Wear 580-581 (2025) 206296

Contents lists available at ScienceDirect

Wear

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

Determination of mechanical properties of physical vapor deposition tool coatings using machine learning

C. Kalscheuer, K. Bobzin, X. Liu *

Surface Engineering Institute, RWTH Aachen University, Kackertstr. 15, 52072, Aachen, Germany

ARTICLE INFO

Keywords:
Physical vapor deposition
Nanoindentation
Machine learning
Finite element method

ABSTRACT

The wear resistance of physical vapor deposition (PVD) coatings is heavily influenced by their elastic and plastic properties. These properties serve as essential inputs for finite element method (FEM) simulations of the thermomechanical load experienced by the coating during the cutting process to predict tool wear, including the elastic modulus for the characterization of elastic properties and parameters of the Ludwik-Hollomon model for plastic behavior. In this study, machine learning models are developed to directly map load-depth curves from nanoindentation to elastic modulus and Ludwik-Hollomon parameters of the coating. A FEM simulation model of nanoindentation is employed to generate a dataset comprising load-depth curves from a wide range of input mechanical properties. For each definition of mechanical properties, simulations of nanoindentation at two different indentation forces are run to generate the dataset. Several machine learning models including support vector regression (SVR), multilayer perceptron (MLP), long short-term memory (LSTM) and gated recurrent unit (GRU) are then trained, validated and compared using this dataset. The inputs to these models consist of simulated load-depth curves, with the target being mechanical properties of coatings. Among these machine learning models, SVR achieves the best accuracy for predicting elastic modulus and GRU achieves the best accuracy for predicting plastic properties. Ultimately, a hybrid model combining SVR and GRU is used to predict mechanical properties of TiAlCrN coatings using experimental load-depth curves. FEM simulations using the predicted mechanical properties show good alignment with nanoindentation experiments at two different forces. The determined properties can serve as input parameters for FEM models simulating thermomechanical load during the cutting process.

1. Introduction

Coatings deposited via physical vapor deposition (PVD) are extensively used to protect cutting tools from high thermomechanical load during the cutting process [1]. A commonly used approach to predict tool wear during the cutting process involves first simulating thermomechanical load experienced by the coating using finite element method (FEM), followed by tool wear calculation based on the simulated load [2,3]. Such simulations can also help deeper insights into the underlying tool wear mechanisms. A prerequisite for accurately simulating the thermomechanical load is a precise evaluation of the mechanical properties of coatings, including both elastic and plastic behavior. Conventional tensile and compression tests commonly used to obtain stress-strain curves of bulk materials are difficult to apply to PVD coatings, which are at a quite different length scale. Therefore, developing a universally applicable approach for accurately characterizing stress-strain relationship of coatings is significant.

Nanoindentation is commonly used to assess coating mechanical behavior under external loads. During nanoindentation, the indenter penetrates deeper into the coating under increasing loads and will be withdrawn during an unloading phase. Throughout nanoindentation, sensors continuously monitor the load and indentation depth in real-time. Load-depth curves can be generated from measurements for further analysis. In the initial phase, at low load, the coating undergoes elastic deformation. Young's modulus E serves as an indicator for evaluating the elastic deformation behavior during this phase. As the external load increases, ceramic nitride PVD coatings, though not metals, exhibit plastic deformation through a grain boundary sliding (GBS) mechanism [4]. While the mechanisms of plastic deformation differ between metals and ceramic coatings, models describing the

This article is part of a special issue entitled: Grey Box Wear Prediction_invited only published in Wear.

* Corresponding author.
E-mail address: liu@iot.rwth-aachen.de (X. Liu).

https://doi.org/10.1016/j.wear.2025.206296
Received 10 June 2025; Received in revised form 12 August 2025; Accepted 14 August 2025
Available online 16 August 2025

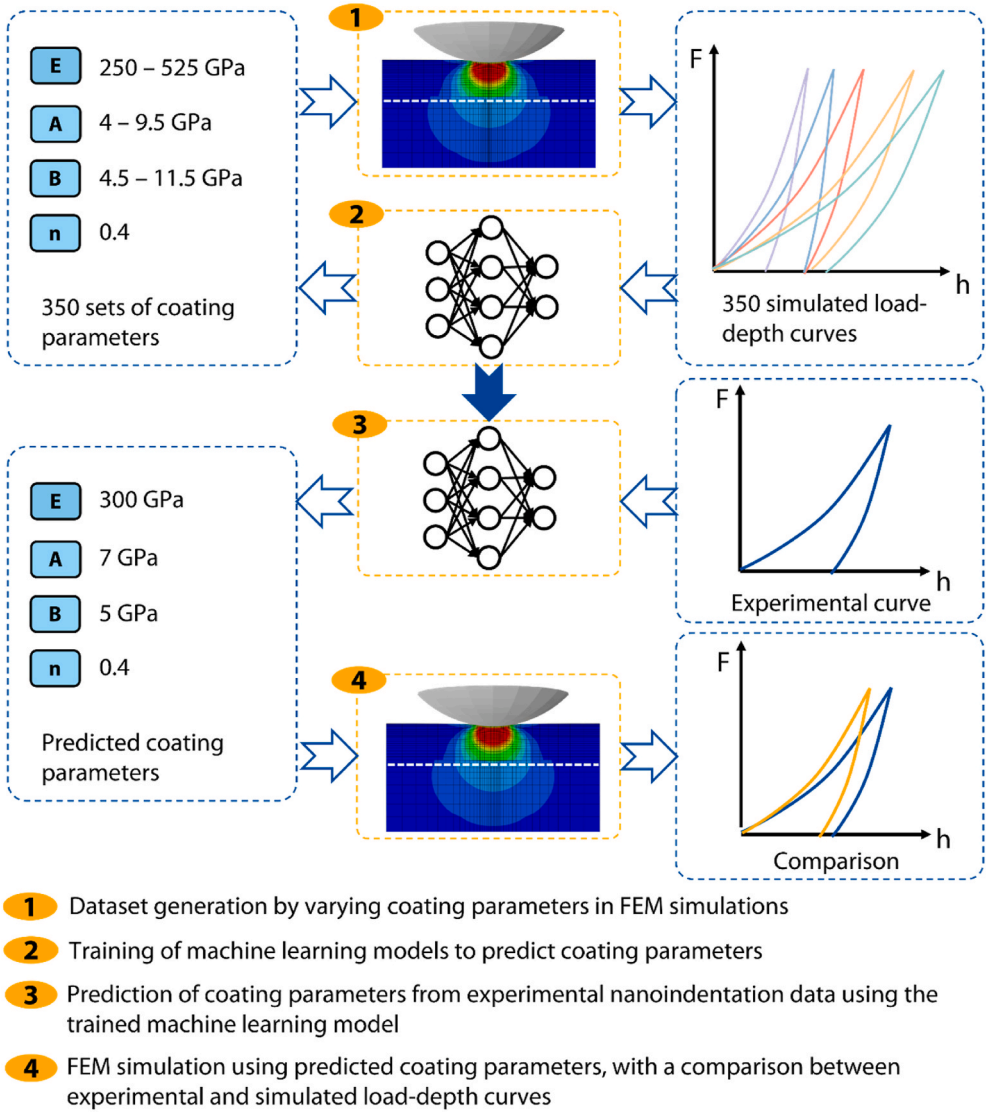

Fig. 1. The illustration of holistic methodology.

plastic deformation behavior of metals, e.g., Voce model and Ludwik-Hollomon model remain potential candidates for describing the plastic behavior of ceramic coatings [5]. The main challenge is that, unlike Young's modulus E, which can be directly calculated from load-depth curves in nanoindentation using Oliver-Pharr method [6], parameters of the plastic Voce and Ludwik-Hollomon models cannot be derived from load-depth curves.

To address this issue, several analytical methods based on experimental nanoindentation data have been developed to extract the stress-strain curve from load-depth curves of nanoindentation. In Ref. [7] the stress-strain curve is derived from the nanoindentation load-depth data by refining the definition of indentation strain and correcting the effective zero-point of contact, based on Hertz's theory [8] and additional geometrical approximations. This approach involves two steps: first determining the effective zero-point and then estimating the contact radius. A major constraint of this approach is the significant additional experimental workload. To obtain each point on the stress-strain curve, a complete loading-unloading cycle should be performed at a specific load level, requiring multiple nanoindentation tests at various loads to construct the full curve.

Apart from analytical methods, several inverse approaches using FEM have also been developed. In these approaches, the nanoindentation process is simulated using FEM with manually initialized coating parameters based on literature. These parameters are then iteratively adjusted to achieve a good alignment between the experimental and simulated results. The criteria for a good match could be based on either load-depth curves or the residual indent profiles [5,9]. The iterative adjustment of coating parameters can be performed manually, guided by the analysis of the relation between load-depth and stress-strain. In Ref. [10] the plastic tangent moduli M is determined step by step, at each small increment of indenter penetration, the

Table 1
Process parameters used for depositing the TiAlCrN coatings.

