Computer Methods in Applied Mechanics and Engineering 418 (2024) 116453

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

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

Data-driven and physics-based modelling of process behaviour and deposit geometry for friction surfacing

Frederic E. Bock [a,*], Zina Kallien [a], Norbert Huber [a,b], Benjamin Klusemann [a,c]

[a] Institute of Materials Mechanics, Helmholtz-Zentrum Hereon, Max–Planck-Straße 1, 21502 Geesthacht, Germany
[b] Institute of Materials Physics and Technology, Hamburg University of Technology, Eißendorfer Straße 42 (M), 21073 Hamburg, Germany
[c] Institute for Production Technology and Systems, Leuphana University Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany

ARTICLE INFO

Dataset link: 10.5281/zenodo.8263026

Keywords:
Machine learning
Feature selection
Numerical modelling
Heat transfer
Design of experiments
Explainable AI

ABSTRACT

In the last decades, there has been an increase in the number of successful machine learning models that have served as a key to identifying and using linkages within the process-structure–property-performance chain for vastly different problems in the domains of materials mechanics. The consideration of physical laws in data-driven modelling has recently been shown to enable enhanced prediction performance and generalization while requiring less data than either physics-based or data-driven modelling approaches independently. In this contribution, we introduce a simulation-assisted machine learning framework applied to the solid-state layer deposition technique friction surfacing, suitable for solid-state additive manufacturing as well as repair or coating applications. The objective of the present study is to use machine learning algorithms to predict and analyse the influence of process parameters and environmental variables, i.e. substrate and backing material properties, on process behaviour and deposit geometry. The effects of maximum process temperatures supplied by a numerical heat transfer model on the predictions of the targets are given special attention. Numerous different machine learning algorithms are implemented, optimized and evaluated to take advantage of their varied capabilities and to choose the optimal one for each target and the provided data. Furthermore, the input feature dependence for each prediction target is evaluated using game-theory related Shapley Additive Explanation values. The experimental data set consists of two separate experimental design spaces, one for varying process parameters and the other for varying substrate and backing material properties, which allowed to keep the experimental effort to a minimum. The aim was to also represent the cross parameter space between the two independent spaces in the predictive model, which was accomplished and resulted in an approximately 44 % reduction in the number of experiments when compared to carrying out an experimental design that included both spaces.

1. Introduction

Friction surfacing (FS) is a solid-state layer deposition technique that is applicable for various similar and dissimilar metallic material combinations [1]. The technique is based on friction and plastic deformation of a consumable material that is deposited onto a substrate [2]. The FS principle is applicable not only as coating technology but also shows potential for repair applications [3] or solid-state additive manufacturing [4]. The FS deposited material presents a homogeneous fine grained and recrystallized microstructure [5], where the heat input to the subjacent structure is low compared to fusion-based layer deposition processes

* Corresponding author.
E-mail address: frederic.bock@hereon.de (F.E. Bock).

https://doi.org/10.1016/j.cma.2023.116453
Received 16 June 2023; Received in revised form 22 August 2023; Accepted 11 September 2023
Available online 30 September 2023

Nomenclature

Acronyms (selection)

$\boldsymbol{MAT}$	Substrate & backing plate material properties c_b, c_s, k_b, k_s, t_b, t_s
$\boldsymbol{PP}$	Process parameters F, RS, TS

Data set features

$\mathscr{A}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, t, w, v_{cr}, M
$\mathscr{B}$	*train:* $\boldsymbol{PP}$, $\boldsymbol{MAT}$, t, w, v_{cr}, M; *test:* $\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{t}$, $\hat{w}$, $\hat{v}_{cr}$, $\hat{M}$
$\mathscr{C}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{t}$, $\hat{w}$, $\hat{v}_{cr}$, $\hat{M}$
$\mathscr{D}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, T_{max}
$\mathscr{E}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{T}_{max}$
$\mathscr{F}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, T_{max}, v_{cr}, M

Symbols

$a_{l,u}$	Layer output function
α	Learning rate
α_i	Lagrange multiplier
α_{ridge}	Scaling factor of penalty term for ridge regression
B	Number of decision trees contained in a random forest
b	Bias term
C	Regularization constant
c_b	Specific heat capacity of backing plate
c_s	Specific heat capacity of substrate
D	Number of features
ϵ	Penalty-free tube width around model function
F	Force
g	Activation function
γ	Scaling parameter of radial basis function
K	Complete feature set
k	Total number of layers
κ	Kernel function
k_b	Thermal conductivity of backing plate
k_s	Thermal conductivity of substrate
L	Total number of regions
l	Region number
M	Torque
$\hat{\mathrm{M}}$	Predicted torque
M_{tree}	Maximum number of trees
N	Number of samples
n	Layer number
P	Number of simplified features
ϕ_0	Predicted mean of all training samples
ϕ_i	Shapley value
$\mathbf{R}$	Set of all regions
RS	Rotational speed
S	Feature subset
s	Threshold value
t	Thickness of deposit layer
$\hat{t}$	Predicted thickness of deposit layer
t_b	Thickness of backing plate

t_{lin}	Linear approximation of t
T_{max}	Maximum temperature from numerical model
$\hat{T}_{max}$	Machine learning predicted maximum temperature
t_s	Thickness of substrate
TS	Travel speed
u	Number of neurons
v_{cr}	Feed rate
$\hat{v}_{cr}$	Predicted feed rate
$v_{cr,lin}$	Linear approximation of v_{cr}
$\mathbf{W}$	Weight matrix
w	Width of deposit layer
$\hat{w}$	Predicted width of deposit layer
$\mathbf{w}$	Weight vector
w_{lin}	Linear approximation of w
x	Input
$\mathbf{x}$	Input vector
x'	Simplified input
ξ	Slack variable of soft margin
x_S	Input feature value in subset S
y	Output
$\hat{y}$	Predicted output
$\mathbf{z}$	Layer arguments
z'	Simplified feature representation in subset S

because the layer deposition is executed below the materials' melting temperature. The three main FS process parameters are rotational speed (RS), axial force (F) and travel speed (TS), which have to be adapted according to the materials [6]. Being a discontinuous process, the layer dimensions are mainly limited by the dimensions of the used stud material; however, the choice of process parameters as well as substrate and backing plate materials affect the deposit geometry, commonly defined by layer thickness and width. Moreover, the process temperature of FS was found to be in direct relation with the deposit geometry [7,8]. For instance, previous studies found that the substrate thickness [8] and cooling during the layer deposition [9] affect deposit geometry. Generally, a fundamental understanding is necessary to exploit the potential of FS, e.g. to foresee process behaviour and layer dimensions based on selected materials and process parameters.

Process simulation via numerical models cannot only minimize experimental effort but also allow insights that are inaccessible during experiments. Nevertheless, the number of available models is scarce, which might be related to several challenges that remain for modelling the FS process [1], e.g. unknown asymmetric material flow, varying thermo-mechanical material properties or unknown and varying friction coefficient. There are some studies that showed reasonable agreements between experiments and simulations of the temperatures during consumable stud plasticizing [10] or layer deposition [11,12]. In further numerical approaches [13,14], temperature and strain rate distributions for the deposit were simulated. One main problem for modelling the deposited layer is the mutual dependence of process temperatures and deposited layer geometry, i.e. temperatures depend on the deposited layer geometry and vice versa. Thus, the knowledge gap on the geometric dimensions cannot be closed unless the experiment has already been performed. Sound physics-based models for the prediction of FS layer geometry, for example, are missing. Overall, there are complex relationships among different phenomena occurring during solid-state layer deposition via FS: Process parameters and material composition govern material flow behaviour and process temperatures, which, in turn, define the resulting deposit in terms of geometry, microstructure and mechanical properties. Most of these complex relations are not yet decrypted.

In this context, machine learning (ML) tools represent key enablers to identify and utilize these relations, i.e. to describe the importance of parameters and to predict process behaviour as well as resulting material structures and properties. ML is a new paradigm in programming as opposed to classical programming. In classical programming, the rules for computing input data into desired outputs are pre-defined by human input. For ML, in contrast, the inputs and outputs are given whereas the rules are generated by the ML-algorithm. Consequently, when these rules are applied to new data, new and original answers can be produced [15].

The applications of ML to the fields of materials mechanics have recently been reviewed by Bock et al. [16], where various examples are discussed, in particular how linkages along the process-structure–property-performance chain can be discovered and exploited through the use of ML. In addition, a recent review on the employment of ML in additive manufacturing (AM) is presented by Meng et al. [17]. For instance, for wire-arc AM (WAAM), Xiong et al. [18] used artificial neural networks (ANNs) that outperformed their benchmark 2nd-order linear regression model with respect to achieving a lower prediction error. Wacker at

al. [19] also used ANNs to predict the geometry and distortion of multiple layers manufactured via WAAM. When Deng et al. [20] compared linear regression, ANNs and extreme gradient boosted (XGB) forests to predict bead geometries, they selected XGB as the best performing model. In addition, Oh [21] built a predictive model via support vector regression (SVR) for a similar task, whereas Barrionuevo et al. [22] predicted the bead geometry for WAAM via a Gaussian process regression model that outperformed their XGB and ANN predictive models. The response surface methodology (RSM) based on Designs of Experiments (DoE) and analysis of variance (ANOVA) was utilized in studies about wire-based and powder-based direct laser deposition to correlate process parameters such as laser power, scanning velocity and powder/wire feed rate with geometrical properties such as thickness, width and penetration depth into the substrate. Models were based on multiple linear regression of 1st-order [23–25], and 2nd-order linear regression [26]; however, other authors used ANNs [27–29], which shows that it cannot be determined in advance, which ML model will produce the lowest prediction errors, as this strongly depends on the specific problem and conditions at hand.

For FS, Vitanov and co-workers [30,31] developed a decision support for the selection of process parameters, where the process parameters are related to coating thickness, width and bond strength via a 2nd-order linear regression model [32]. Sugandhi et al. [33] also developed a 2nd-order linear regression model based on empirical observations to map relationships between process parameters and layer thickness as well as width. For other solid-state materials processing techniques such as cold spraying, which can also be used for AM (cold spray AM), Ikeuchi et al. [34] predicted the deposit geometry via ANNs to obtain complete track profiles based on the process parameters. For Friction Stir Welding (FSW), another solid-state process, studies also focused on DoE-based RSM in companion with ANOVA, again yielding linear regression models in dependence of process parameters for the prediction of mechanical properties [35,36]. In a similar context, Lakshminarayanan and Balasubramanian [37] compared 2nd-order linear regression with ANNs for the prediction of mechanical properties for FSW joints and concluded that ANNs exhibited superior prediction performance in their case. ANNs were also used to predict process state variables of FSW such as maximum temperature [38,39] and torque [38], among others. In short, ML regression models of different complexity are showing best prediction performances, depending on the complexity of the relationships to be mapped and the available data.

In addition to using process parameter inputs, domain-specific physical information about the process behaviour can be used to enhance prediction performance. For example, physical information about the energy occurring during the process of Friction Riveting was additionally used to predict the UTS of produced joints depending on process parameters via RFR [40]. In a similar domain knowledge-informed approach, Ikeuchi et al. [41] modified their initial approach [34] for the prediction of the deposit cross-section shape produced by cold spray AM, with the input from a Gaussian function model that represents the fundamental shape, to increase data-efficiency of the predictive ANN. Furthermore, so-called hybrid modelling was proposed by Chinesta et al. [42], where physics-based and data-driven modelling techniques are combined in a way that the former provides fundamental physical relationships that show discrepancies to the target solution; therefore, is corrected and enhanced by the latter to reach the target solution. This approach has shown superior prediction performances in comparison to exclusive use of either physics-based or data-driven modelling for various problems related to mechanical material behaviour [43–45], especially when data is scarce.

In this study, a number of different ML algorithms are used to predict FS process behaviour and deposit geometry based on input features consisting of two input spaces: process parameters (PP) as well as substrate and backing material properties (MAT). PP consist of F, RS, and TS, while MAT contain thermal conductivity (k_s, k_b), specific heat (c_s, c_b), and thickness (t_s, t_b), for substrates and backing plates, respectively. The target process behaviour variables are maximum temperature T_{max} as well as the mechanical behaviour, which is implicitly represented by feed rate v_{cr} and torque M, whereas the target geometry variables are deposit thickness t and width w. Except for T_{max}, all feature and target values are determined experimentally via two separate designs of experiments (DoE), each with three factors and three levels, to generate samples for training, validation and testing of the ML models. Only T_{max} is obtained from a previously developed physics-based numerical heat transfer model [12] for the corresponding depositions. For each prediction target, different ML regression algorithms is trained, where the respective hyperparameters of all models are optimized and the model with the best prediction performance selected for final predictions and evaluation. The employed ML algorithms were: 1st-, 2nd-, 3rd-order linear regression (LR1, LR2, LR3), 1st-, 2nd-, 3rd-order ridge regression (RR1, RR2, RR3), random forest regression (RFR), extreme gradient boosting (XGB) forests, support vector regression (SVR), and artificial neural networks (ANNs). Besides the effects of different additional features for the prediction of each target, the particular contribution by each feature to the prediction of the best performing model is evaluated via Shapley Additive Explanation (SHAP) values as introduced by Lundberg and Lee [46], which has been successfully applied in other studies [47,48]. The aim is to reduce the number of experiments as well as to exploit physics-based and data-driven modelling; i.e. utilize established knowledge as well as perform and interpret predictions of process behaviour and deposit geometry variables. To the knowledge of the authors, this has not yet been achieved for FS.

2. Methods, materials and data workflow

In this section, the FS process and DoEs for data acquisition of the targets depending on variations of PP and MAT are introduced. The experimentally validated finite element heat transfer model by Kallien and Klusemann [12] is exploited for data mining the corresponding T_{max}. The complete data sets can be found in Appendix D. Since training data is scarce, the predictive models are aimed to be as simple as possible by keeping the number of internal model parameters and features to a minimum. Consequently, the inclusion of additional features such as T_{max}, feed rate v_{cr} and torque M in the input space has been evaluated to find the best performing model for the prediction of each target. The techniques used for supervised ML regression are shortly reviewed as well as measures to evaluate prediction performances and to select the best models are presented. Ultimately, the model interpretation approach via SHAP values is explained.

