder in contact with the laser. The interactions between the laser and the powder are governed by thermodynamics and material properties, while the reward function defines the desired design criteria for the final component.

The agent operates in a finite discrete domain state space S, where each state represents a unique combination of discretized process parameter values. The continuous range of each process parameter, extracted from experimental data (see Table 2), is uniformly divided into n discrete intervals. The state space is then constructed as the Cartesian product of all discretized parameter ranges, resulting in n^N possible states.

The action space A consists of all valid single-step adjustments to the process parameters. At each state, the agent can increase, decrease, or maintain the current value of each process parameter, leading to a total of 3^N possible actions. Consequently, the Q-table which is used to store the Q-value associated with each state-action pair is defined by an $\overbrace{n \times \cdots \times n}^{N} \times 3^N$ matrix.

At each step, the agent is either penalized or rewarded based on multiple design criteria through the following reward function

$$r = r_{\mathrm{depth}_1}\left(\delta_{\mathrm{err}}\right) + r_{\mathrm{depth}_2}\left(\delta_{\mathrm{err}}\right) + r_{\mathrm{keyhole}}\left(\frac{\Delta H}{h_{\mathrm{s}}}\right), \delta_{\mathrm{err}} \geq \delta_{\mathrm{tol}}, \tag{8}$$

$$r = r_{\mathrm{depth}_1}\left(\delta_{\mathrm{err}}\right) + r_{\mathrm{keyhole}}\left(\frac{\Delta H}{h_{\mathrm{s}}}\right),\ \delta_{\mathrm{err}} < \delta_{\mathrm{tol}}, \tag{9}$$

where r_{depth_1} and r_{depth_2} are utilized to achieve the desired melt pool depth, while δ_{err} quantifies the difference between the desired depth and the achieved one. h_{s} is the enthalpy at melting. The tolerance parameter δ_{tol} represents the acceptable margin of error, which is set to 0.03 mm. Furthermore, r_{keyhole} is used to ensure that the melt pool dynamics do not enter the unstable keyhole regime, based on the criteria defined by Behjat et al. [13], where $\frac{\Delta H}{h_{\mathrm{s}}}$ is the normalized enthalpy represented as

$$\frac{\Delta H}{h_{\mathrm{s}}} = \frac{AP}{\pi h_{\mathrm{s}} \sqrt{DVR^3}}, \tag{10}$$

Table 3 Coefficients of the Sigmoid function used for each criterion

Criteria	K_θ	a_θ	b_θ	c_θ
r_{depth_1}	200	–6	δ_{tol}	6
r_{depth_2}	–160	10	δ_{tol}	5
r_{keyhole}	–100	2	6	5

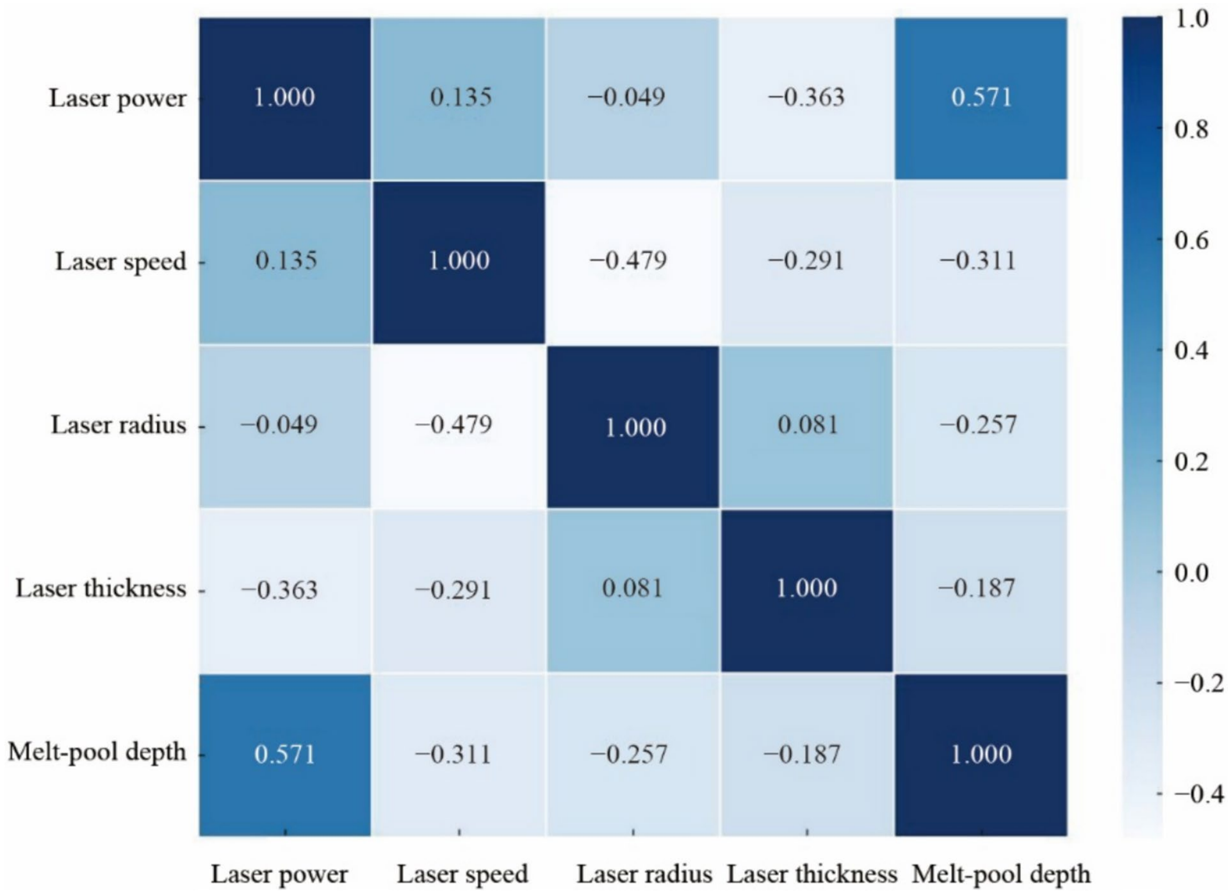


Fig. 1 Heat map plot of the correlation coefficients

where A is the absorption coefficient, and D is the thermal diffusivity.

The reward functions r_θ are defined using the continuous and differentiable sigmoid function σ_θ to facilitate smoother convergence during the training phase, which is expressed as follows

$$r_\theta\left(\delta_{\mathrm{err}}\right) = \sigma_\theta\left(\delta_{\mathrm{err}}\right) = K_\theta \mathrm{e}^{\frac{c_\theta}{\left(1+\mathrm{e}^{-a_\theta\left(\delta_{\mathrm{err}}-b_\theta\right)}\right)}}, \tag{11}$$

where K_θ, c_θ, a_θ and b_θ are the coefficients that determine the shape of the reward function with their values listed in Table 3. These coefficients were derived through a combination of trial-and-error and heuristic methods to achieve

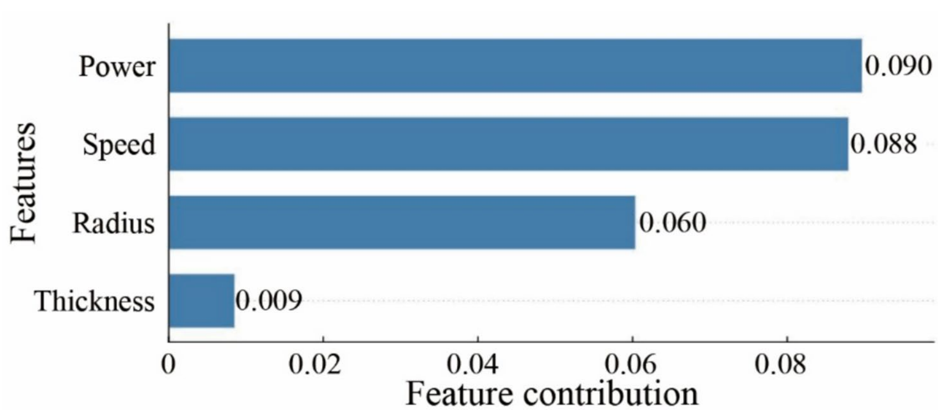


Fig. 2 SHAP featuring an important plot

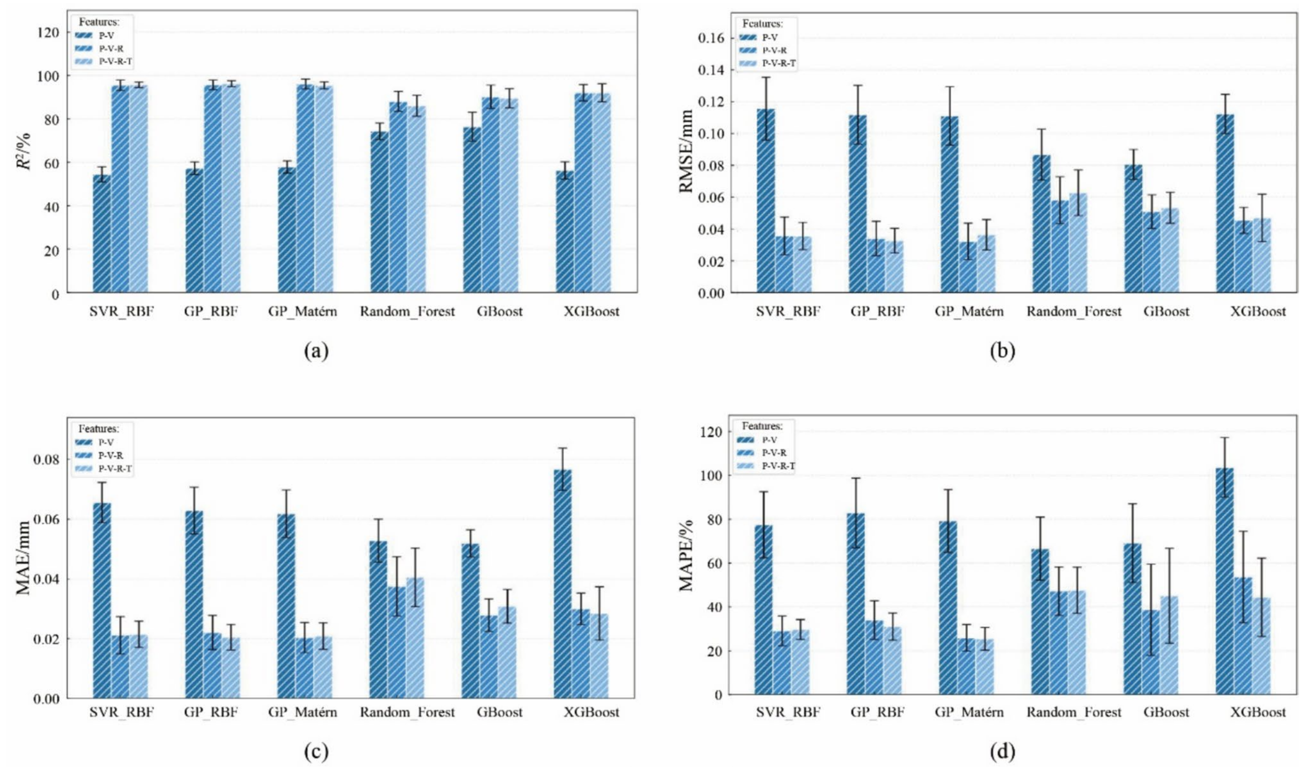


Fig. 3 Performance comparison of various ML models based on **a** R^2, **b** RMSE, **c** MAE, **d** MAPE criteria for different feature sets, where the bar plots show the average and the variation in the value

