ORIGINAL RESEARCH

2S-ML: A simulation-based classification and regression approach for drawability assessment in deep drawing

Tobias Lehrer[1,2] · Arne Kaps[2] · Ingolf Lepenies[3] · Fabian Duddeck[2] · Marcus Wagner[1]

Abstract
New structural sheet metal parts are developed in an iterative, time-consuming manner. To improve the reproducibility and speed up the iterative drawability assessment, we propose a novel low-dimensional multi-fidelity inspired machine learning architecture. The approach utilizes the results of low-fidelity and high-fidelity finite element deep drawing simulation schemes. It hereby relies not only on parameters, but also on additional features to improve the generalization ability and applicability of the drawability assessment compared to classical approaches. Using the machine learning approach on a generated data set for a wide range of different cross-die drawing configurations, a classifier is trained to distinguish between drawable and non-drawable setups. Furthermore, two regression models, one for drawable and one for non-drawable designs are developed that rank designs by drawability. At instantaneous evaluation time, classification scores of high accuracy as well as regression scores of high quality for both regressors are achieved. The presented models can substitute low-fidelity finite element models due to their low evaluation times while at the same time, their predictive quality is close to high-fidelity models. This approach may enable fast and efficient assessments of designs in early development phases at the accuracy of a later design phase in the future.

✉ Tobias Lehrer
tobias.lehrer@oth-regensburg.de

Arne Kaps
arne.kaps@tum.de

Ingolf Lepenies
ingolf.lepenies@scale.eu

Fabian Duddeck
duddeck@tum.de

Marcus Wagner
marcus.wagner@oth-regensburg.de

1   Laboratory Finite-Element-Method, Faculty of Mechanical Engineering, OTH Regensburg, Galgenbergstraße 30, 93053 Regensburg, Bavaria, Germany

2   Computational Solid Mechanics, TUM School of Engineering and Design, Technical University of Munich, Arcisstraße 21, 80333 Munich, Bavaria, Germany

3   SCALE GmbH, Pohlandstraße 19, 01309 Dresden, Saxony, Germany

Introduction

The development of new deep-drawn sheet metal parts is a complex engineering task. Low costs, structural durability, improved crash properties, aesthetics, and sustainability are contrasting requirements. Another constraint in part development is ensuring manufacturability, which has to be guaranteed from the beginning of the early design phase until a part is produced in a press. It needs to be re-assessed for design changes, uncertainties in material and sheet metal properties, and changes in the drawing configuration. To reduce the complexity in development, engineers define the shape and drawing configuration for a part iteratively. This experience-driven process leads to a compromise in the aforementioned requirements and lacks reproducibility. On the other hand, existing methods like optimization are often impractical for manufacturability assessment. This is due to the limitations of the corresponding large-scale simulation models, which are computationally expensive [5].

As we focus on the deep drawing production step, we investigate drawability rather than full manufacturability. There have been ongoing scientific contributions to

improve drawability assessment. Baseline formulas to estimate drawability are given in [12]. The method is based on a cylindrical cup drawing setup. It rather aims to give feedback on whether a part should be deep-drawn or not than to give reliable feedback in part development. A simplified two-dimensional approach of a stamping analysis is presented in [14]. We employ the finite element method (FEM). It is a widely used numerical technique for simulating the behavior of complex systems, such as the deep drawing of sheet metals. The method has become an essential tool in this field, as it allows engineers and scientists to predict the behavior of a system under different loading conditions, material properties, and addendum geometries. The main advantage of FEM is the flexibility and low cost in comparison to real-world experiments.

Machine learning (ML) has shown promising results for hybrid modeling in engineering environments. Ambrogio et al. [1] proposed a kriging meta-model on incremental sheet forming to predict sheet thicknesses. A surrogate model for textile forming applications was presented in [25]. Here, simulation results are evaluated with an image-based data representation to investigate the influence of different base geometries on the surrogate model extrapolation quality. Morand et al. [16] presented a surrogate model that predicts the final geometry and field variables of a deep-drawn cup, focusing on selected process and geometry parameters. Slimani et al. [20] developed an artificial neural network surrogate for the flat rolling process based on a combination of the slab method and FEM to predict the rolling force. A dimension-reduced neural network based on features is used to predict the springback behavior in [11].

In our hybrid approach, we utilize the concept of multi-fidelity (MF) modeling to create surrogate models. Illustrated by Kennedy et al. [13], the goal is to achieve high-fidelity (HF) model accuracy at the computational cost of a low-fidelity (LF) model. Accordingly, a combined MF surrogate with decreased evaluation time is set up. Song et al. [22] mixed an LF and an HF model using radial basis functions to decrease computational cost for surrogate setup. It was tested on numerical and engineering problems. For a comprehensive overview of MF modeling, readers are referred to [24].

To decrease evaluation times and save expensive simulation runs while preserving the prediction quality of sophisticated simulation models, we propose a novel approach named 2S-ML here, an ML-based MF-inspired architecture. It incorporates information from both, LF and HF simulation schemes as well as non-simulated information to train low-dimensional surrogate models based on a data set. Our approach differs from existing MF methods in two ways. First, our surrogate models do not depend solely on simulation parameters but also on additional ML features that contain complementary drawability information. These features improve applicability to a more diverse set of shapes.

Second, unlike MF scaling functions, space mappings, and difference mappings, 2S-ML does not explicitly use mixing or tuning factors, as the mixing is performed implicitly during the training of the ML surrogate models. Third, as our method is intended to work on an established historic data set, we focus on surrogate performance, not computational efficiency in their setup.

Using the 2S-ML architecture, we build a regression surrogate that allows us to continuously assess drawability and rank drawable designs in a large, industry-relevant design space. Subsequently, the same architecture is used to train a classification model that can exclude non-drawable designs. In that regard, we introduce an extension to an existing drawability measure to be able to distinguish different drawable states of drawable designs, which enables continuous ranking of designs. Furthermore, this approach assesses drawability without specific knowledge of tooling. Therefore, it can map the drawability assessment of a later design stage to the early design stage.

The paper is structured in the following manner. In Sect. "2S-ML Machine learning architecture ", two FEM simulation schemes involved in the generation of the data set are presented. Their material model is introduced in Sect. 3. The measure used for drawability assessment is outlined in Sect. 4. Section 5 describes the Design of Experiments (DOE) used for data set generation. It serves as the basis for the development of the regression and classification model presented in Sect. 6. Afterward, the results of the ML models are presented and discussed in Sect. 7. Finally, conclusions are drawn in Sect. 8.

Finite element simulation models

The cross-die geometry used in the simulation schemes is similar to the ones presented in [4, 10]. Its shape contains convex, concave, and planar faces that yield a variety of representative multiaxial stress and strain states.

Low-fidelity one-step scheme

The one-step scheme, also known as the inverse approach [7], is based on the principle of virtual work, Hencky deformation, and kinematic-based geometric mapping [6]. It flattens the final workpiece to the initial flat blank. One-step simulations allow for a fast calculation of the sheet thicknesses, thinning, and plastic strains of a sheet metal. The inverse approach lacks precision, as it does not recreate a deep drawing process itself. Important effects such as incremental plasticity and the influence of tooling contacts cannot be modeled. Therefore, it is considered the LF model here. A simulation configuration for an exemplary LF model is shown in Fig. 1.

Fig. 1 The one-step simulation setup contains the geometry in its final state. To take the influence of drawbeads into account, a subset of nodes on the free edge of the convex hull of the model are selected to apply restraining forces. We enforce a maximum edge length of 2 mm for each element, resulting in roughly 5 elements over the radius. It provides interpretable results like thinning and a preliminary blank shape suggestion
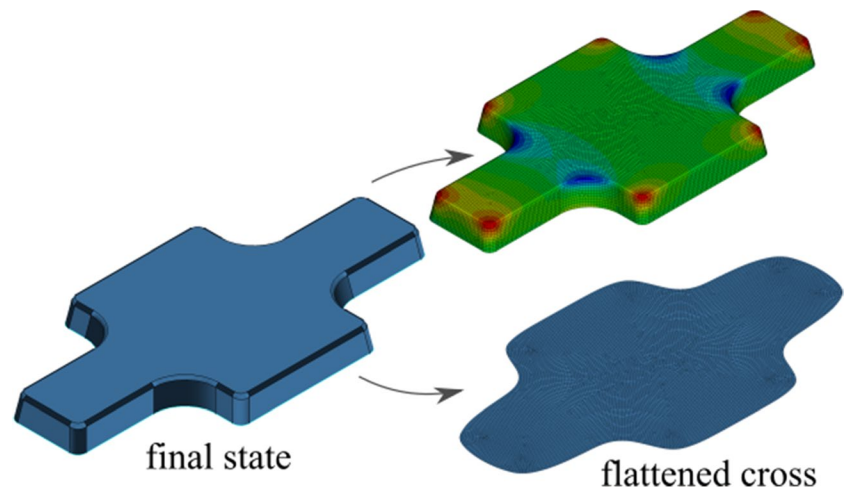
final state

flattened cross

Fig. 2 Four geometries are used in the incremental deep drawing simulation. The tooling, composed of the die, punch, and blankholder, forms the blank into its final geometry state. The initial maximum edge length of blank elements is 4 mm. The minimum edge length of re-meshed elements is set to 0.3 mm, which results in about 15 elements per radius for the formed blank. Drawbeads are considered analytically. Field quantities are evaluated at the last time step when the maximum drawing depth is reached