Process parameter	Unit	Value
Process gas		Ar
Reactive gas		N_2
Argon flow j_{Ar}	ml/min	200
Nitrogen flow j_{N2}	ml/min	Pressure under control
Pulse frequency f	Hz	2000
Pressure p	mPa	520
Substrate bias U_B	V	−60
Heating power P_H	W	8800
Power HPPMS-TiAl48 $P_{HPPMS\text{-}TiAl48}$	W	7000
Power HPPMS-TiAl20 $P_{HPPMS\text{-}TiAl20}$	W	7000
Power dcMS-Cr $P_{dcMS\text{-}Cr}$	W	300

corresponding plasticity tangent moduli is iteratively adjusted until the simulated indentation force matches the experimental result. Using the tangent moduli obtained at each step together with the strain read from the simulation model, the stress for each step can be calculated. Repeating this process yields the complete stress-strain curve in the end. In Ref. [11], parameters of the Ludwik-Hollomon model for direct current magnetron sputtering (dcMS) and high-power pulse magnetron sputtering (HPPMS) CrAlN coatings are iteratively adjusted by focusing on the simulation at two key points: full load and the half of the maximum force during the loading curve. The drawback of this manual approach is that the optimization process lacks rigorous mathematical support. Therefore, the required number of iterations highly depends on the experimenter's experience.

Apart from the manual approach, the parameters could also be automatically optimized utilizing mathematical optimization algorithms. In Ref. [12], the parameters of the Swift unsaturated hardening model and the Voce saturated hardening model for the TiN coating are optimized using the Levenberg-Marquardt algorithm, with the optimization target set to minimize the mean squared error between experimental and simulated load-depth curves. Another algorithm is Nelder-Mead optimization algorithm, which is used in Ref. [5] to determine parameters of the Ludwik-Hollomon model for as-received Cu and annealed Cu. The limitation of these methods lies in the lack of prior knowledge in material science. When using the initialized parameters, the first 30–40 iterations exhibit a significant deviation between the experimental results and the simulation. The fine-tuning begins after 40 iterations. In the end, it takes more than 50 iterations for the iterative adjustment to converge, making the process time-consuming [5,12].

In the past decade, machine learning, particularly deep learning based on large-scale deep neural networks, has made significant progress and achieved remarkable advancements in applications e.g., autonomous driving [13] and medical diagnosis [14]. Machine learning can generate an accurate map from input features to target outcomes, provided there is sufficient training data. The machine learning method has also been combined with the inverse FEM-approach to determine plastic properties of the coating. In Ref. [15], a Support vector regressor (SVR) is trained to predict load-depth curves from the parameters of the first term of the Johnson-Cook model, using FEM-simulated data for training. Given a load-depth curve, the Johnson-Cook parameters can then be optimized via Particle Swarm Optimization (PSO) to make the SVR-predicted load-depth curve closely match the given one. However, this approach is limited by the low precision of the load-depth curve prediction. In Ref. [15], the reported prediction error ranges from −22 mN to 18 mN for nanoindentation at F_{max} = 40 mN. There are several

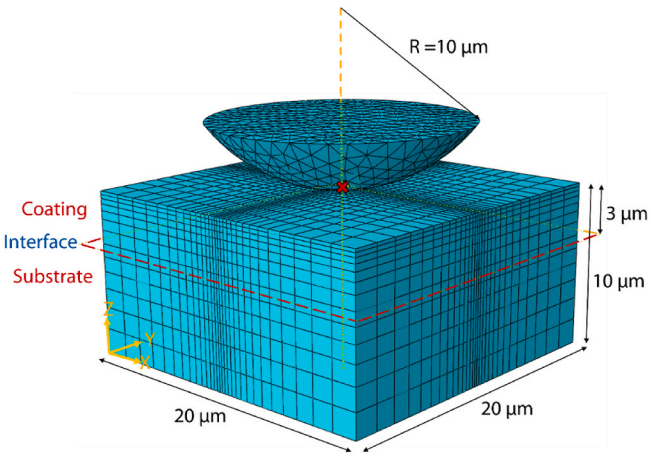

Fig. 3. FEM simulation of the nanoindentation process, incorporating the spherical indenter, the coating and the substrate.

similar studies testing models only on the simulation data, without the application on real experimental data [16,17].

Despite progress in combining machine learning with inverse FEM for plastic parameters determination, a research gap remains in developing models with sufficient accuracy for real-world applications. Bridging this gap requires addressing several key questions. The first question is to choose the most appropriate model among the wide range of classical machine learning algorithms and deep learning neural networks. Moreover, it remains an open question to improve model performance from the feature level. For practical deployment, the minimum amount of training data required to ensure reliable predictions should be known to minimize development time. Finally, the model with good prediction accuracy on simulation data should be validated on real experimental data and its generalization ability should be tested.

This work aims to develop a model which can accurately map experimental load-depth curves to stress-strain relationship of PVD tool coatings. In this work, Hooke's stress-strain relation is employed to define the elastic stress-strain relationship of coatings:

$$\sigma = E\varepsilon \tag{1}$$

where E is the Young's modulus.

Additionally, the Ludwik-Hollomon model is used to define the plastic stress-strain relationship of coatings:

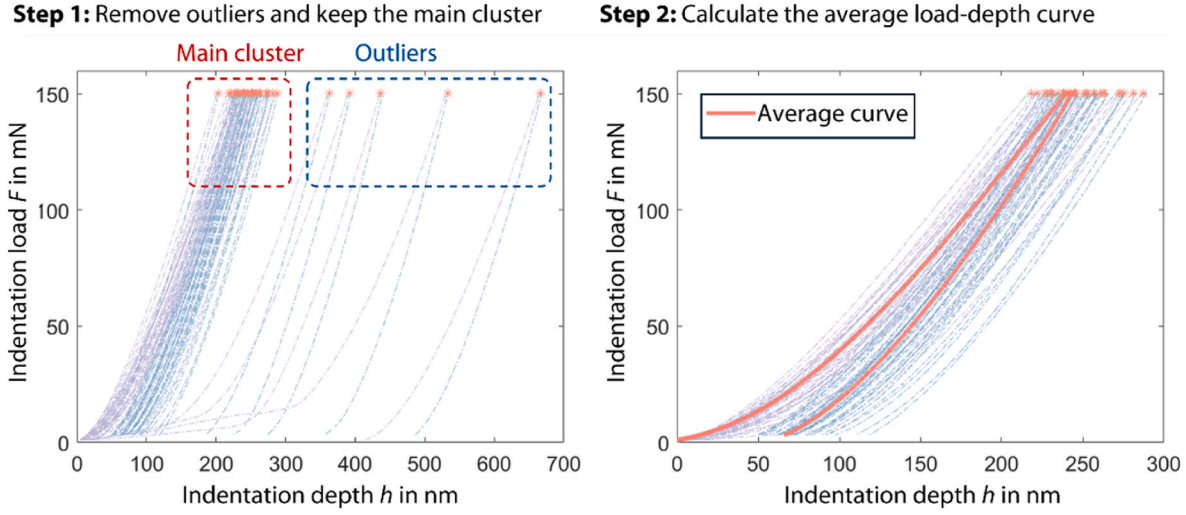

Fig. 2. Preprocessing of experimental nanoindentation data: outlier removal and computation of the average load-depth curve from the remaining measurements.

$$\sigma = A + B\varepsilon^{n} \quad (2)$$

where A represents the yield stress, B is the coefficient, and n is the exponent.

The task is then to directly predict the three parameters E, A and B from load-depth curves using machine learning, with n fixed as n = 0.4. The rationale for setting n = 0.4 is explained in Section 2.4. The holistic methodology is illustrated in Fig. 1.

The first step involves generating the dataset by setting various combinations of three parameters, i.e., E, A and B in a FEM model and running simulations of nanoindentation. Using the simulated dataset, machine learning models are trained, validated and tested in the second step. This study presents a comprehensive comparison of various models, including classical machine learning algorithms and deep learning neural networks with different structures, to develop a model with high prediction accuracy. The models evaluated include the Support vector regressor (SVR), multilayer perceptron (MLP) and two recurrent neural networks (RNN) variants: long-short term memory (LSTM) and gated recurrent unit (GRU). Their performance is further assessed on datasets with varying sizes to examine the effect of data quantity on model performance. A hybrid model combining models showing best performance in evaluation is chosen for application.

Furthermore, feature study is done to investigate how to improve model performance from the feature level and shows that using simulated load-depth curves under two maximum forces as input data can increase the prediction accuracy of B compared to using only a single load-depth curve. The feature importance of each depth value is also investigated using least absolute shrinkage and selection operator (Lasso) algorithm [18].

Finally, the hybrid model trained on simulation data is applied to experimental nanonindentation data. The predicted coating parameters are given to the FEM simulation model to see the alignment between experimental and simulated load-depth curves. In this study, the developed model is used to investigate mechanical properties of several TiAlCrN coatings with varying coating thicknesses, validating the generalization of the approach.

