Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-guided mechanistic residual learning for sparse-condition force prediction in S45C end milling

Sangjin Maeng

sjmaeng@hongik.ac.kr

Hongik University https://orcid.org/0000-0001-9166-7297

Research Article

Keywords: End milling, Cutting force prediction, Mechanistic model, Physics-guided machine learning, Sparse condition holdout

Posted Date: July 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10156326/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Physics-guided mechanistic residual learning for sparse-condition force prediction in S45C end milling

Sangjin Maeng[a, *]

[a] Mechanical & System Design Engineering, Hongik University, Seoul, 04066, South Korea

* Corresponding author

E–mail: sjmaeng@hongik.ac.kr

Abstract

Reliable prediction of milling-force waveforms under unseen machining conditions is essential for data-efficient process modeling and intelligent machining systems. Mechanistic force models provide physically interpretable predictions with limited calibration data, but they often cannot fully capture systematic waveform deviations caused by non-ideal cutting behavior, measurement dynamics, engagement uncertainty, and axis-dependent residual effects. Conversely, purely data-driven neural networks can represent nonlinear force patterns, but their generalization may become unstable when experimental coverage of machining conditions is sparse. To address these limitations, this study proposes a physics-guided mechanistic residual learning framework for angle-domain three-axis milling-force prediction. The proposed framework uses an equivalent single-active-edge mechanistic model as a physics-based baseline force map. The mechanistic component incorporates engagement geometry, instantaneous chip thickness, helical axial slicing, phase lag, and coordinate transformation. A neural network is then trained to learn only a bounded residual correction to the mechanistic prediction, rather than replacing the physical model with an unconstrained black-box predictor. The training objective combines force-domain data loss, active-cut physics loss, and residual regularization to preserve the stability of the mechanistic model while compensating for systematic prediction errors. The framework was evaluated using dry up-milling force data from S45C steel under sparse condition-level holdout protocols designed to test generalization to entirely unseen machining conditions. The proposed model consistently improved prediction accuracy and waveform stability compared with both a purely data-driven ANN baseline and a conventional mechanistic baseline. Additional interpolation, extrapolation, limited-coverage, and noise-robustness tests further confirmed that the bounded residual formulation enhances generalization under sparse and imperfect training conditions. These results demonstrate that the proposed approach provides a robust and data-efficient strategy for milling-force waveform prediction when reliable extrapolation to unseen cutting conditions is required.

Keywords: End milling; Cutting force prediction; Mechanistic model; Physics-guided machine learning; Sparse condition holdout

1. Introduction

In milling processes, cutting force is a key process response arising from the combined effects of instantaneous chip thickness, cutting coefficients, tool geometry, cutter–workpiece engagement, the dynamic characteristics of the workpiece–tool–machine structural system, and force-signal acquisition conditions. Because cutting force is directly associated with tool and workpiece deformation, tool failure, chatter occurrence, surface quality, spindle and feed-drive loads, and cutting power, its prediction has long been regarded as a fundamental component of process planning, monitoring, optimization, and digital twin-based machining systems [1–4]. In particular, in practical production environments where it is difficult to experimentally examine all possible machining conditions, it is essential to develop a model that can reliably predict cutting forces for unseen combinations of spindle speed, feed rate, axial depth of cut, and radial depth of cut using only limited experimental data.

Traditional mechanistic milling models calculate the force waveform within a spindle revolution based on instantaneous uncut chip thickness, engagement angle, tangential, radial, and axial cutting coefficients, edge forces, helix geometry, and runout [5–13]. These models have the advantages of high physical interpretability and the ability to describe the major trends in cutting force with a relatively small amount of calibration data. However, experimentally measured force waveforms inevitably contain non-ideal effects, such as tool runout, edge radius, tool wear, fixture compliance, sensor dynamics, engagement errors, entry and exit transients, and variations in material and frictional states. If these factors are not explicitly modeled, calibration of cutting coefficients alone is often insufficient to eliminate axis-dependent bias, local waveform distortion, residual forces in non-cutting regions, or systematic deviations under high-load cutting conditions.

In contrast, purely data-driven models, including artificial neural networks (ANNs), can directly learn complex nonlinear mappings between machining conditions and force responses. These approaches have evolved in various forms, ranging from early ANN-based force models to dynamic neural networks and CNN–LSTM-based cutting-force prediction models [14–19]. Their main advantage is that they can learn patterns embedded in experimental data even when explicit cutting-geometry or coefficient models are incomplete. However, manufacturing data are expensive to obtain for each combination of machining conditions, and even when the apparent number of samples is large, the actual diversity of machining conditions is often limited. In particular, when angle-domain waveforms are divided using a sample-wise random split, angle samples belonging to the same machining condition may be included in both the training and test sets. This can lead to an overestimation of the generalization performance for truly unseen conditions. From a practical application perspective, a model should be able to predict the force waveform of a new machining condition under a condition-level holdout setting, in which the entire waveform corresponding to a given condition is excluded from training.