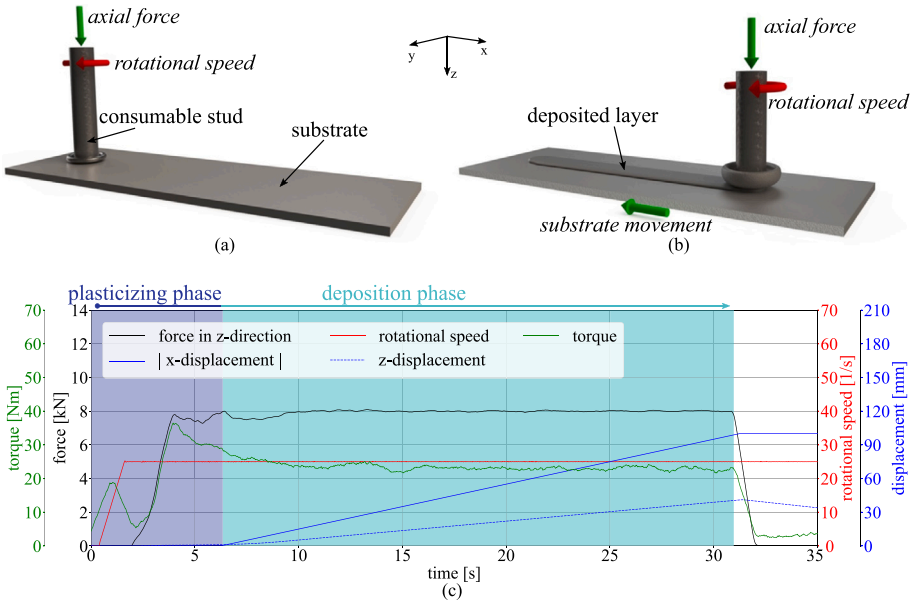


Fig. 1. Schematic of the FS process during plasticizing phase (a) and deposition phase (b) [49], where the rotating stud material is fixed in x-/y-direction and the substrate is moved by the underlying table into negative x-direction. In (c), exemplary process records for FS process performed at 8 kN, 25 1/s and 4 mm/s for F, RS, TS, respectively, of AA5083 over AA7050 are shown.

2.1. Friction surfacing process principle and experimental setup

The solid-state layer deposition via FS always follows the same principle and the process can be divided in two phases. During the plasticizing phase (i), the consumable stud material experiences a rotational speed and an axial force. As a result, the stud is pressed onto the substrate and frictional heat occurs at the materials' interface. The tip of the consumable stud deforms and plasticizes, Fig. 1(a). The deposition phase (ii) is initiated by superimposing a relative translational movement, which enables the deposition of the plasticized stud material onto the substrate as a layer, Fig. 1(b). The process ends with the retraction of the remaining stud when the desired length is achieved.

The experiments were performed using a custom-designed friction welding machine (RAS, Henry Loitz Robotik, Germany), which allows maximum rotational speed of 100 1/s, maximum force of 60 kN and maximum torque of 200 N m. All depositions were performed at room temperature and force-controlled. The deposition path was programmed via computer numerical control (CNC), where the pre-programmed length was 100 mm. AA5083 H112 consumable stud material (20 mm diameter, 125 mm length) was deposited over different aluminium substrates (300 mm length, 130 mm width, 8 to 12 mm thickness) for the present study. A backing plate was used between substrate and machine table, where the effects of backing materials (AA7050, Ti–6Al–4V and Steel GL-A36) as well as substrate materials (AA2050, AA5083 and AA7050) and substrate thickness (8 mm, 10 mm, 12 mm) were subject of investigation. The records of the welding equipment, i.e. forces, displacement, rotational speed and torque, are used for analysis of the process behaviour. An example is given in Fig. 1(c). For the force-controlled deposition processes, the stud consumption rate v_{cr} defined by the average velocity in axial force direction and average toque M during deposition phase are in particular interest for the following process analysis. The deposited layers were cut at half of the deposition path, embedded and prepared following common grinding and polishing practices. The cross sections were analysed in terms of deposit thickness and width as implemented by Kallien et al. [8] using an optical light microscope (VHX-6000, Keyence, Germany).

2.2. Experimental data acquisition

The experimental acquisition of data was achieved according to pre-defined variations of PP and MAT, which were varied individually via two DoEs based on Box–Behnken with three factors and three levels for each factor, leading to a total of 30 training samples, including six centre point replicas out of three for each space. A combined Box–Behnken DoE consisting of both spaces simultaneously would have had six factors with three levels each, amounting to a total number of 54 required samples, including five centre point replicas. In this comparison, the amount of experimental samples is reduced by approximately 44% through this approach. Data sets referred to as training, validation and test sets in the respective separate variations of PP and MAT variables are combined into three global training, validation and test sets, respectively.

Table 1
Process parameter space: High, mid and low levels of factors: F, RS and TS.

Factor	F [kN]	RS [$\frac{1}{s}$]	TS [$\frac{mm}{s}$]
High level	10	25	8
Mid level	8	20	6
Low level	6	15	4

Table 2
Substrate and backing material space [52–56]: High, mid and low levels of factors: k_S, k_B and t_S.

Factor	k_S [$\frac{W}{mK}$]	k_B [$\frac{W}{mK}$]	t_S [mm]
High level	157	157	12
Mid level	117	49.6	10
Low level	75	6.7	8

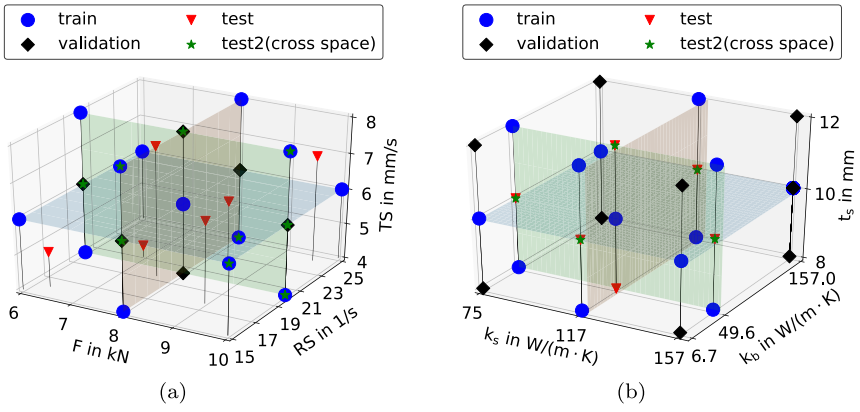


Fig. 2. Parameter spaces spanned with their three respective variables that are each assigned to different value levels in training, validation, test and test2 data sets. In (a), process parameter variation of force F in kN, rotation speed RS in s^{-1} and travel speed TS in mm s^{-1}. (b) Material parameter variation in terms of thermal conductivity of the substrate k_S in W/(mK), thickness of the substrate t_S in mm and the thermal conductivity of the backing plate k_B in W/(mK).

2.2.1. Variation space of PP

In the first series of experiments, the $\boldsymbol{PP} = [RS, TS, F]$ are varied using an AA7050 T7451 substrate (10 mm thickness) and an AA7050 T7451 backing plate (8 mm thickness). A process window for FS of AA5083 H112 onto AA7050 substrates, where feasible combinations of process and material parameters are determined, was published by Kallien et al. [8]. Each parameter is varied to three different equidistant values, i.e. to high, mid and low values that correspond to normalized values [1,0,−1], respectively, see Table 1. As shown in Fig. 2(a), the 13 samples contained in this DoE represent the edge centres and the centre point of a cube spanned by the three parameters and were used as the training data set for the predictive models. For the validation set, the cube's 6 facial centre points were used to account for the same value ranges but different combinations of parameters. To obtain a suitable test set whereupon the generalization of the models can be evaluated, six additional random samples within the parameter space were randomly created via latin hypercube sampling [50] as implemented in the *surrogate modelling toolbox (SMT)* [51] as part of the *pyDOE* library package.

2.2.2. Variation space of MAT

In the second series of experiments, systematic variation of $\boldsymbol{MAT} = [k_b, k_s, t_s]$ has been performed, where three properties are altered in a similar way than $\boldsymbol{PP}$ with three levels each. To evaluate the effect on deposit geometry from a variation among $\boldsymbol{MAT}$, which has principally been shown in [8], substrate thermal conductivity k_S, backing plate thermal conductivity k_B and substrate thickness t_S were varied according to values listed in Table 2, keeping the process parameters constant at 8 kN, 20 1/s and 6 mm/s. As shown in Fig. 2(b), the data sets used for training, validation and testing consist of 13, 8 and 6 samples, respectively. To reach the given thermal material properties, substrate materials were varied from AA2050 over AA5083 to AA7050 with thermal conductivities k_S ranging in values from low to high levels, respectively. The backing material was altered by using titanium Ti–6Al–4V, steel GL-A36, and aluminium alloy AA7050 that also posses thermal conductivities ranging within values from low to high levels, respectively. The thickness of the substrate t_S was also changed, ranging from low to high levels. The thickness of the backing plate t_B was aimed to be constant at 10.0 mm, which was the case for AA7050 and GL-A36 but amounted to 10.2 mm for Ti–6Al–4V. Due to the specific materials used for substrates and backing plates and their corresponding thermal conductivities, the factor levels could not be strictly defined as equidistant to each other; thus, are not according to Box–Behnken.

2.2.3. Ultimate test space of PP and MAT

Generalization of all predictive models is tested and evaluated based on a test2 set where ***PP*** and ***MAT*** variables have been varied simultaneously, unlike samples in training, validation and test sets. This test2 data set serves as a benchmark to justify the data reduction approach introduced earlier. For that, 12 samples were randomly selected out of the six-factor DoE with 54 samples that was avoided to be acquired completely for training. Those samples are used for ultimate predictions with all trained models to evaluate prediction performance and assess the generalization ability of the utilized ML models.

2.2.4. Heat transfer modelling via Finite Element Method

As experimental temperature measurements, e.g. via thermocouple, incur significant experimental effort and provide only a limited resolution, the finite element heat transfer model presented by Kallien and Klusemann [12], which was validated with FS experimental temperature data, was used in this study to obtain FS process temperatures. Process temperature allows valuable insights into the FS process and was found to affect the deposited layer geometry as already discussed. However, the approach of the simulation requires knowledge about the deposition geometry and feed rate in advance, which reasons the aim to predict those via ML. Initially, the model was used to calculate the process temperatures for the performed FS layer depositions with values obtained by experiments. The obtained maximum process temperature is used as an additional input for the predictions via ML, whereupon its impact on prediction performance as well as its importance as a feature for the prediction of process behaviour and deposit geometry is assessed.

2.2.5. Data-mining of heat transfer model

To further exploit the relationships contained in the heat transfer model introduced in the previous Section 2.2.4 and to be able to build a surrogate model via ML, data-mining is performed. For that, the additional value combinations of model input variables t, w and v_{cr} are required to be physically feasible; thus, additional samples are augmented by using linear approximations of those variables depending on T_{max}, which were established in [8] based on experiments. Accordingly, the approximation of the thickness t is computed as

$$\hat{t}_{lin}(T_{max}) = -0.0111 \cdot T_{max} + 6.2799, \tag{1}$$

the width w as

$$\hat{w}_{lin}(T_{max}) = 0.0121 \cdot T_{max} + 13.519, \tag{2}$$

and the feed rate v_{cr} through

$$\hat{v}_{cr,lin}(T_{max}) = -0.0032 \cdot T_{max} + b. \tag{3}$$

The values of the bias i.e. the intercept b for predicting the feed rate $\hat{v}_{cr,lin}(T_{max})$ were adjusted to experimental data points related to F being either 6 kN, 8 kN or 10 kN, leading to constant b values 2.6358, 3.2051 or 3.777, respectively[1]. Note that since Eqs. (1)–(3) depend only on maximum process temperature T_{max}, more significant prediction errors in comparison to the experimental (exp.) errors are expected, especially when variables are changed that are not considered by these approximations, i.e. anything else besides temperature, such as process parameters as well as substrate and backing material properties. Overall, due to the simplicity of this model, high precision cannot be expected; however, moderate accuracy is achieved, particularly in comparison to the alternative of random value guessing. The prediction errors for t, w, and v_{cr} achieved with Eqs. (1)–(3) on the data sets used in this study are listed in Table A.1 in Appendix A. Physical data augmentation was achieved by the variation of T_{max} with a Δ of $\pm$60 °C in 20 °C steps starting from each sample in training and validation sets to use the resulting values of $\hat{t}_{lin}$, $\hat{w}_{lin}$ and $\hat{v}_{cr,lin}$ from Eqs. (1)–(3) as inputs t, w and v_{cr} for the heat transfer model and compute the related T_{max} numerically. It is investigated whether an increase in the number of training and validation samples, to 140 and 75, respectively, based on this physical data augmentation can enhance the ML prediction performance of T_{max}. Note that the implementation of this data augmentation via the exploitation of the FE-model in companion with Eqs. (1)–(3) is not required for deployment of the ML models built with these enriched data sets.

2.3. Variation of feature space for each target

A fundamental chain of relations among physical quantities of interest shown in Fig. 3, where ***PP*** and ***MAT*** variables lead to a certain T_{max}, v_{cr} and M, which then altogether lead to specific deposit geometry variables t and w. For the prediction of each output target along this chain, step-wise prediction of those quantities is performed where previously predicted physical quantities are used as features for the prediction of the following target outputs, as illustrated in Fig. 4, which represents the proposed prediction framework of this study. First, the prediction of T_{max} is achieved by building a surrogate model of the heat transfer FE model introduced in Section 2.2.4 via ML; similar inputs needed for the FE model are also required for the ML model. Because of this, v_{cr}, M, t and w need to be predicted first just based on ***PP*** and ***MAT*** as input features in order to be available as additional features besides ***PP*** and ***MAT*** variables for the prediction of T_{max}, see Fig. 4(a). Second, the usage of T_{max} as additional feature to ***PP*** and ***MAT*** variables for the prediction of v_{cr} and M is evaluated. And third, it is assessed whether using T_{max}, v_{cr} and M as additional features besides ***PP*** and ***MAT*** variables leads to enhanced prediction performance for t and w, respectively.

[1] Physical SI-units are omitted in Eqs. (1)–(3) for simplicity.

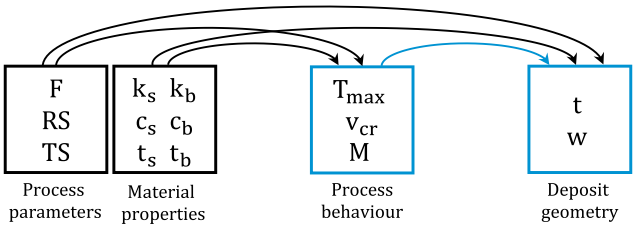


Fig. 3. Schematic of fundamental relationships: Process parameters as well as material properties of substrate and backing plate lead to a certain process behaviour, which altogether determine the resulting deposit geometry. Lines in black denote true variable values that are initially available, whereas lines in blue denote variable values that need to be predicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

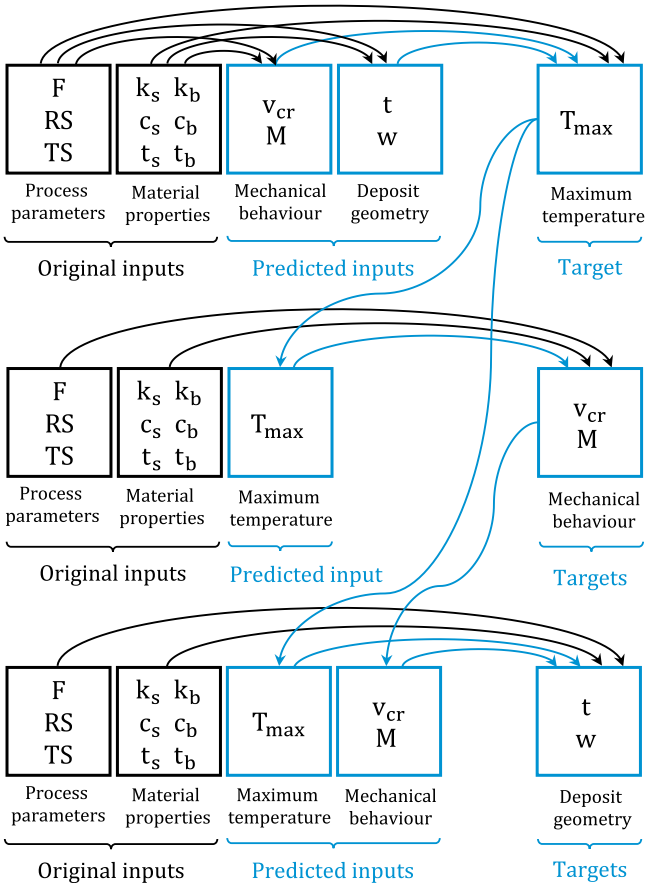


Fig. 4. Chain of different prediction targets and their respective utilization as input features to approximate maximum temperature T_{max}, process behaviour and ultimately the deposit geometry. Blue lines indicate predicted variable values, e.g. in the top case for the prediction of T_{max}, the variables representing process behaviour and deposit geometry are required as input besides the variables that represent process parameters and material properties; thus, they need to be predicted first to be available. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Additionally, once additional features have been shown to be beneficial for the prediction performance, the corresponding true values need to be replaced by their predictions since they are unknown in a real use-case application. In this regard, T_{max}, v_{cr}, M, t and w represent true values of features, whereas $\hat{T}_{max}$, $\hat{v}_{cr}$, $\hat{M}$, $\hat{t}$ and $\hat{w}$ embody predicted values. It is investigated if utilizing true or predicted feature values in the training and validation sets results in better predictions since the test set only contains predicted values. For the prediction of T_{max}, it is also evaluated whether data-mining of the heat transfer model, as explained in Section 2.2.5, leads to improved prediction performance via ML. An overview of all feature sets with their respective names that are used in this study is provided in Table 3.