multiple design targets. The use of various sigmoid functions shapes the reward function to ensure that the agent is not penalized when the error remains within an acceptable tolerance range to allow for smoother convergence. Additionally, the coefficients are carefully tuned to balance the effort between minimizing melt pool depth error and addressing the keyhole formation criteria, without saturating the reward or allowing one objective to dominate at the expense of the other.

Table 4 Optimized hyperparameters for the ML models

Models	Optimized hyperparameters
GPR with RBF kernel	$\alpha = 7.41 \times 10^{-4}, l = 1\,346.3$
GPR with Matérn kernel	$\alpha = 4 \times 10^{-4}, \nu = 2.5, l = 1 \times 10^{-5}$
SVR with RBF kernel	$C = 10, \epsilon = 0.01, \gamma = 0.1$
Random forest	$\min_{\text{Samples\_Leaf}} = 1, \min_{\text{Samples\_Split}} = 2, n_{\text{estimators}} = 175, \max_{\text{depth}} = 10$
GBoost	$\min_{\text{Samples\_Leaf}} = 4, \min_{\text{Samples\_Split}} = 5, \max_{\text{depth}} = \text{none}$
	$n_{\text{estimators}} = 175, \text{learning\_rate} = 0.4, \text{subsample} = 0.8$
XGBoost	$\min_{\text{child\_weight}} = 3, \max_{\text{depth}} = 2, \text{colsample\_bytree} = 0.7, n_{\text{estimators}} = 175, \text{learning\_rate} = 0.4, \text{subsample} = 0.7$

After calculating the reward at each step, the Q-values are updated using the Q-learning update rule, which is derived from the Bellman optimality equation.

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(r + \gamma \max_{\acute{a}} Q(\acute{s},\acute{a}) - Q(s,a)\right), \tag{12}$$

where $Q(s,a)$ denotes the action-value function, representing the expected cumulative discounted reward associated with taking actions a in states s. In addition, $\acute{s}$ denotes the next state reached after executing action a, and $\acute{a}$ represents a candidate action from the action space. Meanwhile, α represents the learning rate and controls how much the agent updates its knowledge after receiving new information, and γ determines the importance of future rewards compared to immediate ones.

The training proceeds for a predetermined number of episodes N, where each episode represents a complete trial beginning from a randomized initial state and ending when the melt-pool depth enters the target range or the maximum step limit is reached. In each episode, training continues until one of the termination conditions is met, at which point the episode count increments and both the environment and the agent's internal state are reset to a new random configuration. Once the episode counter surpasses a defined limit,

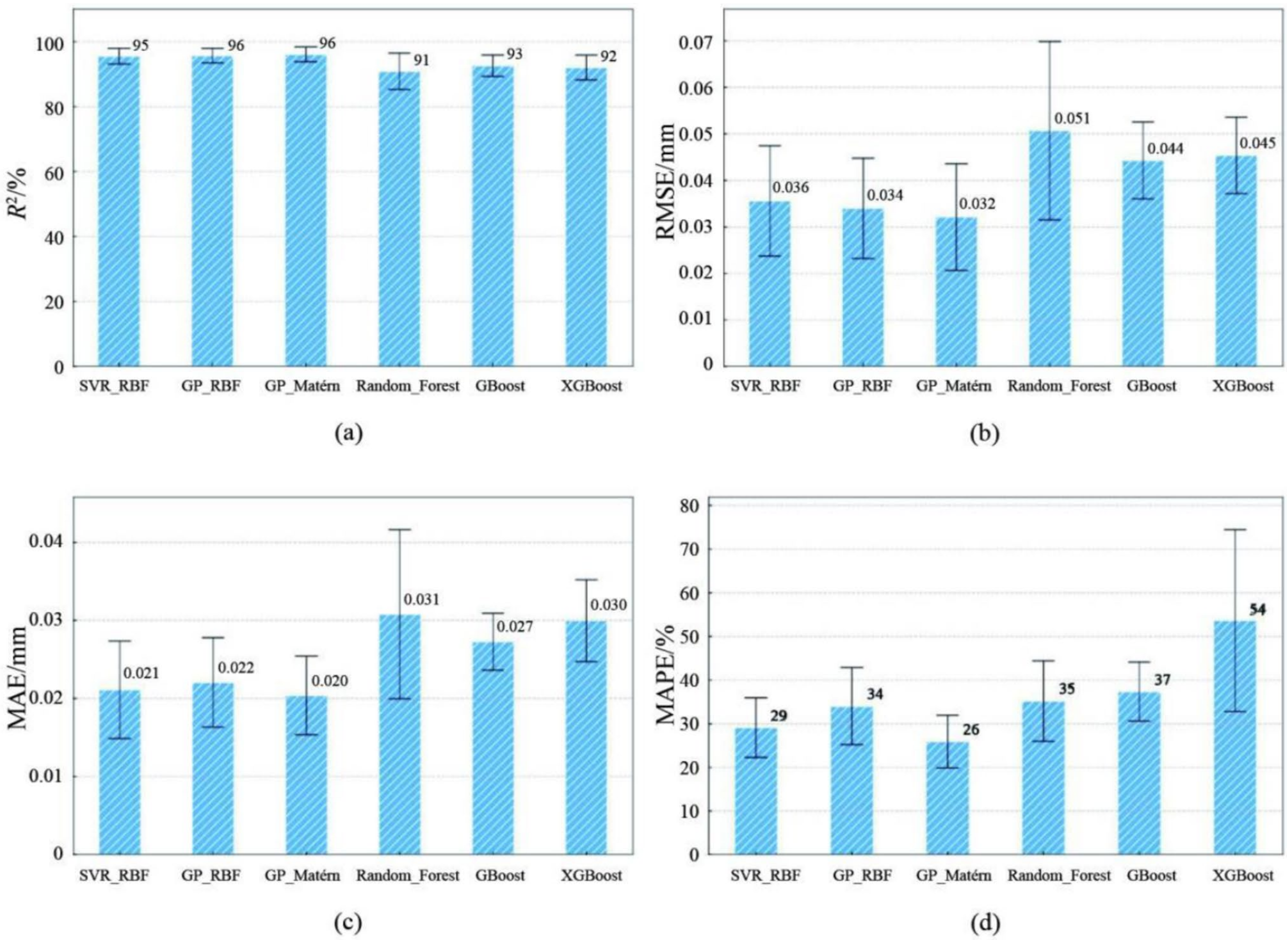


Fig. 4 Performance comparison of various ML models based on **a** R^2, **b** RMSE, **c** MAE, and **d** MAPE criteria for a train-test split coefficient of 25% and the P-V-R feature set

the final Q table is extracted, which contains the learned value estimates for every visited state-action pair.

To identify the optimal process parameters, the Q table is filtered to include only those entries corresponding to actions that leave every parameter unchanged, ensuring that the selected process parameters reflect stable operation rather than ongoing exploration. The state with the highest Q value is chosen as the optimal set of process parameters, balancing target melt-pool depth accuracy with the keyhole formation criteria.

Since Q-learning is an off-policy learning model, it learns the optimal policy by following a non-optimal policy during the initial portion of the learning steps. This allows the agent to explore different state-action pairs to fully explore the domain. Therefore, it is important to strike a balance between exploration and exploitation to guarantee convergence while preventing the agent from getting stuck in a local optimum.

In this study, the epsilon-greedy strategy was employed to achieve this goal, where the agent selects actions randomly with a probability of ϵ and chooses the action with the highest Q-value with a probability of $1-\epsilon$ at each state. To allow for increased exploration during initial training phases and more exploitation toward the conclusion, the parameter ϵ varies linearly with the episode number.

The hyperparameters of this model are composed of the exploration-exploitation trade-off factor ϵ, learning rate α, discount factor γ, discretization of the state space n, number of episodes N, and the maximum allowable epoch N_{epoch} and their effects on convergence and accuracy are analyzed.