To alleviate these limitations, physics-informed, physics-guided, informed machine learning, and hybrid physics–machine learning approaches, which combine physical knowledge with data-driven learning, have been actively investigated [20–23]. Physics-informed neural networks were originally established as a framework for learning solutions by incorporating the residuals of governing partial differential equations into the loss function [20]. In engineering and manufacturing applications, however, physical knowledge is integrated in broader forms, including physics-based features, constraints, regularization terms, simulator outputs, mechanistic baselines, and residual-correction structures [21–25]. For milling-force prediction, rather than directly enforcing a governing partial differential equation in the loss function, it may be more practical to employ a mechanistic force map as a

reference function and allow the learning model to correct only the systematic residuals that cannot be explained by the mechanistic model.

This study aims to mitigate the problem that purely ANN-based models exhibit large variations in predictive performance depending on the selection of training conditions in sparse machining-condition environments, where only a limited number of machining conditions are available. To this end, an equivalent mechanistic milling-force model is first used to generate a baseline cutting-force waveform in the cutter-rotation-angle domain. A neural network is then designed to correct only the residual component that cannot be explained by this physics-based baseline within a bounded range. In other words, the proposed physics-guided mechanistic residual learning (PGMRL) framework does not directly learn the total cutting force in a black-box manner. Instead, it performs bounded residual correction with respect to a mechanistic force map.

In addition, this study proposes a reproducible evaluation protocol that integrates a residual bound, a mechanistic physics loss applied to the active-cutting region, scaling based only on the fitting split, sparse holdout at the machining-condition level, and phase-alignment sensitivity analysis. These components enable a systematic assessment of the model's sparse-condition generalization capability. In this regard, unlike conventional hybrid physics–machine learning or residual-learning approaches, the present study provides both architectural and evaluative contributions specifically tailored to the angle-domain waveform prediction problem in milling-force modeling.

2. Physics-Guided Mechanistic Residual Learning Formulation

2.1 Mechanistic Force Representation

The preprocessed cutting-force data were aligned with respect to one spindle revolution for each machining condition and were then re-expressed as force maps in the 0–360° spindle-angle domain with an angular resolution of 1°. In the cutting experiments, a two-flute carbide flat end mill (SEM810080, YG-1) was used. However, the flute-to-flute tip protrusion difference measured after tool mounting was larger than the nominal feed-per-tooth range for all machining conditions, namely 0.0143–0.0476 mm/tooth. Therefore, force generation was considered to be closer to single-active-edge behavior dominated by the protruding cutting edge, rather than to ideal two-flute engagement in which both cutting edges remove material uniformly. Accordingly, an equivalent active-edge mechanistic model with $Z_{\mathrm{eq}} = 1$was adopted in this study. Hereafter, this model is referred to simply as the active-edge model.

The configuration of the active-edge model was defined using a tool diameter of $D = 8$mm, an entry-angle offset of $\theta_{\mathrm{entry}} = 15^{\circ}$, a helix angle of $\lambda = 45^{\circ}$, the number of axial slices $N_z = 10$, and the equivalent number of cutting edges $Z_{\mathrm{eq}} = 1$. In other words, although the model assumes a single active edge, this edge is represented as a multi-slice helical cutting edge with a helix angle of 45°. The equivalent feed per active edge and the nominal feed per tooth based on the actual two-flute tool are distinguished as follows:

$$\begin{aligned} f_z^{\mathrm{eq}} &= \frac{v_f}{SZ_{\mathrm{eq}}} = \frac{v_f}{S}, Z_{\mathrm{eq}} = 1, \\ f_{z,\mathrm{tool}} &= \frac{v_f}{SZ_{\mathrm{tool}}}, Z_{\mathrm{tool}} = 2. \end{aligned} \tag{1}$$

Where, v_f is the feed rate in mm/min, and S is the spindle speed in rev/min. The variable f_z^{eq} in Eq. (1) is not the actual tooth-level feed. Rather, it is a model-side chip-thickness variable introduced to represent the active-edge assumption, in which the protruding cutting edge dominantly participates in material removal. Therefore, the identified coefficient vector K should not be interpreted as a flute-resolved intrinsic coefficient set of the actual two-flute cutter. Instead, it should be regarded as an effective angle-domain coefficient vector that depends on the measured single-active-edge dominance and the angle-domain alignment procedure.