Table 3
Input feature sets consisting of different input variables used for training, validation and testing (unless indicated otherwise) to find best predictive models. All feature values represent true values except for $\hat{\imath}, \hat{w}, \hat{v}_{cr}, \hat{M}, \hat{T}_{max}$, which denote predicted values, respectively.

Space	Inputs variables
$\boldsymbol{PP}$	F, RS, TS
$\boldsymbol{MAT}$	k_s, c_s, t_s, k_b, c_b, t_b
(–)	$\boldsymbol{PP}$, $\boldsymbol{MAT}$
$\mathscr{A}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\imath$, w, v_{cr}, M
$\mathscr{B}$	*train*: $\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\imath$, w, v_{cr}, M *test*: $\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{\imath}$, $\hat{w}$, $\hat{v}_{cr}$, $\hat{M}$
$\mathscr{C}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{\imath}$, $\hat{w}$, $\hat{v}_{cr}$, $\hat{M}$
$\mathscr{D}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, T_{max}
$\mathscr{E}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, $\hat{T}_{max}$
$\mathscr{F}$	$\boldsymbol{PP}$, $\boldsymbol{MAT}$, T_{max}, v_{cr}, M

2.4. Predictive model building via machine learning regression

In this study, supervised ML is used to perform regression analyses where the output is known and represents a continuous value [57]. To attain good predictive models for the target outputs (T_{max}, v_{cr}, M, t and w), a number of ML models with different 'learning' concepts and properties, ranging from simple to comparatively complex algorithms in terms of the number of model parameters, are evaluated based on their prediction performance on training, validation and test data sets. The *Python*-based *scikit-learn* [58] library was used for the implementation of LR, RR, RFR and SVR, whereas for the *xgboost* [59] library was used for XGB and the *Keras* [60] library with the *Tensorflow* [61] backend for ANNs. Each of the employed algorithms is briefly explained in the following, where a sample collection is denoted as $(\mathbf{x}_i, y_i)_{i=1}^N$ with N number of samples, the input feature vector $\mathbf{x}_i$ of dimension D and the true output y_i of a sample i in the range $[1, \dots, N]$ depends on every feature $x_i^{(j)}$ with j in the range $[1, \dots, D]$, unless stated otherwise.

2.4.1. Linear regression

Via LR, a predictive model can be built out of a linear combination of input features. The model is built as follows:

$$f_{\text{lin}}(\mathbf{x}) = \mathbf{w}\mathbf{x} + b, \tag{4}$$

where f_{lin} is parameterized by a weight vector $\mathbf{w}$ of D dimensions and a bias term b. The model is used in the form $y \leftarrow f_{\text{lin}}(\mathbf{x})$ to predict an unknown y for given $\mathbf{x}$, to determine the optimal values $\mathbf{w}^*$ and b^* of Eq. (4) via minimization of the squared error loss function:

$$\min \frac{1}{N} \sum_{i=1}^{N} \left[f_{\text{lin}}(\mathbf{x}_i) - y_i \right]^2. \tag{5}$$

For 2nd and 3rd order linear regression, the input features are combined using a polynomial of the respective order, while the weights remain linear. However, the significant increase in the number of features by using polynomials can easily surpass the number of training samples provided, which then can prevent sufficient training.

2.4.2. Ridge regression

RR is also known as L2 regularization and similar to linear regression except that overfitting can be prevented because the cost function (Eq. (5)), the sum of squared errors, is minimized with the addition of a penalty term via:

$$\min \frac{1}{N} \sum_{i=1}^{N} \left[f_{\text{lin}}(\mathbf{x}_i) - y_i \right]^2 + \alpha_{ridge} \mathbf{w}^2, \tag{6}$$

where α_{ridge} represents the hyperparameter to be optimized during tuning as its particular value scales the penalty term. The weights in RR can also be of 2nd and 3rd polynomial order. In this work, optimal values of α were determined via grid search in the range [0,20] in 0.01 steps with leave-one-out cross-validation (LOOCV) on the training set.

2.4.3. Random forest regression

An RFR is an ensemble method where a model consists of numerous decision tree regression (DTR) models and returns the averaged value of all predictions from the constituting trees. DTR models partition the input space recursively into discrete and non-overlapping regions, wherein a local model is defined for each region, which can be represented with one leaf. The particular choice of leaf that defines the prediction output for a given input depends on a hierarchical structure of nodes representing decision rules where values are differentiated by being above or below a threshold value s. The regions can be described through $R_l(j, s) = \{\mathbf{x}_i | x_i^{(j)} \leq s\}$ and $R_{l+1}(j, s) = \{\mathbf{x}_i | x_i^{(j)} > s\}$ with l as region number. The predictive function is given by:

$$f_{tree}(\mathbf{x}_i \in R_l) = \text{mean}(y_i | \mathbf{x}_i \in R_l) =: \hat{y}_i \tag{7}$$

where $\mathbf{x} = [x_1, x_2, \dots, x_j]$, $\mathbf{R} = [R_1, R_2, \dots, R_L]$ with L as total number of regions and $\hat{y}_i$ as the average of the target variables contained in region R_l [57]. The regions are defined through the minimization of the residual sum of squares:

$$\min \sum_{l=1}^{L} \sum_{i \in R_l} [y_i - \hat{y}_{R_l}]^2. \tag{8}$$

To prevent overfitting, regularization can be performed, for example by restricting the maximum depth of a tree or the minimum region size.

In RFR, each DTR is build based on a different randomly selected subset of input features and random sample subsets from the training data set, respectively. Those distinct random feature selections are used to split each node of each DTR during training, which can reduce the effect of artefacts such as noise, outliers, and under- and overfitting. In particular, a RFR prediction for a given sample $\mathbf{x}$ is computed by the averaged predictions of B number of decision trees contained in the forest as follows:

$$y \leftarrow f_{\text{RFR}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} f_{\text{tree}}(\mathbf{x}), \tag{9}$$

where f_{tree} is the B'th predictive decision tree model. During training, the generalization error of a forest is aimed to converge to a lower limit with increasing number of trees and depends on the prediction performance of individual trees. Ultimately, RFR can yield a predictor where only a few hyperparameters need to be determined (number of maximum trees and the size of the random feature subsets that are considered at each split). The particular procedure for hyperparameter optimization in this work was achieved via a grid search with LOOCV on the training set through the maximum number of trees, ranging from 1 to 400, whereas the maximum number of input features equals the number of relevant features for the respective target.

2.4.4. Extreme gradient boosted forests

A gradient boosted forest is another ensemble method where, in contrast to RFR, additional trees that are added to the model, partly compensate the errors made by previous trees until the maximum number of trees is reached [62]. For that, sample targets in the training set are replaced by residuals in the form:

$$\hat{y}_i \leftarrow y_i - f(\mathbf{x}_i) \tag{10}$$

where $\hat{y}_i$ is the residual based on the input $\mathbf{x}_i$. Once all samples in the training set are modified with residuals, it is used to build a new DTR model f_1 that is then part of the boosting model through $f = f_0 + \alpha f_1$ with α as the learning rate and f_0 as the initial DTR in the ensemble. Then, the residuals in the training set are updated again and the model redefined via $f = f_0 + \alpha f_1 + \alpha f_2$; consequently, the model form reaches:

$$f = f_0 + \alpha f_1 + \cdots + \alpha f_{M_{tree}} \tag{11}$$

where M_{tree} is the maximum number of trees. Extreme gradient boosted (XGB) forests represent a significant upgrade from gradient boosting because in XGB, Newton–Raphson optimization with a second order Taylor approximation is used in the loss function [63], as opposed to gradient descent in gradient boosting. In combination with a regularization function, increased prediction precision and enhanced computational efficiency can be achieved. Since XGB is a highly regularized variant of gradient boosting, a comparatively high number of hyperparameters need to be optimized during tuning, where grid or random search have been shown to be unsuitable; therefore, replaced by Bayesian optimization, in particular with *Hyperopt* [64].

2.4.5. Support vector regression

The SVR algorithm performs a non-linear mapping of input feature vectors into a higher dimensional space (via the so-called kernel trick), where a description of all data points via a linear function becomes possible [65]. Starting out from the original feature space, predictions are computed similarly to Eq. (4), except that $\mathbf{w} = \sum_i \alpha_i \mathbf{x}_i$ is used and the inner product $\mathbf{x}_i^T \mathbf{x}$ is replaced by a kernelized solution $\kappa(\mathbf{x}_i, \mathbf{x})$ to yield predictions with:

$$y \leftarrow f_{\text{SVR}}(\mathbf{x}) = b + \sum_{i=1}^{N} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}), \tag{12}$$

where α_i is a Lagrange multiplier and κ represents a kernel function such as a radial basis, linear, or polynomial function, among others. During training, the objective function

$$J = C \sum_{i=1}^{N} [\xi_i^+ + \xi_i^-] + \frac{1}{2} \|\mathbf{w}\|^2 \tag{13}$$

with the regularization constant C is minimized through the adjustment of weight vector $\mathbf{w}$ with the following constraints:

$$y_i \leq f_{\text{SVR}}(\mathbf{x}_i) + \epsilon + \xi_i^+ \tag{14}$$

$$y_i \geq f_{\text{SVR}}(\mathbf{x}_i) - \epsilon - \xi_i^-, \tag{15}$$

where an ϵ-tube around the prediction function represents an error margin below of which errors are not penalized; thus, for each training output y_i, a quantitative representation on the degree of being outside the ϵ-tube is given. In addition to those constraints,

positivity constraints $\xi_i^+ \geq 0$ and $\xi_i^- \geq 0$ for points outside the ϵ-tube as well as $0 \leq \alpha_i \leq C$ are satisfied [66,67]. Consequently, predictions only depend on a subset of the training data which are the so-called support vectors [62].

In this work, SVR was implemented with a radial basis function as a kernel, ϵ was kept at 0.1, γ adjusted according to the *scale* technique within the *scikit-learn* library and the optimal value of C determined via grid search with LOOCV on the training set within a logarithmic spaced range of [−1,3] with 1000 increments.

2.4.6. Neural network regression

ANNs can be used to map any non-linear function between provided inputs and outputs [68]. The core concept of a perceptron was to imitate neural cell behaviour in the nervous system of the human brain [69]. Combining multiple perceptrons into one or more layers of neurons represent the architectural basis for feed forward neural networks, where each neuron in each layer computes an output via an inherent non-linear activation function and the input provided from the previous layer. Neurons are connected between layers via weight-edges. When all neurons from one layer are connected via edges to all neurons of the following layer and this applies to the complete network, one obtains a fully-connected or dense neural network. The signal is processed forward from inputs through the network to outputs in unidirection, progressively transforming the input signal into the output signal. The overall function of an ANN with k number of layers can be expressed through:

$$y = f_{\mathrm{NN}}(\mathbf{x}) = f_k(...(\boldsymbol{f}_2(\boldsymbol{f}_1))), \tag{16}$$

where the output y depends on inputs $\mathbf{x} = [x_1, x_2, \dots, x_D]^T$ of size D. The outer function f_k of k number of layers represents a scalar function. The nested function $f_{\mathrm{NN}}(\mathbf{x})$ consists of individual layer functions $[\boldsymbol{f}_1, \boldsymbol{f}_2, \dots, \boldsymbol{f}_{k-1}, f_k]$, where the vector functions for layer n are defined as follows:

$$\boldsymbol{f}_n(\mathbf{z}) = \boldsymbol{g}_n(\mathbf{W}_n\mathbf{z} + \mathbf{b}_n), \tag{17}$$

with activation function $\boldsymbol{g}_n$, weight matrix $\mathbf{W}_n$, bias weight $\mathbf{b}_n$ and layer arguments $\mathbf{z}$. For the individual number of nodes u per layer, the respective output of $\boldsymbol{f}_n(\mathbf{z})$ can be represented by the vector $[g_n(a_{n,1}), g_n(a_{n,2}), \dots, g_n(a_{n,u})]^T$, with $a_{n,u} = \mathbf{w}_{n,u}\mathbf{z} + b_{n,u}$ and the scalar activation function g_n.

During training of ANNs, back propagating error minimization via gradient descent is performed, where the error caused by the discrepancy between current network output and true output is minimized by gradually modifying the individual weights of connecting edges between neurons; thus, altering the behaviour of those neurons. In particular, the mean-squared-error loss function is defined as:

$$\mathbf{E} := \underset{\boldsymbol{\mathcal{W}}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left[y^{(i)} - f_{\mathrm{NN}}\left(\mathbf{x}^{(i)}; \boldsymbol{\mathcal{W}}\right) \right]^2, \tag{18}$$

with $y^{(i)}$ as the true output, $f_{\mathrm{NN}}\left(\mathbf{x}^{(i)}; \boldsymbol{\mathcal{W}}\right)$ as the predicted value and the network weights $\boldsymbol{\mathcal{W}} = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{n+1}\}$.

The weights are initially set to random values and the step size per weight update during gradient descent is determined by the learning rate. In this work, an adaptive learning rate was implemented via the Adam optimizer [70]. For determination of the optimal number of neurons u and number of layers n, the *Keras tuner* tool [71] was used to execute a random search in ranges of $[1, D]$ and $[1, 3]$ for u and n, respectively, with 100 randomly selected combinations and three executions per variation, as the same model could reach a different local minimum depending on the particular random weight initialization. Hyperparameters that remained constant were the *Adam* optimizer, a sigmoid activation function on hidden layers, and a linear activation function on output layer. In addition, input and output values were normalized to ranges [−1,1] and [1,5], respectively. The batch size was set to one and the MSE was used as loss function. To avoid overfitting on the training data, a maximum of 5000 epochs is set together with the *earlystopping* criterion, which is met when the prediction performance, i.e. the MSE loss on the validation data set (outside the training data set), no longer improves during training, even if the error on the training set continues to decrease; thus, training is terminated and the last best model used for further deployment.