3 Results and discussion

In the following section, data pre-processing and feature selection are performed on the collected datasets before being utilized to train various predictor models with optimized hyperparameters. The best-performing model is determined and used as a digital twin of the process, serving as the foundation for training the RL agent and extracting optimal process parameters. The data points are imported as DataFrames, and features are normalized to ensure that they

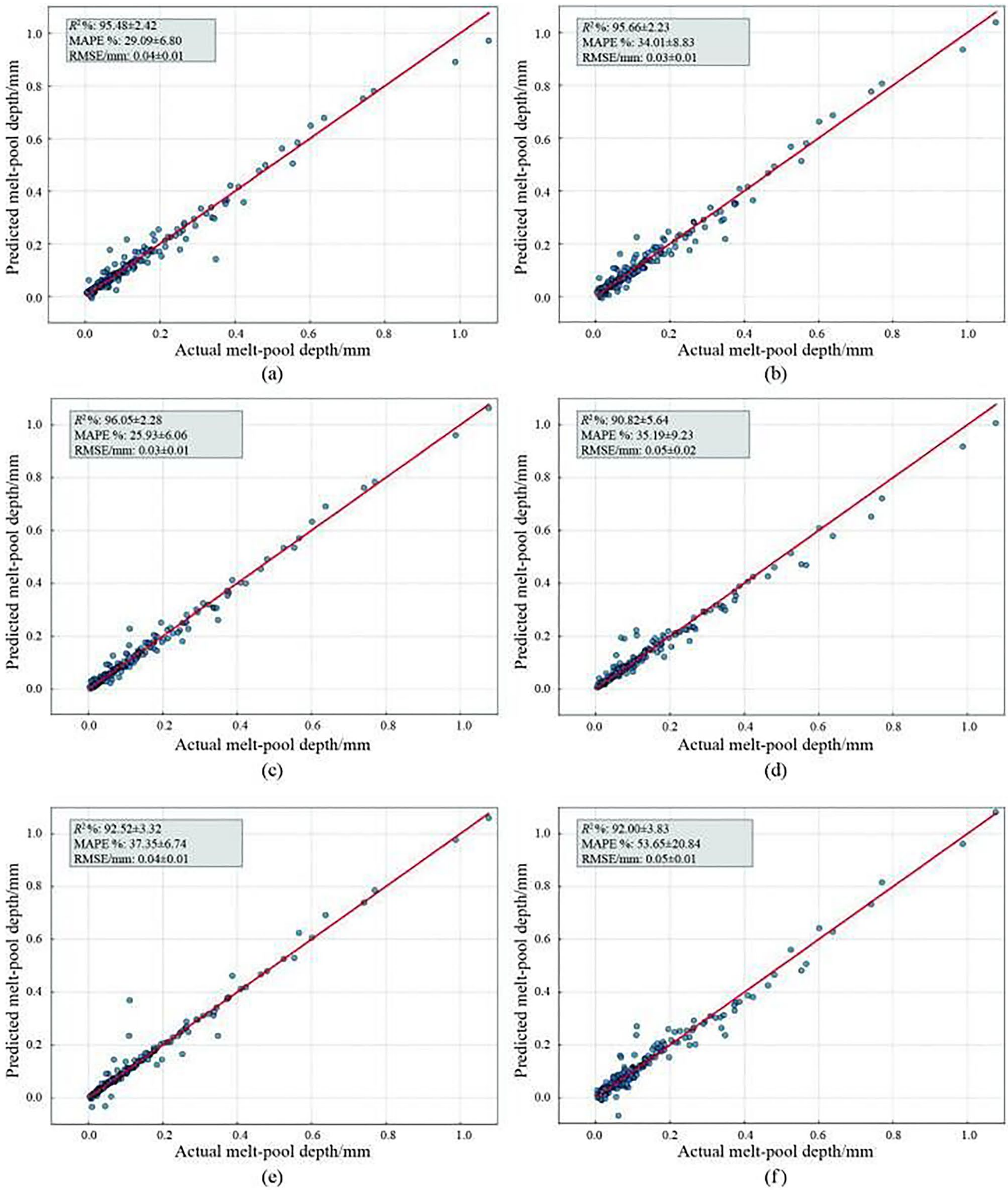


Fig. 5 Parity plots of **a** SVR_RBF, **b** GP_RBF, **c** GP_Matérn, **d** Random_Forest, **e** GBoost and **f** XGBoost

are centered around zero with unit variance. This allows for the comparison of the feature importance on a similar scale and faster convergence in training the models. All developed models and agents were trained and simulated in Python utilizing libraries such as scikit-learn [69], pandas, and OpenAI/gym [70] , and they were visualized using Seaborn and matplotlib.

Table 5 *Q*-learning agent training hyperparameters and the range of values considered in the analysis

Parameter	Default value	Range
Learning rate, α	0.35	[0.25, 0.30, 0.35, 0.40]
Discount factor, γ	0.95	[0.85, 0.90, 0.95, 0.99]
Exploration-exploitation trade-off factor, ϵ	0.4	[0.2, 0.4, 0.6, 0.8]
Discretization size, n	10	[5, 10, 15, 20]
Number of episodes, N	3 000	3 000
Max steps per episode, N_{epoch}	150	150

3.1 Feature selection

To improve the efficiency and accuracy of the prediction models, it is important to analyze and remove irrelevant and redundant data from the datasets. This can be achieved by selecting the most important process parameters as features, which reduces the training time and simplifies the models.

A correlation matrix, as shown in Fig. 1, can be constructed by statistically evaluating correlation coefficients between the feature variables. Positive and negative values indicate direct and inverse correlations, respectively, and the maximum magnitude can reach unity. It can be noticed that the laser power has the highest impact on melt pool depth with a coefficient of 0.57, while layer thickness has the lowest with a coefficient of –0.19.

SHAP values can be utilized to understand each feature's influence on the model predictions. Figure 2 illustrates the SHAP values for all process parameters, revealing that layer thickness minimally impacts the prediction, while laser speed and laser power show the highest contribution.

To further illustrate the impact of feature selection on model training, the performance of the trained models was analyzed using the criteria introduced in the machine learning section. Three distinct sets of input features—P-V, P-V-R, and P-V-R-T—were selected, and a train-test split ratio of 25% with 5-fold cross-validation was used for this analysis. As shown in Fig. 3, the poorest performance occurs when the features consist solely of laser power and laser speed. Adding laser radius to the feature set results in a noticeable improvement in performance across all criteria, particularly for the GPR models. Finally, incorporating layer thickness slightly enhances the prediction accuracy of ensemble methods, though it has no positive effect on the other models.

Based on these observations, the P-V-R feature set is identified as the optimal solution in the feature selection process and used to derive the final prediction model.

3.2 Optimal ML model fitting and selection

In this section, various machine learning models were trained using the pre-processed data and selected feature sets. To ensure a fair comparison, hyperparameters were optimized for all models. Bayesian optimization was applied to GPR models, while the remaining models were optimized using grid search.

Grid search exhaustively evaluates all combinations of a predefined set of hyperparameters, offering simplicity but at a high computational cost. In contrast, Bayesian optimization uses probabilistic models to guide hyperparameter selection, balancing exploration and exploitation by constructing a surrogate model to predict model performance. The optimized hyperparameters are shown in Table 4.

Once hyperparameters were optimized, the models were evaluated using 5-fold cross-validation with a 25% train-test split. The accuracy of each model was recorded, and the average and standard deviation of each criterion were

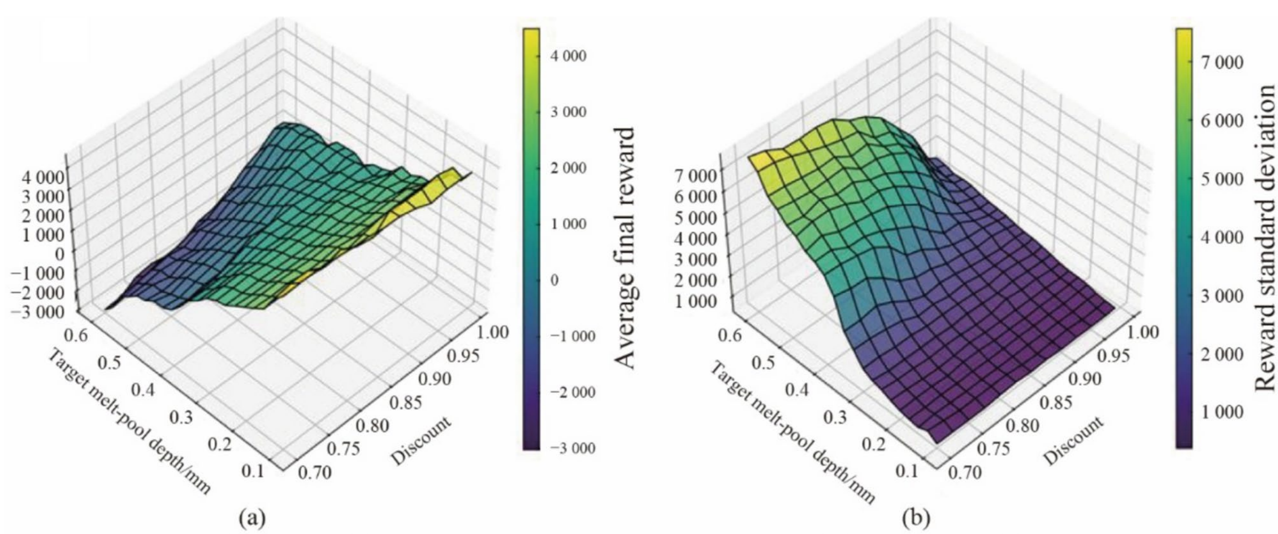


Fig. 6 Effects of the discount factor on **a** the accumulated final reward and **b** its standard deviation for various targets

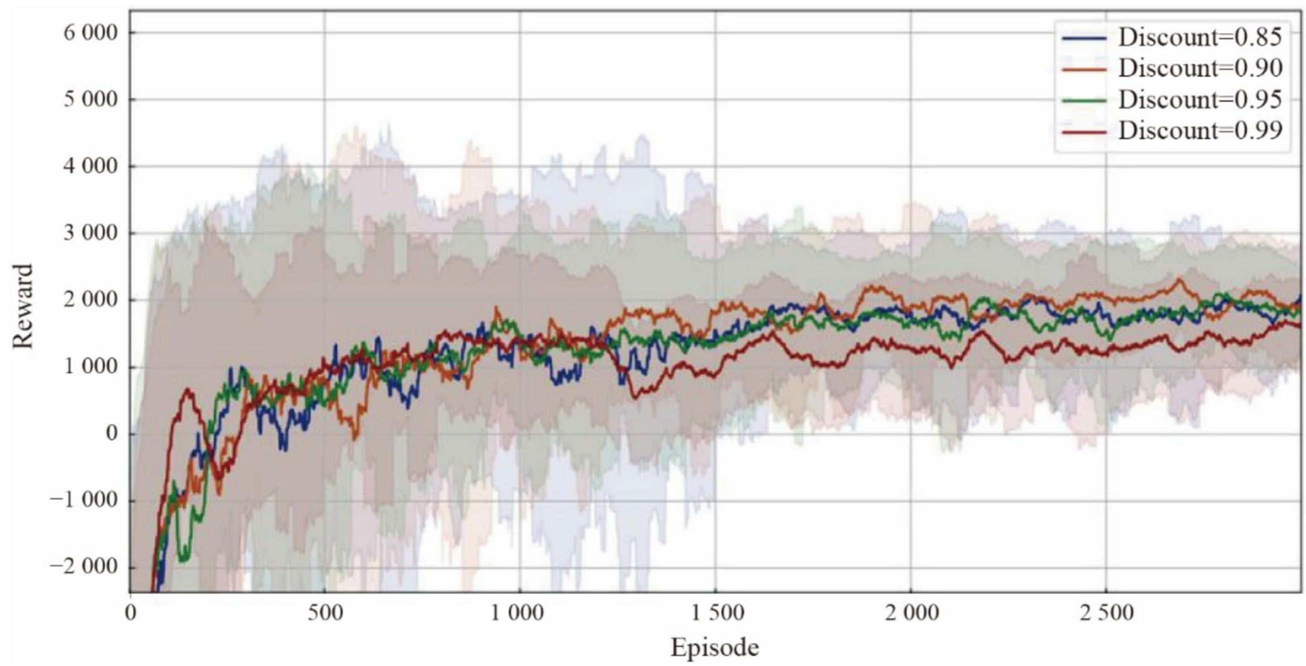


Fig. 7 Effects of the discount factor on the accumulated final reward for a target melt pool depth of 0.3 mm

reported, as shown in Figs. 4, 5. While all the models demonstrate acceptable performances, it is important to choose the model with the highest accuracy that also satisfies generalization requirements. Therefore, both the average values and the standard deviations are considered.