initial configuration

formed blank
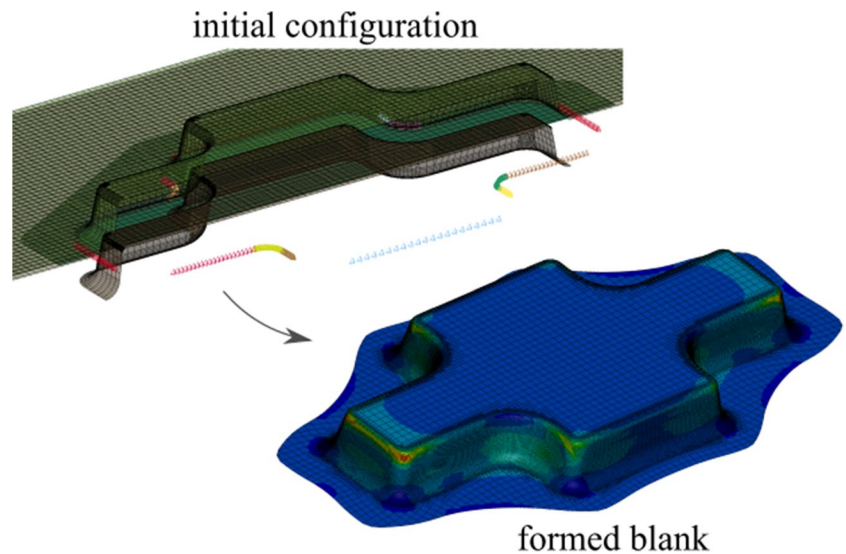
One-step simulations are suitable for early design stage evaluation since they do not require tooling. We use implicit time integration[1] to calculate the flattened state of equilibrium. The short simulation times are beneficial for practical part-to-part comparisons and iterative methods like optimization. For example, to define the initial blank size as was investigated in [18].

High-fidelity incremental deep drawing scheme

The incremental deep drawing simulations scheme returns high precision results by calculating the deep drawing process time-dependently, including tooling, contacts, and manufacturing boundary conditions. On the downside, the simulation scheme is computationally expensive. Tooling including drawbead design has to be created and embedded before the simulation can be run. Therefore, it can only be implemented at the later design stage. As deep drawing simulations represent the most sophisticated results, they are the last virtual step before the experimental tryout process. The incremental deep drawing scheme is defined here as the HF simulation. A sample configuration is shown in Fig. 2.

We use reduced integrated bilinear Belytschko-Tsay shell elements and limit the hourglass energy to 5% of the internal energy. For the HF scheme, we use explicit time integration[1]. This has several reasons. Nonlinearities induced by the contacts or the material model require small time steps. This counteracts the advantage of unlimited time step sizes of implicit time integration. Also, the explicit scheme here benefits the convergence behavior for discontinuous boundary conditions. Deviations from the exact solution are accepted, as the overall process time is

---

[1] The commercial solver is the shared memory parallel (smp) R13 LS-DYNA™ solver with single precision for HF and double precision for LF. Computation times depend on the geometry, but are approximately 10 min and 3 min on 10 cores.

small (considering time scaling). In addition, we employ four techniques to reduce computational costs. First, an initially coarse mesh for the blank is refined adaptively over the simulation time based on two heuristics. We re-mesh regions where neighboring elements are deformed beyond an angle of 25° and when tooling surfaces are approaching to cope with contact interaction. Second, we apply time scaling to reduce the overall simulation time. Third, the tooling is considered rigid. Fourth, we utilize element-wise mass scaling to limit the minimum time step size to $\Delta t_{min} = 1.0 \cdot 10^{-7}s$. To obtain a thorough summary of deep drawing FEM models, readers are directed to [2].

Hockett-Sherby hardening model

We use an elasto-plastic material model with Hill48 [8] as yield criterion, assuming transverse anisotropy and plane stress condition in the thin-walled sheet. To embed a variety of different hardening properties in our upcoming surrogates, we need a tunable and representative hardening law for deep drawing steels. The isotropic hardening is calculated based on the Hockett-Sherby saturation law [9]

$$\sigma = \sigma_s - (\sigma_s - \sigma_y)e^{-(N\epsilon)^p} \tag{1}$$

with σ_s as saturation and σ_y as yield stress, N and p as material constants, and ϵ as the plastic strain. Strain rate-dependent effects are considered negligible. To generate only monotonically increasing hardening curves and to represent different hardening trends, a stress ratio

$$i_{\sigma_c} = \frac{R_m \cdot (1 + A_g)}{\sigma_y} \tag{2}$$

between the true stress equivalent of the tensile strength R_m, converted with the elongation without necking A_g and yield stress σ_y is introduced. By setting σ_y and i_{σ_c}, the computation of the hardening curve can be conducted.

By adjusting the parameter ranges (see Table 4), we generate representative material properties. Figure 3 shows a comparison of our sampling range to standardized cold rolled (CR) forming steels. All flow curves used afterward are drawn from the sampling range.

Drawability measure

The HF simulation model itself calculates nodal deformations, but no direct measure for drawability. Therefore, there is a need to further evaluate the FEM model to describe and calculate drawability. A commonly used method for this is using the forming limit diagram
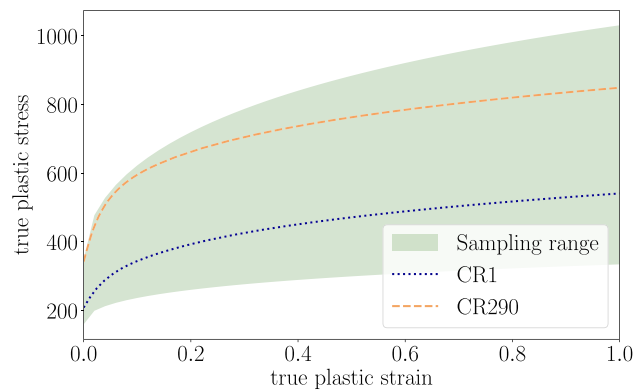

Fig. 3 The parameter ranges are chosen so that sampled stress–strain curves (within green sampling range) represent typically used forming steels. This is exemplified by the two depicted standardized CR steels

(FLD). It allows for element-wise subdivision of the minor and major true principal strains of a formed sheet metal into regions of cracks, wrinkles, and good (drawable) points. Sun et al. [23] calculate a weighted sum of element-wise vertical distances from crack points to the forming limit curve (FLC) and wrinkle points to the wrinkling limit curve (WLC). In [15], a strain path-dependent distance between the FLC and each point in the crack region is presented.

We select the function proposed by Sun et al. [23] as an outset because of its sophisticated drawability representation. It is originally defined in a positive semidefinite manner, where non-drawable configurations can be distinguished from one another. Drawable setups cannot, because their distance inherently is 0. Therefore, we extend the drawability measure

$$f_{\text{drawability}} = \begin{cases} f_{\text{n-d}}, & \text{if} f_{\text{n-d}} > f_{\text{dt}} \\ f_{\text{d}}, & \text{otherwise} \end{cases} \tag{3}$$

with $f_{\text{n-d}}$ for non-drawable configurations and f_{d} for drawable ones. This allows ranking of different levels of drawable designs. As most drawing setups tend to have some minor wrinkling, especially near the blankholder region where the mesh is relatively coarse, we introduce a setup-specific, user-adjustable drawability threshold of $f_{\text{dt}} = 1.395 \cdot 10^{-6}$. Setups with values below the threshold are considered drawable. The calculation of the two domains

$$f_{\text{n-d}} = \sum_{c=1}^{n_c} A_c^e \cdot d_{c-\text{flc}}^e + w_w \cdot \sum_{w=1}^{n_w} A_w^e \cdot d_{w-\text{wlc}}^e \tag{4}$$

$$f_{\text{d}} = -w_g \cdot \sum_{g=1}^{n_g} A_g^e \cdot \frac{d_{g-\text{flc}}^e + w_w \cdot d_{g-\text{wlc}}^e}{1 + w_w} \tag{5}$$

is conducted with A^e as the surface area of an element normalized with the total surface area of the formed blank, d^e as an element's minimum Euclidean distance to the FLC or WLC, respectively, n as the number of elements, and w as weights. The indices c, w, and g denote the affiliation with the crack, wrinkle, or good (drawable) regions in the FLD. We set the weights for wrinkling at $w_w = 0.1$ and for good points at $w_g = 0.2$.

We make four modifications to the original function to improve robustness and simulation-specific mechanics. First, we use the minimum Euclidean distance because it more accurately represents the deformation process, especially for wrinkles. Second, by element-wise area weighting, we mitigate the effect of different element sizes in the mesh to minimize overall mesh dependence. Third, as we do not delete cracked elements during simulation, each cracked element's surface area is set to the surface area at the time step at which the crack occurred. Otherwise, the weighted distances calculation would be affected by non-realistically deformed crack elements with a high surface area. Fourth, a maximum element-wise distance for crack elements $d^e_{c-flc,max} = 2.35$ is introduced. This limits the effects of heavily distorted or failed elements on distance calculation. The cumulative distance approach adopted here is preferred over a perhaps more intuitive maximum distance approach for two reasons. First, it mitigates the impact of single heavily distorted elements on the drawability measure. Second, cumulative distances provide more consistent feedback, which benefits surrogate training and optimization tasks conducted with them.

The resulting weighted distribution in the drawable domain is depicted in Fig. 4.

Design of experiments

To embed a learnable drawability threshold, a data set containing drawable and non-drawable parts is needed. To avoid imbalanced learning, the ratio between drawable and non-drawable setups should not exceed about 85% instances of one label. We use a Sobol sequence, originally published in [21], to on the one hand sample material parameters and on the other hand sample the remaining parameters separately. Parameters are then combined randomly. The parameters and their ranges are shown in Table 4. The resulting parameter combinations are depicted in Fig. 5.