2.5. Model performance evaluation and best model selection

All trained models are evaluated based on three prediction performance measures: maximum absolute values of the relative error *err*, MSE and determination coefficient R^2. The absolute value of the relative error *err* is given by:

$$err := \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{19}$$

with true values y_i and predicted values $\hat{y}_i$, where $\{i \in \mathbb{R} : 1 \leq N\}$. The determination coefficient R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left[y_i - \hat{y}_i \right]^2}{\sum_{i=1}^{N} \left[y_i - \bar{y} \right]^2}, \tag{20}$$

where $\bar{y}$ represents the mean of the true values. And the MSE is determined via:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left[y_i - \hat{y}_i \right]^2. \tag{21}$$

To obtain models with optimal and similar prediction performance on both training and validation sets, i.e. to prevent overfitting and underfitting; thus, achieve good generalization, the best model with the lowest maximum of the relative error on the validation set, where the score on the training set is similar, is selected for further data processing and to evaluate the ability to generalize on the test sets.

2.6. Explanation of feature dependence via Shapley Additive Explanation Values

In general, there is a trade-off in machine learning between model interpretability and model complexity. The higher the complexity in the given input data and underlying patterns, the higher is the required number of model parameters that leads to increased model complexity, which, in turn, obstruct any direct approach to explain the model's decision making process. Explainability refers to the human ability to understand the importance of individual features and the quantitative dependence between relevant features that lead to a model prediction. Additive feature contribution methods are based on an approximation of complex models by a simpler and explainable model to obtain interpretable decisions and are applicable to a vast range of different machine learning techniques. In this work, SHAP values [46] are used to explain the decisions made by the predictive models. These explanations are then related to human observations reported in literature to allow further understanding of the FS process. The approach is based on the game theory concept of Shapley values [72], where reproducing the predicted outcomes through the model represents the rules of a game and the model's features act as the players. SHAP values quantify the contribution of each feature to the model predictions, similar to evaluating each player's contribution to a game, where one prediction is equivalent to one game.

The importance of an individual feature is computed based on its mean marginal contribution, which is calculated by considering the differences between results produced by including and by excluding that feature among all possible combination of features. That way, interaction effects between features are considered. Since the permutation of all features would require computational costs that are currently considered unfeasible, SHAP values are based on approximations with significantly reduced number of samples. Essentially, SHAP values are summed up over all samples for each feature and features with large absolute SHAP values are important.

In particular, the approximation of the original model f is used to yield a simpler explanation model g that can be interpreted. For that, the input x is simplified to x' via a mapping function $x = h_x(x')$ where it is ensured that $g(z') \approx f(h_x(z'))$ when $x' \approx z'$. Then, the approximation of any model in the form of an explanatory model can be achieved via:

$$g(z') = \phi_0 + \sum_{j=1}^{P} \phi_j z'_j \tag{22}$$

where $z' \in \{0, 1\}$ is the simplified feature, being either 0 or 1 when the related feature is excluded or included, respectively, P the number of simplified features, $\phi_j \in \mathbb{R}$ represents the Shapley values, i.e. the attribution value of each feature, and the constant ϕ_0 consists of the predicted mean value of all training samples. The estimation of the Shapley value ϕ_i for feature i is achieved by computing the weighted average of all possible differences with:

$$\phi_i = \sum_{S \subseteq K \setminus \{i\}} |S|!(|K| - |S| - 1)![f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)]\frac{1}{|K|!}, \tag{23}$$

where K is the set of all features, $S \subseteq K$ all feature subsets, x_S the input feature values in subset S and ! the factorial. Essentially, the effect of a feature on the models prediction, i.e. the feature importance, is computed by training one model $f_{S\cup\{i\}}$ where that feature was included and another model f_S where it was excluded. The computation of SHAP values not only depends on the computation of Shapley values but also on the nature of the original model to be interpreted, since modifications in the specific computation are made. For more details, the reader is referred to [46]. In principal, there are different *explainers* available such as linear, tree, kernel and deep *explainers* that were used here in accordance to the nature of the model to be interpreted. In this work, the approximation of SHAP values has been executed via the SHAP library proposed by Lundberg et al. [73]. It is essential to point out that SHAP values only allow for interpretations of the predictive models and should not be used for the inference of causal relationships.

3. Results & discussion

In this section, prediction performance measures of the best selected models for each target are presented and discussed (hyperparameters and all performance measures can be found in Appendix B). Predicted and true values are compared and model explanations provided for two sets of targets: the process behaviour variables T_{max}, v_{cr} and M, as well as the deposit geometry variables t and w. Performance measures of all trained models where upon the best models were selected can be found in Appendix C. Overall, the prediction of each target is based on different relationships towards the inputs; therefore, the best selected model can vary in complexity to represent those relations to the best degree. It will be assessed, whether this approach is suitable for the given use case.

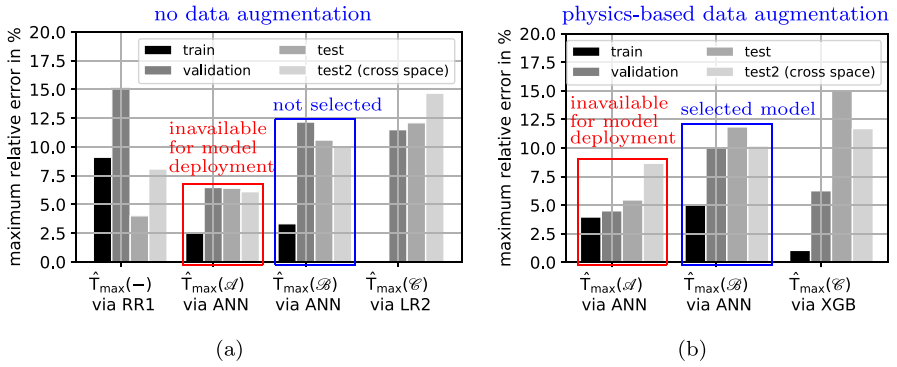


Fig. 5. Predicted maximum temperature $\hat{T}_{max}$ based on feature space (–), where t, w, v_{cr} and M are not additional features, ($\mathscr{A}$) with true values of t, w, v_{cr} and M, which are not available for real application; thus, need to predicted based on $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ features for test predictions in ($\mathscr{B}$) and for training and testing in ($\mathscr{C}$), respectively. Results based on original data sets (a) and with exploitation of numerical heat transfer model via physics-based data augmentation (b).

3.1. Prediction of process behaviour

3.1.1. Maximum temperature

For the prediction of T_{max}, the influence of different sets of features is evaluated. Those feature sets are denoted with (–), ($\mathscr{A}$), ($\mathscr{B}$) and ($\mathscr{C}$), as described in Section 2.3 and listed in Table 3. In this context, using true values of the variables v_{cr}, M, t and w in feature set ($\mathscr{A}$), serves as a reference benchmark to evaluate how far the prediction errors can be reduced when the truth is used as input. Nevertheless, for a real use-case application, these true variables are unknown and inavailable for model deployment; thus, the models where true values are used for evaluation cannot be selected as best models. However, those models can be used further, evaluated and deployed when predicted values (of v_{cr}, M, t and w) based on the input space (–), consisting of $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ variations, are used as input, which is equivalent to feature set $\mathscr{B}$. Consequently, true values of input variables v_{cr}, M, t and w in feature set ($\mathscr{A}$) are replaced by their best available predictions $\hat{v}_{cr}$, $\hat{M}$, $\hat{t}$ and $\hat{w}$ for $\hat{T}_{max}(\mathscr{B})$ in Fig. 5(a), whereby the errors are increased by 0.8%, 5.7%, 4.2% and 4.0% on the training, validation, test and test2 (cross space) data sets, respectively. However, using predicted variables also for training in feature set $\mathscr{C}$ leads to overfitting as the training error is close to zero, whereas the errors on validation, test and test2 sets range from 11.5% to 14.7%, respectively, see $\hat{T}_{max}(\mathscr{C})$ in Fig. 5(a). In general, the usage of v_{cr}, M, t and w as features for the prediction of T_{max} is beneficial with respect to an enhanced prediction performance, even when noise is added to those variables by replacing their values with their predicted equivalents that contain errors.

For the exploitation of the numerical heat transfer model, the variables v_{cr}, M, t and w are required inputs as described in Section 2.2.5. The usage of the corresponding additional samples in training and validation data sets enabled predictions with errors shown in Fig. 5(b). The lowest error on the validation set can be observed on $\hat{T}_{max}(\mathscr{A})$, which is slightly increased by replacing true with predicted variables after training in $\hat{T}_{max}(\mathscr{B})$. Using predicted values also for training in $\hat{T}_{max}(\mathscr{C})$ leads to overfitting on the training set and relatively poor generalization, i.e. high errors on test and test2 sets, which is similar to observations for $\hat{T}_{max}(\mathscr{C})$ predictions without numerical samples in training and validation data sets, see Fig. 5(a). As a result, the predictive model for $\hat{T}_{max}(\mathscr{B})$ based on numerically enriched training and validation sets was selected as best model, which is an ANN that shows similar error on training and validation sets. The comparison of true and predicted values, see Fig. 6(a), exhibits good agreement on the training, validation and test data sets as well as on the test2 (cross-space) data set where variables in both $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ spaces were varied simultaneously, indicating good generalization. Further improvement of the prediction performance could be limited due to the simplifications contained in the heat transfer models as well as in Eqs. (1)–(3), respectively.

The feature dependence of the ANN used to compute $\hat{T}_{max}$ is shown in Fig. 6(b). The highest impact is attributed to the thickness t, where low values of t correlate to high values of $\hat{T}_{max}$ and vice versa, which is in agreement with experimental observations [8]. The 2nd and 4th most important impacts are achieved by the thermal conductivity of the backing plate k_b and the substrate k_s, respectively, where low values of k_b and k_s correspond to high values of $\hat{T}_{max}$, possibly caused by the reduced heat flow. Furthermore, high values of specific heat capacities c_b and c_s, ranked as 6th and 10th most important feature, respectively, correlate to high values of $\hat{T}_{max}$, likely to be caused by the related heat accumulation. For the feed rate v_{cr}, ranked as the 3rd most important feature, high values also correlate with high values of $\hat{T}_{max}$ that facilitates deformation and flow of stud material, which can be enforced by increased force F and rotation speed RS that both relate to high values of $\hat{T}_{max}$, which is also in agreement with experimental results [7,8]. The first process parameter TS is ranked as the 5th most significant feature, where high values of TS correlate with low values of $\hat{T}_{max}$, as a faster travel speed leads to a shorter duration of the process and lower process temperatures are reached compared to a slower travel speed. The impacts of the variables c_s, t_s, t_b and w based on the SHAP values are only minor, in comparison to the impact of the other variables. For w, this weaker impact on $\hat{T}_{max}$ is also in agreement with experimental measurements performed by Kallien et al. [8], where correlations between w and T_{max} as well as t and T_{max} have been approximated linearly and the correlation of w to T_{max} was also weak in comparison to the correlation between t and T_{max}. Overall, the feature dependence appears to be consistent with results reported in literature; thus, the interpretations can be considered trustworthy and the presented results allow

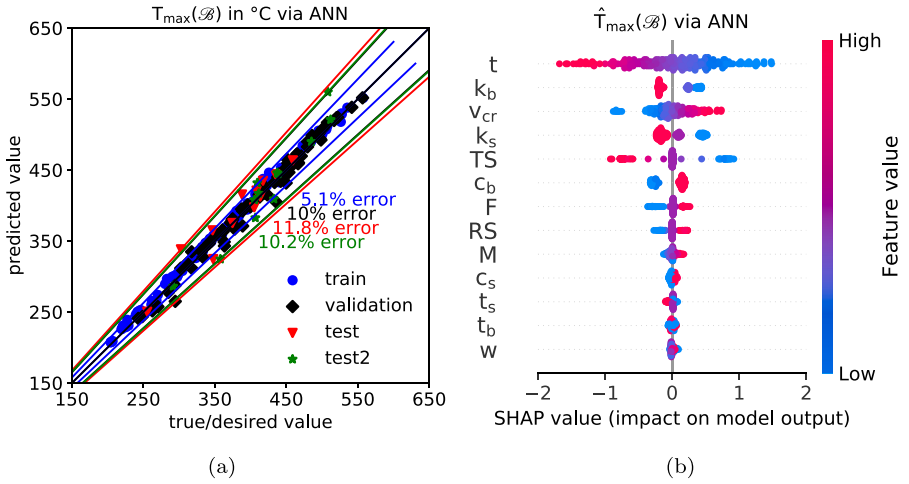


Fig. 6. Predicted and true values of T_{max} enriched with additional samples generated via physics-based data-mining (a) and the corresponding feature dependence of the ANN regression model based on SHAP values (b).

a qualitative assessment of important features that influence $\hat{T}_{max}$ during FS. In particular, the impacts of t, k_b, v_{cr} and k_s on $\hat{T}_{max}$ outperforming the importance of process parameters TS, F and RS can enhance the understanding of the FS process. Nevertheless, the feature dependence is based on correlations made by predictive models showing acceptable prediction performance and they do not represent causal relationships.

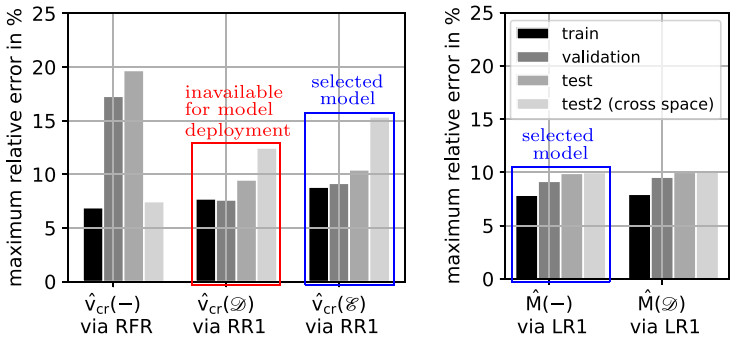
3.1.2. Feed rate

In search of the best predictive model for the feed rate v_{cr}, the effect of using T_{max} as additional input feature, next to the features contained in $\boldsymbol{PP}$ and $\boldsymbol{MAT}$, is investigated with feature set ($\mathscr{D}$). It can be observed that T_{max} as additional feature leads to a lower error in comparison to excluding it from the input space, see Fig. 7(a). For feature set (–) the errors amount to 6.9%, 17.3%, 19.7% and 7.5% on training, validation, test and test2 sets, respectively. This indicates overfitting to the training samples and poor generalization on validation and test sets. Furthermore, when using T_{max} as feature, the replacement of the FE-computed true values of T_{max} by predicted values $\hat{T}_{max}$ in feature set ($\mathscr{E}$) only leads to a minor error increase on training, validation, test and test2 (cross space) data sets by 1.1%, 1.6%, 0.9% and 2.9%, leading to absolute maximum errors of 8.8%, 9.2%, 10.4% and 15.4%, respectively. Considering the experimental error for v_{cr} of 3.5%, the agreement between the true and predicted values of v_{cr} is acceptable. The benefit of using feature set $\mathscr{E}$ is that the need of computationally expensive FE results is circumvented by obtaining those values through ML predictions that exhibit higher efficiency.