At a spindle angle θ, the local immersion angle of the k-th axial slice was defined as

$$\begin{aligned} z_k &= \left(k+\frac{1}{2}\right)\frac{a_p}{N_z}, dz = \frac{a_p}{N_z}, k = 0,1,\dots,N_z-1, \\ \Delta\phi_k &= \frac{2z_k \tan\lambda}{D}\frac{180}{\pi}, \\ \phi_k &= \left[\theta-\theta_{\mathrm{entry}}-\Delta\phi_k\right] \bmod 360. \end{aligned} \tag{2}$$

Here, $\Delta\phi_k$ represents the axial phase lag induced by the helix angle. Because $Z_{\mathrm{eq}} = 1$ was used in this study, no tooth-to-tooth angular shift was included. However, the axial phase lag associated with $\lambda = 45^\circ$ and $N_z = 10$ was reflected in the slice-wise immersion angle ϕ_k.

The engagement function and instantaneous uncut chip thickness according to the radial immersion ratio a_e/D were calculated as follows:

$$\begin{aligned} \phi_{\mathrm{eng}}\left(\frac{a_e}{D}\right) &= \cos^{-1}\left(1-\frac{2a_e}{D}\right), \\ g\left(\phi_k;\frac{a_e}{D}\right) &= \begin{cases} 1, & 0 \le \phi_k \le \phi_{\mathrm{eng}}, \\ 0, & \text{otherwise}, \end{cases} \\ h(\phi_k) &= \max\left(f_z^{\mathrm{eq}}\sin\phi_k, 0\right). \end{aligned} \tag{3}$$

The active-cut mask was determined by the geometric engagement function $g(\phi_k; a_e/D)$ in Eq. (3). For each axial slice, the tangential, radial, and axial differential force components were calculated as

$$\begin{aligned} dF_{t,k} &= (K_{tc}h(\phi_k) + K_{te})dz\, g\left(\phi_k;\frac{a_e}{D}\right), \\ dF_{r,k} &= (K_{rc}h(\phi_k) + K_{re})dz\, g\left(\phi_k;\frac{a_e}{D}\right), \\ dF_{a,k} &= (K_{ac}h(\phi_k) + K_{ae})dz\, g\left(\phi_k;\frac{a_e}{D}\right). \end{aligned} \tag{4}$$

The transformation to the machine-axis force components was applied to each axial slice, and the contributions from all slices were then summed:

$$
\begin{aligned}
F_x &= \sum_{k=0}^{N_z-1} -\left(dF_{t,k}\cos\lambda\cos\phi_k + dF_{r,k}\sin\phi_k\right), \\
F_y &= \sum_{k=0}^{N_z-1} \left(dF_{t,k}\cos\lambda\sin\phi_k - dF_{r,k}\cos\phi_k\right), \\
F_z &= \sum_{k=0}^{N_z-1} \left(dF_{a,k} + dF_{t,k}\sin\lambda\right).
\end{aligned} \tag{5}
$$

Thus, under the present setting of $\lambda = 45°$, the axial force component includes not only $dF_{a,k}$, but also the helix-induced projection of the tangential component, $dF_{t,k}\sin\lambda$. By incorporating the engagement function, chip-thickness calculation, axial slicing, helical phase lag, and coordinate transformation described above, the mechanistic force can be expressed in the form of a design matrix that is linear with respect to the coefficient vector K:

$$
\begin{aligned}
F_{\text{mech}}(\theta, S, v_f, a_p, a_e) &= A(\theta, S, v_f, a_p, a_e; \lambda, N_z, Z_{\text{eq}})K, \\
K &= [K_{tc}, K_{te}, K_{rc}, K_{re}, K_{ac}, K_{ae}]^T.
\end{aligned} \tag{6}
$$

Where, A is the design matrix of the active-edge model configured with $Z_{\text{eq}} = 1$, $\lambda = 45°$, $N_z = 10$, and $\theta_{\text{entry}} = 15°$. Because this formulation is linear with respect to K, the mechanistic baseline can be stably estimated through positivity-constrained coefficient fitting.