The mixing of lines to the left of the r axis shows the random combinations of material parameters with respect to the remaining parameters. Otherwise, the uniformity of the Sobol sequence can be seen from evenly intersecting
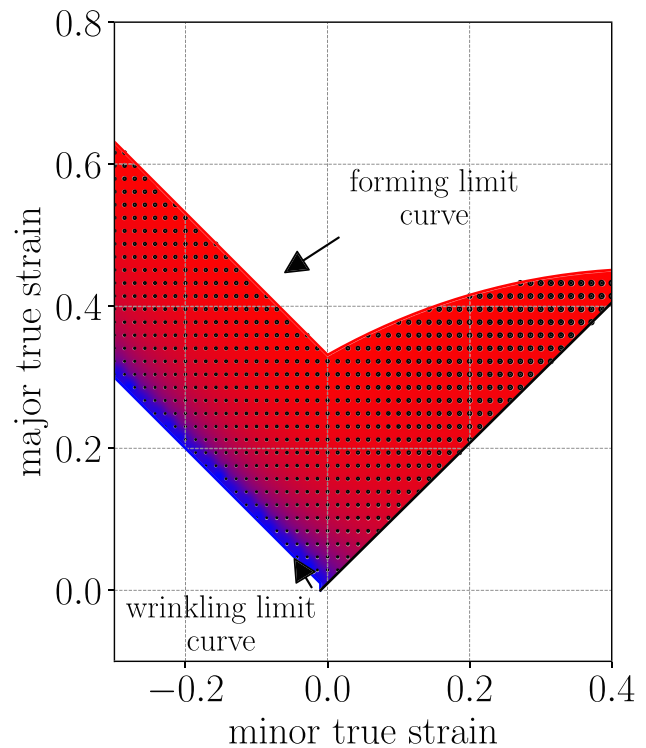

Fig. 4 The drawable domain is defined within the area bounded by the FLC (red), the WLC (blue) and the line of undefined deformation state (black). The color represents the influence of the failure mode (red for cracks and blue for wrinkles). The predominant red filling is caused by the implementation of the wrinkling weight w_w. The more cracks influence a region, the larger the black dots

parallel lines. Regions of thicker blue lines indicate that there are more feasible combinations of geometric parameters (e.g. without undercuts). Regions with fewer lines are prone to less accurate predictions. As a result of feasible parameter combinations and error-free simulation runs, we end up generating a data set of 2541 samples. This relatively large design space represents a variety of setups, making it harder to obtain high-quality results with a limited number of samples.

2S-ML Machine learning architecture

Noteworthy industrial appearance of complex deep drawn parts with multiple development iterations is mostly found in car structures. A chassis contains approximately 200—500 sheet metal parts, of which only a varying fraction is deep drawn due to cost implications [3]. Same-part strategies, symmetric structures, inconsistent documentation, and changes in the state-of-the-art of the deep drawing process further narrow the number of usable sheet metals for ML. In this limited data environment, deep learning is unlikely to provide the best model
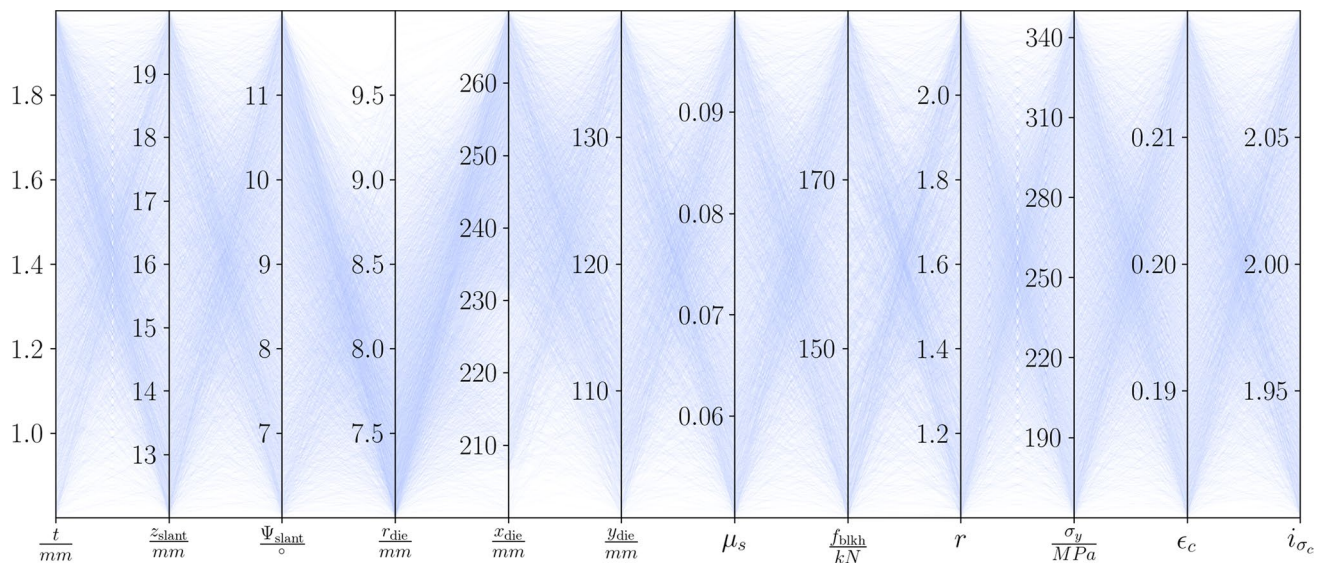
Fig. 5 There are 12 independent parameters for cross-die geometry, process simulation conditions, and material models. A comprehensive overview of their meaning can be obtained from Table 4 and Fig. 15. Sheet thickness t is chosen in a range where the FLD is known to give reliable results. Geometry parameters are set to produce dissimilar cross-die shapes. Every parameter combination represents a data point in the data set (blue lines)
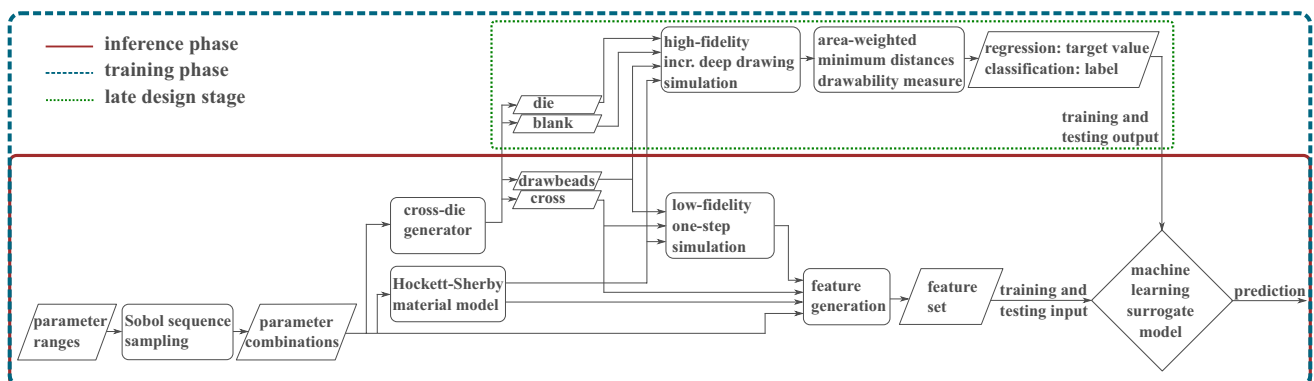

Fig. 6 The 2S-ML architecture can be used to train both, regression and classification models. The training phase is enclosed by the dashed blue line. The solid red line marks the inference phase. Steps enclosed by the dotted green line are only available from the late design stage onwards

performance. To promote applicability, we use domain knowledge to set up the 2S-ML architecture with application-dependent feature and label engineering.

We interpret drawability assessment in two ways:

- The ML model distinguishes whether a setup is drawable or not using binary classification.
- The ML model assigns a target value representing drawability using regression to rank drawability of different configurations.

We use supervised learning to ensure the best possible model performance. An overview of the proposed 2S-ML architecture is given in Fig. 6.

In the training phase, the first step involves setting parameter ranges. Next, the DOE is employed using the Sobol sequence to sample parameter combinations. For each parameter set representing a drawing configuration, all cross-die-related geometry is created using the geometric parameters.[2] The punch geometry here is derived by offsetting the die geometry. Analytical drawbead sets are used directly in the incremental simulation to apply restraining forces to nearby blank nodes. The material parameters are used to specify the Hockett-Sherby hardening and the anisotropic elasto-plastic material model.

---

[2] For more complex shapes, sophisticated tooling geometry is derived via method planning and not parametrically. The effects of different tooling shapes per part are further neglected.