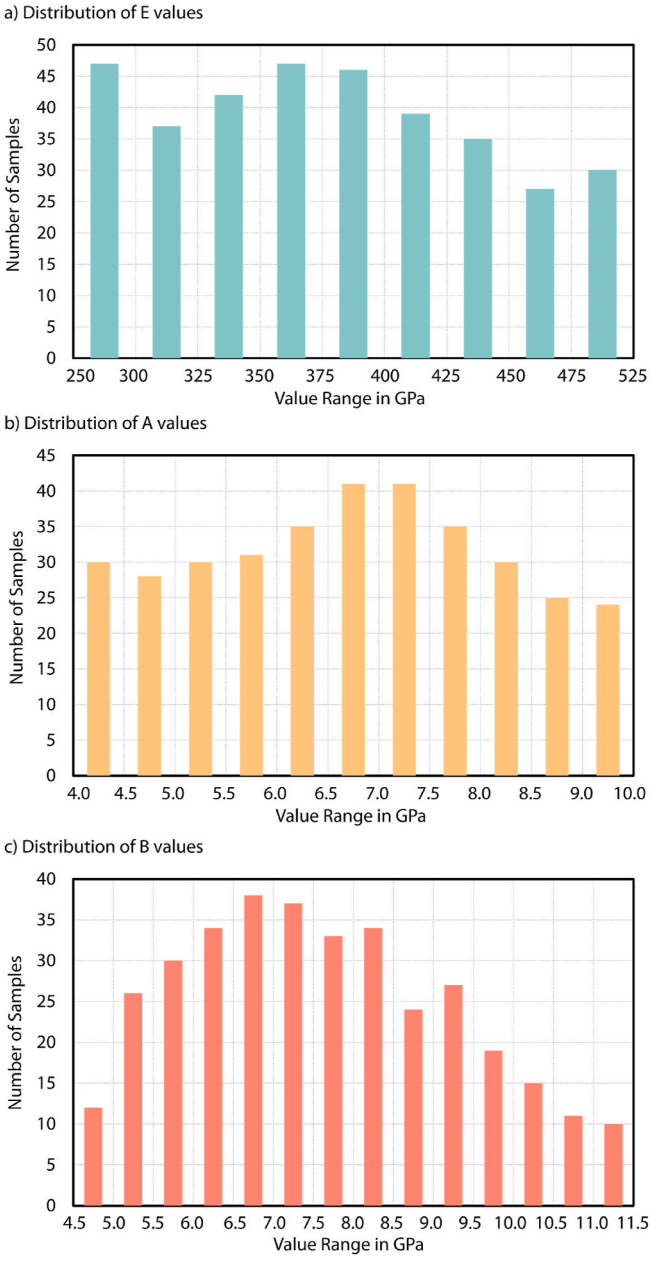

Fig. 4. Distribution of values used to select E, A and B as inputs for FEM simulations to generate the dataset.

2. Experimental details

2.1. Coating deposition

The TiAlCrN coating requiring investigation is deposited on the cemented carbide substrate (WC-Co with 6 wt% Co) supplied by CERATIZIT GmbH, Luxemburg, using an industrial CC800/9 HPPMS coating unit from Cemecon AG, Würselen, Germany. Two HPPMS power supplies are connected to TiAl48 and TiAl20 targets, while a dc power supply is connected to the Cr target during the coating deposition. The process parameters are shown in Table 1. By adjusting the coating deposition time, coatings with three different thicknesses are achieved for further investigation of the model's generalization ability to coatings with varying thicknesses.

2.2. Coating characterization

The morphology of the coating is evaluated using scanning electron microscopy (SEM) at Central Facility for Electron Microscopy GFE, RWTH Aachen University, where the chemical composition is also determined using energy dispersive X-ray spectroscopy (EDX). The experimental load-depth curves from nanoindentation are obtained using a TI950 Triboindenter from Bruker Corporation, Billerica, Massachusetts, USA, equipped with a spherical indenter of tip radius r = 10 µm. Nanoindentations are conducted on the TiAlCrN coating using the maximum forces of F_{max} = 70 mN as well as F_{max} = 150 mN. The selected force in this range is intended to minimize the effect of the surface roughness and crack behavior during nanoindentation. The nanoindentation process comprises sequential loading process and unloading process. During loading, the indentation force gradually increases from F = 0 mN to F_{max} over 5 s. This is followed by a 3-s holding period before the unloading phase, in which the force is continuously reduced from F_{max} to F = 0 mN over 5 s. Fifty indentations should be performed on the coating. The processing of these fifty indentations as experimental data is illustrated in Fig. 2. Among these indentations, several outliers are identified and excluded. The remaining curves, representing the main cluster, were retained for further analysis. An average load-depth curve is then calculated from the remaining indentation curves of the main cluster to statistically characterize the elastic and plastic behavior of the coating.

2.3. Finite element simulation

In addition to real nanoindentation experiments, the nanoindentation process is also simulated using FEM with Abaqus 6.14 from Dassault Systèmes, France. The FEM simulation model is depicted in

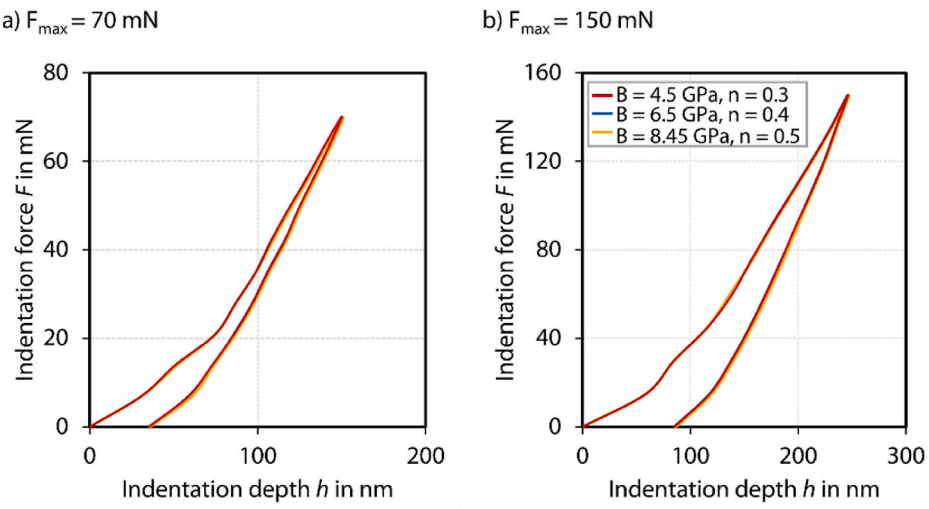

Fig. 5. Nearly identical simulated load-depth curves generated from different (B, n) pairs with identical E and A in simulations (E = 360 GPa, A = 6.5 GPa).

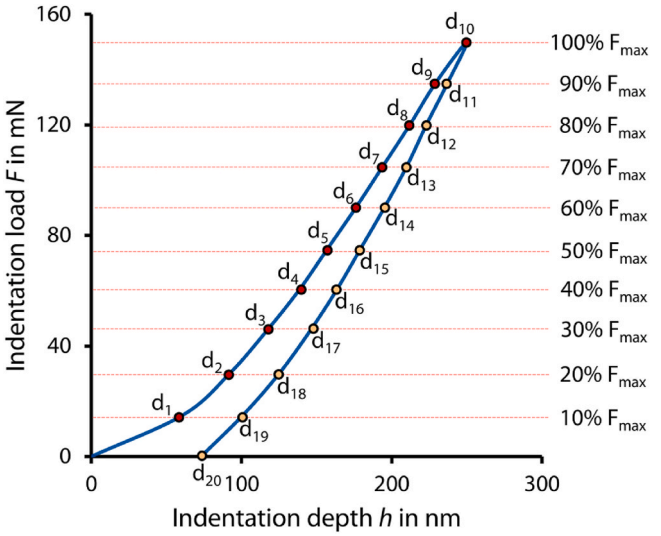

Fig. 6. Selection of depth values corresponding to specific indentation loads during both the loading and the unloading phase as input feature for machine learning models.

Fig. 3. To reduce computational demands, only a small specimen of the coating and the substrate having contact with the indenter is modeled. The length-to-width ratio of the modeled specimen remains consistent with the original sample. The specimen measures 20 µm in both length and width and 10 µm in height. As shown in Fig. 3, the coating thickness is initially set to s = 3 µm, a typical value for PVD coatings used in cutting tools. This study first focuses on predicting the mechanical properties of PVD coatings with thickness close to this value. Subsequently, an investigation explores whether a model trained using simulation data with this specific thickness can be generalized to coatings with different thicknesses.

To simplify the model, both the "substrate" and "coating" segments are assumed to have uniform material properties throughout. The elastic properties of both the coating and the substrate are assumed to adhere to Hooke's stress-strain relation shown in Eq. (1). Elastic modulus E is required to define elastic properties. And the model used to describe the plastic properties of the coating and the substrate is the Ludwik-Hollomon model shown in Eq. (2). The yield stress A, coefficient B and exponent n are then required to define the plastic properties. Poisson's ratio of the coating is set to $\nu = 0.25$ and Poisson's ratio of the substrate is set to $\nu = 0.21$. The elastic modulus of the substrate in the simulation model is set to be E = 600 GPa provided by the producer. The Ludwik-Hollomon model parameters of the substrate is set to be A = 1.3 GPa, B = 12.0 GPa and n = 0.5 following [19]. The density of the coating is calculated by measuring the samples' mass before coating, after etching, and after coating. Based on the calculated value the density is set to 6.2 **g/cm^3**. The density of the substrate comes from the producer and is set to 14.88 **g/cm^3**.