2.2 Bounded Residual Model, Scaling, and Training Objective

The proposed PGMRL model does not predict cutting forces directly using a fully black-box neural network. Instead, it predicts the force by adding a bounded residual correction to the mechanistic prediction of the active-edge model. The final raw-force prediction and input feature vector are defined as follows:

$$
\begin{aligned}
\hat{F} &= F_{\text{mech}}(K) + \Delta Y_{\text{fit}} \odot c_r \tanh(N_w(x)), \\
x &= \left[\cos\theta, \sin\theta, S, v_f, a_p, a_e\right].
\end{aligned} \tag{7}
$$

Here, ΔY_{fit} denotes the axis-wise target min–max denominator, that is, the target range for each force axis, computed only from the fitting split. The parameter $c_r = 0.20$ is the residual bound in the normalized target space, and $\odot$ denotes axis-wise element-wise multiplication. Therefore, $c_r = 0.20$ should not be interpreted as an absolute force bound of ± 0.20N. Instead, it corresponds to a bounded correction equal to 20% of the target range of each force axis. The residual network N_w is a tanh multilayer perceptron, or MLP, with three hidden layers and a hidden dimension of 512. Its final output is a three-dimensional vector corresponding to the residuals of the three force components.

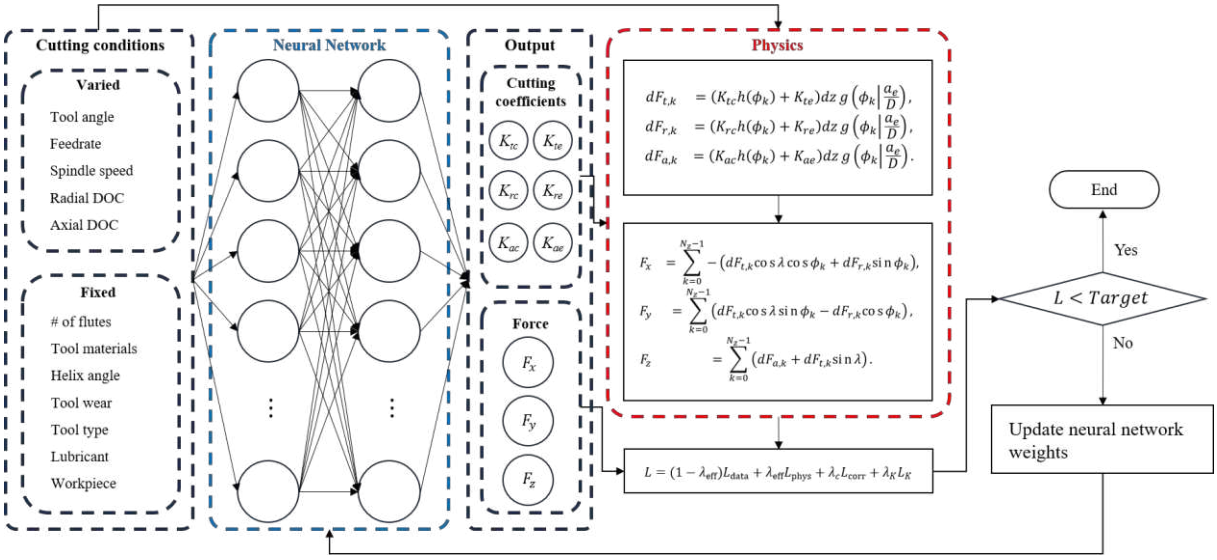


Figure 1. Conceptual schematic of the proposed PGMRL framework.

The active-edge model provides the physics-based baseline force prediction, whereas the neural network learns a bounded residual correction from the spindle angle and machining-condition inputs. The final prediction is obtained by adding the bounded residual term to the mechanistic force component. Therefore, the neural network does not directly replace the mechanistic force map. Rather, it plays the role of correcting the systematic deviations of the mechanistic model within a constrained range.

The final training objective consists of a normalized force-domain data loss, an active-cut mechanistic physics loss, a residual-correction regularization term, and a coefficient regularization term:

$$\begin{aligned} L &= (1-\lambda_{\text{eff}})L_{\text{data}} + \lambda_{\text{eff}}L_{\text{phys}} + \lambda_c L_{\text{corr}} + \lambda_K L_K, \\ L_{\text{data}} &= \frac{1}{N}\sum_i \| \hat{F}_{\text{norm},i} - F_{\text{norm},i} \|_2^2, \\ L_{\text{phys}} &= \frac{1}{\max(\sum_i m_i,\, 1)} \sum_i m_i \sum_j \left[\frac{F_{\text{mech},ij} - F_{ij}}{q_j}\right]^2, \\ L_{\text{corr}} &= \text{mean}[\tanh(N_w(x))^2]. \end{aligned} \tag{8}$$

Here, m_i is the active-cut mask, and q_j is the axis-wise 95th-percentile absolute force scale computed from the fitting samples. In the multi-slice mechanistic representation, a sample was defined as an active-cut sample when at least one axial slice was in engagement. The data loss L_{data} was calculated in the normalized force domain, whereas the physics loss L_{phys} was calculated in the raw-force domain so that the mechanistic component itself could explain the measured force in the active-cut region. The effective physics-loss weight λ_{eff} was linearly increased from 0 to $\lambda_p = 0.10$ during the first 100 epochs and was then kept constant at 0.10. The residual-correction regularization weight was set to $\lambda_c = 0.001$. Although $\lambda_K = 0.001$ was specified in the configuration, L_K was deactivated in the final implementation because the global coefficient mode was used. Consequently, its actual contribution to the total loss was zero.