Fig. 7 The features can be related to four different domains. 16 (40.0%) of them belong to the geometry source (brown), 9 (22.5%) to material (green), 4 (10.0%) to manufacturing (orange), and 11 (27.5%) to simulation (blue)
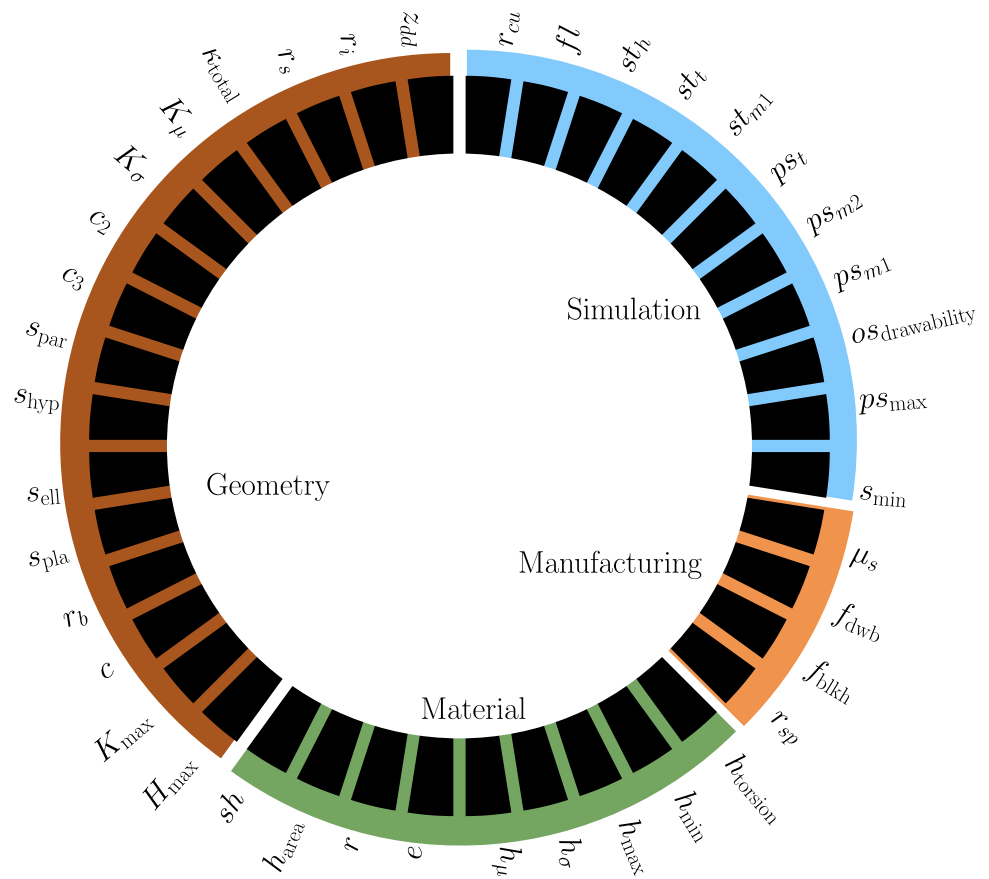
An HF simulation is then set up using the die, drawbead, and blank geometry, along with the material model and process parameters. The drawability measure evaluates the incremental simulation to generate a label for classification or a target value for regression. Additionally, an LF simulation is set up and run using the cross, the material model, and process parameters. Here, restraining forces are mapped from the drawbeads to the boundary nodes of the cross. Features are generated based on the material model, the one-step simulation, and the parameter combinations. These features are then used as input for a classification or regression ML model, while the label or target value is used for the supervision of the training.

In the inference phase, the first step involves defining the parameter set of interest. Next, a representative feature set is generated as described above. Finally, the label or target value is predicted using the feature set on the trained ML model. It should be noted that no incremental forming simulation needs to be run during the inference phase.

We establish a set of 40 initial features that contain baseline simulation parameters and supplementary features. An in-depth overview of their calculation is given in the Appendix 2. The goal of the features is to capture all necessary information for drawability assessment.

Features associated with the shape of the part are inspired by differential geometry. Most material features are intended to embed the hardening behavior. The simulation-based features are application-dependent and should contain relevant drawability information. We do not propose features to characterize tooling geometry, as it is generated parametrically here.

Using this combination of information, a surrogate model captures the representative underlying drawability problem, which helps to generalize to new configurations. Additionally, the supplementary features are independent of the cross-die geometry. They can therefore be applied to a variety of shapes. The assignment of all computed features to their source domain is shown in Fig. 7.

Computing the feature set for every configuration in the whole data set includes five aberrations. These are further excluded from the upcoming training and testing process. The features of the resulting data set are shown in Fig. 17.

Further investigation of the data set is done by calculating the Pearson correlation coefficient

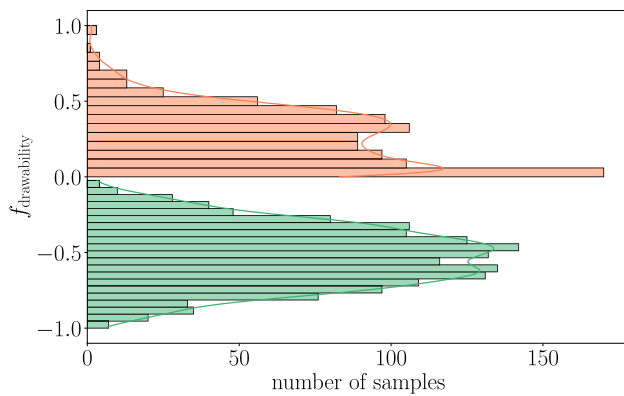$$r_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \tag{6}$$

Fig. 8 The distribution of the non-drawable (red) and the drawable (green) domains is separated by the abscissa. There are 955 (37.7%) non-drawable configurations and 1579 (62.3%) drawable ones

between two features X and Y, $cov(X, Y)$ as the covariance between the features, and the standard deviation of the features σ_X and σ_Y. A high correlation of features can lead to worse prediction quality. The features h_{min}, h_σ, h_μ, h_{max}, K_{max}, and ps_{m1} are neglected due to one or more Pearson correlation coefficients greater than 0.9.

Despite feature deletion, there remains redundant information (colinearities) in the feature set, which can limit surrogate quality. Still, further deletion would reduce relevant information and likewise decrease surrogate quality. To address this issue, principal component analysis (PCA) is utilized. While losing the interpretability of a feature's influence, PCA enhances the model's generalization ability and here guarantees the applicability of the proposed approach to non-parametric data sets. To find the most optimal compromise between small dimensionality and variance loss, we make the number of principal components available for hyperparameter optimization.

For regression, the drawability domains are min–max-normalized each. The resulting distribution is shown in Fig. 8.

Due to the piecewise definition of the drawability measure, there exist two separable histograms. The drawable setups are quite evenly distributed, while the non-drawable setups show an agglomeration near the drawability threshold. The data set is sparsely filled for values greater than 0.75. The resulting label ratio for classification is within the limits of imbalanced learning.

Because there is no prior knowledge of which ML algorithm will build the most suitable surrogate for drawability assessment, we heuristically select a subset of training algorithms (compare Figs. 9 and 12) published in [19]. For an in-depth overview of low-dimensional ML algorithms, interested readers are referred to [17]. As ensemble methods usually provide gains in prediction quality, we train bagging meta-models that use

the selected algorithms as base estimators. Also, a stacking and a voting ensemble model are built based on the aforementioned meta-models. We tune the hyperparameters of the PCA, the base estimators, the bagging meta-model, and of the voting and stacking models via cross-validated grid search optimization. The same ensemble approach is applied to both, regressors and classifiers.

To measure surrogate performance, appropriate metrics must be defined. For classification, the most costly misjudgments are false positive predictions, as this means continuing to develop a part that is not drawable. Therefore, we use

$$F_\beta = \frac{(1 + \beta^2 \cdot TP)}{(1 + \beta^2) \cdot TP + \beta^2 \cdot FN + FP} \tag{7}$$

with the weight factor $\beta = 0.5$, TP for true positive, TN for true negative, FP for false positive, and FN for false negative predictions as our binary classification metric for training.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

is used subsidiarily to measure overall performance. For regression, we utilize the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{9}$$

where n is the number of data points, y_i is the true value, $\widehat{y}_i$ is the predicted value, and $\overline{y}_i$ is the observed mean value to investigate the overall regression quality. The relative maximum absolute error

$$RMAE = \frac{\max\limits_{i=1}^{n}(|y_i - \widehat{y}_i|)}{\max\limits_{i=1}^{n}(|y_i|)} \tag{10}$$

is used to indicate local accuracy.

Results and discussion

For regression and classification, the data is randomly divided into 80% training data and 20% test data. Cross-validation with five iterations eliminates the need for validation data.

Regression

Independent of the regression training algorithm, we initially evaluate two different modeling approaches. The first approach includes setting up a regressor to predict drawability for both, drawable and non-drawable setups.

Table 1 The prediction quality of the best regression surrogates of the combined and separated approaches

Metrics	Combined	Separated	
		Drawable	Non-drawable
R^2	0.815	0.918	0.859
$RMAE$	60.1%	21.8%	23.6%

is caused by the piecewise definition of the drawability measure. Yet, using two surrogates in parallel raises the necessity for a model that decides which regressor should be used for prediction. This can be done by the classifier introduced in Sect. 7.2.

We further concentrate on the separated approach because of its higher prediction quality. A comparison between different models for the regression is shown in Fig. 9.