The R^2 values show that the GPR model with a Matérn kernel achieved the highest average generalization ability (96.05%), and a low standard deviation (2.28%). Moreover, the GPR model with an RBF kernel ranks second, with an average of 95.66% and a standard deviation of 2.23%. Random forest, however, showed the lowest R^2 with an average of 90.82% and a standard deviation of 5.64%.

In terms of RMSE, the GPR model with the Matérn kernel had the lowest average error of 30 μm and a standard deviation of 10 μm. Conversely, random forest exhibited the highest average RMSE of 50 μm, with a standard deviation of 20 μm.

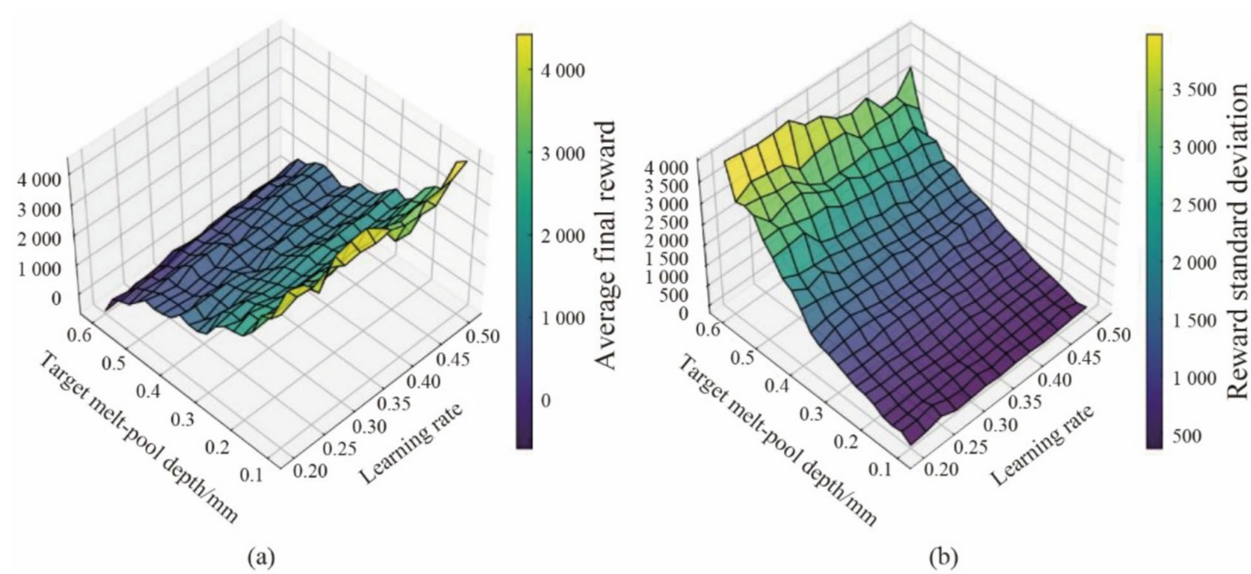


Fig. 8 Effects of the learning rate on **a** the accumulated final reward and **b** its standard deviation for various targets

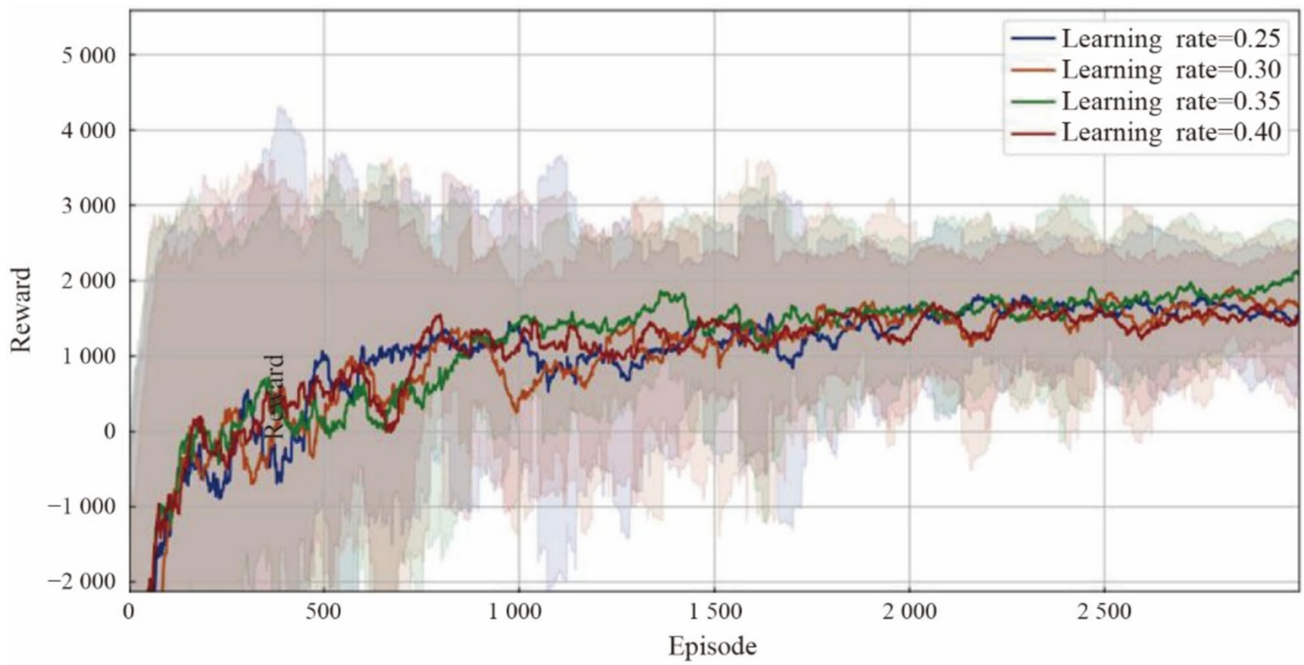


Fig. 9 Effects of the learning rate on the accumulated final reward for a target melt pool depth of 0.3 mm

The MAPE results aligned with the previous findings, where GPR with the Matérn kernel achieved the lowest average MAPE of 25.92% and a standard deviation of 6.06%, while XGBoost recorded the highest average MAPE at 53.64% and a standard deviation of 20.84%.

Based on the obtained results, the GPR model with the Matérn kernel demonstrated the best accuracy and generalization performance and was selected as the predictor model to assess the impact of these models on training an RL agent and its capability in extracting the optimal process parameters.

3.3 RL agent training and process parameter optimization

3.3.1 Extracting optimal hyperparameters of the agent

The selected GPR model was utilized to train the *Q*-learning-based RL agents for various target melt pool depths and normalized enthalpy threshold requirements. A performance analysis was employed to extract the optimal hyperparameters on a range of possible hyperparameter values. Table 5 lists the hyperparameters used for the GPR model along with their value range used in the search algorithm.

The learning rate and discount factor parameters were chosen based on three custom performance criteria, which were calculated after applying a moving average filter with a window size of 75 episodes to smooth the accumulated rewards. These criteria included the average reward of the final 1 000 episodes, which was used to represent the model's competence in satisfying the reward function requirements. The second proposed criterion is the convergence episode, defined as the episode number after which the accumulated reward falls within a 5% range of the average final reward and is used as a reference for comparing the training time. The final criterion is reward stability, which measures the average standard deviation of the accumulated reward in the final 1 000 episodes of the training process. The values for these three criteria were obtained for various discount factors and learning rates, and the results are depicted in Figs. 6, 7, 8, 9 and 10.

Based on the results illustrated in Fig. 10, a discount factor of 0.95 yields the best results across all three efficiency criteria. However, the same conclusion cannot be made when choosing the learning rate parameter. Specifically, as shown in Fig. 10, a learning rate of 0.4 yielded the highest accumulated reward and fastest convergence speed, but a value of 0.35 was chosen as the optimal parameter because it also ensured adequate reward stability, striking a balance in performance.

While the proposed criteria offer valuable insights into selecting the learning rate and discount factor, they are not suitable for determining other parameters, as these criteria fail to account for all the factors influenced by their selection. For instance, as shown by the *Q*-tables in Fig. 11, a higher exploration-exploitation coefficient (ϵ) encourages a broader search. This enables a more comprehensive analysis of the state-action space to avoid sub-optimal outcomes and convergence to local optima.