2.3 Comparative Models

In this study, three basic models were compared: the mechanistic baseline, the ANN baseline, and the

proposed PGMRL model. The mechanistic baseline uses only the active-edge model in Eq. (6), and the global coefficient vector K is estimated by performing positivity-constrained coefficient fitting using the active fitting samples. The ANN baseline is a fully data-driven model that directly predicts the normalized force without using any mechanistic term. It uses the same legacy_raw input features and the same tanh MLP architecture with three hidden layers and a hidden dimension of 512 as the PGMRL model. Therefore, the difference between the ANN baseline and PGMRL does not lie in the neural-network capacity, but in whether the mechanistic prior and bounded residual formulation are incorporated.

The ANN baseline has 530,435 trainable parameters, whereas PGMRL has 530,441 trainable parameters, with six additional global mechanistic coefficients. For the mechanistic baseline, ANN baseline, and PGMRL model, all normalization and scaling parameters were computed only from the fitting split. Information from the validation condition and the final-test condition was not used in the calculation of the scalers.

3. Experimental Setup, Data Acquisition, Preprocessing, and Evaluation

3.1 Experimental Setup and Force Measurement

The cutting-force data used in this study were acquired from top-surface end-milling passes performed on a vertical machining center (MYNX 400). The cutting-force measurement chain consisted of a multicomponent piezoelectric dynamometer (9119A, Kistler) and a charge amplifier/DAQ (LabAmp 5167A, Kistler). The dynamometer is capable of simultaneously measuring the three orthogonal force components, F_x, F_y, and F_z, and is suitable for milling-force measurements using small workpiece specimens, as in the present study. The LabAmp provides a laboratory charge amplifier for piezoelectric sensors with integrated 24-bit data acquisition. In this study, the nominal sampling rate was set to 50 kHz per channel.

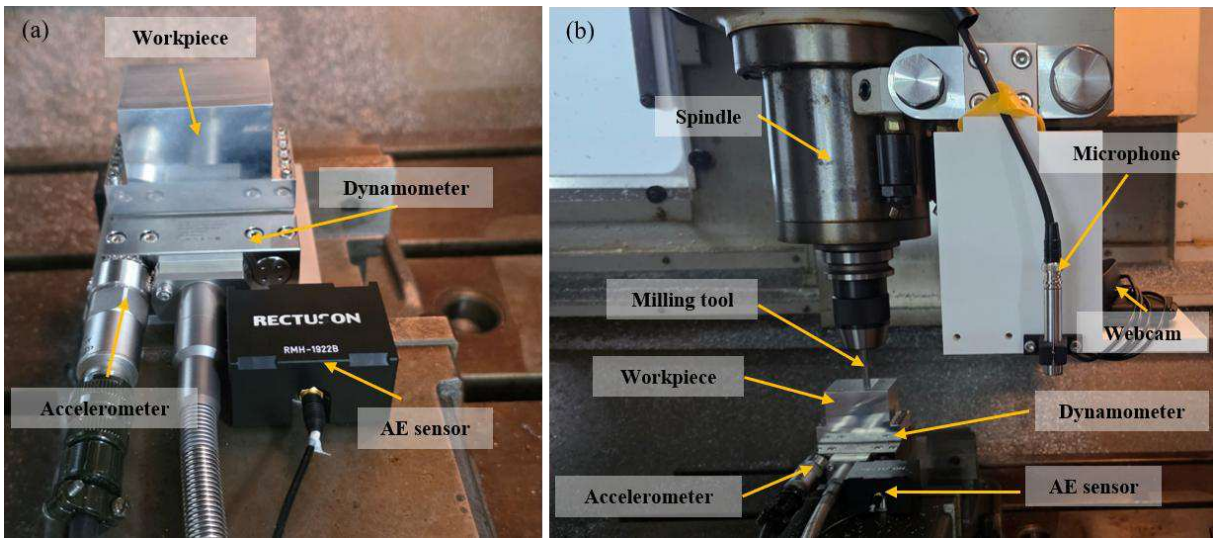


Figure 2. Experimental setup and measurement configuration for cutting-force acquisition.