The feature dependence of this ridge regression model to compute $\hat{v}_{cr}$, see Fig. 8(b), reveals that $\hat{T}_{max}$ is the most influential feature, where high values of $\hat{T}_{max}$ correlate with high values of $\hat{v}_{cr}$, which was also shown in Fig. 6(a) and discussed in the previous Section 3.1.1, indicating that high temperature facilitates materials deformation and flow [6,8]. There is a strong influence by the substrate and backing material properties k_s, k_b and c_b, which are ranked 2nd, 3rd and 4th most important features, respectively. In particular, high values of k_s and k_b correspond to high values of $\hat{v}_{cr}$, whereas high values of c_b correlate to low $\hat{v}_{cr}$ values. In other words, increased thermal conductivity and decreased specific heat can enable low process temperatures, which, in turn, facilitate materials deformation and flow resulting in larger feed rates. Minor impacts on high values of $\hat{v}_{cr}$ are also given by high values of t_s and t_b. The process parameters RS, TS and F are ranked 5th, 6th and 7th most important features, respectively, where high values of RS relate to low values of $\hat{v}_{cr}$, whereas high values of TS and high values of F correlate with high values of $\hat{v}_{cr}$, which agrees with [1,30]. Intuitively, one could think that F has a higher impact on $\hat{v}_{cr}$ than it does here, but such a small contribution of F to $\hat{v}_{cr}$ was also shown within the numerical simulations in [74]. The almost negligible impact of c_s on $\hat{v}_{cr}$ can be attributed to the rather simple ridge regression prediction model where the simplicity in the mapping between features and target might be in disagreement with the more complex physical reality that may be insufficiently represented in the scarce data sets; thus, by the model. In addition, the effect of changes in c_s could have been represented indirectly through considering k_s since values of both variables are only changed simultaneously due to the chosen materials. Overall, the predictive model for the feed rate $\hat{v}_{cr}$ is useful with its acceptable prediction errors and its feature dependence that reinforces the high impact of T_{max} as well as the thermal properties of substrate and backing materials.

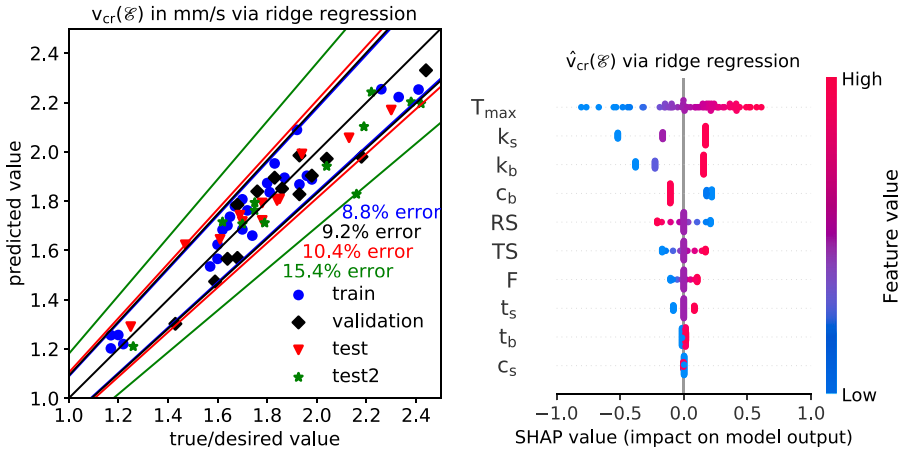
3.1.3. Torque

For the prediction of the torque M, using T_{max} as additional feature in feature set ($\mathscr{D}$) does not lead to any improvement of prediction performance in comparison to excluding it as feature in (–), see Fig. 7(b). One reason for this can be the comparatively simple linear regression model that was selected as best predictor in both cases, where the addition of T_{max} as a 10th feature was ranked the least important feature by the model and did not provide any more useful information for a linear approximation. For this



(a) (b)

Fig. 7. Prediction errors on training, validation, test and test2 data sets for the predicted feed rate $\hat{v}_{cr}$ in (a) with (–), i.e. without T_{max}, with ($\mathscr{D}$) that includes T_{max} and with ($\mathscr{E}$) that contains $\hat{T}_{max}$ as additional feature as well as for the predicted torque $\hat{M}$ in (b) with (–), i.e. without T_{max} and with ($\mathscr{D}$) that includes T_{max} as additional feature.



(a) (b)

Fig. 8. Predicted and true values for $v_{cr}(\mathscr{E})$ where the predicted maximum temperature $\hat{T}_{max}$ is an additional feature besides ***PP*** and ***MAT*** variables (a) and the corresponding feature dependence of the ridge regression model based on SHAP values (b).

reason, there was no need to replace T_{max} with $\hat{T}_{max}$, since there would have been no further improvement in prediction performance. The agreement between true and predicted values is acceptable since the errors on training, validation and test data sets are very similar in a range between 7.9% and 9.9%, see Fig. 9(a), and the experimental error amounts to 5.8%. In addition, the error on the test2 data set amounts to 10%, exhibiting similar and acceptable generalization.

The feature dependence of the predictive model to compute $\hat{M}$ in Fig. 9(b) shows that RS and F are most important features, where low values of RS and high values of F correlate to high values of $\hat{M}$. Additionally, high values of TS correspond to high values of $\hat{M}$. Furthermore, high values of material properties k_s and k_b correspond to high values of $\hat{M}$, whereas the impact of other ***MAT*** properties on $\hat{M}$ appears to be negligible. This minor variation of M is in agreement with the experimental measurements of M in the ***MAT*** space DoE where values remained within the range between 28.2 N m and 31.2 N m for all ***MAT*** variable variations, which supports the less significant influence of ***MAT*** features on M. Overall, the linear regression model can be used for predictions of M despite its high simplicity because it represents fundamental relationships with acceptable prediction performance and provides useful feature interpretations.

3.2. Prediction of deposit geometry

3.2.1. Thickness

For the prediction of the deposit thickness t, the lowest error on the validation set was achieved based on inputs (–) that neither contained T_{max}, v_{cr} nor M as features, see Fig. 10. Thus, for the best predictive model to compute $\hat{t}$, an ANN in this case, the T_{max} included in ($\mathscr{D}$) as feature was not beneficial for its performance despite the high correlation between $\hat{T}_{max}$ and t that was identified

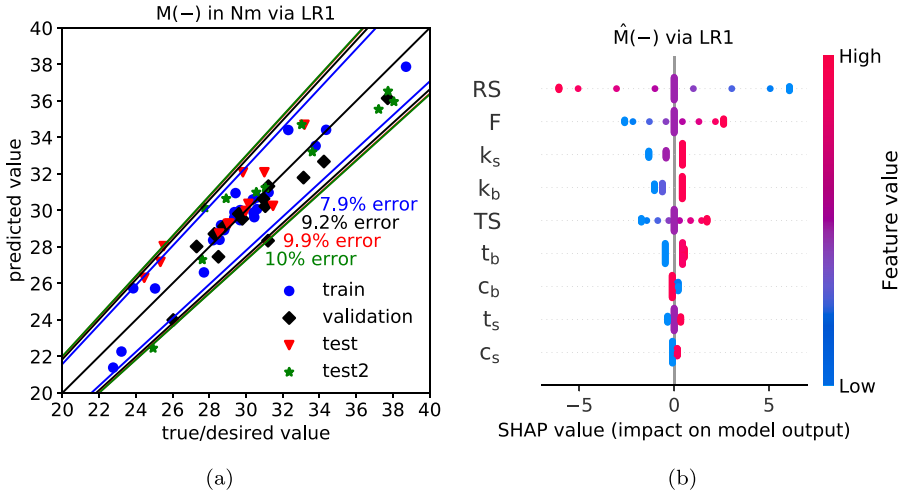


Fig. 9. Predicted and true values for $M(-)$ where only $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ variables were used as features (a) and the corresponding feature dependence of the linear regression model based on SHAP values (b).

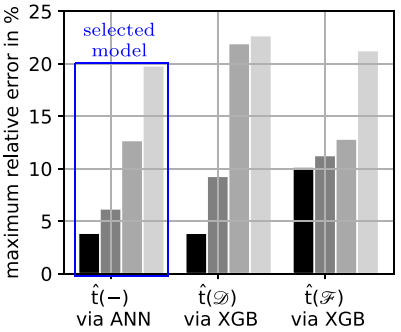


Fig. 10. Prediction errors on training, validation, test and test2 data sets for the predicted thickness $\hat{t}$ without T_{max} in $(-)$, with T_{max} in $(\mathscr{D})$ and with T_{max}, v_{cr} and M in $(\mathscr{F})$ as additional features, respectively.

by the model in Section 3.1.1 for the prediction of T_{max}. Likewise, the further addition of v_{cr} and M as features in $(\mathscr{F})$ did not improve the prediction performance either. A possible explanation for this could lie in the scarcity of the data whereupon a model with increased complexity, i.e. more internal parameters, due to the increased number of features, can only insufficiently map relevant relationships to achieve satisfactory predictions. Essentially, the ANN predictor for $\hat{t}(-)$ achieves very good agreement between predicted and true values on the training, validation and test sets, as the errors amount to 3.9%, 6.2% and 12.7%, respectively, which is below an experimental error of 15.4%. This is accompanied with an acceptable generalization since the predictive error on the cross space test2 set amounts to 19.8%, see Fig. 11(a).

According to the feature dependence shown in Fig. 11(b), k_s is considered most important, where high values relate to high values of $\hat{t}$. This is similar to the feature dependence of v_{cr}, where k_s was ranked 2nd most important feature, see Fig. 8(b), as the corresponding increased material flow can contribute to high values of $\hat{t}$. Moreover, this is supported by an additional dependence of high values of $\hat{t}$ on low values of RS, which is also the case for high values of v_{cr}. Ranked 2nd and 3rd most important features are process parameters RS and TS, with low values corresponding to high values of $\hat{t}$, respectively. This is also true for F, whose importance rank is second last out of 9 features. Those relations between $\hat{t}$ and process parameters is in agreement with several experimental studies [1,8,30–32]. For the other substrate and backing material properties, high values of k_b as well as high values of t_b and t_s correlate to high values of $\hat{t}$, whereas low values of c_b and c_s are associated with high values of $\hat{t}$. These impacts of properties indirectly correspond to lower temperatures, which is in agreement with [1,8] and the feature dependence of the predictive model that computes $\hat{T}_{max}$ in Section 3.1.1, see Fig. 6(b). In general, there is a significant impact by thermal material properties of substrate and backing plate on specific values of $\hat{t}$, besides the influence of the process parameters.

3.2.2. Width

The best models selected for the prediction of the deposit width w are compared in Fig. 12. It can be observed that using only features based on $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ variables in $(-)$ is most beneficial for prediction performances with respect to reaching low values of maximum relative errors on the validation set, in comparison to considering T_{max} in $(\mathscr{D})$ or $\hat{T}_{max}$ together with v_{cr} and M in

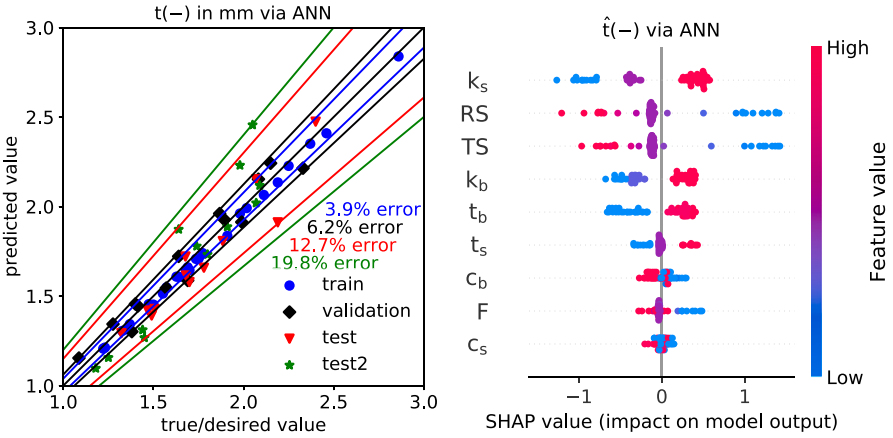


Fig. 11. Predicted and true values for $t(-)$ where only ***PP*** and ***MAT*** variables comprised the input features (a) and the corresponding feature dependence of the ANN regression model based on SHAP values (b).

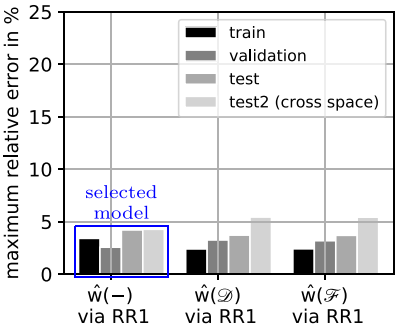


Fig. 12. Prediction errors on training, validation, test and test2 data sets for the predicted width $\hat{w}$ without T_{max} in $(-)$, with T_{max} in $(\mathscr{D})$ and with T_{max}, v_{cr} and M in $(\mathscr{F})$ as additional features, respectively.

$(\mathscr{F})$, respectively. This is similar to the results for computing $\hat{t}$ and can also be explained by the higher number of features leading to increased model complexity, i.e. internal parameters that need to be adjusted, where the scarce data only enables insufficient training. The predicted values of the best selected ridge regression model are compared against the true values in Fig. 13, where errors amount to 2.6%, 3.4%, 4.2% on training, validation and test data sets as well as to 4.3% on the generalization test2 data set. Those predictions are in very good agreement with experimental measurements considering the greater experimental maximum relative error obtained from the three replicas that amounts to 6.2%.

The feature dependence of the ridge regression model, see Fig. 13(b), shows that k_s is the most important feature, similar to predictions of t; however, while low values of k_s relate to low values of $\hat{t}$, they correspond to high values of $\hat{w}$, here. Similar inverse relations can be observed for the other material properties t_b, k_b, c_s and t_s, which are ranked as 3rd, 4th, 8th and least (9th) most important features, respectively, where low values agree with high values of $\hat{t}$. This could indicate that those variables facilitate rather low temperature, which is in agreement with [8]. The process parameter RS is 2nd most important feature for predictions of w, again with low values of RS correlating with high values of $\hat{w}$, which is in parallel to predictions of t. F and TS are ranked 5th and 6th most important features, respectively, where high values of F and low values of TS relate to high values of $\hat{w}$, respectively, which is in agreement with [8,31,32]. Overall, the comparatively simple ridge regression model for the prediction of w achieves prediction error below the experimental error and provides qualitative assessments of feature impacts, reconfirming the important relations of process temperatures and temperature-related properties of substrate and backing plate materials besides the influence of process parameters; therefore, this model can also serve as additional building block to enable further understanding of the FS process.