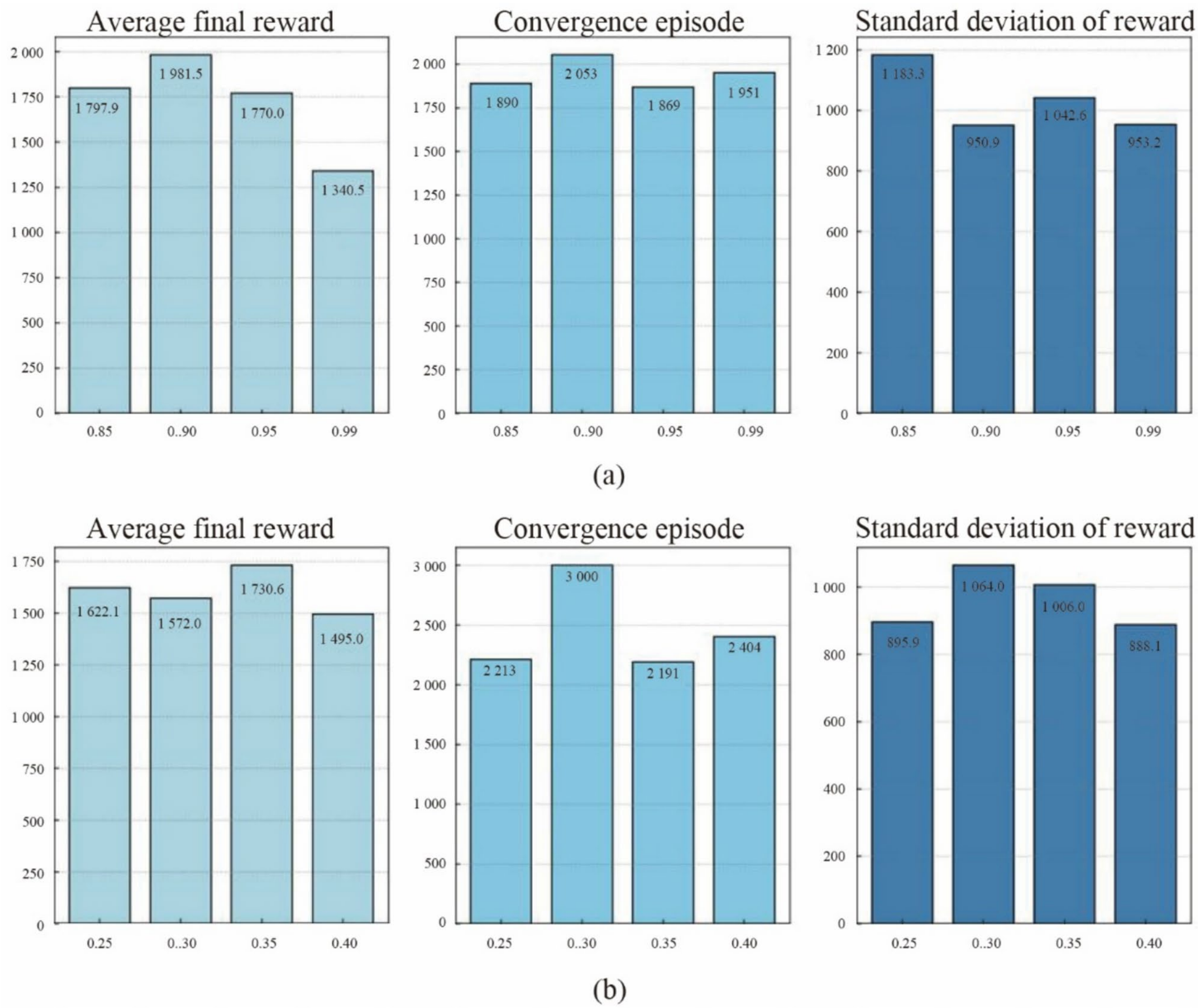


Fig.10 Comparison of the effects of **a** the discount factor and **b** the learning rate on the performance criteria for a target melt pool depth of 0.3 mm

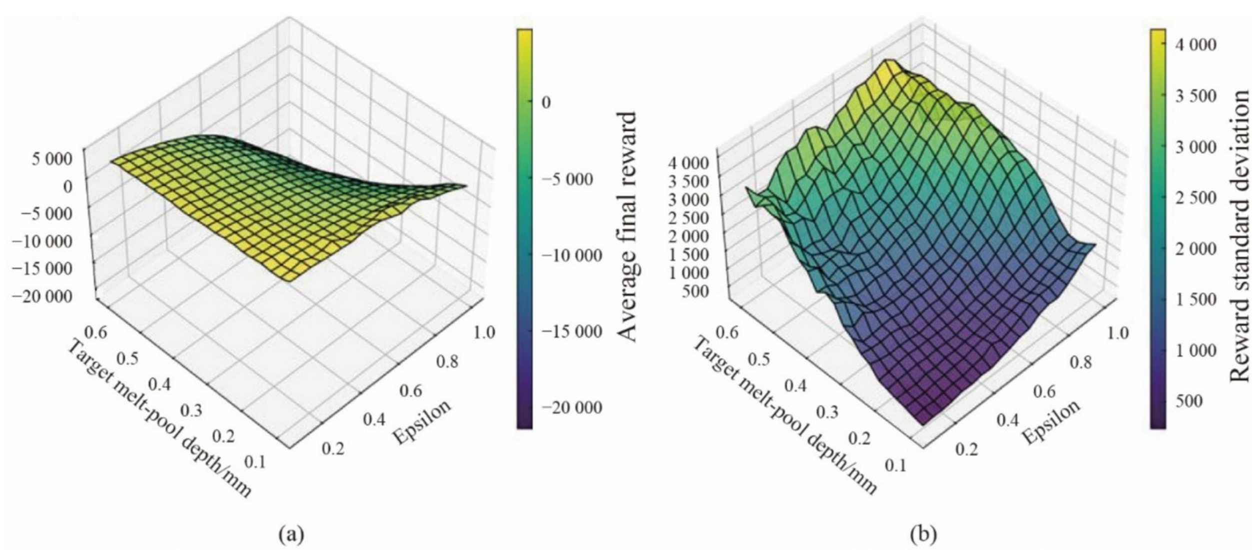


Fig.11 Effects of the exploration-exploitation coefficient on **a** the accumulated final reward and **b** its standard deviation for various targets

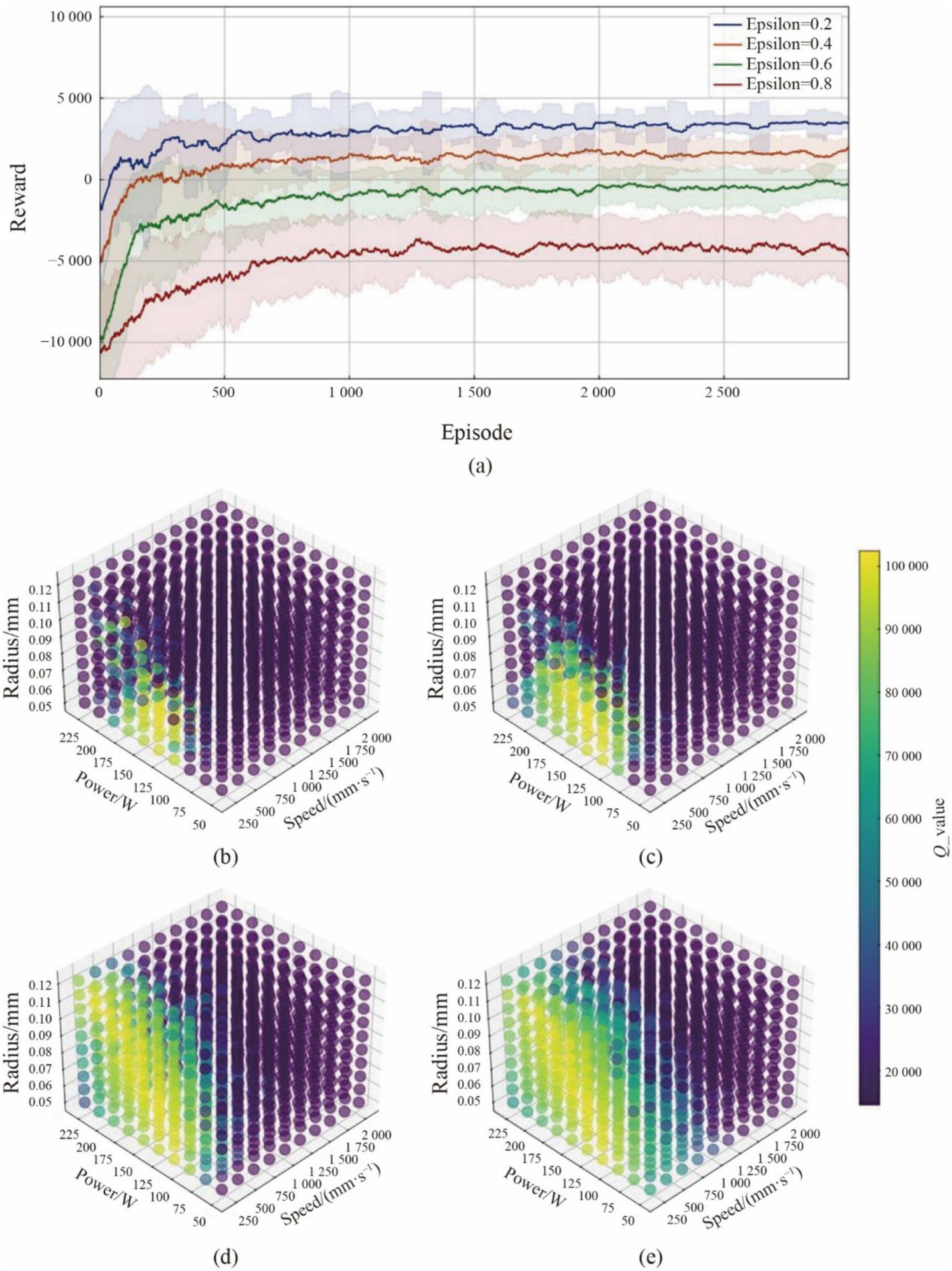


Fig. 12 Effects of the exploration coefficient on the **a** accumulated reward and Q-tables where ϵ is **b** 0.2, **c** 0.4, **d** 0.6, and **e** 0.8 with a target melt pool depth of 0.3 mm