The workpiece material was S45C steel, and the specimens were prepared as rectangular blocks with dimensions of 60 mm in length, 40 mm in width, and 50 mm in height. A mounting protrusion was machined at the bottom of each specimen so that it could be directly bolted to the dynamometer cover plate. Before the cutting experiments, the upper surface and reference side surface of the specimen were finish-machined to ensure sufficient flatness and perpendicularity. The specimen was directly fixed to the upper cover plate of the dynamometer, and the cutting location was positioned close to the center of the cover plate to reduce the influence

of moment loading.

The cutting tool was a two-flute carbide flat end mill(SEM810080, YG-1). The tool diameter was $D = 8$mm, the shank diameter was 8 mm, the flute length was 20 mm, and the overall length was 70 mm. The tool overhang was kept constant throughout all experiments. After tool mounting, the tool-tip runout and flute-to-flute tip protrusion difference were measured using a dial gauge. Although the absolute runout was within 2% of the tool diameter, the measured tip protrusion difference between the two cutting edges was larger than the nominal feed per tooth calculated for all machining conditions considered in this study. In this dataset, the range of $f_{z,\mathrm{tool}}$ was approximately 0.0143–0.0476 mm/tooth, whereas the measured protrusion about 85μm difference exceeded this range. Therefore, this study assumed that the protruding edge dominated force generation, and the mechanistic term was formulated using an active-edge model with $Z_{\mathrm{eq}} = 1$.

The PGMRL input features do not include the tool-wear state. Therefore, to maintain the wear state as constant as possible during data acquisition, new or minimally worn tools were used, and the tool-wear state was inspected after a specified number of passes.

The machining conditions were constructed as a full-factorial Cartesian combination of spindle speed S, feed rate v_f, axial depth of cut a_p, and radial depth of cut a_e:

$$\begin{aligned} C &= \{2100,2800\} \times \{80,110,140,170,200\} \times \{0.5,1.0\} \times \{1.0,1.5,\ldots,7.5\}, \\ |C| &= 2 \times 5 \times 2 \times 14 = 280. \end{aligned} \tag{9}$$

Here, S is given in rpm, v_f in mm/min, and both a_p and a_e in mm. All cutting tests were conducted under dry-cutting conditions, and the dataset used for analysis was organized based on the up-milling direction. Each machining condition was represented by 360 wrapped-angle samples at 1° intervals. Accordingly, the final learning dataset consisted of $280 \times 360 = 100{,}800$ angle-domain samples.

For each run, spindle warm-up was performed before cutting, after which the dynamometer zero offset was reset and a no-load baseline signal was recorded. Each test-cutting file was saved so as to include the air-cut interval before tool entry into the workpiece, the steady-state cutting interval, and the trailing air-cut interval. In addition, sufficient lead-in and overtravel regions were secured to prevent entry and exit transients from contaminating the central steady-state waveform. Only the central steady-state waveform, in which full cutting engagement was maintained, was used for model training.