3.2.3. Final remarks

The heat-transfer model was successfully exploited to circumvent any need to perform experimental temperature measurements. The computational costs required by the numerical FE model were saved by the building a surrogate model via ML. Even though, T_{max} as feature was only beneficial for the prediction of v_{cr}, the indirect influence of the temperature via thermal conductivity of

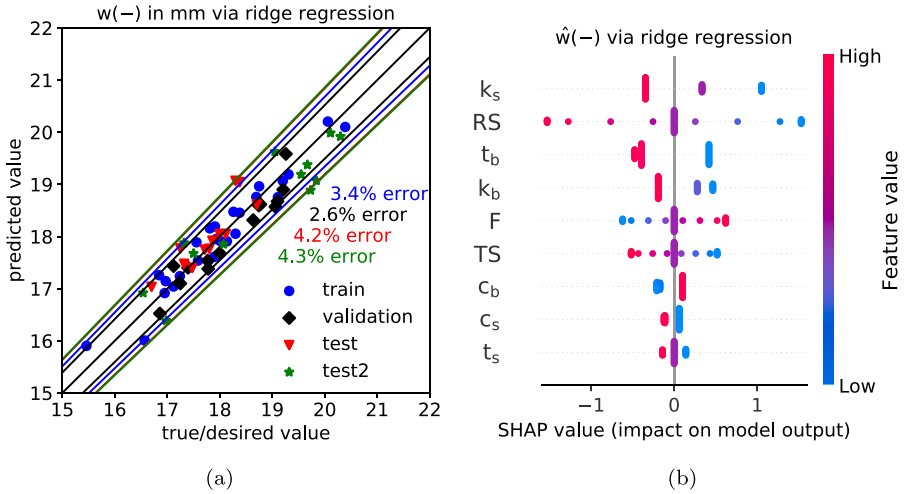


Fig. 13. Predicted and true values for $w(-)$ where only $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ variables comprised the input features (a) and the corresponding feature dependence of the ridge regression model based on SHAP values (b).

substrate k_s and backing plate k_b is clearly shown as they rank among the top four most important features for the prediction of v_{cr}, M, t and w, respectively. In particular for similar and opposite control of t and w, the influences of $\boldsymbol{PP}$ and $\boldsymbol{MAT}$ variables should be tailored according to their parallel and contrasting feature dependencies. For $\boldsymbol{PP}$, low values of RS and TS increase both t and w values, whereas high values of F correlate to decreased t but increased w. For $\boldsymbol{MAT}$, k_s, k_b, t_s, t_b and c_b are controversial with high values correlating with increased t but decreased w, respectively, because all enable increased process temperatures.

4. Conclusion

In this study, the identification and utilization of impacts from process parameters as well as substrate and backing material properties on the resulting process behaviour and deposit geometry for FS has been achieved. The prediction of T_{max} is in good agreement with target values obtained from the FE heat transfer model. In this regard, the physical data-mining of the FE model was successfully used to increase prediction performance in comparison to excluding this data. Due to the defined inputs of this FE model, the features consisting of the feed rate v_{cr} and torque M as well as deposit geometries t and w are also required and first need to be predicted for a sufficient prediction of T_{max} via ML. In general, the predictions of all targets T_{max}, v_{cr}, M, t and w showed acceptable agreement with their true solution despite the highly scarce amount of available data and the inherent experimental scatter. All ML models are trained on data where process parameters are varied while substrate and backing material properties are kept constant and vice versa but also exhibit acceptable generalization on the cross-space where both are varied simultaneously. Based on those ML models, explanations via feature importance and dependence were largely found to be in agreement with experimental observations in literature but also provide insights that can enable further understanding of the FS process. On the one hand, the use of T_{max} as additional feature led to a decrease in prediction errors only for the prediction of v_{cr} and not for the prediction of M, t and w, respectively, despite the known impact of T_{max} on those targets. In this regard, the scarce data available for training could have limited the increased model complexity required to sufficiently map relevant relationships with the increased number of features. On the other hand, an indirect influence of T_{max} was identified and utilized by attributing high importance to the thermal conductivity of the substrate k_s and backing plate k_b for the predictive models of v_{cr}, M, t and w as well as for the prediction of T_{max} itself, as they rank among the top four most important features (out of a total of 9 to 12 features) for the prediction of T_{max}, v_{cr} and M, as well as the most important feature for t and w, respectively. Ultimately, this qualitative assessment of feature dependence on process behaviour and deposit geometry targets can serve as additional key to unlock further understanding of the FS process.

Declaration of competing interest

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

Data availability

The data and *Python*-codes used for this research are publicly available on the *Zenodo* online platform and can be found with DOI: 10.5281/zenodo.8263026.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 101001567).

Appendix A

The prediction performance achieved for thickness $t_{lin}(T_{max})$, width $w_{lin}(T_{max})$ and feed rate $v_{cr,lin}(T_{max})$ via the linear approximations stated in Eqs. (1), (2) and (3), respectively, are listed in Table A.1.

Table A.1
Mean absolute error (MAE) achieved via Eqs. (1), (2) and (3) for thickness t, width w and feed rate v_{cr}, respectively, on training, validation, test and test2 data sets. The experimental (exp.) error is based on three replica samples, i.e. the centre points of the Designs of Experiments (DoE).

MAE	t_{lin} [mm]	w_{lin} [mm]	$v_{cr,lin}$ [mm/s]
Training	0.47	1.01	0.26
Validation	0.33	0.58	0.10
Test	0.34	0.74	0.21
Test2	0.51	1.00	0.15
exp. error	0.27	1.06	0.06

Appendix B

The hyperparameters of the best models selected for the prediction of T_{max}, v_{cr}, M, t and w are provided below, respectively.

Maximum temperature T_{max}

The employed ANN consists of an input layer with 13 neurons (according to the number of features), one hidden layer with 14 neurons and one neuron on the output layer, yielding a total number of 393 trainable model parameters. Prediction performance measures of this model are listed in Table B.1.

Table B.1
Performance measures on training, validation, test and test2 sets: Determination coefficient (R^2), mean absolute error (MAE) and the maximum of the relative error (*err*). of the best model (ANN) to predict T_{max}.

ANN → T_{max}	R^2 [–]	MAE [°C]	Maximum *err* [%]
Training	0.993874	5.630336	5.079982
Validation	0.984805	7.234900	10.042205
Test	0.903396	13.569292	11.846732
Test2	0.864374	16.419428	10.179453

Feed rate v_{cr}

The 1st-order ridge regression model with an $\alpha = 8.93$ to yield $\hat{v}_{cr}$ is described with:

$$\begin{aligned}\hat{v}_{cr} = {} & 5.22664062 \cdot 10^{-2} \cdot F - 4.17564606 \cdot 10^{-2} \cdot RS + 8.50486799 \cdot 10^{-2} \cdot TS \\ & + 8.40810129 \cdot 10^{-3} \cdot k_s - 1.35502479 \cdot 10^{-10} \cdot c_s + 4.1297693 \cdot 10^{-2} \cdot t_s \\ & + 3.56202548 \cdot 10^{-3} \cdot k_b - 8.43630798 \cdot 10^{-10} \cdot c_b + 1.38157328 \cdot 10^{-2} \cdot t_b \\ & + 5.09642138 \cdot 10^{-3} \cdot T_{max} \\ & - 1.65124793\end{aligned} \tag{B.1}$$

The prediction performance measures of the models described by Eq. (B.1) above are listed in Table B.2 below.

Table B.2
Performance measures on training, validation, test and test2 sets: Determination coefficient (R^2), mean absolute error (MAE) and the maximum of the relative error (*err*). of the best model (RR) to predict v_{cr}.

RR → v_{cr}	R^2 [–]	MAE [°C]	Maximum *err* [%]
Training	0.936734	0.066619	8.807957
Validation	0.834654	0.092736	9.165473
Test	0.927930	0.058686	10.419877
Test2	0.828632	0.102524	15.354523

Torque M

The 1st-order linear regression model that yields $\hat{M}$ is described via:

$$\begin{aligned}\hat{M} &= 1.304375 \cdot F - 1.215 \cdot RS + 8.64375 \cdot 10^{-1} \cdot TS + 2.16567448 \cdot 10^{-2} \cdot k_s \\ &+ 6.10763765 \cdot 10^{-9} \cdot c_s + 1.725 \cdot 10^{-1} \cdot t_s + 9.7565104 \cdot 10^{-3} \cdot k_b \\ &- 8.16240131 \cdot 10^{-10} \cdot c_b + 4.56909341 \cdot 10^{-1} \cdot t_b + 23.88753081.\end{aligned} \tag{B.2}$$

The prediction performance measures of the models described by Eq. (B.2) above are listed in Table B.3 below.

Table B.3
Performance measures on training, validation, test and test2 sets: Determination coefficient (R^2), mean absolute error (MAE) and the maximum of the relative error (*err*). of the best model (RR) to predict M.

LR → M	R^2 [–]	MAE [°C]	Maximum *err* [%]
Training	0.948422	0.673526	7.854310
Validation	0.814933	0.968677	9.157729
Test	0.693845	1.100085	9.905198
Test2	0.882765	1.413958	9.985394

Thickness t

The employed ANN consists of an input layer with 9 neurons (according to the number of features), two hidden layers with 10 and 9 neurons, respectively, as well as one neuron on the output layer, yielding a total number of 299 trainable model parameters. The prediction performance measures of this model are listed in Table B.4 below.

Table B.4
Performance measures on training, validation, test and test2 sets: Determination coefficient (R^2), mean absolute error (MAE) and the maximum of the relative error (*err*). of the best model (RR) to predict t.

ANN → t	R^2 [–]	MAE [°C]	Maximum *err* [%]
Training	0.991774	0.032971	3.856503
Validation	0.954308	0.070950	6.176544
Test	0.867477	0.093444	12.690104
Test2	0.688787	0.132149	19.808895

Width w

The 1st-order ridge regression model with $\alpha = 0.61$ to yield $\hat{w}$ is described with:

$$\begin{aligned}\hat{w} &= 3.10334253 \cdot 10^{-1} \cdot F - 3.06066497 \cdot 10^{-1} \cdot RS - 2.58816314 \cdot 10^{-1} \cdot TS \\ &- 1.70556367 \cdot 10^{-2} \cdot k_s - 4.32295718 \cdot 10^{-9} \cdot c_s - 7.03092837 \cdot 10^{-2} \cdot t_s \\ &- 4.37306152 \cdot 10^{-3} \cdot k_b + 8.03448114 \cdot 10^{-10} \cdot c_b - 4.06880347 \cdot 10^{-1} \cdot t_b \\ &+ 33.59430335\end{aligned} \tag{B.3}$$

The prediction performance measures of the models described by Eq. (B.3) above are listed in Table B.5 below.

Table B.5
Performance measures on training, validation, test and test2 sets: Determination coefficient (R^2), mean absolute error (MAE) and the maximum of the relative error (*err*). of the best model (RR) to predict w.

RR → w	R^2 [–]	MAE [°C]	Maximum *err* [%]
Training	0.943655	0.209642	3.406364
Validation	0.860429	0.280106	2.564726
Test	0.567643	0.238336	4.209272
Test2	0.855633	0.439249	4.270765

Appendix C

Performance measures (maximum of relative error *err*) of all predictive models used for best model selection. One target per table; thus, values for T_{max} based on DoE data sets are listed in Table C.1 and for T_{max} based on data enriched via physics-based data-mining in Table C.2. Relative errors for v_{cr} can be found in Table C.3, for M in Table C.4, for t in Table C.5 and for w in Table C.6.

Table C.1
Maximum of relative error *err* for all models to predict T_{max} based on the original DoE data sets without sample enrichment through physics-based data-mining.

	$\hat{T}_{max}(-)$			$\hat{T}_{max}(\mathscr{A})$			$\hat{T}_{max}(\mathscr{B})$			$\hat{T}_{max}(\mathscr{C})$		
	train	val	test	train	val	test	train	val	test	train	val	test
LR1	9.1	15.1	4.0	4.4	3.9	3.5	3.3	12.3	17.8	3.3	12.3	17.8
LR2	4.5	9.1	9.5	0.0	5.1	5.4	0.0	11.5	12.1	0.0	11.5	12.1
LR3	0.0	10.5	11.8	0.0	5.8	5.6	0.0	10.7	12.5	0.0	10.7	12.5
RR	9.1	15.1	4.0	4.4	3.9	3.5	5.9	11.8	9.5	3.5	12.6	16.0
RR2	55.5	23.5	34.8	52.4	25.2	32.5	52.0	24.6	31.7	51.9	24.6	31.6
RR3	55.5	23.5	34.8	52.4	24.5	33.4	52.2	24.1	32.8	53.9	23.6	33.5
RFR	10.7	13.2	28.3	15.8	22.3	24.8	16.1	14.1	26.5	11.0	14.0	25.9
XGB	0.3	10.6	25.1	6.8	6.4	19.6	13.9	15.2	19.0	0.6	4.4	24.1
SVR	14.7	16.9	8.5	3.1	13.2	8.0	4.3	18.6	12.1	3.1	18.3	12.1
ANN	56.5	31.4	53.7	2.5	6.5	6.4	3.3	12.2	10.6	3.8	11.4	14.3

Table C.2
Maximum of relative error *err* for all models to predict T_{max} based on the inclusion of additional samples generated via physics-based data-mining, where values above 1000% have been replaced by infinity ∞.

	$\hat{T}_{max}(\mathscr{A})$			$\hat{T}_{max}(\mathscr{B})$			$\hat{T}_{max}(\mathscr{C})$		
	train	val	test	train	val	test	train	val	test
LR1	25.4	6.4	11.7	25.2	12.8	12.1	25.2	12.8	12.1
LR2	3.6	∞	∞	3.2	∞	∞	3.2	∞	∞
LR3	0.6	∞	∞	0.6	∞	∞	0.6	∞	∞
RR	25.0	6.6	13.0	25.0	10.9	10.4	24.7	11.9	10.8
RR2	16.7	8.2	11.8	16.7	9.8	8.5	20.6	13.0	12.6
RR3	72.8	45.5	36.8	72.8	45.5	36.4	63.6	37.7	31.6
RFR	30.5	21.2	13.2	32.0	11.7	21.5	19.4	16.6	16.7
XGB	1.4	6.7	20.3	13.9	14.6	16.7	1.0	6.3	15.1
SVR	4.1	18.2	9.2	5.5	19.8	11.0	4.1	19.6	10.7
ANN	4.0	4.5	5.5	**5.1**	**10.0**	**11.8**	5.9	10.4	9.5

Table C.3
Maximum of relative error *err* for all models to predict v_{cr} based on ***PP*** and ***MAT*** variables as features as well as with true values T_{max} and predicted values $\hat{T}_{max}$.