Fig. 9 The prediction quality of all investigated regressors, measured by R^2 and $RMAE$
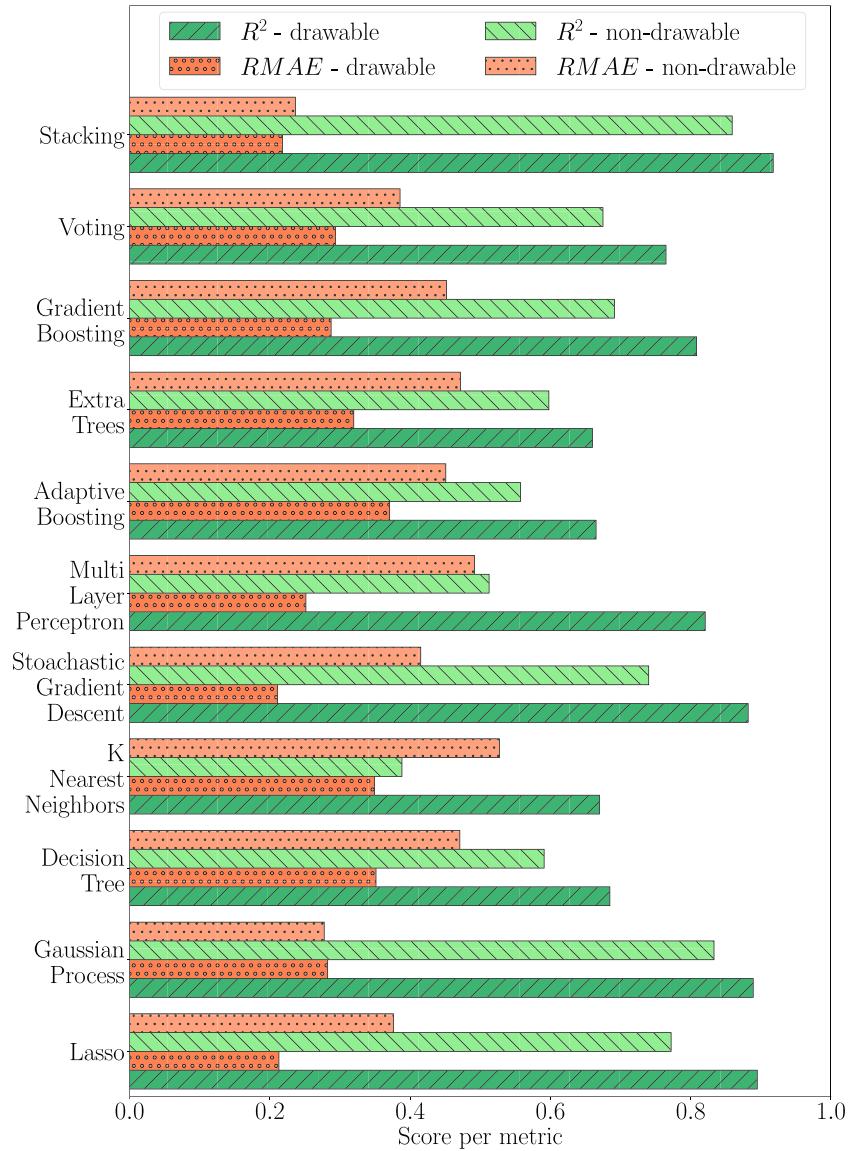
In the second approach, two separate regressors are trained, one for the drawable and one for the non-drawable domain. Results of the best-performing surrogates are listed in Table 1.

It can be noted that the overall high prediction scores illustrate the applicability of our method. For regression, splitting the data set into drawable and non-drawable subsets significantly improves prediction quality. This

It can be observed that all models achieve higher R^2-scores for the drawable domain. The authors attribute this to the fact that the failure composed of cracks and wrinkles is mechanically more complex compared to the good points in the drawability measure calculation. The stacking regressor shows the best R^2-scores of 0.918 for the drawable (green) and 0.859 for the non-drawable domain (red).
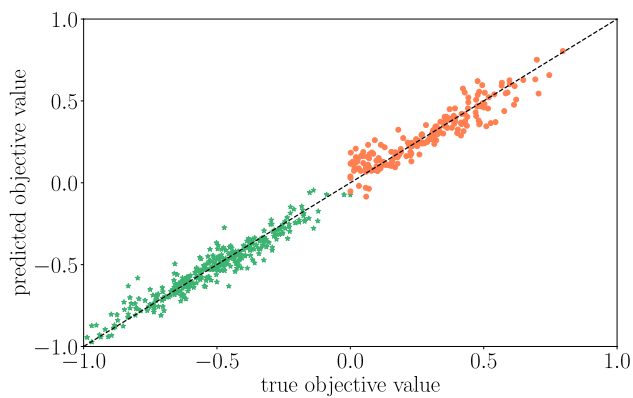
Fig. 10 The joint predictions for the drawable (green stars) and non-drawable (red circles) domains of the stacking regressor on the test data. The dashed black line represents ideal prediction, where true and predicted values would coincide

Fig. 11 The R^2-scores for the drawable (green) and non-drawable (red) domain for the stacking regressor over the number of training samples. The drawable domain contains 1581 data points (1264 for training). The non-drawable domain 955 (764 for training). The black dashed line marks the threshold of around 10% relative error to maximum R^2 for both regressors, which is at 240 training samples each

It also has the lowest *RMAE* s of 21.8% and 23.6%. The voting regressor's prediction quality suffers from being based on all depicted models, including worse-performing models.

In Fig. 10, predicted target values are compared with the true ones for the best-performing regressor.

The overall true-predicted comparison is homogeneous. There are no severe outliers affecting the prediction scores and errors, which can also be deduced from the low *RMAE* s. This allows for a drawability ranking of different drawing setups in the whole design space. Drawable configurations are overall very accurately predicted. In the non-drawable domain, there is a slight trend towards predicting worse drawability for nearly drawable designs. This agglomeration of slightly too poor drawability estimation near the drawability threshold could be influenced by the defined drawability threshold.

Furthermore, there are sources of noise in the features and labels. As mass scaling is only applied to shell elements below a certain size, there is uneven mass induction for different cross-die geometries, which influences the calculation of the drawability measure. Also, the coarse blank mesh in the region of the addendum affects the regression. Those elements often are wrinkled elements and due to their element size relatively strongly weighted in the drawability computation. A finer mesh produces distinctive wrinkled geometry. Therefore, more realistic element surface areas are combined with more precise strain results, which results in more accurate drawability values. Also, low-dimensional feature-based approaches work on the basis of feature sets that condense the necessary information to describe the problem at hand. Adding features that contain supplementary information benefits the training process.
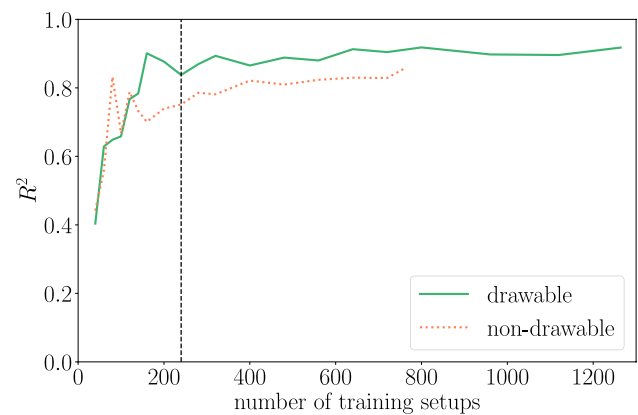
Another possibility for improved surrogate performance is to increase the number of training instances. To get insights into how many data points are needed for what model quality, training is conducted for different data set sizes for the stacking regressor. The results are shown in Fig. 11.
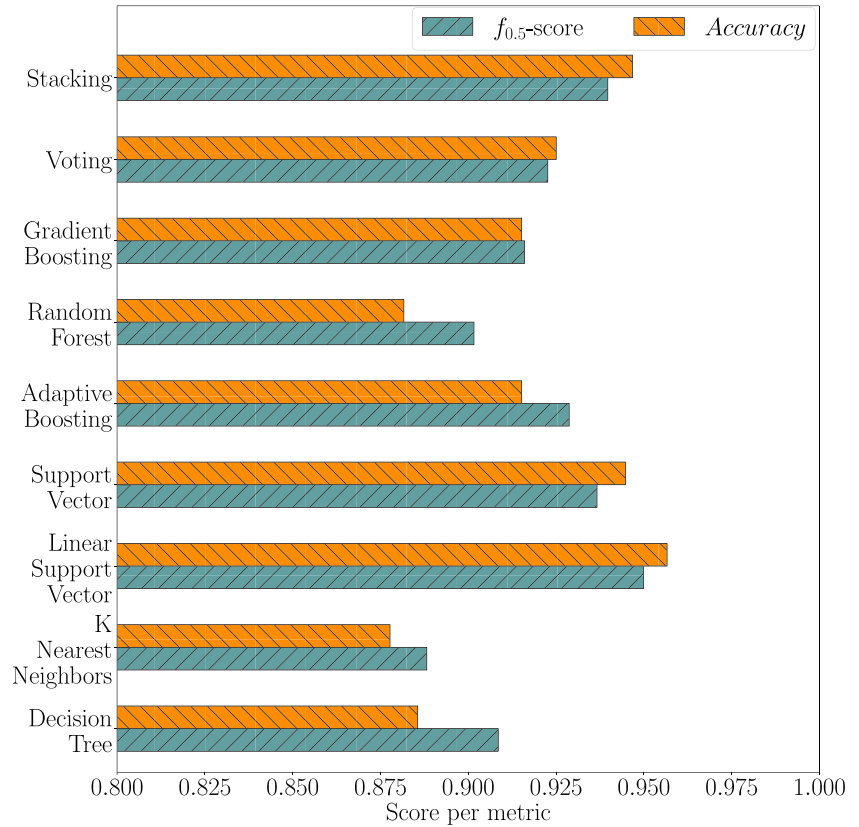
It can be noted that the R^2-scores rise with additional training data. Below 240 training samples, the drawable and non-drawable curves alter significantly. A notable convergence in R^2-score for the stacking regressors starts at about 240 training samples. Therefore, a minimum ratio of about $20 \cdot d$ with d as the number of parameters is required for reliable surrogate performance. Above 240 training samples, there is still a slight trend of increasing R^2-scores. Therefore, to achieve maximum R^2-scores, a higher ratio than $20 \cdot d$ is necessary for our 12-parameter problem (compare Fig. 5). From the convergence trend of the drawable surrogate (green line in Fig. 11), we conclude that to achieve even higher R^2-scores, additional features would be necessary. Based on the convergence trend, we expect a lasting gap between drawable and non-drawable curves, at least for an industrial stock of training samples. Besides the more complex non-drawable state, more training samples are likely to still decrease the gap between R^2-scores of the domains.