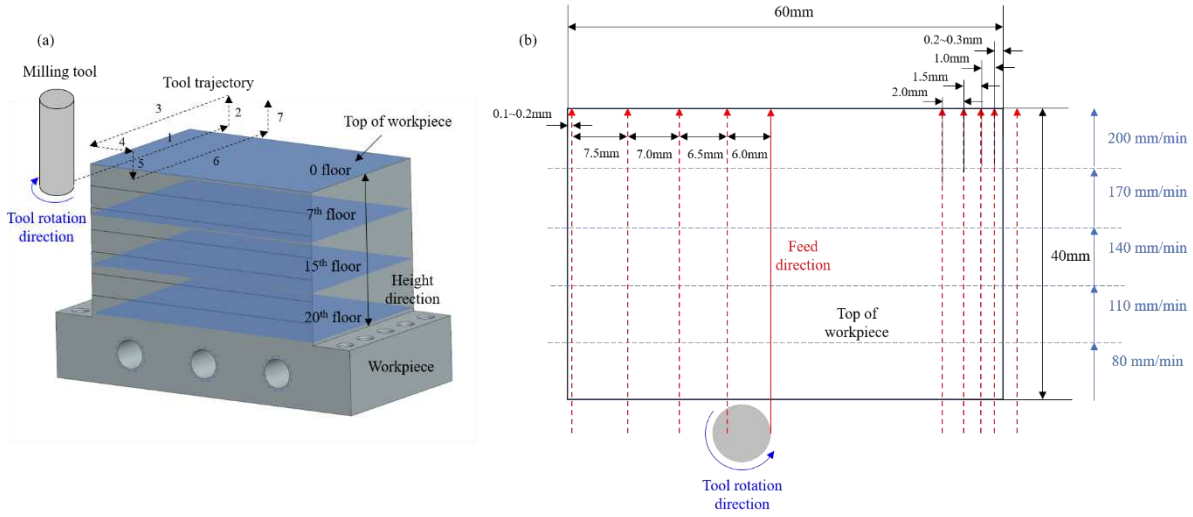


Figure 3. Workpiece geometry and top-surface end-milling pass layout. (a) Schematic of the workpiece geometry and the stacked top-surface milling passes arranged along the height direction. (b) Plan view of the straight milling passes performed on the S45C specimen, showing the feed direction, tool-rotation direction, and radial-engagement offsets.

Table1. Experiment conditions

Parameter	Value
Workpiece material	S45C
Tool diameter	8 mm
Radial depth of cut	1.0, 1.5, 2.0, ..., 7.5 mm (14 levels)
Axial depth of cut	0.5, 1 mm
Feed rate	80, 110, 140, 170, 200 mm/min
Spindle speed	2100, 2800 RPM
Cutting direction	Up milling
Lubricant condition	Dry (Not applied)

3.2 Preprocessing

The angle-domain force maps used for training and evaluation were generated from the time-domain cutting-force signals measured by the dynamometer. When necessary, linear drift correction was applied based on the initial and final air-cut intervals. For automatic detection of the cutting interval, a 100 Hz block-RMS envelope of the resultant force was used. In the preprocessing run reported in this study, 20-sample centered smoothing, a 10 N threshold, a 1.5 s gap-merge rule, and a 0.5 s minimum-duration rule were applied. The detected steady-state cutting segment was then used for subsequent angle-domain preprocessing.

Because an absolute spindle encoder was not used in the present experiments, a provisional accumulated spindle angle was first calculated using the spindle command speed and the time axis. Subsequently, the initial phase delay was estimated by performing FFT-based cross-correlation between the measured resultant-force waveform and an Altintas-style reference waveform. The estimated phase was further adjusted so that the dominant force peak was located in the first-flute sector.

Finally, the accumulated angle was linearly interpolated onto a 1° grid, and among complete rotations, stable cycles whose resultant waveforms were closest to the mean template were selected. The final learning files were organized into 360 wrapped-angle samples for each machining condition.

3.3 Evaluation

The main evaluation was performed using a sparse train-condition holdout protocol. Among the total $N_c = 280$ machining conditions, fitting conditions were randomly selected according to sparse fractions of $p = \{5,10,15,20,25,30\}\%$. Then, additional conditions corresponding to approximately 20% of the number of fitting conditions were separated as the internal validation set. The remaining conditions were completely excluded and used as the final test set.

The selection unit was not an individual angle sample, but a complete machining condition. Therefore, the 360 angle-domain samples belonging to a given machining condition were included in only one of the fitting, internal validation, or final test subsets. Each sparse fraction was repeated ten times using different random seeds. This protocol was designed to ensure that the final test conditions were not used for model fitting, early stopping, or scaler fitting.

The primary performance metric for all results was the active-cut three-axis RMSE. The active-cut set Ω_{active} was defined based on the mechanistic engagement mask described in Section 2, where angle-domain samples were considered active when at least one axial slice was in engagement:

$$\Omega_{\text{active}} = \left\{ i \mid \max_k g\left(\phi_{i,k}; \frac{a_{e,i}}{D}\right) = 1 \right\},$$

$$\text{RMSE}_{\text{active,3axis}} = \sqrt{\frac{1}{3 \mid \Omega_{\text{active}} \mid} \sum_{i \in \Omega_{\text{active}}} \sum_{j \in \{x,y,z\}} \left(\hat{F}_{ij} - F_{ij}\right)^2}. \quad (10)$$

In Eq. (10), F_{ij} denotes the measured force, and $\hat{F}_{ij}$ denotes the predicted force. This metric is computed by concatenating the F_x, F_y, and F_z error components for all active-cut samples and then calculating a single scalar RMSE.