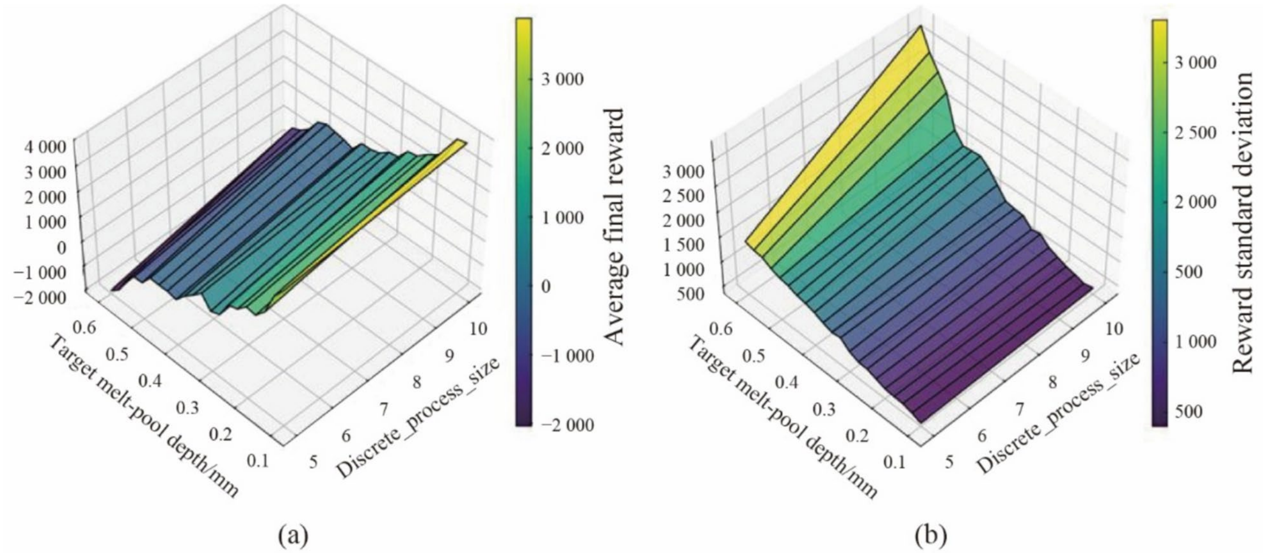


Fig.13 Effects of the discretization size on **a** the accumulated final reward and **b** its standard deviation for various targets

However, a smaller coefficient prompts the model to focus on regions with higher value estimates, resulting in larger accumulated rewards and faster convergence, as shown in the surface plots in Fig. 12. Therefore, a value of 0.4 was chosen to achieve a large reward accumulation while also exploring a sufficient portion of the state-action space.

The previous criteria are also not applicable for selecting the discretization size (n), as many coefficient values result in the model failing to converge within the predefined number of episodes. A discretization size of 10 was chosen because it led to a higher accumulated reward compared to a size of 5 while still ensuring convergence within the specified number of episodes, as shown in Figs. 13, 14.

3.3.2 *Effect of the keyhole criteria*

The performances of the RL agent with and without the keyhole criterion were compared to illustrate the criterion's effectiveness in encouraging the model to avoid choosing parameters that could generate unstable keyhole regions. The output of the RL agent consisted of the estimated process parameters for the LPBF process aimed at achieving a specific melt pool depth. These parameters were then used in the ML model, GPR model with the Matérn kernel, to assess whether the resultant RL outputs accurately produced the desired melt pool depths.

As shown in Fig. 15, it is evident that the RL agent can achieve acceptable accuracy up to target melt pool depths of 0.6 mm with both criteria. However, according to Fig. 16, the model that incorporates the keyhole criteria into its reward function outperforms the alternative by selecting process parameters that not only achieve accurate target melt pool depth generation but also result in reduced normalized enthalpy. This demonstrates the framework's effectiveness in simultaneously addressing multiple objectives within the manufacturing process. Besides, Fig. 16 suggests that in certain cases, two different sets of process parameters may lead to the same melt pool depth but different normalized enthalpy values, casting doubt over the accuracy of the criteria for following the melt pool depth trend. This can be explained by the fact that normalized enthalpy Eq. (10) relies on the ratio of the influential process parameters, P, V, and R, which in turn may fail to adequately capture the actual dynamics of the melt pool under the two different processing conditions.

4 Conclusions

This study has developed a novel data-driven framework combining ML and RL methods to optimize process parameters for the L-PBF manufacturing process to achieve a desired melt pool depth. The following remarks can be made.

(i) The findings demonstrated the feasibility and advantages of the adopted multi-objective data-driven approach for process parameter optimization over traditional methods such as DOE and pure simulation-based techniques.
(ii) Relying on the experimental data extracted from the published literature on LPBF Ti6Al4V alloy, the GPR model with a Matérn kernel was identified as the optimal surrogate model, offering high prediction accuracy with minimal training data. By fine-tuning the GPR with a Matérn kernel, a low average RMSE of 34 μm (standard deviation 10.8 μm) and an R^2 of 96.05%

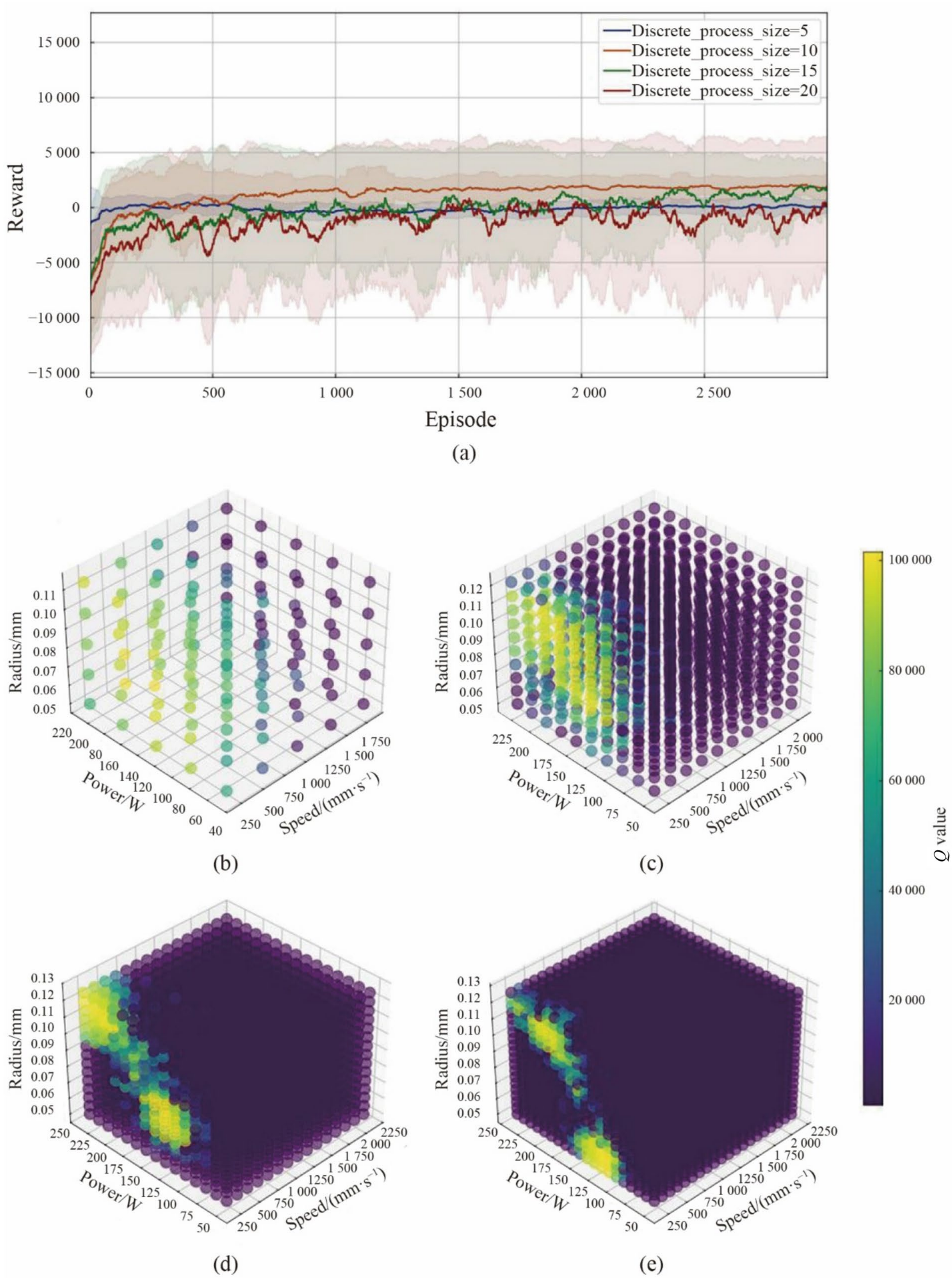


Fig. 14 Effects of the discretization size on **a** the accumulated final reward, and the Q-tables where n is **b** 5, **c** 10, **d** 15, and **e** 20 for a target melt pool depth of 0.3 mm

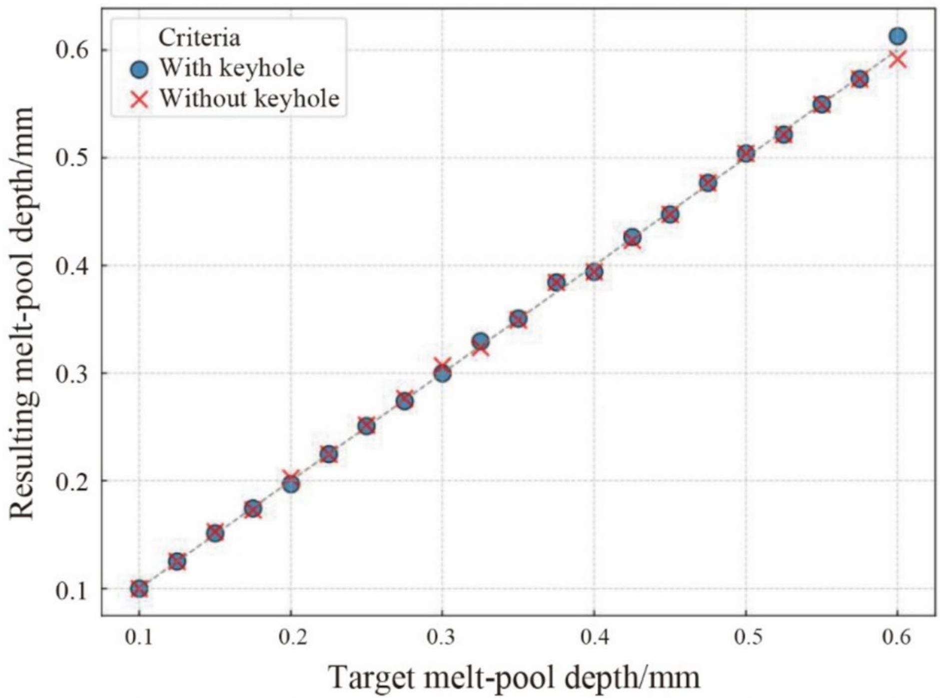


Fig. 15 Effect of the keyhole criteria on the accuracy of predicting the target melt pool depth

(standard deviation 2.28%) were achieved, confirming both high accuracy and generalization capabilities.