	$\hat{v}_{cr}(-)$			$\hat{v}_{cr}(T_{max})$			$\hat{v}_{cr}(\hat{T}_{max})$		
	train	val	test	train	val	test	train	val	test
LR1	16.1	16.2	7.1	9.7	7.6	7.4	9.5	13.5	5.1
LR2	16.1	16.2	7.1	6.0	18.1	9.1	6.1	16.8	7.1
LR3	16.1	16.2	7.1	6.1	29.9	18.3	6.1	24.4	14.6
RR	15.4	16.2	7.1	7.7	7.6	9.5	**7.6**	**11.1**	**8.5**
RR2	50.7	27.7	41.1	26.6	15.3	20.9	27.4	18.0	24.2
RR3	50.8	27.7	41.2	26.3	15.3	20.8	27.5	17.8	24.5
RFR	6.9	17.3	19.7	18.0	16.4	13.9	18.0	17.3	11.6
XGB	25.4	4.8	25.4	6.7	11.6	25.0	9.7	11.6	31.5
SVR	2.7	23.6	8.5	2.7	19.9	9.6	2.7	22.9	9.7
ANN	6.0	15.1	10.3	2.3	14.9	10.3	2.8	17.6	9.5

Table C.4
Maximum of relative error *err* for all models to predict M based on ***PP*** and ***MAT*** variables as features as well as with true values T_{max}.

	$\hat{M}(-)$			$\hat{M}(T_{max})$		
	train	val	test	train	val	test
LR1	**7.9**	**9.2**	**9.9**	7.9	9.5	10.0
LR2	7.9	9.2	9.9	8.9	12.7	11.1
LR3	7.9	9.2	9.9	9.0	12.7	13.5
RR	12.2	7.7	10.6	11.2	5.0	11.1
RR2	32.4	20.0	23.3	34.3	19.5	23.2
RR3	32.4	20.0	23.3	34.2	19.5	23.1
RFR	9.0	11.0	19.4	8.5	13.5	19.8
XGB	14.7	5.2	22.7	11.6	4.9	20.2
SVR	1.7	13.2	12.7	1.7	12.5	13.0
ANN	1.8	11.3	12.5	2.3	7.3	12.9

Table C.5
Maximum of relative error *err* for all models to predict t based on ***PP*** and ***MAT*** variables as features as well as with true values T_{max} and with true values T_{max}, v_{cr} and M.

	$\hat{t}(-)$			$\hat{t}(T_{max})$			$\hat{t}(T_{max}, v_{cr}, M)$		
	train	val	test	train	val	test	train	val	test
LR1	12.9	17.8	8.1	9.8	14.3	7.0	7.4	16.0	10.7
LR2	12.9	17.8	8.1	10.5	7.6	9.4	7.1	43.3	13.4
LR3	12.9	17.8	8.1	10.4	27.7	31.6	7.1	17.4	28.6
RR	13.0	18.1	8.0	11.5	16.3	7.5	7.2	15.5	10.4
RR2	47.1	41.8	35.1	44.5	35.6	30.1	29.0	24.2	19.3
RR3	47.1	41.8	35.1	44.3	35.6	30.0	31.5	26.3	21.3
RFR	21.6	17.5	14.9	11.7	18.5	18.7	16.5	14.8	11.8
XGB	5.4	9.0	18.4	3.9	9.3	21.9	10.2	11.3	12.8
SVR	3.2	16.5	10.0	3.8	16.1	9.6	3.2	19.8	9.2
ANN	**3.9**	**6.2**	**12.7**	1.5	7.1	6.4	1.2	5.5	7.1

Table C.6
Maximum of relative error *err* for all models to predict w based on ***PP*** and ***MAT*** variables as features as well as with true values T_{max} and with true values T_{max}, v_{cr} and M.

	$\hat{w}(-)$			$\hat{w}(T_{max})$			$\hat{w}(T_{max}, v_{cr}, M)$		
	train	val	test	train	val	test	train	val	test
LR1	3.4	2.5	4.3	2.3	3.4	3.8	2.4	3.1	3.9
LR2	3.4	4.4	4.3	1.7	4.4	6.7	1.9	21.5	6.8
LR3	3.4	3.2	4.3	1.7	5.3	3.9	1.9	8.1	12.5
RR	**3.4**	**2.6**	**4.2**	2.4	3.3	3.7	2.4	3.2	3.7
RR2	16.5	6.8	7.8	16.0	7.4	8.4	6.7	6.1	4.5
RR3	16.4	6.7	7.8	16.0	7.5	8.4	6.4	6.4	4.1
RFR	3.8	3.8	6.0	3.8	3.7	9.6	4.1	4.3	2.3
XGB	0.7	3.1	6.8	1.9	2.3	5.2	2.1	1.4	5.2
SVR	0.7	3.8	5.8	0.7	3.5	5.4	4.3	3.5	4.3
ANN	1.5	2.5	5.0	0.4	2.5	4.6	7.0	2.8	3.7

Appendix D

Below, the data sets used for training, validation and testing with variations of ***PP*** as well as ***MAT*** variables are listed below. The global training, validation and test data sets consisted of local one originating from respective ***PP*** and ***MAT*** variable variations. For ***PP*** variables, local train, validation and test sets can be found in Tables D.1–D.3, respectively. For ***MAT*** variables, local train, validation and test sets can be found in Tables D.4–D.6, respectively. The test2 (cross-space) set used for ultimate evaluation of model generalization is listed in Table D.7.

Table D.1
Train set of ***PP*** variation.

Sample	F [kN]	RS [$\frac{1}{s}$]	TS [$\frac{mm}{s}$]	k_s [$\frac{N}{s\,K}$]	c_s [$\frac{mm^2}{s^2\,K}$]·10^7	t_s [mm]	k_b [$\frac{N}{s\,K}$]	c_b [$\frac{mm^2}{s^2\,K}$]·10^7	t_b [mm]	T_{max} [°C]
1	8.0	20.0	6.0	157	86	10	157	86	8	377.53
2	8.0	25.0	4.0	157	86	10	157	86	8	417.36
3	6.0	15.0	6.0	157	86	10	157	86	8	221.32
4	8.0	15.0	4.0	157	86	10	157	86	8	342.91
5	10.0	20.0	8.0	157	86	10	157	86	8	383.28
6	6.0	20.0	4.0	157	86	10	157	86	8	282.91
7	8.0	15.0	8.0	157	86	10	157	86	8	292.55
8	10.0	20.0	4.0	157	86	10	157	86	8	451.14
9	8.0	25.0	8.0	157	86	10	157	86	8	375.14
10	10.0	25.0	6.0	157	86	10	157	86	8	425.69
11	10.0	15.0	6.0	157	86	10	157	86	8	370.85
12	6.0	20.0	8.0	157	86	10	157	86	8	227.41
13	6.0	25.0	6.0	157	86	10	157	86	8	292.94

Table D.2
Validation set of ***PP*** variation.

Sample	F [kN]	RS $[\frac{1}{s}]$	TS $[\frac{mm}{s}]$	k_s $[\frac{N}{s\,K}]$	c_s $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_s [mm]	k_b $[\frac{N}{s\,K}]$	c_b $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_b [mm]	T_{max} [°C]
1	8.0	25.0	6.0	157	86	10	157	86	8	420.51
2	8.0	15.0	6.0	157	86	10	157	86	8	317.45
3	8.0	20.0	8.0	157	86	10	157	86	8	363.38
4	8.0	20.0	4.0	157	86	10	157	86	8	436.19
5	10.0	20.0	6.0	157	86	10	157	86	8	449.73
6	6.0	20.0	6.0	157	86	10	157	86	8	294.57

Table D.3
Test set of ***PP*** variation.

Sample	F [kN]	RS $[\frac{1}{s}]$	TS $[\frac{mm}{s}]$	k_s $[\frac{N}{s\,K}]$	c_s $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_s [mm]	k_b $[\frac{N}{s\,K}]$	c_b $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_b [mm]	T_{max} [°C]
1	9.7	24.2	7.0	157	86	10	157	86	8	408.32
2	8.3	22.5	5.7	157	86	10	157	86	8	387.96
3	7.0	20.8	4.3	157	86	10	157	86	8	347.66
4	9.0	17.5	6.3	157	86	10	157	86	8	346.78
5	7.7	19.2	7.7	157	86	10	157	86	8	302.54
6	6.3	15.8	5.0	157	86	10	157	86	8	255.30

Table D.4
Train set of ***MAT*** variation.

Sample	F [kN]	RS $[\frac{1}{s}]$	TS $[\frac{mm}{s}]$	k_s $[\frac{N}{s\,K}]$	c_s $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_s [mm]	k_b $[\frac{N}{s\,K}]$	c_b $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_b [mm]	T_{max} [°C]
1	8	20	6	117	90	10	49.6	47.3	10.0	416.75
2	8	20	6	157	86	10	6.7	52	10.2	380.07
3	8	20	6	75	90.23	8	49.6	47.3	10.0	464.26
4	8	20	6	75	90.23	10	6.7	52	10.2	478.99
5	8	20	6	117	90	12	157.0	86	10.0	403.01
6	8	20	6	117	90	8	6.7	52	10.2	458.48
7	8	20	6	75	90.23	10	157.0	86	10.0	460.01
8	8	20	6	117	90	12	6.7	52	10.2	417.13
9	8	20	6	157	86	10	157.0	86	10.0	360.05
10	8	20	6	157	86	12	49.6	47.3	10.0	352.69
11	8	20	6	75	90.23	12	49.6	47.3	10.0	442.91
12	8	20	6	117	90	8	157.0	86	10.0	406.53
13	8	20	6	157	86	8	49.6	47.3	10.0	385.59

Table D.5
Validation set of ***MAT*** variation.

Sample	F [kN]	RS $[\frac{1}{s}]$	TS $[\frac{mm}{s}]$	k_s $[\frac{N}{s\,K}]$	c_s $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_s [mm]	k_b $[\frac{N}{s\,K}]$	c_b $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_b [mm]	T_{max} [°C]
1	8	20	6	75	90.23	8	6.7	52	10.2	497.25
2	8	20	6	75	90.23	8	157.0	86	10.0	451.19
3	8	20	6	75	90.23	12	6.7	52	10.2	470.90
4	8	20	6	75	90.23	12	157.0	86	10.0	456.98
5	8	20	6	157	86	8	6.7	52	10.2	423.32
6	8	20	6	157	86	8	157.0	86	10.0	377.74
7	8	20	6	157	86	12	6.7	52	10.2	388.85
8	8	20	6	157	86	12	157.0	86	10.0	363.05

Table D.6
Test set of ***MAT*** variation.

Sample	F [kN]	RS $[\frac{1}{s}]$	TS $[\frac{mm}{s}]$	k_s $[\frac{N}{s\,K}]$	c_s $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_s [mm]	k_b $[\frac{N}{s\,K}]$	c_b $[\frac{mm^2}{s^2\,K}]\cdot 10^7$	t_b [mm]	T_{max} [°C]
1	8	20	6	117	90	10	6.7	52	10.2	434.71
2	8	20	6	117	90	10	157.0	86	10.0	413.46
3	8	20	6	75	90.23	10	49.6	47.3	10.0	459.19
4	8	20	6	157	86	10	49.6	47.3	10.0	374.62
5	8	20	6	117	90	8	49.6	47.3	10.0	417.34
6	8	20	6	117	90	12	49.6	47.3	10.0	405.02

Table D.7
Test set 2 of ***PP*** and ***MAT*** variation.

Sample	F [kN]	RS [$\frac{1}{s}$]	TS [$\frac{mm}{s}$]	k_s [$\frac{N}{s\,K}$]	c_s [$\frac{mm^2}{s^2\,K}$]·10^7	t_s [mm]	k_b [$\frac{N}{s\,K}$]	c_b [$\frac{mm^2}{s^2\,K}$]·10^7	t_b [mm]	T_{max} [°C]
1	8	15	6	75	90.23	10	157.0	86	10.0	411.43
2	10	20	4	117	90	10	157.0	86	10.0	485.19
3	10	20	8	117	90	10	157.0	86	10.0	438.54
4	8	15	8	75	90.23	10	49.6	47.3	10.0	406.66
5	10	20	6	75	90.23	8	49.6	47.3	10.0	508.17
6	10	15	6	117	90	12	49.6	47.3	10.0	409.42
7	8	25	4	157	86	10	49.6	47.3	10.0	437.21
8	10	20	6	157	86	12	49.6	47.3	10.0	405.51
9	6	20	6	157	86	8	49.6	47.3	10.0	291.62
10	10	20	4	117	90	10	6.7	52	10.2	511.98
11	8	20	8	117	90	8	6.7	52	10.2	432.59
12	8	15	6	157	86	10	6.7	52	10.2	357.71

References

[1] J. Gandra, H. Krohn, R.M. Miranda, P. Vilaça, L. Quintino, J.F. Dos Santos, Friction surfacing—A review, J. Mater Process. Technol. 214 (5) (2014) 1062–1093, http://dx.doi.org/10.1016/j.jmatprotec.2013.12.008.

[2] H. Klopstock, A.R. Neelands, An improved method of joining or welding metals, 1941.

[3] R. Damodaram, P. Rai, S. Cyril Joseph Daniel, R. Bauri, D. Yadav, Friction surfacing: A tool for surface crack repair, Surf. Coat. Technol. 422 (2021) 127482, http://dx.doi.org/10.1016/j.surfcoat.2021.127482.

[4] J.J.S. Dilip, S. Babu, S.V. Rajan, K.H. Rafi, G.D. Janaki Ram, B.E. Stucker, Use of friction surfacing for additive manufacturing, Mater. Manuf. Process. 28 (2) (2013) 189–194, http://dx.doi.org/10.1080/10426914.2012.677912.

[5] U. Suhuddin, S. Mironov, H. Krohn, M. Beyer, J.F. Dos Santos, Microstructural evolution during friction surfacing of dissimilar aluminum alloys, Metall. Mater. Trans. A 43 (13) (2012) 5224–5231, http://dx.doi.org/10.1007/s11661-012-1345-8.

[6] S. Hanke, J.F. Dos Santos, Comparative study of severe plastic deformation at elevated temperatures of two aluminium alloys during friction surfacing, J. Mater Process. Technol. 247 (2017) 257–267, http://dx.doi.org/10.1016/j.jmatprotec.2017.04.021.

[7] F.Y. Isupov, O. Panchenko, L. Zhabrev, I. Mushnikov, E. Rylkov, A.A. Popovich, Finite element simulation of temperature field during friction surfacing of Al-5Mg consumable rod, Key Eng. Mater. 822 (2019) 737–744, http://dx.doi.org/10.4028/www.scientific.net/KEM.822.737.

[8] Z. Kallien, L. Rath, A. Roos, B. Klusemann, Experimentally established correlation of friction surfacing process temperature and deposit geometry, Surf. Coat. Technol. 397 (6) (2020) 126040, http://dx.doi.org/10.1016/j.surfcoat.2020.126040.

[9] H. Krohn, S. Hanke, M. Beyer, J.F. Dos Santos, Influence of external cooling configuration on friction surfacing of AA6082 T6 over AA2024 T351, Manuf. Lett. 5 (2015) 17–20, http://dx.doi.org/10.1016/j.mfglet.2015.04.004.

[10] X. Liu, J. Yao, X. Wang, Z. Zou, S. Qu, Finite difference modeling on the temperature field of consumable-rod in friction surfacing, J. Mater Process. Technol. 209 (3) (2009) 1392–1399, http://dx.doi.org/10.1016/j.jmatprotec.2008.03.067.