None of the models presented in Fig. 9 show a noteworthy different convergence trend—including the limited data environment with less than 240 training samples. Still, it is possible that other surrogate techniques or ML algorithms (e.g. naive Bayes, cokriging) could achieve higher R^2-scores in the region smaller than $20 \cdot d$. If there is no data set or trained model for drawability assessment, a compromise between computational investment to set up a surrogate and running the original model has to be considered.

Table 2 The best regressor without simulation features and only on geometry features both is the stacking regressor. The omitted features are affiliated to certain domains (see Fig. 7)

Metrics	Only geometry features		Without simulation features	
	Drawable	Non-drawable	Drawable	Non-drawable
R^2	0.712	0.179	0.916	0.843
RMAE	35.7%	53.1%	21.5%	26.4%

Considering the benchmark values shown in Fig. 10, it is apparent that relying only on geometric features has a substancial negative impact on the R^2-scores. The non-drawable domain cannot be assessed correctly, while the prediction quality in the drawable domain is only moderate. Consequently, the approach solely with geometry features is not a feasible way to assess and rank drawability. On the other hand, neglecting simulation features leads to practically no difference in prediction quality, but results in a significant reduction in evaluation time

Fig. 12 The prediction quality, measured by $F_{0.5}$-score and *Accuracy*, on the test data set of the investigated classifiers

When 2S-ML regression surrogates are used to rank multiple drawing configurations in an optimization scheme, there is an increased requirement for short evaluation times. As part of 2S-ML, a one-step simulation is run to retrieve the full feature set required for prediction (compare Fig. 6). To determine whether the loss in prediction quality is justifiable by the decrease in evaluation time, we investigate the surrogate model performance by omitting one-step features in the training of a new regression surrogate. Furthermore, in early-stage part design, sheet metal geometry is often the only existing decision basis for the drawability assessment. Therefore, we challenge our 2S-ML surrogate to be trained solely on geometry-based features. The results of the reduced feature set investigations are listed in Table 2.

(instantaneous feedback compared to several seconds to minutes). Forgoing one-step simulation in the architecture eliminates the need for a commercial license, as there is currently no open-source one-step code available. Additionally, it reduces the risk of unstable simulation runs and subsequent infeasible predictions. If there is a need to generate new samples, this can be done without conducting a one-step simulation. For existing data sets, simulation-based features can be used to slightly improve prediction quality. We emphasize though that one-step results have the potential to improve prediction quality for other data sets that are less uniformly filled and that contain less similar setups.
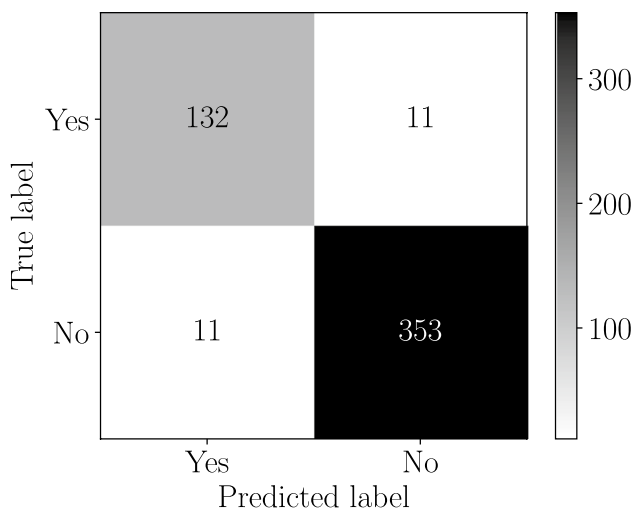
Fig. 13 The confusion matrix for the binary classification of the bagging LSVC model. The test subset contains 507 (20%) data points in total
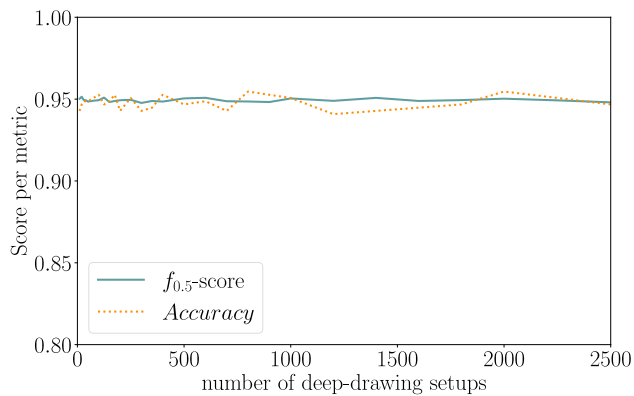

Fig. 14 The trend of the test scores of $F_{0.5}$-score and *Accuracy* for the trained bagging LSVC over an increasing number of training samples

Classification

For classification, the model needs to distinguish between drawable and non-drawable configurations. We use the same feature set as for regression. Figure 12 shows the test scores for each trained model.

Overall, most models perform similarly in the drawability assessment. The best predictor is the bagging linear support vector classification (LSVC) model with an $F_{0.5}$-score of 0.950 and an *Accuracy* of 95.7%. Although the bagging LSVC model shows the best test scores in the data set, the stacking model can also be considered for further evaluation. Due to its inherently robust predictions, it might perform better on data sets with less similar geometries.

Table 3 Scores for models trained solely on geometry features (bagging adaptive boosting classifier) and without LF features (bagging LSVC classifier). The feature affiliation as the basis for neglection can be seen in Fig. 7

Metrics	Only geometry	Without simula- tion
$F_{0.5}$-score	0.823	0.943
Accuracy	75,5%	94.5%

The test scores for the bagging LSVC model are based on the predictions displayed in Fig. 13.

The bagging LSVC classifier's hyperparameters are optimized regarding the $F_{0.5}$-score. A low number of FP values is desirable, in addition to good overall prediction scores. It can be seen that there are 11 FP (2.2%) predictions, which represent exactly half of the false predictions. It can be noted that there exist both, drawable and non-drawable predictions, proofing balanced learning.

To check how many data points are needed to build a sophisticated binary classifier, we train and test the bagging LSVC model for different data set sizes. The results are shown in Fig. 14.

It is evident that the prediction quality of the bagging LSVC is nearly constant over different data set sizes. Adding more samples has no effect on the scores. Therefore, the classifier is suited to limited data environments. However, less uniform data sets are likely to start with lower prediction scores and an increase in prediction quality over larger data set sizes. Similarly to the regression model, we expect further potential in adding features.

As described in Sect. 7.1, there exist scenarios where there is a lack of features. To investigate the performance of classifiers in this environment, we train equivalent classification models. Table 3 lists the most accurate geometry-only and LF-free features models.

There is a noticeable decline in the prediction scores for the geometry model, considering the reference values of $F_{0.5} - \text{score} = 0.950$ and *Accuracy* $= 95.7\%$. Key information is missing for proper assessment. Still, if no other reproducible methods exist, this approach can be further investigated to address early-stage drawability assessment from the perspective of a worst-case assumption. The prediction quality of the classifier without simulation features is slightly lower compared to the full model, but still comparable. Neglecting simulation features only leads to a small reduction in prediction quality. Furthermore, there are no noteworthy differences regarding convergence behavior.

Conclusion

In this paper, an ML approach to developing a regression and a classification model for drawability assessment in the early design stage in deep drawing is presented. To generate the data sets for their training, a DOE with a Sobol sequence is followed by a one-step and incremental forming simulation. Feature computation is performed to obtain a drawability representation. An evaluation using an enhanced area-weighted minimum distance drawability measure is applied to the HF model to obtain supervision for the training.

The purpose of the classifier is to distinguish between drawable and non-drawable configurations to detect scrap configurations. It is shown that high prediction scores of $F_{0.5}$-score of 0.943 and an *Accuracy* of 94.5% can be achieved using LSVC. A major finding is that even for a much smaller data set, almost identical prediction quality can be achieved for the classifier. Small potential lies in adding one-step-based features at the cost of running the simulation. Very accurate, late design phase FEM models will remain irreplaceable in a limited data environment. However, the calculation of many small to medium-sized simulations during the early design process can be significantly reduced. Another benefit of our method is that it can be used by component developers who are not simulation engineers. Reproducible evaluations by the classifier allow decision pathways to be traced for process improvements. Saving calculation time and licensing costs is another advantage.

The regressor can be used to rank the drawing setups at hand. The goal to accelerate assessment while keeping its prediction quality is achieved by the regressor, which is split into a regression model for the drawable and one for the non-drawable domain for reasons of higher prediction quality. The stacking regressors have R^2-scores of 0.916 for drawable and 0.843 for non-drawable configurations for a data set size of 2536 samples. If the data set size is limited, R^2-scores of 0.835 and 0.744 can be achieved for about 240 samples ($20 \cdot d$) each. The classification model can be used to decide which model to choose for prediction.

The use of classification and regression for drawability assessment allows saving and speeding up design iteration loops, decoupling part design and simulation to some extent. We can conclude that our cross-die showcase illustrates the applicability of the method to sheet metal deep drawing. Moreover, the 2S-ML architecture is not limited to deep drawing. The application-dependent features and labels can be adapted to assess other manufacturing processes if a hierarchy between different sources (e.g. simulation, measurement) of information can be defined. Other manufacturing process surrogates trained with this architecture, especially forming applications, can benefit from the presented features.