(iii) The *Q*-learning framework was successfully integrated with a trained ML model to obtain optimal process parameters. The RL model converged to an optimal solution that not only matched the desired melt pool depth but also maintained the normalized enthalpy below the critical threshold and prevented keyhole defects. It was further suggested that in some instances, different combinations of process parameters could achieve the same melt pool depth while producing varying normalized enthalpy values, raising concern over the reliability of the RL agent, as the only measure for evaluation, to track the melt pool depth variations.

The proposed framework offers several practical advantages. Firstly, by identifying and training a data-efficient ML surrogate model using fewer than 200 literature-sourced

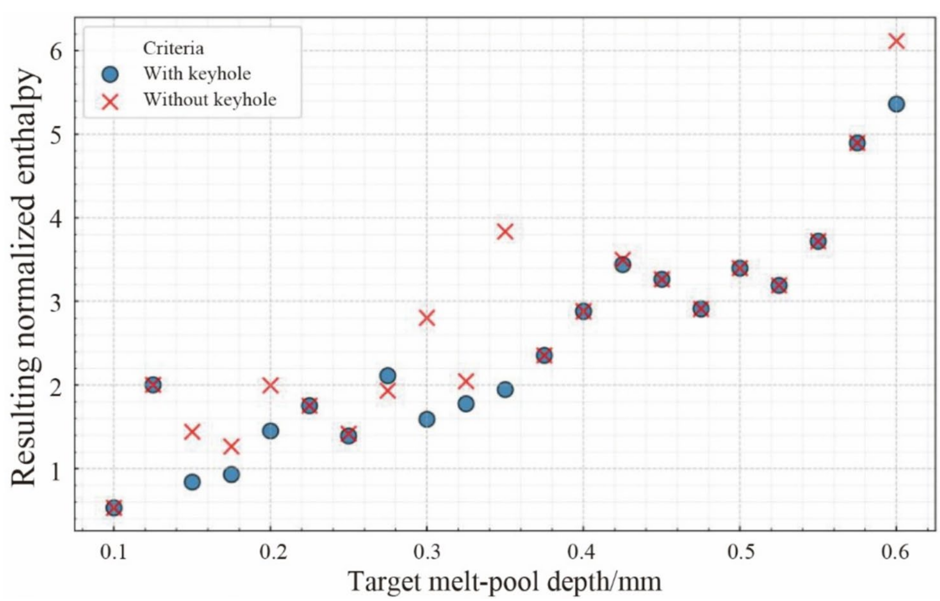


Fig. 16 Effect of considering the keyhole criteria on the resultant normalized enthalpy

experiments, the reliance on large and costly AM datasets was reduced, making the approach viable in resource-constrained settings. Secondly, by embedding this ML model as the RL agent's digital-twin environment, the need for physics-based simulations was eliminated, and dynamic exploration of parameter combinations across the discrete process space was enabled. Finally, the RL agent employed a flexible reward scheme to automatically balance competing objectives, such as achieving precise melt-pool depth while avoiding keyhole formation, resulting in process parameters that deliver both geometric accuracy and improved material integrity. Together, these features provide a cost-effective and adaptable optimization tool for practical additive manufacturing applications.

Future work aims to improve the model by incorporating advanced RL algorithms like proximal policy optimization and actor-critic methods to enhance efficiency. Expanding the dataset to include more alloys and parameters and refining the reward function through tuning coefficients and adding metrics will optimize outcomes. On the other hand, real-time sensor feedback will enable dynamic adjustments, and experimental validation on a lab-scale L-PBF system will compare predictions to measurements, providing data to refine the framework further. These steps will bridge simulations with real-world applications, advancing ML and RL integration for practical AM use.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

References

1. DebRoy T, Mukherjee T, Milewski JO et al (2019) Scientific, technological and economic issues in metal printing and their solutions. Nat Mater 18(10):1026–1032. https://doi.org/10.1038/s41563-019-0408-2
2. Saboori A, Aversa A, Marchese G et al (2019) Application of directed energy deposition-based additive manufacturing in repair. Appl Sci 9(16):3316. https://doi.org/10.3390/app9163316
3. Frazier WE (2014) Metal additive manufacturing: a review. J Mater Eng Perform 23(6):1917–1928. https://doi.org/10.1007/s11665-014-0958-z
4. Mosallanejad MH, Abdi A, Karpasand et al (2023) Additive manufacturing of titanium alloys: processability, properties, and applications. Adv Eng Mater 25:2301122. https://doi.org/10.1002/adem.202301122
5. King WE, Anderson AT, Ferencz RM et al (2015) Laser powder bed fusion additive manufacturing of metals; physics, computational, and materials challenges. Appl Phys Rev 2(4):041304. https://doi.org/10.1063/1.4937809
6. Wei HL, Mukherjee T, Zhang W et al (2021) Mechanistic models for additive manufacturing of metallic components. Prog Mater Sci 116:100703. https://doi.org/10.1016/j.pmatsci.2020.100703
7. Mosallanejad MH, Niroumand B, Aversa A et al (2021) In-situ alloying in laser-based additive manufacturing processes: a critical review. J Alloys Compd 872:159567. https://doi.org/10.1016/j.jallcom.2021.159567
8. Dadkhah M, Tulliani JM, Saboori A et al (2023) Additive manufacturing of ceramics: advances, challenges, and outlook. J Eur Ceram Soc 43(15):6635–6664. https://doi.org/10.1016/j.jeurceramsoc.2023.07.033
9. Behjat A, Shamanian M, Sadeghi F et al (2024) Additive manufacturing of a novel in-situ alloyed AISI316L-Cu stainless steel: microstructure and antibacterial properties. Mater Lett 355:135363. https://doi.org/10.1016/j.matlet.2023.135363
10. Dadkhah M, Mosallanejad MH, Iuliano L et al (2021) A comprehensive overview on the latest progress in the additive manufacturing of metal matrix composites: potential, challenges, and feasible solutions. Acta Metall Sin (Eng Lett) 34(9):1173–1200. https://doi.org/10.1007/s40195-021-01249-7
11. Galati M, Defanti S, Saboori A et al (2022) An investigation on the processing conditions of Ti-6Al-2Sn-4Zr-2Mo by electron beam powder bed fusion: microstructure, defect distribution, mechanical properties and dimensional accuracy. Addit Manuf 50:102564. https://doi.org/10.1016/j.addma.2021.102564
12. Zhang Y, Wu L, Guo X et al (2018) Additive manufacturing of metallic materials: a review. J Mater Eng Perform 27(1):1–13. https://doi.org/10.1007/s11665-017-2747-y
13. Behjat A, Shamanian M, Taherizadeh A et al (2022) Enhanced surface properties and bioactivity of additively manufactured 316L stainless steel using different post-treatments. Mater Today Proc 70:188–194
14. Khairallah SA, Anderson AT, Rubenchik A et al (2016) Laser powder-bed fusion additive manufacturing: physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. Acta Mater 108:36-45. https://doi.org/10.1016/j.actamat.2016.02.014
15. King WE, Barth HD, Castillo VM et al (2014) Observation of keyhole-mode laser melting in laser powder-bed fusion additive manufacturing. J Mater Process Technol 214(12):2915–2925
16. Gong H, Rafi K, Gu H et al (2014) Analysis of defect generation in Ti-6Al-4V parts made using powder bed fusion additive manufacturing processes. Addit Manuf 1:87–98. https://doi.org/10.1016/j.addma.2014.08.002
17. Mower TM, Long MJ (2016) Mechanical behavior of additive manufactured, powder-bed laser-fused materials. Mater Sci Eng A 651:198–213
18. Sames WJ, List FA, Pannala S et al (2016) The metallurgy and processing science of metal additive manufacturing. Int Mater Rev 61(5):315–360
19. Viale V, Stavridis J, Salmi A et al (2022) Optimisation of downskin parameters to produce metallic parts via laser powder bed fusion process: an overview. Int J Adv Manuf Technol 123(7):2159–2182
20. Taghian M, Mosallanejad MH, Lannunziata E et al (2023) Laser powder bed fusion of metallic components: latest progress in