Different friction coefficients between the coating and the substrate ranging from 0.05 to 0.2 are tested, and the resulting indentation depths are compared. The results show no significant difference between various friction coefficients. However, a noticeable difference is observed when the interface is assumed to be ideal, with no friction. Therefore, the friction coefficient between the coating and the indenter is set to be µ = 0.1.

The substrate and the coating are discretized using cubic meshes, incorporating gradient meshing to have finer mesh near the indenter, as shown in Fig. 3. A mesh convergence study is first conducted to assess the impact of mesh size on indentation depths during the simulation. According to the results of the mesh convergence analysis, the gradient mesh size is finally set to be between 0.1 µm and 2 µm to achieve a balance between the simulation accuracy and the computation speed. The indenter is made of diamond, whose elastic modulus is set to E = 1200 GPa. The indenter is discretized using irregular meshes.

Explicit methods are used to solve the equations iteratively with time steps in Abaqus. By solving these equations, states of the system at time steps during the loading and unloading of nanoindentation can be simulated.

2.4. Dataset generation

As the first step of the holistic methodology shown in Fig. 1, the dataset is generated using the built FEM simulation model. The research first focuses on properties of coatings with a thickness of s = 3 µm, which is a common coating thickness of PVD. Therefore, all simulation models used to generate the dataset have a coating thickness of s = 3 µm. The dataset is generated by altering the mechanical properties of the coating. Three parameters including elastic modulus E, yield stress A, and coefficient B are randomly selected from a wide range with the exponent n fixed as n = 0.4. Distributions of selected values are shown in Fig. 4. For each parameter set, simulations at two load levels F_{max} = 70 mN and F_{max} = 150 mN are run.

The reason for fixing n = 0.4 is that the coefficient B and exponent n are strongly coupled, jointly defining how stress increases with strain

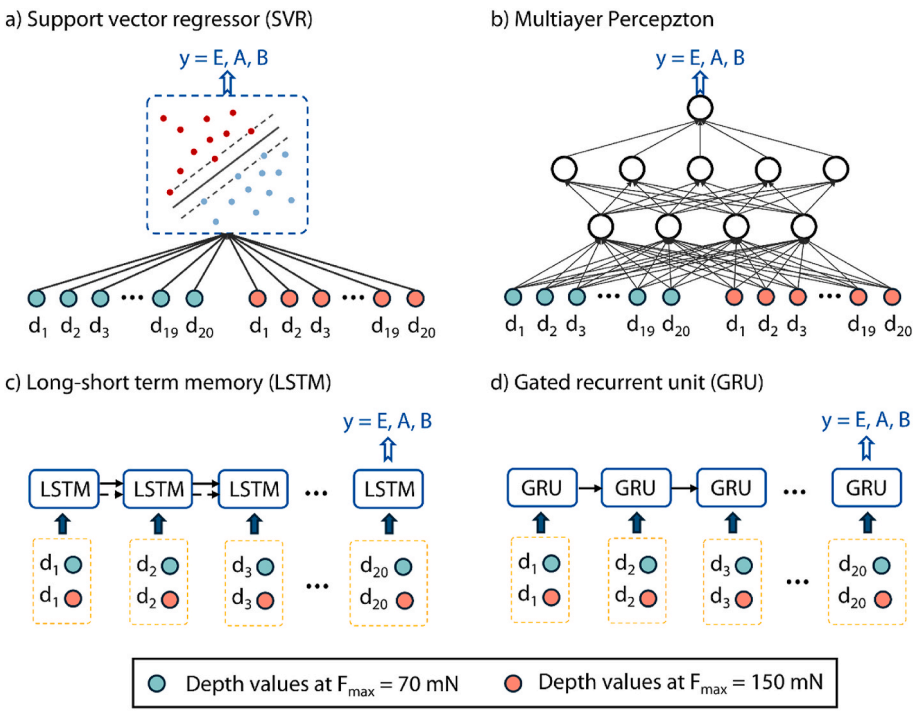

Fig. 7. Model structures of the four machine learning models, using depth values as input features and predicting E, A and B as output targets.

Table 2
Coating characterization of TiAlCrN coatings.

Coating	ID5280	ID5320	ID5403
Coating thickness s [μm]	2.1	2.9	3.8
Ti content [at %]	36.1	36.9	36.3
Al content [at %]	58.9	57.9	58.5
Cr content [at %]	5.0	5.2	5.2
Average line roughness Ra [μm]	0.02	0.02	0.02
Indentation modulus E_{IT} [GPa]	348.3 ± 15.9	351.3 ± 22.4	368.8 ± 20.6
Indentation hardness H_{IT} [GPa]	33.8 ± 2.8	30.9 ± 3.6	34.4 ± 3.34

after the stress reaches the yield stress. For a given pair (B, n), a change in n can be compensated by a corresponding change in B, resulting in a similar plastic stress-strain behavior. Consequently, load-depth curves generated from these different (B, n) pairs, despite having individual parameter values, appear nearly identical. For example, as shown in Fig. 5, three simulated load-depth curves generated using different combinations of (B, n) are almost indistinguishable. This implies that multiple (B, n) pairs can plausibly fit the same load-depth curve, making a simultaneous prediction of (B, n) almost impossible.

Since the ultimate goal is to capture the plastic stress-strain behavior beyond the yield stress rather than precisely predict individual (B, n) pairs, and diverse plastic stress-strain behaviors of the coating can already be represented by variability of B values, decoupling B and n to focus on a precise prediction of B is a more practical solution. The fixed value of n = 0.4 is chosen based on our previous works [20,21], where n values for nitride PVD coatings are determined using inverse FEM-simulations to be between 0.3 and 0.5.

The load-depth curve for each simulation with coating properties defined by three randomly chosen parameters is recorded. Each simulation takes around 20 min. Finally, the dataset consisted of 350 pairs of "material parameters: load-depth curve".

2.5. Machine learning model

As the second step of the holistic methodology shown in Fig. 1, machine learning models are trained to predict the material parameters from load-depth curves. Therefore, the output data of machine learning models are material parameters, i.e., three parameters E, A and B. The input data consists of depths at specific loads from both load-depth curves using the maximum indentation forces of F_{max} = 70 mN and

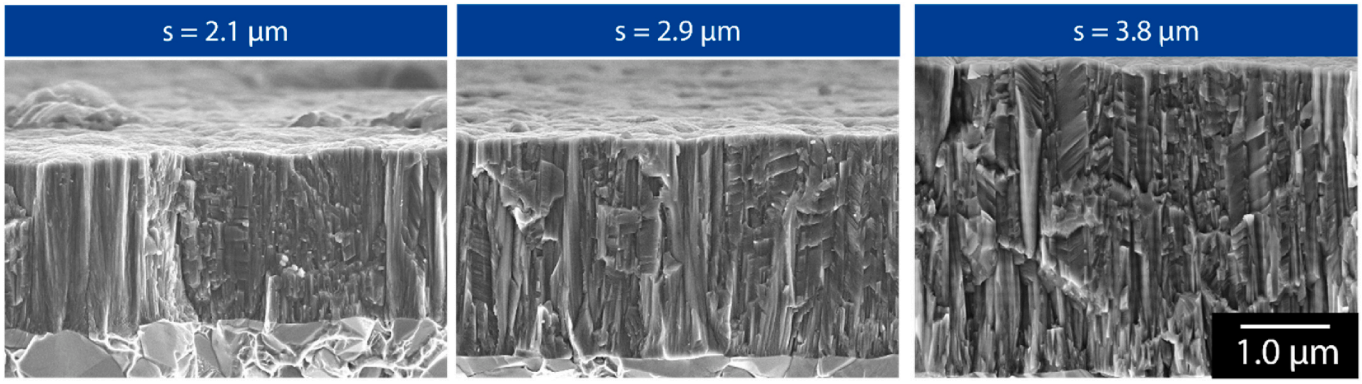

Fig. 8. Morphology of TiAlCrN coatings on the WC-Co substrate with coating thicknesses of s = 2.1 μm, s = 2.9 μm and s = 3.8 μm.

Table 3
Determined optimal hyperparameters for each model.

Model	Hyperparameters
SVR	Prediction of E: radial basis function (RBF) as the kernel function, Regularization parameter C = 1000, Kernel coefficient γ = 'scale', Penalty coefficient: ε = 0.001 Prediction of A: radial basis function (RBF) as the kernel function, Regularization parameter C = 10, Kernel coefficient γ = 'auto', Penalty coefficient: ε = 0.001 Prediction of B: radial basis function (RBF) as the kernel function, Regularization parameter C = 5, Kernel coefficient γ = 'scale', Penalty coefficient: ε = 0.001
MLP	Prediction of E: 2 hidden layers, 128 units in the first layer, 64 units in the second layer, dropout rate 0.3, Adam optimizer, learning rate 0.001, weight decay 0.00001, training epochs 200, batch size: 32 Prediction of A: 2 hidden layers, 128 units in the first layer, 64 units in the second layer, dropout rate 0.3, Adam optimizer, learning rate 0.0005, weight decay 0.0001, training epochs 150, batch size: 8 Prediction of B: 2 hidden layers, 128 units in the first layer, 64 units in the second layer, dropout rate 0.3, Adam optimizer, learning rate 0.001, weight decay 0.0001, training epochs 200, batch size: 32
LSTM	Prediction of E: 2 hidden layers, each layer 128 units, dropout rate 0.3, Adam optimizer, learning rate 0.0005, weight decay 0.00001, training epochs 200, batch size: 32 Prediction of A: 3 hidden layer, each layer 128 units, dropout rate 0.3, Adam optimizer, learning rate 0.0005, weight decay 0.00001, training epochs 150, batch size: 32 Prediction of B: 1 hidden layer, each layer 32 units, dropout rate 0.3, Adam optimizer, learning rate 0.001, weight decay 0.00001, training epochs 150, batch size: 16
GRU	Prediction of E: 2 hidden layers, each layer 32 units, dropout rate 0.3, Adam optimizer, learning rate 0.001, weight decay 0.00001, training epochs 150, batch size: 8 Prediction of A: 1 hidden layer, each layer 64 units, dropout rate 0.3, Adam optimizer, learning rate 0.01, weight decay 0.00001, training epochs 100, batch size: 32 Prediction of B: 2 hidden layer, each layer 128 units, dropout rate 0.3, Adam optimizer, learning rate 0.001, weight decay 0.00001, training epochs 150, batch size: 16

Table 4
MAPE for the prediction of three parameters using four machine learning models.