To apply the proposed method, a data set of previously investigated drawing configurations has to exist. Given these setups have diverse shapes, cross-die-specific features z_{slant}, Ψ_{slant}, r_{die}, x_{die}, and y_{die} (12.5% of proposed features) cannot be utilized. Also, there typically is more than one feasible tooling configuration per part. To account for this influence, we note that further features to describe the tooling geometry have to be complemented to the feature set to ensure the applicability of our proposed method. As most documented configurations are likely to be drawable, there is a risk of imbalanced learning. Another aspect of our method is the loss of local space-resolved feedback. There is no way to determine where, and to what extent drawability is given on a part. The surrogate models only allow for a global assessment. As no measured data from a press is included in the architecture, the prediction can only be as accurate as the incremental deep drawing simulation. Further influences like tailor-rolled or welded blanks, simultaneous multiple-part drawing, edge crack sensitivity, and springback are not considered in the presented workflow. Furthermore, additional criteria for drawability such as dimensional accuracy and surface irregularities have to be taken into account once automotive body parts are assessed. The 2S-ML architecture has the flexibility in its feature set and labeling process to consider these influences.

Future investigations could focus on overcoming the current limitations. Furthermore, we identified the following promising aspects:

- Investigating the sensitivities of the surrogate models will allow insight into which models rely on what features. This has the potential to detect further features, estimate the possible risk of overfitting by reliance on one or a few features, and reveal a nonsensical correlation of features to a prediction that will not generalize well to unseen data.
- Assuming there is an adequate number of industry-standard geometries available, deep learning approaches are likely to further improve the drawability assessment.
- If not directly applicable to optimization, the regression surrogate could be the source domain for transfer learning to a fine-tuned part-specific model. This approach is likely to decrease the number of necessary samples for part-specific surrogates. It could also be used in an MF scheme.
- Alternative training methods such as semi-supervised learning can be applied to decrease the number of computationally intensive HF simulation runs.
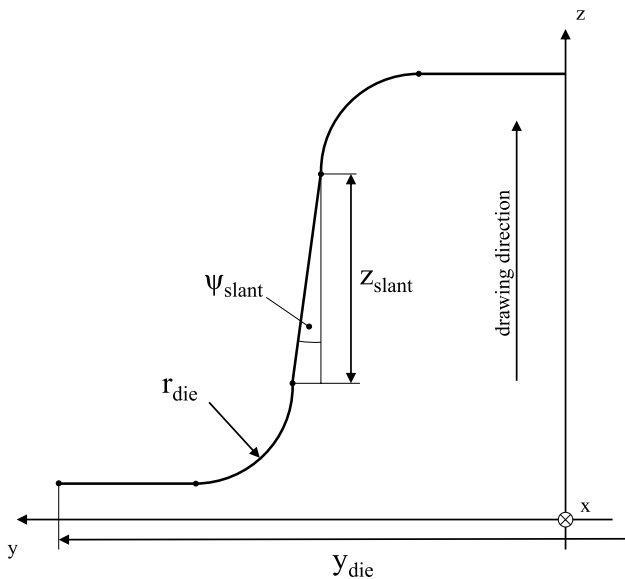
Appendix 1. Parameters for the design of experiments

Fig. 15 Schematic cross-section of the cross-die in the yz-plane. Parameter x_{die} is analogous to y_{die}, but in the xz-plane

Table 4 The parameters used in the DOE (compare Fig. 5). The true strain equivalent of the elongation without necking A_g is computed as follows: $\varepsilon_c = \log(1 + A_g)$

Parameter	Range	Meaning
t	(0.8–2.0) mm	sheet thickness
z_{slant}	(12.0–20.0) mm	height of the cross-die sidewall
Ψ_{slant}	(6.0–12.0)°	angle of the cross-die sidewall
r_{die}	(7.0–10.0) mm	die radius of the cross-die
x_{die}	(200.0–270.0) mm	x-dimension of the cross-die
y_{die}	(100.0–140.0) mm	y-dimension of the cross-die
μ_s	(0.05–0.10)	static coefficient of friction
f_{blkh}	(130.0–190.0) kN	blankholder force
r	(1.0–2.2)	Lankford coefficient
σ_y	(160.0–350.0) MPa	yield strength of the blank
ε_c	(0.18–0.22)	true strain equivalent of the elongation without necking
i_{σ_c}	(1.9–2.1)	stress ratio (s. Section 3)

Appendix 2. Calculation of the features

The feature set contains parameters and supplementary features. Here, the calculation of the non-parameter features that were not neglected from the training process is presented. Table 5 gives an overview of the meaning of the features.

Table 5 The features are inspired by differential geometry, material models, and the forming process itself

Feature	Meaning
z_{dd}	measure of draw depth
r_i	inertia ratio to embed blank geometry
r_s	surface ratio to indicate forming energy
c_2	compactness for two dimensions to one dimension
c_3	compactness for three dimensions to two dimensions
c	crinkliness
r_b	ratio to embed blank geometry
κ_{total}	total curvature to indicate shape
$H_{\max}$	maximum mean curvature to capture intricate geometries
h_{area}	area below hardening curve to indicate inner energy
h_{torsion}	torsion to differ between hardening types
sh	strain hardening
r_{sp}	speed ratio for strain-rate effects
$s_{\min}$	minimal thinning to indicate crack risk
$ps_{\max}$	maximum plastic strain to indicate crack risk
fl	distance measure to cumulate total deformation
r_{cu}	ratio to reveal challenging configurations

Geometry features

$$z_{dd} = r_{\text{die}} + z_{\text{slant}} + r_{\text{punch}}$$
$$r_i = \frac{f_{\text{blkh}}}{I_x + I_y}$$
$$r_s = \frac{A^{\text{blank}}}{A^{\text{cross}}}$$
$$c_2 = \frac{A^{\text{blank}}}{l^{\text{cross}}_{\text{convexhull}}}$$
$$c_3 = \frac{V^{\text{cross}}_{\text{boundingbox}}}{A^{\text{blank}}}$$
$$c = \frac{A^{\text{blank}}}{A_{\text{sphere}}(V^{\text{cross}}_{\text{boundingbox}})}$$

$$r_b = \frac{A^{\text{blank}}}{A^{\text{punch}}}$$
$$\kappa_{\text{total}} = \sum_{e=1}^{n_e} A^e \cdot H^e$$
$$H_{\max} = \max_{e=1}^{n_e} (H^e)$$

Fig. 16 The bounding box and convex hull edge curve (red) of the cross geometry used in the calculation of features c_2 and c_3

edge of convex hull

bounding box
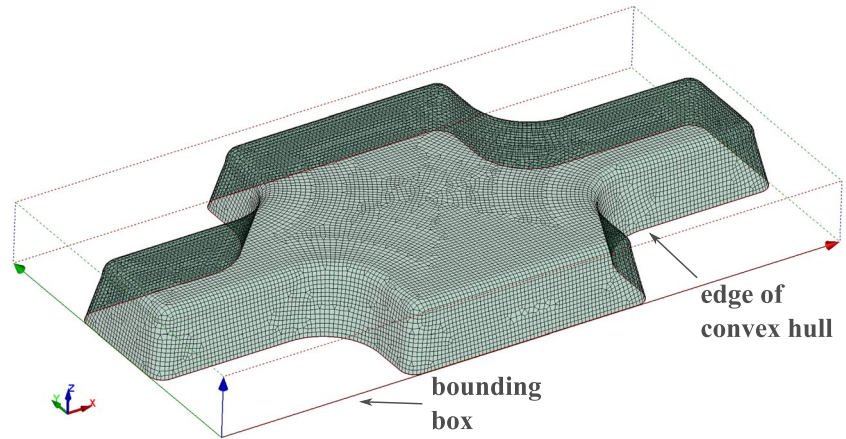
are calculated with r_{die} as the die radius, z_{slant} as the slant depth of the cross-die, and the punch radius r_{punch}. It is calculated here based on geometry parameters $r_{\mathrm{punch}} = r_{\mathrm{die}} - t \cdot (1 + t_s)$ and a safety gap of $t_s = 0.05$. The blankholder force is f_{blkh} and the geometrical moments of inertia about the x and y-axis (compare Fig. 1616) of the cross geometry are I_x, I_y. Additionally, A denotes the surface area and V the volume of either the cross, punch, or blank geometry. $l_{\mathrm{convexhull}}^{\mathrm{cross}}$ denotes the arc length of the boundary edge of the convex hull of the cross. An exemplary bounding box and convex hull edge curve involved in the feature computation is shown in Fig. 16.