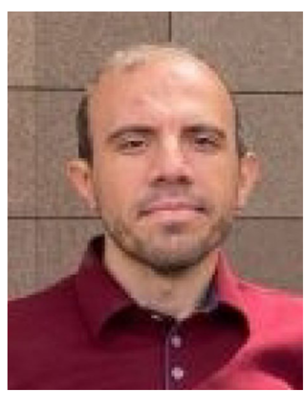
productivity, quality, and cost perspectives. J Mater Res Technol 27:6484–6500
21. Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. Struct Multidiscip Optim 26(6):369–395. https://doi.org/10.1007/s00158-003-0368-6
22. Deb K, Jain H (2013) An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans Evol Comput 18(4):577–601
23. Chia HY, Wu J, Wang X et al (2022) Process parameter optimization of metal additive manufacturing: a review and outlook. J Mater Inform 2:16. https://doi.org/10.20517/jmi.2022.18
24. Moradi A, Tajalli S, Mosallanejad MH et al (2025) Intelligent laser-based metal additive manufacturing: a review on machine learning for process optimization and property prediction. Int J Adv Manuf Technol 136(2):527–560. https://doi.org/10.1007/s00170-024-14858-0
25. Schleder GR, Padilha AC, Acosta CM et al (2019) From DFT to machine learning: recent approaches to materials science—a review. J Phys Mater 2(3):032001. https://doi.org/10.1088/2515-7639/ab084b
26. Kamath C (2016) Data mining and statistical inference in selective laser melting. Int J Adv Manuf Technol 86(5):1659–1677
27. Mosallanejad MH, Gashmard H, Javanbakht M et al (2024) A novel feature engineering approach for predicting melt pool depth during LPBF by machine learning models. Addit Manuf Lett 10:100214. https://doi.org/10.1016/j.addlet.2024.100214
28. Wang C, Tan XP, Tor SB et al (2020) Machine learning in additive manufacturing: state-of-the-art and perspectives. Addit Manuf 36:101538. https://doi.org/10.1016/j.addma.2020.101538
29. DebRoy T, Mukherjee T, Wei HL et al (2021) Metallurgy, mechanistic models and machine learning in metal printing. Nat Rev Mater 6(1):48–68. https://doi.org/10.1038/s41578-020-00236-1
30. Meng L, McWilliams B, Jarosinski W et al (2020) Machine learning in additive manufacturing: a review. JOM 72(6):2363–2377. https://doi.org/10.1007/s11837-020-04155-y
31. Himanen L, Geurts A, Foster AS et al (2019) Data-driven materials science: status, challenges, and perspectives. Adv Sci 6(21):1900808. https://doi.org/10.1002/advs.201900808
32. Wang AYT, Murdock RJ, Kauwe SK et al (2020) Machine learning for materials scientists: an introductory guide toward best practices. Chem Mater 32(12):4954–4965
33. He P, Liu Q, Kruzic JJ et al (2022) Machine-learning assisted additive manufacturing of a TiCN reinforced AlSi10Mg composite with tailorable mechanical properties. Mater Lett 307:131018. https://doi.org/10.1016/j.matlet.2021.131018
34. Barrionuevo GO, Walczak M, Ramos-Grez J et al (2023) Microhardness and wear resistance in materials manufactured by laser powder bed fusion: machine learning approach for property prediction. CIRP J Manuf Sci Technol 43:106–114. https://doi.org/10.1016/j.cirpj.2023.03.002
35. Minkowitz L, Arneitz S, Effertz PS et al (2023) Laser-powder bed fusion process optimisation of AlSi10Mg using extra trees regression. Mater Des 227:111718. https://doi.org/10.1016/j.matdes.2023.111718
36. Theeda S, Jagdale SH, Ravichander BB et al (2023) Optimization of process parameters in laser powder bed fusion of SS 316L parts using artificial neural networks. Metals 13(5):842. https://doi.org/10.3390/met13050842
37. Hassanin H, Zweiri Y, Finet L et al (2021) Laser powder bed fusion of Ti-6Al-2Sn-4Zr-6Mo alloy and properties prediction using deep learning approaches. Mater 14:2056. https://doi.org/10.3390/ma14082056
38. 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 123(7):2733–2756. https://doi.org/10.1007/s00170-022-10331-y
39. Muhammad W, Kang J, Ibragimova O et al (2023) Experimental investigation and development of a deep learning framework to predict process-induced surface roughness in additively manufactured aluminum alloys. Weld World 67(4):897–921. https://doi.org/10.1007/s40194-022-01445-8
40. Lee CH, Kühn U, Lee SC et al (2021) Optimizing laser powder bed fusion of Ti-5Al-5V-5Mo-3Cr by artificial intelligence. J Alloys Compd 862:158018. https://doi.org/10.1016/j.jallcom.2020.158018
41. Chen C, Wong SJL, Raghavan S et al (2022) Design of experiments informed deep learning for modeling of directed energy deposition process with a small-size experimental dataset. Mater Des 222:111098. https://doi.org/10.1016/j.matdes.2022.111098
42. Zhu X, Jiang F, Guo C et al (2023) Prediction of melt pool shape in additive manufacturing based on machine learning methods. Opt Laser Technol 159:108964. https://doi.org/10.1016/j.optlastec.2022.108964
43. Akbari P, Ogoke F, Kao NY et al (2022) MeltpoolNet: melt pool characteristic prediction in metal additive manufacturing using machine learning. Addit Manuf 55:102817. https://doi.org/10.1016/j.addma.2022.102817
44. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285. https://doi.org/10.1613/jair.301
45. Sousa J, Brandau B, Darabi R et al (2025) Artificial intelligence for control in laser-based additive manufacturing: a systematic review. IEEE Access 13:30845–30860. https://doi.org/10.1109/ACCESS.2025.3537859
46. Mattera G, Caggiano A, Nele L (2025) Optimal data-driven control of manufacturing processes using reinforcement learning: an application to wire arc additive manufacturing. J Intell Manuf 36(2):1291–1310. https://doi.org/10.1007/s10845-023-02307-w
47. Knaak C, Masseling L, Duong E et al (2021) Improving build quality in laser powder bed fusion using high dynamic range imaging and model-based reinforcement learning. IEEE access 9:55214–55231. https://doi.org/10.1109/ACCESS.2021.3067302
48. Wu H, Bordatchev E, Tutunea-Fatan OR (2024) Training reinforcement learning-based controller using performance simulation of the laser remelting process. Procedia Comput Sci 232:1849–1858. https://doi.org/10.1016/J.PROCS.2024.02.007
49. Petrik J, Bambach M (2024) RLTube: reinforcement learning based deposition path planner for thin-walled bent tubes with optionally varying diameter manufactured by wire-arc additive manufacturing. Manuf Lett 40:31–36. https://doi.org/10.1016/j.mfglet.2024.01.007
50. Petrik J, Bambach M (2023) Reinforcement learning and optimization based path planning for thin-walled structures in wire arc additive manufacturing. J Manuf Process 93:75–89. https://doi.org/10.1016/j.jmapro.2023.03.013
51. Vagenas S, Al-Saadi T, Panoutsos G (2026) Multi-layer process control in selective laser melting: a reinforcement learning approach. J Intell Manuf 37:281–298. https://doi.org/10.1007/s10845-024-02548-3
52. Dharmadhikari S, Menon N, Basak A (2023) A reinforcement learning approach for process parameter optimization in additive manufacturing. Addit Manuf 71:103556. https://doi.org/10.1016/j.addma.2023.103556

53. Watkins CJCH, Dayan P (1992) Technical note: *Q*-learning. Mach Learn 8:279–292. https://doi.org/10.1023/A:1022676722315
54. Vaghefi E, Hosseini S, Afsharinejad AH et al (2024) Additive manufacturing process parameter design for variable component geometries using reinforcement learning. Addit Manuf 84:104121. https://doi.org/10.1016/J.ADDMA.2024.104121
55. Haarnoja T, Zhou A, Abbeel P et al (2018) Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. https://doi.org/10.48550/arXiv.1801.01290
56. Sousa J, Darabi R, Sousa A et al (2023) Enhancing sample efficiency for temperature control in DED with reinforcement learning and MOOSE framework. In: proceedings of ASME 2023 international mechanical engineering congress and exposition. October 29–November 2, 2023, New Orleans, US. https://doi.org/10.1115/IMECE2023-113629
57. Shi S, Liu X, Wang Z et al (2024) An intelligent process parameters optimization approach for directed energy deposition of nickel-based alloys using deep reinforcement learning. J Manuf Process 120:1130–1140. https://doi.org/10.1016/j.jmapro.2024.05.001
58. Liu S, Shin YC (2019) Additive manufacturing of Ti6Al4V alloy: a review. Mater Des 164:107552. https://doi.org/10.1016/j.matdes.2018.107552
59. Dilip JJS, Zhang S, Teng C et al (2017) Influence of processing parameters on the evolution of melt pool, porosity, and microstructures in Ti-6Al-4V alloy parts fabricated by selective laser melting. Prog Addit Manuf 2(3):157–167. https://doi.org/10.1007/s40964-017-0030-2
60. Ye J, Khairallah SA, Rubenchik AM et al (2019) Energy coupling mechanisms and scaling behavior associated with laser powder bed fusion additive manufacturing. Adv Eng Mater 21(7):1900185. https://doi.org/10.1002/adem.201900185
61. Yang J, Han J, Yu H et al (2016) Role of molten pool mode on formability, microstructure and mechanical properties of selective laser melted Ti-6Al-4V alloy. Mater Des 110:558–570. https://doi.org/10.1016/j.matdes.2016.08.036
62. Vaglio E, Totis G, Lanzutti A et al (2024) A novel thermo-geometrical model for accurate keyhole porosity prediction in laser powder-bed fusion. Prog Addit Manuf 9(2):247–261. https://doi.org/10.1007/s40964-023-00446-3
63. Rafi HK, Karthik NV, Gong H et al (2013) Microstructures and mechanical properties of Ti6Al4V parts fabricated by selective laser melting and electron beam melting. J Mater Eng Perform 22(12):3872–3883. https://doi.org/10.1007/s11665-013-0658-0
64. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
65. Rasmussen CE (2004) Gaussian processes in machine learning. In: Bousquet O, von Luxburg U, Rätsch G (eds) Advanced lectures on machine learning. ML 2003. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_4
66. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
67. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 1189–1232. https://doi.org/10.1214/AOS/1013203451
68. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 13–17 August, San Francisco. 785–794. https://doi.org/10.1145/2939672.2939785
69. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
70. Brockman G, Cheung V, Pettersson L et al (2016) Openai gym. https://doi.org/10.48550/arXiv.1606.01540

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Arash Seifoddini holds a master's degree in Mechanical Engineering from Politecnico di Torino and is currently an R&D engineer. His research interests include artificial intelligence, optimal control, and robotics, with a strong focus on AI-driven automation and intelligent control systems for advanced manufacturing processes.

Arefeh Azad is a passionate robotics researcher pursuing a master's degree in Mechanical Engineering at Politecnico di Torino. Her research interests include artificial intelligence, multi-agent systems, and robotics, with a strong focus on artificial intelligence-driven robotics, autonomous navigation, and uncertainty reduction in dynamic environments, particularly for mobile robots and connected vehicles.



Mohammad Hossein Mosallanejad is an assistant professor of Advanced Materials Processing and Metal Additive Manufacturing in Department of Materials Engineering at Isfahan University of Technology (IUT), Iran. He was a postdoctoral research fellow at the Integrated Additive Manufacturing Center at Politecnico di Torino, Italy. He has published over 20 papers in academic journals and national and international conference proceedings. His research deals with the correlation between the properties of metallic materials and the development of novel alloys via metal additive manufacturing processes.

Abdollah Saboori is currently an associate professor of Advanced Materials Processing at Department of Management and Production Engineering and a member of the Integrated Additive Manufacturing Center at Politecnico di Torino. He has published over 125 papers in academic journals and national and international conference proceedings. He is an editorial board member of several international journals and a scientific committee member of several international conferences in the field of metal additive manufacturing. His research deals with the correlation between the properties of metallic materials and metal additive manufacturing processes.