	E	A	B
SVR	0.12 %	2.54 %	12.18 %
MLP	0.90 %	3.84 %	16.08 %
LSTM	1.35 %	4.12 %	20.90 %
GRU	0.87 %	1.87 %	6.81 %

F_{max} = 150 mN. The way to choose data from load-depth curve is shown in Fig. 6. During the loading process, depths are selected at 10 %, 20 %, 30 % and so on, up to 100 % of the maximum forces. During the unloading process, depths are chosen at 90 %, 80 %, 70 % and so forth, down to 0 % of the maximum force. For the load-depth curve at each maximum indentation force 20 depths are chosen as input features. From two load-depth curves 40 features are achieved.

Depending on how these depths are used as features in the model, different machine learning model approaches are explored. Fig. 7 summarizes all these model structures and corresponding data feeding methods. One approach involves feeding all depth values directly into the model simultaneously. For this approach SVR and MLP are tested. However, since nanoindentation is a sequential process and load-depth curves represent sequential data, an alternative approach is to input the depths sequentially. At each recurrent step, two depth values are entered into the model, each corresponding to depths at a specific proportion of the maximum force on two different load-depth curves. Based on this method, two variants of RNN, i.e., LSTM and GRU could be used. These RNN models with gated structure could mitigate the vanishing gradient problem during the training of the traditional RNN models [22].

3. Results

3.1. Coating characterization

The characterized properties of TiAlCrN coatings are outlined in Table 2. The morphology of coatings with different coating thicknesses under SEM is shown in Fig. 8. Coating thicknesses are determined under SEM, yielding measurement values of s = 2.1 μm, s = 2.9 μm and s = 3.8 μm. The indentation modulus E_{IT} and the indentation hardness H_{IT} of the TiAlCrN coating polished prior to nanoindentation are measured using the nanoindentation with Berkovich indenter under F_{max} = 8 mN.

3.2. Comparison of machine learning models

Machine learning models are implemented using Python programming language. The machine learning models, whose structures are illustrated in Fig. 7, are trained, tested and compared using simulation data. Among the 350 samples, 10 % are first randomly selected as test data. The remaining 90 % are used to determine the optimal hyperparameters via 5-fold cross-validation. In this process, the dataset is divided into 5 equal folds: in each iteration, four folds of the data are used for training the model and one fold for validation. This is repeated five times, with each fold used as the validation set. The average validation accuracy across all folds is then calculated, and the hyperparameter setting that yields the highest average accuracy is selected as the final hyperparameter. The determined hyperparameters for each model are shown in Table 3.

After selecting the best hyperparameters for four models, these four models with their best hyperparameters are then applied to the test data and compared. The loss function used as evaluation criteria is mean absolute percentage error (MAPE) to assess relative error. Greater MAPE indicates bigger prediction errors. The comparison of the four models on the test set is presented in Table 4.

Elastic modulus E is clearly the easiest parameter to predict. All models achieve a MAPE of less than 1 %, with SVR performing best at 0.12 %. Given the complexity of the prediction task and the size of the dataset, this result is reasonable, as SVR is well-suited for small datasets and robust against overfitting [23]. The strong influence of modulus E on depth variations during elastic deformation likely provides distinct features that models can effectively utilize.

Yield stress A is the second easiest parameter to predict, with all models achieving a MAPE below 5 %. GRU works best for the prediction of the yield stress A with an accuracy of 1.87 %. Yield stress A strongly affects the proportion of the plastic deformation, which could be shown by the residual indentation depth [24]. And the maximum indentation depth is also strongly affected by the yield stress. These features could give valuable information that aids the models in making accurate predictions.

However, coefficient B is the most difficult to predict. All models exhibit a MAPE of over 5 % when predicting B, with MLP and LSTM exceeding 15 %. Similar to the prediction of the yield stress A, SVR and GRU perform better in predicting the coefficient B. Among all models, GRU achieves the highest prediction accuracy in predicting B. The reason could be that GRU efficiently uses sequential information, which is likely beneficial for predicting B, as this parameter B depends on complex, deeply embedded patterns in the input features.

3.3. Data size effect

To investigate the impact of dataset size on model performance, the models are trained on progressively smaller subsets of the data, and their prediction accuracy is evaluated on the test set. The smaller datasets, consisting of 63, 126, 189 and 252 samples, are randomly selected from the original train-validation set. Models are then trained on these smaller subsets, and their corresponding test MAPE are shown in Fig. 9.

For SVR, the prediction accuracy of all three parameters steadily

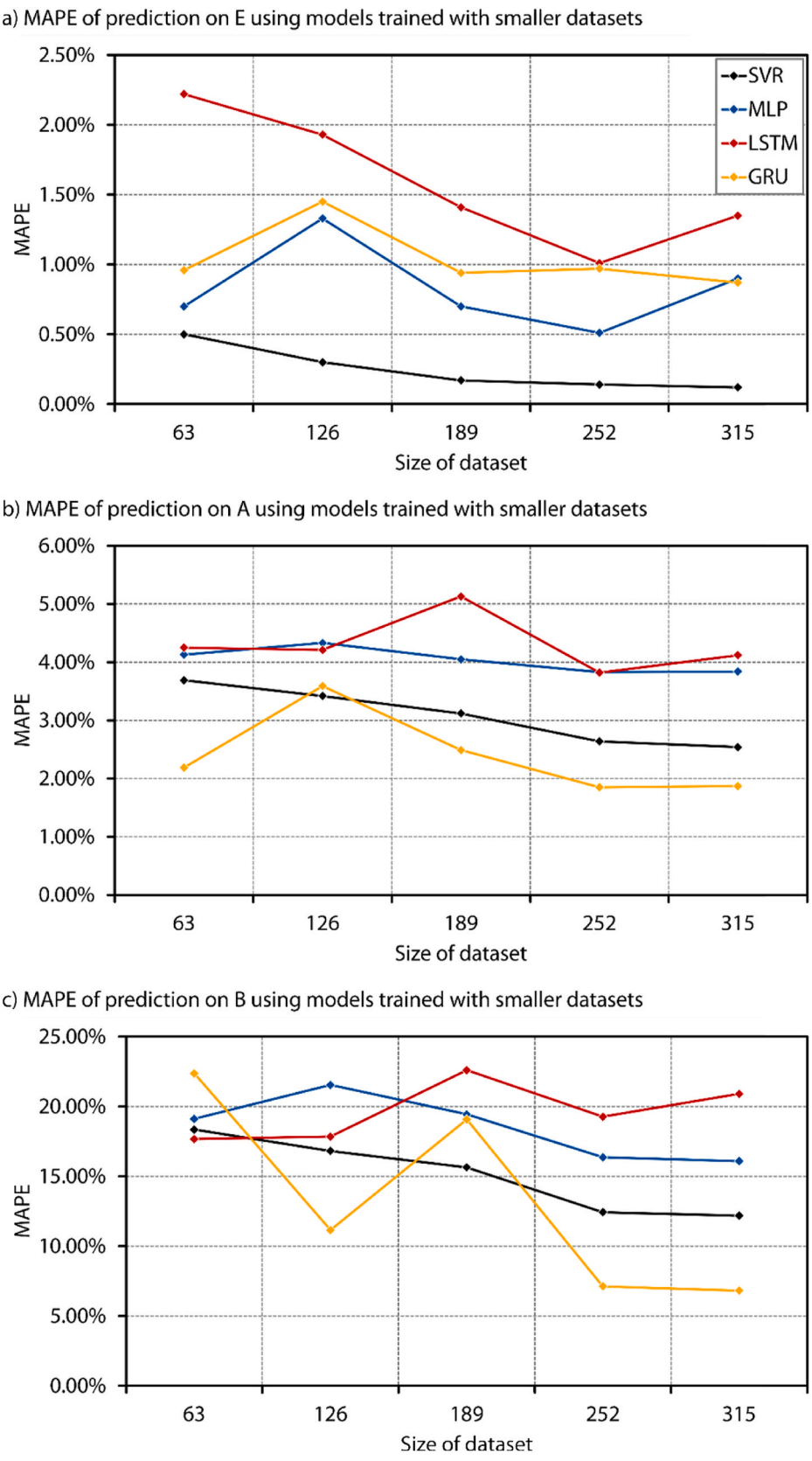

Fig. 9. Prediction MAPE of models when they are trained using a smaller dataset.