$A_{\mathrm{sphere}}(V_{\mathrm{boundingbox}}^{\mathrm{cross}})$ refers to the surface area of a sphere whose volume is equivalent to the volume of the cross's bounding box $V_{\mathrm{boundingbox}}^{\mathrm{cross}}$. n_e represents the number of elements with H^e and A^e the element-wise mean curvature and surface area, respectively. The features K_μ and K_σ are the mean and standard deviation of a probability density function of all element's Gaussian curvatures. The features s_{par}, s_{hyp}, s_{ell}, and s_{pla} are the percentages of the cross's nodes that meet the
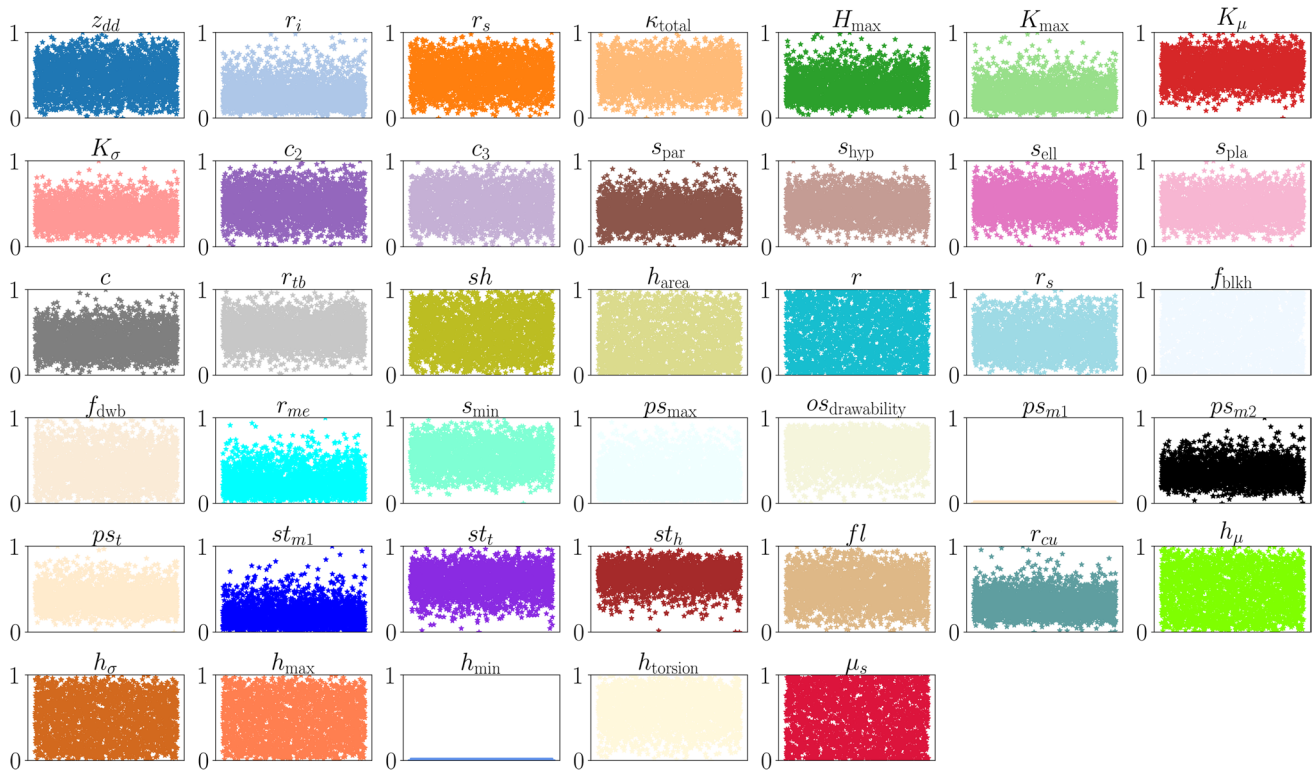

Fig. 17 The proposed 40 initial features to train the regression and classification models. After removing aberrations, there is a distribution of feature values that fills the feature space, except for ps_{m1} and h_{min}. The values of these features have no variability and are therefore not useful for training surrogates. To be suitable for training with all algorithms, all features are min–max normalized

differential geometry criteria parabolic, hyperbolic, elliptic, and planar, respectively.

Material features are

$$h_{\mathrm{area}} = \int_{0.0}^{1.0} \sigma(\varepsilon) d\varepsilon$$
$$h_{\mathrm{torsion}} = \frac{\partial^2 \sigma}{\partial \varepsilon^2}$$
$$sh = \frac{\max(\sigma_{\mathrm{hardeningcurve}})}{\sigma_y}$$

where σ is the true stress and ε the true strain of the hardening curve.

The manufacturing source contains the feature

$$r_{sp} = \frac{v_{\mathrm{punch}}}{z_{dd}}$$

with v_{punch} as the velocity of the punch, if the strain-rate influence is considered relevant. f_{dwb} is the drawbead force proposed by the one-step simulation scheme.[17]

Its features are

$$s_{\min} = \min_{e=1}^{n_e} \left(\frac{(t_0^e - t_f^e)}{t_0^e} \right)$$
$$ps_{\max} = \max_{e=1}^{n_e} (\varepsilon_e) \cdot \frac{E}{\sigma_y}$$
$$fl = \sum_{e=1}^{n_e} d_0^e \cdot A^e$$
$$r_{cu} = \frac{A_c^e + A_w^e}{A_g^e}$$

with t_0^e as the initial and t_f^e as the final sheet thickness. E is the Young's modulus. d_0^e is an elements Euclidean distance from its final state f to its unformed state 0. A_c^e, A_w^e, and A_g^e denote a cracked, wrinkled, or good element's surface area. $os_{\mathrm{drawability}}$ is the drawability measure (see Eq. 3) applied to the LF simulation. Features ps_{m2} and ps_t are coefficients of a fit with a polynomial of degree nine and st_{m1}, st_t and st_h are coefficients of fit using a polynomial of degree three.

Authors' contributions TL: Conceptualization, Methodology, Software, Formal analysis and investigation, Writing—original draft, Writing—review and editing; AK: Software, Writing—review and editing; IL: Software, Writing—review and editing; FD: Writing—review and editing, Supervision; MW: Writing—review and editing, Supervision.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no conflict of interest.

References

1. Ambrogio G, Ciancio C, Filice L et al (2017) Innovative meta-modelling-based process design for manufacturing: an application to incremental sheet forming. Int J Mater Form 10(3):279–286. https://doi.org/10.1007/s12289-015-1276-1

2. Banabic D (2010) Sheet metal forming processes: Constitutive modelling and numerical simulation. Springer, Berlin and Heidelberg

3. Birkert A, Haage S, Straub M (2013) Umformtechnische Herstellung komplexer Karosserieteile. Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-34670-5

4. Carvalho-Resende T, Balan T, Bouvier S et al (2013) Numerical investigation and experimental validation of a plasticity model for sheet steel forming. Model Simul Mater Sci Eng 21(1):015,008. https://doi.org/10.1088/0965-0393/21/1/015008

5. Fleischer M, Sitz S, Asen F, et al (2022) Simulaton of sheet metal forming and current developments. In: 16. LS-DYNA Forum 2022, URL https://www.dynamore.de/de/fortbildung/konferenzen/vergangene/16-ls-dyna-forum-2022/fleischer_bmw-group.pdf. Accessed 24 May 2023

6. Gerdeen JC, Chen P (1989) Geometric mapping method of computer modeling of sheet metal forming. NUMISHEET 89:437–444

7. Guo YQ, Batoz JL, Detraux JM et al (1990) Finite element procedures for strain estimations of sheet metal forming parts. Int J Numer Methods Eng 30(8):1385–1401. https://doi.org/10.1002/nme.1620300804

8. Hill R (1948) A theory of the yielding and plastic flow of anisotropic metals. Proc R Soc Lond A 193(1033):281–297 https://doi.org/10.1098/rspa.1948.0045

9. Hockett JE, Sherby OD (1975) Large strain deformation of polycrystalline metals at low homologous temperatures. J Mech Phys Solids 23(2):87–98. https://doi.org/10.1016/0022-5096(75)90018-6

10. Hoque SE, Duddeck F (2021) Characterization of plasticity and fracture behavior of aluminum 6061-t4 sheet for deep drawing simulation. 16th Edition Int Conf Comput Plasticity CIMNE https://doi.org/10.23967/complas.2021.013

11. Hou CKJ, Behdinan K (2023) Neural networks with dimensionality reduction for efficient springback prediction in deep drawing of multi-material cylindrical cups. J Exp Theor Artif Intell 1–20. https://doi.org/10.1080/0952813X.2023.2183271

12. Kugler H (2009) Umformtechnik: Umformen metallischer Konstruktionswerkstoffe. Fachbuchverlag Leipzig im Carl Hanser Verlag, München

13. Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87(1):1–13. http://www.jstor.org/stable/2673557. Accessed 16 Nov 2022

14. Marciniak Z, Duncan JL, Hu SJ (2002) Mechanics of sheet metal forming, 2nd edn. Butterworth-Heinemann, Oxford

15 Merten M, Liebold K, Haufe A (2021) Robustness analysis with ls-opt and ls-dyna for sheet metal forming simulations. IOP Conf Ser: Mater Sci Eng 1157(1):012,088. https://doi.org/10.1088/1757-899X/1157/1/012088

16. Morand L, Helm D, Iza-Teran R et al (2019) A knowledge-based surrogate modeling approach for cup drawing with limited data. IOP Conf Ser: Mater Sci Eng 651(012):047. https://doi.org/10.1088/1757-899X/651/1/012047

17. Müller AC, Guido S (2017) Introduction to machine learning with Python: A guide for data scientists, 1st edn. O'Reilly Media, Sebastopol, CA. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1361381

18. Naceur H, Guo YQ, Batoz JL (2004) Blank optimization in sheet metal forming using an evolutionary algorithm. J Mater Process Technol 151(1–3):183–191. https://doi.org/10.1016/j.jmatprotec.2004.04.036

19. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in python. JMLR 12(85):2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html. Accessed 16 Nov 2022

20. Slimani K, Zaaf M, Balan T (2023) Accurate surrogate models for the flat rolling process. Int J Mater Form 16(2). https://doi.org/10.1007/s12289-023-01744-5

21 Sobol' I (1967) On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput Mathematics Mathematical Phys 7(4):86–112. https://doi.org/10.1016/0041-5553(67)90144-9

22. Song X, Lv L, Sun W et al (2019) A radial basis function-based multi-fidelity surrogate model: exploring correlation between high-fidelity and low-fidelity models. Struct Multidiscip Optim 60(3):965–981. https://doi.org/10.1007/s00158-019-02248-0

23. Sun G, Li G, Zhou S et al (2011) Multi-fidelity optimization for sheet metal forming process. Struct Multidiscip Optim 44(1):111–124. https://doi.org/10.1007/s00158-010-0596-5

24. Zhou Q, Zhao M, Hu J, et al (2023) Multi-fidelity Surrogates, vol 12. Springer Nature, Singapore https://doi.org/10.1007/978-981-19-7210-2

25. Zimmerling C, Fengler B, Kärger L (2022) Formability assessment of variable geometries using machine learning - analysis of the influence of the database. Key Eng Mater 926:2247–2257. https://doi.org/10.4028/p-1o0007