[11] V.I. Vitanov, N. Javaid, Investigation of the thermal field in micro friction surfacing, Surf. Coat. Technol. 204 (16–17) (2010) 2624–2631, http://dx.doi.org/10.1016/j.surfcoat.2010.02.003.

[12] Z. Kallien, B. Klusemann, Combined experimental-numerical analysis of the temperature evolution and distribution during friction surfacing, Surf. Coat. Technol. 437 (8) (2022) 128350, http://dx.doi.org/10.1016/j.surfcoat.2022.128350.

[13] P. Pirhayati, H. Jamshidi Aval, An investigation on thermo-mechanical and microstructural issues in friction surfacing of Al–Cu aluminum alloys, Mater. Res. Expr. 6 (5) (2019) http://dx.doi.org/10.1088/2053-1591/ab0635.

[14] Z. Rahmati, H. Jamshidi Aval, S. Nourouzi, R. Jamaati, Modeling and experimental study of friction surfacing of AA2024 alloy over AA1050 plates, Mater. Res. Expr. 6 (8) (2019) http://dx.doi.org/10.1088/2053-1591/ab255a.

[15] F. Chollet, Deep Learning with Python, Manning, 2017.

[16] F.E. Bock, R.C. Aydin, C.J. Cyron, N. Huber, S.R. Kalidindi, B. Klusemann, A review of the application of machine learning and data mining approaches in continuum materials mechanics, Front. Mater. 6 (2019) 443, http://dx.doi.org/10.3389/fmats.2019.00110.

[17] L. Meng, B. McWilliams, W. Jarosinski, H.-Y. Park, Y.-G. Jung, J. Lee, J. Zhang, Machine learning in additive manufacturing: A review, JOM 72 (6) (2020) 2363–2377, http://dx.doi.org/10.1007/s11837-020-04155-y.

[18] J. Xiong, G. Zhang, J. Hu, L. Wu, Bead geometry prediction for robotic GMAW-based rapid manufacturing through a neural network and a second-order regression analysis, J. Intell. Manuf. 25 (1) (2014) 157–163, http://dx.doi.org/10.1007/s10845-012-0682-1.

[19] C. Wacker, M. Köhler, M. David, F. Aschersleben, F. Gabriel, J. Hensel, K. Dilger, K. Dröder, Geometry and distortion prediction of multiple layers for wire arc additive manufacturing with artificial neural networks, Appl. Sci. 11 (10) (2021) 4694, http://dx.doi.org/10.3390/app11104694.

[20] J. Deng, Y. Xu, Z. Zuo, Z. Hou, S. Chen, Bead geometry prediction for multi-layer and multi-bead wire and arc additive manufacturing based on XGBoost, in: S. Chen, Y. Zhang, Z. Feng (Eds.), Transactions on Intelligent Welding Manufacturing, in: Transactions on Intelligent Welding Manufacturing, Vol. 39, Springer Singapore, Singapore, 2019, pp. 125–135, http://dx.doi.org/10.1007/978-981-13-8668-8_7.

[21] W.-J. Oh, C.-M. Lee, D.-H. Kim, Prediction of deposition bead geometry in wire arc additive manufacturing using machine learning, J. Mater. Res. Technol. 20 (2022) 4283–4296, http://dx.doi.org/10.1016/j.jmrt.2022.08.154.

[22] G.O. Barrionuevo, P.M. Sequeira-Almeida, S. Ríos, J.A. Ramos-Grez, S.W. Williams, A machine learning approach for the prediction of melting efficiency in wire arc additive manufacturing, Int. J. Adv. Manuf. Technol. 120 (5–6) (2022) 3123–3133, http://dx.doi.org/10.1007/s00170-022-08966-y.

[23] J. Paulo Davim, C. Oliveira, A. Cardoso, Predicting the geometric form of clad in laser cladding by powder using multiple regression analysis (MRA), Mater. Des. 29 (2) (2008) 554–557, http://dx.doi.org/10.1016/j.matdes.2007.01.023.

[24] S. Milhomme, J. Lartigau, C. Brugger, C. Froustey, Bead geometry prediction using multiple linear regression analysis, Int. J. Adv. Manuf. Technol. 117 (1–2) (2021) 607–620, http://dx.doi.org/10.1007/s00170-021-07697-w.

[25] M. Biyikli, T. Karagoz, M. Calli, T. Muslim, A.A. Ozalp, A. Bayram, Single track geometry prediction of laser metal deposited 316L-Si via multi-physics modelling and regression analysis with experimental validation, Metals Mater. Int. 43 (10) (2022) 292s, http://dx.doi.org/10.1007/s12540-022-01243-3.

[26] E.M. Lee, G.Y. Shin, H.S. Yoon, D.S. Shim, Study of the effects of process parameters on deposited single track of M4 powder based direct energy deposition, J. Mech. Sci. Technol. 31 (7) (2017) 3411–3418, http://dx.doi.org/10.1007/s12206-017-0239-5.

[27] F. Caiazzo, A. Caggiano, Laser direct metal deposition of 2024 Al alloy: Trace geometry prediction via machine learning, Mater. (Basel, Switzerland) 11 (3) (2018) http://dx.doi.org/10.3390/ma11030444.
[28] H. Liu, X. Qin, S. Huang, L. Jin, Y. Wang, K. Lei, Geometry characteristics prediction of single track cladding deposited by high power diode laser based on genetic algorithm and neural network, Int. J. Precis. Eng. Manuf. 19 (7) (2018) 1061–1070, http://dx.doi.org/10.1007/s12541-018-0126-8.
[29] D.R. Feenstra, A. Molotnikov, N. Birbilis, Utilisatioan of artificial neural networks to rationalise processing windows in directed energy deposition applications, Mater. Des. 198 (2021) 109342, http://dx.doi.org/10.1016/j.matdes.2020.109342.
[30] V. Vitanov, I. Voutchkov, G. Bedford, Decision support system to optimise the Frictec (friction surfacing) process, J. Mater Process. Technol. 107 (1–3) (2000) 236–242, http://dx.doi.org/10.1016/S0924-0136(00)00710-X.
[31] V. Vitanov, I. Voutchkov, G. Bedford, Neurofuzzy approach to process parameter selection for friction surfacing applications, Surf. Coat. Technol. 140 (3) (2001) 256–262, http://dx.doi.org/10.1016/S0257-8972(01)01128-8.
[32] V.I. Vitanov, I.I. Voutchkov, Process parameters selection for friction surfacing applications using intelligent decision support, J. Mater Process. Technol. 159 (1) (2005) 27–32, http://dx.doi.org/10.1016/j.jmatprotec.2003.11.006.
[33] V. Sugandhi, V. Ravishankar, Optimization of friction surfacing process parameters for aa1100 aluminum alloy coating with mild steel substrate using response surface methodology (RSM) technique, Mod. Appl. Sci. 6 (2) (2012) 69, http://dx.doi.org/10.5539/mas.v6n2p69.
[34] D. Ikeuchi, A. Vargas-Uscategui, X. Wu, P.C. King, Neural network modelling of track profile in cold spray additive manufacturing, Mater. (Basel, Switzerland) 12 (17) (2019) http://dx.doi.org/10.3390/ma12172827.
[35] R. Kadaganchi, M.R. Gankidi, H. Gokhale, Optimization of process parameters of aluminum alloy AA 2014-T6 friction stir welds by response surface methodology, Def. Technol. 11 (3) (2015) 209–219, http://dx.doi.org/10.1016/j.dt.2015.03.003.
[36] M. Srinivasa Rao, N. Ramanaiah, Optimization of process parameters for FSW of Al-Mg-Mn-Sc-Zr alloy using CCD and RSM, Strojnícky Cas. – J. Mech. Eng. 68 (3) (2018) 195–224, http://dx.doi.org/10.2478/scjme-2018-0035.
[37] A.K. Lakshminarayanan, V. Balasubramanian, Comparison of RSM with ANN in predicting tensile strength of friction stir welded AA7039 aluminium alloy joints, Trans. Nonferr. Met. Soc. China 19 (1) (2009) 9–18, http://dx.doi.org/10.1016/S1003-6326(08)60221-6.
[38] V.D. Manvatkar, A. Arora, A. De, T. DebRoy, Neural network models of peak temperature, torque, traverse force, bending stress and maximum shear stress during friction stir welding, Sci. Technol. Weld. Join. 17 (6) (2012) 460–466, http://dx.doi.org/10.1179/1362171812Y.0000000035.
[39] M.H. Shojaeefard, M. Akbari, P. Asadi, Multi objective optimization of friction stir welding parameters using FEM and neural network, Int. J. Precis. Eng. Manuf. 15 (11) (2014) 2351–2356, http://dx.doi.org/10.1007/s12541-014-0600-x.
[40] F.E. Bock, L.A. Blaga, B. Klusemann, Mechanical performance prediction for friction riveting joints of dissimilar materials via machine learning, Procedia Manuf. 47 (2020) 615–622, http://dx.doi.org/10.1016/j.promfg.2020.04.189.
[41] D. Ikeuchi, A. Vargas-Uscategui, X. Wu, P. King, Data-efficient neural network for track profile modelling in cold spray additive manufacturing, Appl. Sci. 11 (4) (2021) 1654, http://dx.doi.org/10.3390/app11041654.
[42] F. Chinesta, E. Cueto, E. Abisset-Chavanne, J.L. Duval, F.E. Khaldi, Virtual, digital and hybrid twins: A new paradigm in data-based engineering and engineered data, Arch. Comput. Methods Eng. 18 (1) (2018) 1, http://dx.doi.org/10.1007/s11831-018-9301-4.
[43] F.E. Bock, S. Keller, N. Huber, B. Klusemann, Hybrid modelling by machine learning corrections of analytical model predictions towards high-fidelity simulation solutions, Mater. (Basel, Switzerland) 14 (8) (2021) 1883, http://dx.doi.org/10.3390/ma14081883.
[44] R. Ibáñez, E. Abisset-Chavanne, D. González, J.-L. Duval, E. Cueto, F. Chinesta, Hybrid constitutive modeling: Data-driven learning of corrections to plasticity models, Int. J. Mater. Form. 326 (1567) (2018) 565, http://dx.doi.org/10.1007/s12289-018-1448-x.
[45] D. González, F. Chinesta, E. Cueto, Learning corrections for hyperelastic models from data, Front. Mater. 6 (2019) 752, http://dx.doi.org/10.3389/fmats.2019.00014.
[46] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc, 2017, pp. 4765–4774.
[47] S. Chen, T. Kaufmann, Development of data-driven machine learning models for the prediction of casting surface defects, Metals 12 (1) (2022) 1, http://dx.doi.org/10.3390/met12010001.
[48] I. Baturynska, K. Martinsen, Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms, J. Intell. Manuf. 32 (1) (2021) 179–200, http://dx.doi.org/10.1007/s10845-020-01567-0.
[49] V. Fitseva, S. Hanke, J.F. Dos Santos, Influence of rotational speed on process characteristics, material flow and microstructure evolution in friction surfacing of Ti-6Al-4V, Mater. Manuf. Process. 32 (5) (2016) 557–563, http://dx.doi.org/10.1080/10426914.2016.1257799.
[50] R. Jin, W. Chen, A. Sudjianto, An efficient algorithm for constructing optimal design of computer experiments, J. Statist. Plann. Inference 134 (1) (2005) 268–287, http://dx.doi.org/10.1016/j.jspi.2004.02.014.
[51] M.A. Bouhlel, J.T. Hwang, N. Bartoli, R. Lafage, J. Morlier, J.R.R.A. Martins, A Python surrogate modeling framework with derivatives, Adv. Eng. Softw. (2019) 102662, http://dx.doi.org/10.1016/j.advengsoft.2019.03.005.
[52] Constellium, Airware 2050-T84 plate, 2017.
[53] MatWeb - Material Property Data, Aluminum 5083-H112, 2022, http://www.matweb.com/search/DataSheet.aspx?MatGUID=bd6317b19dd94faf8bff851e4f339e88, Accessed 28.05.2021.
[54] MatWeb - Material Property Data, Aluminum 7050-T7451 (7050-t73651), 2022, http://www.matweb.com/search/DataSheet.aspx?MATGUID=142262cf7fbc4c83917ca5c3d17df1ed, Accessed 08.03.2021.
[55] MatWeb - Material Property Data, Overview of materials for Low Carbon Steel, 2022, https://www.matweb.com/search/DataSheet.aspx?MatGUID=034970339dd14349a8297d2c83134649. Accessed 15.09.2022.
[56] MatWeb - Material Property Data, Titanium Ti-6Al-4V (Grade 5), STA, 2022, http://www.matweb.com/search/DataSheet.aspx?MatGUID=b350a789eda946c6b86a3e4d3c577b39. Accessed 08.03.2021.
[57] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, third ed., in: Morgan Kaufmann Series in Data Management Systems, Elsevier/Morgan Kaufmann, Amsterdam, 2011.
[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830, URL http://arxiv.org/pdf/1201.0490v4.
[59] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, 2016, pp. 785–794, http://dx.doi.org/10.1145/2939672.2939785, URL http://doi.acm.org/10.1145/2939672.2939785.
[60] F. Chollet, et al., Keras, 2015, URL https://github.com/fchollet/keras.
[61] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, URL https://www.tensorflow.org/. Software available from tensorflow.org.
[62] K. Murphy, Machine Learning - a Probabilistic Perspective, in: Adaptive Computation and Machine Learning, MIT Press, Cambridge, 2012.

[63] C. Wade, Hands-on Gradient Boosting with XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting with Python, first ed., Packt Publishing, 2020.
[64] J. Bergstra, D. Yamins, D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in: S. Dasgupta, D. McAllester (Eds.), Proceedings of the 30th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, Vol. 28, (1) PMLR, Atlanta, Georgia, USA, 2013, pp. 115–123.
[65] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer New York, New York, NY, 1995.
[66] V. Vapnik, S.E. Golowich, A. Smola, Support vector method for function approximation, regression estimation and signal processing, in: Proceedings of the 9th International Conference on Neural Information Processing Systems, NIPS '96, MIT Press, Cambridge, MA, USA, 1996, pp. 281–287.
[67] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, in: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, Vol. 9, MIT Press, 1996.
[68] S.S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice Hall, Upper Saddle River, NJ, 1998.
[69] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (6) (1958) 386–408, http://dx.doi.org/10.1037/h0042519.
[70] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, 2014.
[71] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, et al., Keras Tuner, 2019, https://github.com/keras-team/keras-tuner.
[72] L.S. Shapley, 17. A value for n-person games, in: H.W. Kuhn, A.W. Tucker (Eds.), Contributions To the Theory of Games (AM-28), Volume II, Princeton University Press, 1953, pp. 307–318, http://dx.doi.org/10.1515/9781400881970-018.
[73] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 2522–5839, http://dx.doi.org/10.1038/s42256-019-0138-9.
[74] H. Li, Z. Xu, Z. Zhou, Numerical simulation of the plunge stage of friction surfacing of AA5083 aluminum alloy, IOP Conf. Ser.: Mater. Sci. Eng. 758 (1) (2020) 012043, http://dx.doi.org/10.1088/1757-899X/758/1/012043.