Table 5
Comparison between MAPE using load-depth curves from different forces.

Used load-depth curves	E	A	B
Single force F_{max} = 70 mN	0.13 %	3.15 %	19.97 %
Single force F_{max} = 150 mN	0.15 %	4.25 %	18.56 %
Two forces	0.12 %	1.87 %	6.81 %

improves with increases in dataset size. For MLP, the prediction accuracy of E is already high with 63 samples and does not improve further with larger datasets, whereas the prediction of A and B benefits from more data, showing a sharp increase once the dataset exceeds 252 samples. For LSTM, increasing the dataset yields little additional advantages. For GRU, accurate prediction of E is achieved with 63 samples, while a substantial improvement in predicting of A and B occurs when the dataset size reaches 252 samples. Notably, 252 appears to be a critical point: at this dataset size, the prediction accuracy of SVR, MLP and GRU for A and B rises sharply and then stabilizes.

3.4. Feature engineering

Feature engineering is an important step in understanding which input features are used by the model for prediction. As a first step, the necessity of using nanoindentation data obtained under two different maximum forces should be validated. Since SVR yields the best performance for the prediction of E and GRU excels in predicting A and B, a hybrid model is further investigated, with SVR used for E prediction and GRU for A and B. The hybrid model with optimal hyperparameters is trained using datasets containing load-depth curves under a single force (F_{max} = 70 mN), a single force (F_{max} = 150 mN) and both forces. The prediction accuracies on the test set are then compared. The results are presented in Table 5. The results show that the prediction accuracy for E does not change significantly when using data from only a single force. This suggests that a single load-depth curve provides sufficient information for predicting E. In contrast, accurately determining A and B from a single curve appears to be more challenging. Prediction accuracies for the yield stress A and coefficient B greatly increases when data from an additional load-depth curve is included.

Table 6
Predicted parameters for the TiAlCrN coating.

	E [GPa]	A [GPa]	B [GPa]
TiAlCrN, s = 2.9 µm	343	7.6	6.8

To gain deeper insight into which features are selected by the model when data from two load-depth curves are provided, the least absolute shrinkage and selection operator (Lasso) algorithm is employed. Lasso uses L1 regularization parameters within a linear SVR framework to identify the most influential features [18]. In this study, Lasso is used to evaluate the importance of each feature, i.e., the depth value, when in total 40 depth values from two load-depth curves are input to the model. The results are shown in Fig. 10.

The results indicate that features from both curves are significant for the prediction. The features used to predict E focus on the initial deformation phase, depths at maximum forces and the residual indentation depth at F_{max} = 70 mN, particularly the initial deformation phase, where elastic deformation plays an important role. The prediction of the yield stress A relies more on the unloading phase, especially the unloading phase at F_{max} = 70 mN. The prediction of B incorporates an even greater number of depth values from the unloading phase, notably the last four depth values at both maximum loads. The initial loading phase, which is dominated by elastic deformation, contributes not much to the prediction of B.

3.5. Model performance on experimental data

After investigating the models on simulation data, the model should be applied to experimental data, as the third step of the holistic methodology shown in Fig. 1. This task is more challenging, as real coatings are not as ideal as those in the simulations. The non-ideal deformation behavior in real experiments could result in irregular depth values [25], which, in turn, could affect the predictions of machine learning models. For the real coating, nanoindentation data with two maximum forces,

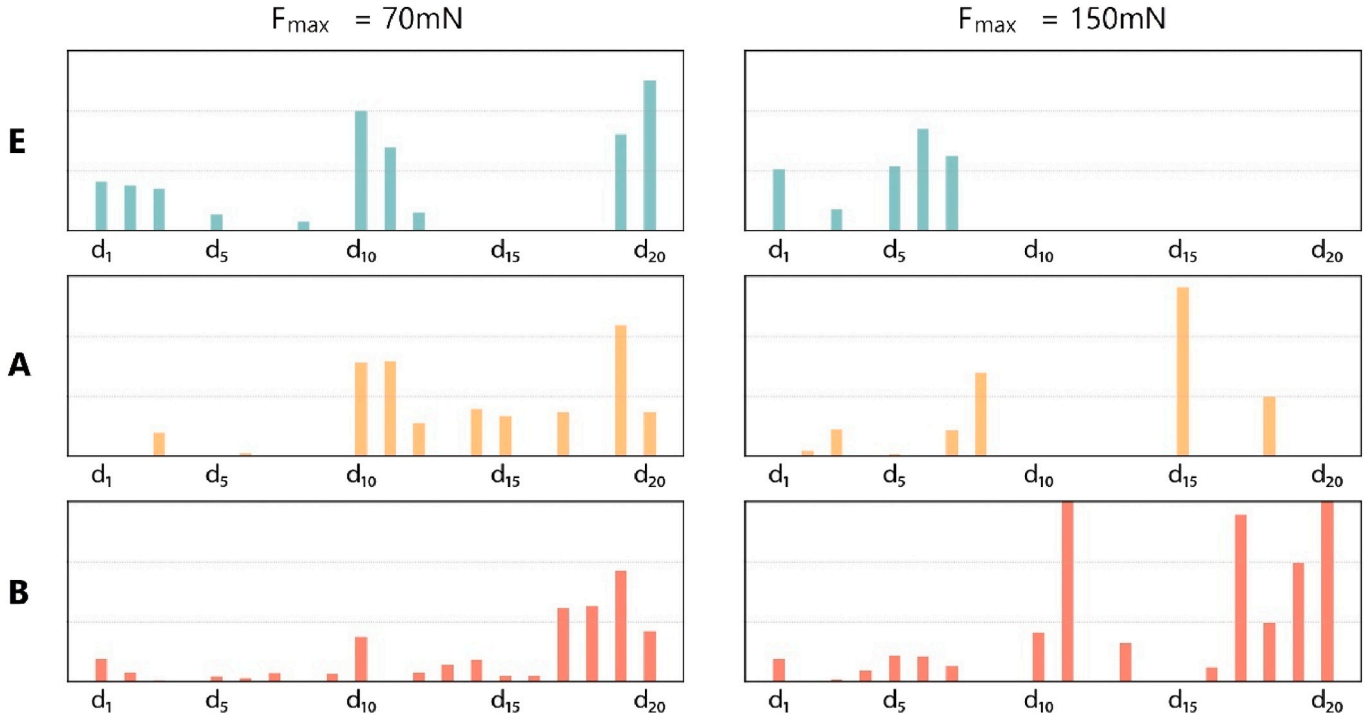

Fig. 10. Feature importance for predicting each target variable (E, A and B) using the Lasso algorithm to the SVR model, where taller bars indicating greater importance.

F_{max} = 70 mN and F_{max} = 150 mN, are used as input for the machine learning models, same as the procedure using simulation data, as shown in Fig. 6. The input experimental load-depth curves are mean curves from 50 measurements, as illustrated in Fig. 2. The trained hybrid model combining SVR and GRU is first applied to a TiAlCrN coating with a coating thickness of s = 2.9 μm, which is close to the coating thickness of s = 3 μm in the simulation model to generate the dataset. The predicted parameters for the TiAlCrN coating with the coating thickness of s = 2.9 μm are shown in Table 6.

As the final step of the holistic methodology shown in Fig. 1, simulations are conducted using the predicted parameters, and the comparison between experimental and simulated load-depth curves is shown in Fig. 11. Simulated curves align well with experimental curves under F_{max} = 70 mN and F_{max} = 150 mN. The shaded area in the figure indicates the confidence intervals of experimental and simulated curves. For experimental load-depth curve, the confidence interval is calculated based on the standard deviation of depth values at specific loads. The confidence region for the simulated curve considers the prediction error of the hybrid model, especially the prediction error of coefficient B.

To evaluate if the model can generalize to coatings with different thicknesses, experimental nanoindentation data under two maximum forces, F_{max} = 70 mN and F_{max} = 150 mN, from TiAlCrN coatings with thicknesses of s = 2.1 μm and s = 3.8 μm is further provided to the trained hybrid model. The parameters predicted by the model are presented in Table 7.

Using the predicted parameters, the simulation model is first run still with the coating thickness of s = 3 μm to assess whether experimental data from a different coating thickness affects the prediction due to potential different deformation behavior. Subsequently, the coating thickness in the simulation model is adjusted to the actual values of s = 2.1 μm and s = 3.8 μm. The alignment between the experimental load-depth curve and the simulated load-depth curves at s = 3 μm and the actual coating thicknesses is shown in Fig. 12.

For the coating with a thickness of s = 2.1 μm, the simulation at F_{max} = 70 mN shows good agreement with the experimental load-depth curve. At F_{max} = 150 mN, the simulation using a thickness of s = 3 μm also aligns well with the experimental data, indicating that the machine learning model provides an accurate prediction. However, the simulation using the actual coating thickness exhibits a generally good match but shows a larger deviation toward the end of the unloading phase. At F_{max} = 150 mN, the maximum indentation depth reaches 242 nm, exceeding 10 % of the coating thickness. In this case, the coating thickness could have an influence on the prediction. When the coating thickness is increased to s = 3.8 μm, the maximum indentation depth remains below 10 % of the coating thickness, even at F_{max} = 150 mN. Under these conditions, there is almost no difference between the simulation model using s = 3 μm and s = 3.8 μm, and both curves closely follow the experimental data. In summary, the machine learning model trained on simulation data with a single thickness can be generalized to predict properties for coatings with different thicknesses, particularly for those with a larger thickness.

Table 7
Predicted parameters for TiAlCrN coatings with different thicknesses.

	E [GPa]	A [GPa]	B [GPa]
TiAlCrN, s = 2.1 μm	376	8.4	6.6
TiAlCrN, s = 3.8 μm	349	7.3	8.4

The holistic approach is sensitive to the quality of the experimental data provided. In real nanoindentation experiments, load-depth curves can be influenced by various factors, such as the surface roughness, grain size variations, indentations near grain boundaries, and three-dimensional coating defects. These factors can lead to abnormal curve features, such as pop-in or pop-out events [26]. In this study, data preprocessing is performed, and outliers are removed. Although a few curves exhibiting pop-in or pop-out events remained. The influence of these events appeared negligible after calculating the mean load-depth curve from over 40 remaining curves.

To further evaluate the robustness of the model in cases where only limited load-depth curves are available and the input curve contains pop-in or pop-out events, we test the trained hybrid model with an experimental load-depth curve exhibiting a pop-in event at F_{max} = 70 mN, alongside the calculated mean load-depth curve under F_{max} = 150 mN for the coating with a thickness of s = 2.9 μm. The predicted parameters are shown in Table 8.

The simulated load-depth curves obtained using the predicted parameters are compared with the experimental curves in Fig. 13. In this case, the residual indentation depth under F_{max} = 70 mN is even larger than that under F_{max} = 150 mN, which is a contradiction that can confuse the model. Although the model responds reasonably by reducing the yield stress A and increasing the coefficient B, the overall agreement between experiment and simulation remains unsatisfactory. This case demonstrates that the model is sensitive to inconsistencies in experimental data. Therefore, careful preprocessing is essential for the successful application of the approach, including performing sufficient repetitions of nanoindentation and rigorously removing outliers. Moreover, incorporating controlled disturbances into the simulated training data could be a potential solution to enhance the robustness of the trained model.

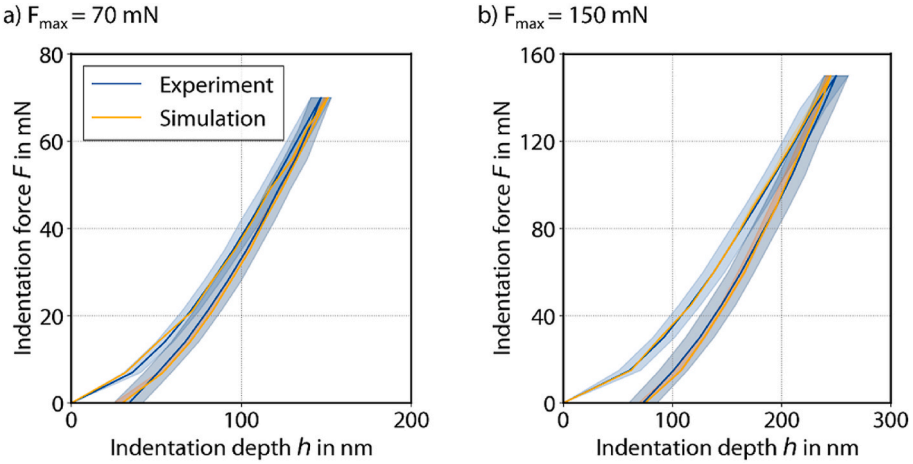

Fig. 11. Comparison between experimental and simulated load-depth curves using predicted values of E, A and B for the coating properties in the simulation model.

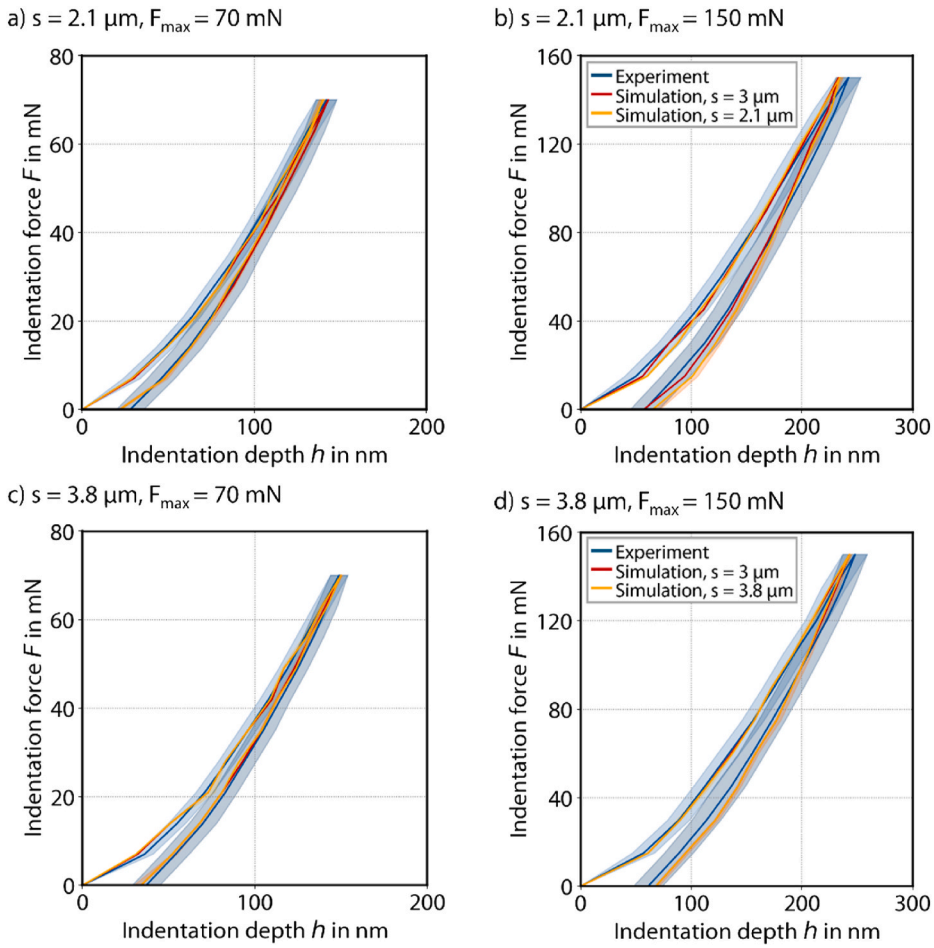

Fig. 12. Comparison between the experimental load-depth curve and the simulated curves at a coating thickness of s = 3 µm and at the actual coating thickness.

Table 8
Predicted parameters given a load-depth curve with pop-in events.

	E [GPa]	A [GPa]	B [GPa]
TiAlCrN, s = 2.9 µm	367	4	11

4. Discussion

In our work, machine learning models demonstrate their ability to directly predict mechanical properties from nanoindentation load-depth curves. For predicting the elastic modulus E, a simple and overfitting-resistant machine learning model SVR, can already achieve good performance. GRU works better for more challenging prediction tasks of A and B.

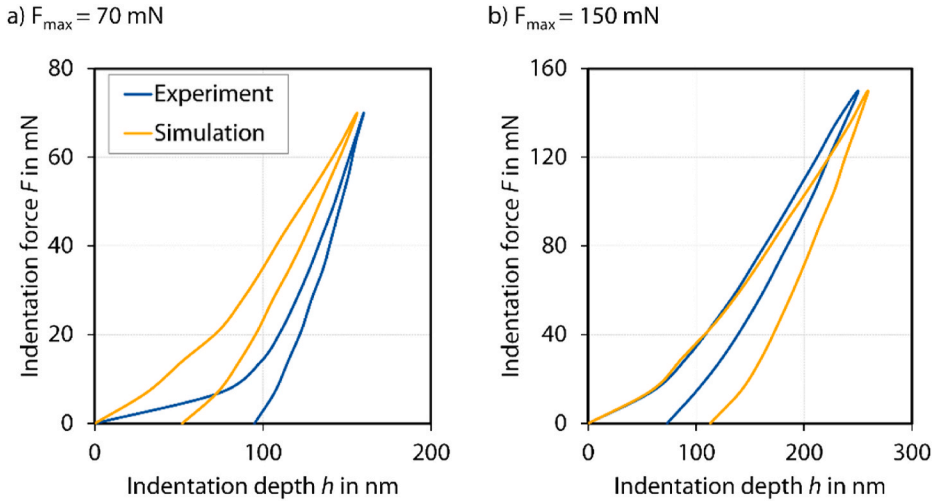

Fig. 13. Comparison between simulated and experimental load-depth curves when the model is given the data with abnormal events.

The advantages and limitations of this approach should also be discussed. Compared with the analytical approach like in Ref. [7], the experimental work load of the approach in this study is relatively low, requiring just nanoindentation at two force levels and minimal manual calculations and analysis. Compared with inverse FEM simulations using manual parameter adjustment, the proposed approach is fully automatic and does not rely on user experience.

Compared with inverse FEM simulations based on automatic optimization, each method has its own advantages and limitations depending on the application scenario. For example, in Ref. [5], 90 iterations are required to achieve convergence. In Ref. [12] convergence is reached after 60 iterations. The initial phase of the machine learning approach involves running hundreds of simulations to generate a dataset for the model training, which is time-consuming. Therefore, if the goal is to determine properties of only a single coating. Inverse FEM-simulation based on automatic optimization is preferable. However, for more than 10 coatings, the average number of simulations per coating in the machine learning approach can be reduced to around 20–30, which is much lower than more than 50 simulations typically required for automatic optimization approach. A promising future direction is to combine both approaches. Machine learning approach can provide better initial parameter estimates for automatic optimization, greatly reducing the number of iterations. Another limitation of the proposed approach is its sensitivity to load-depth curves exhibiting abnormal events, which restricts its applicability to coatings with rough surfaces or many 3D defects.

In the present study, the exponent n is fixed at $n = 0.4$ during dataset generation. The effect of this fixed value on prediction accuracy should be examined to determine whether an optimal n exists for maximizing model performance. Furthermore, current study focuses on coatings deposited on a single WC-Co substrate. Future research should aim to extend the approach to coatings on different substrates to enhance its applicability in real-world scenarios.

5. Conclusion

In this study, we aim to develop a model capable of accurately predicting the elastic modulus and Ludwik-Hollomon model parameters from nanoindentation load-depth curves. Four machine learning models are compared, leading to a hybrid model combining SVR and GRU, achieving the highest prediction accuracy across all target parameters. It is also validated in this study that using load-depth curves under two maximum forces is significant for improving the prediction accuracy. The developed model is validated with experimental data from three coatings with different thicknesses, confirming its generalization capability. However, approach exhibits reduced robustness when applied to experimental data with abnormal curve features, a limitation that should be addressed in future research. The proposed approach offers a promising means of obtaining plastic parameters of coatings, which are often missed in thermomechanical load simulations in cutting process required for tool wear prediction. Incorporating this approach into such simulations is expected to improve the accuracy of tool wear predictions.

CRediT authorship contribution statement

C. Kalscheuer: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **K. Bobzin:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **X. Liu:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgement

The authors gratefully acknowledge the financial support of the German Research Foundation, Deutsche Forschungsgemeinschaft (DFG), within KA 4899/1-1, project number 521383048.

Data availability

Data will be made available on request.

References

[1] K. Bobzin, High-performance coatings for cutting tools, CIRP J. Manuf. Sci. Technol. 18 (2017) 1–9, https://doi.org/10.1016/j.cirpj.2016.11.004.
[2] V. Schulze, F. Zanger, Development of a simulation model to investigate tool wear in Ti-6Al-4V alloy machining, AMR (Adv. Magn. Reson.) 223 (2011) 535–544. https://dx.doi.org/10.4028/www.scientific.net/AMR.223.535.
[3] S. Beblein, B. Breidenstein, B. Denkena, C. Pusch, H. Hoche, M. Oechsner, Thermomechanical coating load in dependence of fundamental coating properties, Proced. CIRP 58 (2017) 25–30, https://doi.org/10.1016/j.procir.2017.03.184.
[4] K. Bobzin, T. Brögelmann, N.C. Kruppe, M. Arghavani, J. Mayer, T.E. Weirich, On the plastic deformation of chromium-based nitride hard coatings deposited by hybrid dcMS/HPPMS: a fundamental study using nanoscratch test, Surf. Coating. Technol. 308 (2016) 298–306, https://doi.org/10.1016/j.surfcoat.2016.05.093.
[5] J.E. Campbell, J. Dean, R.P. Thompson, T.W. Clyne, Comparison between stress-strain plots obtained from indentation plastometry, based on residual indent profiles, and from uniaxial testing, SSRN J. (2018), https://doi.org/10.2139/ssrn.3300693.
[6] W.C. Oliver, G.M. Pharr, An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments, J. Mater. Res. 7 (1992) 1564–1583, https://doi.org/10.1557/jmr.1992.1564.
[7] S. Pathak, J. Shaffer, S. Kalidindi, Determination of an effective zero-point and extraction of indentation stress–strain curves without the continuous stiffness measurement signal, Scr. Mater. 60 (2009) 439–442, https://doi.org/10.1016/j.scriptamat.2008.11.028.
[8] H. Hertz, Miscellaneous Papers, Macmillan, 1896.
[9] K. Bobzin, T. Brögelmann, R.H. Brugnara, M. Arghavani, T.-S. Yang, Y.-Y. Chang, S.-Y. Chang, Investigation on plastic behavior of HPPMS CrN, AlN and CrN/AlN-multilayer coatings using finite element simulation and nanoindentation, Surf. Coating. Technol. 284 (2015) 310–317, https://doi.org/10.1016/j.surfcoat.2015.07.081.
[10] K.-D. Bouzakis, N. Michailidis, G. Erkens, Thin hard coatings stress–strain curve determination through a FEM supported evaluation of nanoindentation test results, Surf. Coating. Technol. 142–144 (2001) 102–109, https://doi.org/10.1016/s0257-8972(01)01275-0.
[11] K. Bobzin, N. Bagcivan, S. Theiß, J. Perne, Flow curve determination on dc-MS and HPPMS CrAlN coatings, J. Phys. D Appl. Phys. 46 (2013) 84006, https://doi.org/10.1088/0022-3727/46/8/084006.
[12] G. He, D. Sun, S. Zang, J. Chen, Z. Fang, Evaluation of the elastic-plastic properties of TiN coating by nanoindentation technologies using FEM-reverse algorithm, Surf. Coating. Technol. 409 (2021) 126855, https://doi.org/10.1016/j.surfcoat.2021.126855.
[13] S. Grigorescu, B. Trasnea, T. Cocias, G. Macesanu, A survey of deep learning techniques for autonomous driving, J. Field Robot. 37 (2020) 362–386, https://doi.org/10.1002/rob.21918.
[14] M. Bakator, D. Radosav, Deep learning and medical diagnosis: a review of literature, MTI 2 (2018) 47, https://doi.org/10.3390/mti2030047.
[15] J. Weng, R. Lindvall, K. Zhuang, J.-E. Ståhl, H. Ding, J. Zhou, A machine learning based approach for determining the stress-strain relation of grey cast iron from nanoindentation, Mech. Mater. 148 (2020) 103522, https://doi.org/10.1016/j.mechmat.2020.103522.
[16] X. Long, X. Ding, J. Li, R. Dong, Y. Su, C. Chang, Indentation reverse algorithm of mechanical response for elastoplastic coatings based on LSTM deep learning, Materials 16 (2023) 2617, https://doi.org/10.3390/ma16072617.
[17] X. Long, C. Lu, Z. Shen, Y. Su, Identification of mechanical properties of thin-film elastoplastic materials by machine learning, Acta Mech. Solida Sin. 36 (2023) 13–21, https://doi.org/10.1007/s10338-022-00340-5.
[18] J. Ranstam, J.A. Cook, LASSO regression, Br. J. Surg. 105 (2018) 1348, https://doi.org/10.1002/bjs.10895.
[19] K. Bobzin, N. Bagcivan, R.H. Brugnara, J. Perne, Flow curve determination of thin films by improved finite element models and different nanoindenter geometries, Thin Solid Films 549 (2013) 313–320, https://doi.org/10.1016/j.tsf.2013.06.037.
[20] K. Bobzin, C. Kalscheuer, M. Carlet, S. Schmauder, V. Guski, W. Verestek, M. Tayyab, 3D deformation modeling of CrAlN coated tool steel compound during nanoindentation, Surf. Coating. Technol. 453 (2023) 129148, https://doi.org/10.1016/j.surfcoat.2022.129148.
[21] J. Perne, Plastic flow behavior of (Cr, Al)N hard coatings in dependence of strain rate and nanostructure, Thin Solid Films 556 (2014) 390–394, https://doi.org/10.1016/j.tsf.2014.01.069.

[22] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, Int. J. Unc. Fuzz. Knowl. Based Syst. 6 (1998) 107–116, https://doi.org/10.1142/S0218488598000094.
[23] Efficient Learning Machines, Apress, Berkeley, CA, 2015.
[24] D. Tonazzi, F. Piva, A. Mondelin, G. Le Jeune, Y. Mahéo, F. Massi, Investigation of the material elasto-plastic response under contact indentation: the effect of indenter material, Lubricants 11 (2023) 438, https://doi.org/10.3390/lubricants11100438.
[25] X. Fang, H. Bishara, K. Ding, H. Tsybenko, L. Porz, M. Höfling, E. Bruder, Y. Li, G. Dehm, K. Durst, Nanoindentation pop-in in oxides at room temperature: dislocation activation or crack formation? J. Am. Ceram. Soc. 104 (2021) 4728–4741, https://doi.org/10.1111/jace.17806.
[26] S. Pathak, J.L. Riesterer, S.R. Kalidindi, J. Michler, Understanding pop-ins in spherical nanoindentation, Appl. Phys. Lett. 105 (2014) 161913, https://doi.org/10.1063/1.4898698.