Auxiliary evaluations were also conducted to complement the main sparse train-condition holdout protocol. The auxiliary tests consisted of interpolation-gap tests, boundary-extrapolation tests, held-out condition ratio sweeps, limited training-condition coverage tests, and force-noise robustness tests. The interpolation-gap and boundary-extrapolation tests were conducted along the feed-rate axis v_f and the radial-depth axis a_e, based on the run reported in this study. In the held-out condition ratio sweep, the random condition holdout ratio was varied from 10% to 90%. In the limited training-condition coverage test, only 100%, 75%, 50%, 25%, and 10% of the available fitting conditions were used to analyze performance variations in sparse-data regimes. In the force-noise robustness test, Gaussian noise was added only to the fitting force samples, whereas the final test force samples were kept clean. In the reported training run, the model input was defined as $[\cos\theta, \sin\theta, S, v_f, a_p, a_e]$.

The residual bound was set to $c_r = 0.20$. The mechanistic model used the active-edge model defined in Section 2, with the main configuration values of $D = 8.0\text{mm}$, $Z_{\text{eq}} = 1$, $\lambda = 45°$, $\theta_{\text{entry}} = 15°$, and $N_z = 10$.

The Adam optimizer was used for the neural model, with a learning rate of 1.0×10^{-4}, a maximum of 1000 epochs, an early-stopping patience of 100, a hidden dimension of 512, and a random seed of 42.

4. Results and Discussion

4.1 Main Sparse Train-Condition Holdout Result

Figure 4 and Table 2 summarize the active-cut three-axis RMSE results obtained using the sparse train-condition holdout protocol. For each fitting-condition fraction, ten condition-level random splits were repeated. The complete main sparse-train results were analyzed based on six fitting fractions and ten splits per fraction, resulting in a total of 60 split-level evaluations. Here, the active-cut RMSE was used as the primary metric to evaluate force-prediction accuracy in the actual cutting-engagement region.

Across all 60 sparse splits, the active-cut three-axis RMSE of PGMRL was 3.964 ± 0.248N. This value was 55.4% lower than that of the ANN model, 8.895 ± 2.567N, and 14.4% lower than that of the mechanistic baseline, 4.629 ± 0.055N. In addition, the RMSE of the mechanistic component alone in PGMRL was 4.679 ± 0.067N, which was higher than that of the final PGMRL output. This indicates that the bounded residual correction did not merely repeat mechanistic coefficient fitting, but additionally corrected the systematic residuals remaining in the mechanistic component.

In the sparsest 5% fitting-condition setting, only 14 conditions among the total 280 machining conditions were used for fitting. Under this setting, the RMSE of the ANN model increased to 13.713 ± 2.961N, whereas PGMRL maintained an RMSE of 4.384 ± 0.097N. This result demonstrates that PGMRL substantially alleviated the split-to-split sensitivity observed in the ANN baseline, even under severely limited condition coverage.

Because the three models were evaluated using the same condition splits, the differences between models were tested using paired split-level comparisons. Across all 60 sparse splits, PGMRL reduced the active-cut RMSE by an average of 4.931 N compared with the ANN model, with a 95% confidence interval of 4.312–5.550 N. The paired t-test yielded $p = 4.87 \times 10^{-23}$, and the Wilcoxon signed-rank test yielded $p = 1.63 \times 10^{-11}$. The improvement of PGMRL over the mechanistic baseline was smaller than that over the ANN model, but it was highly consistent. The mean RMSE reduction was 0.665 N, with a 95% confidence interval of 0.603–0.727 N. The paired t-test yielded $p = 1.33 \times 10^{-29}$, and the Wilcoxon signed-rank test yielded $p = 1.63 \times 10^{-11}$. The unit of analysis in these paired tests was not individual angle-domain samples, but split-level RMSE values. Therefore, the statistical tests were conducted based on performance differences across condition splits, avoiding pseudo-replication that could otherwise inflate statistical significance owing to the large number of force samples.

The same trend was also observed in additional diagnostic metrics. Across all 60 sparse splits, PGMRL reduced the full-cycle RMSE from 7.109 ± 1.788N for the ANN model to 3.431 ± 0.103N. The non-cut RMSE was also reduced from 6.313 ± 1.548N to 3.218 ± 0.061N. The 95th-percentile predicted force magnitude in the non-cut region was substantially lower for PGMRL, 3.826 ± 0.629N, than for the ANN model, 19.534 ± 6.591N. These non-cut metrics were used to examine whether the model predicted unnecessary forces in regions where actual cutting did not occur. When the condition-level failure rate was calculated using an active-cut RMSE threshold of 10 N, the ANN model showed a failure rate of 22.668%, whereas the mechanistic baseline and PGMRL showed failure rates of 0.000% and 0.008%, respectively. Therefore, the residual correction in PGMRL improved active-cut accuracy without inducing the large non-cut false-force behavior and condition-level failures

observed in the ANN baseline.

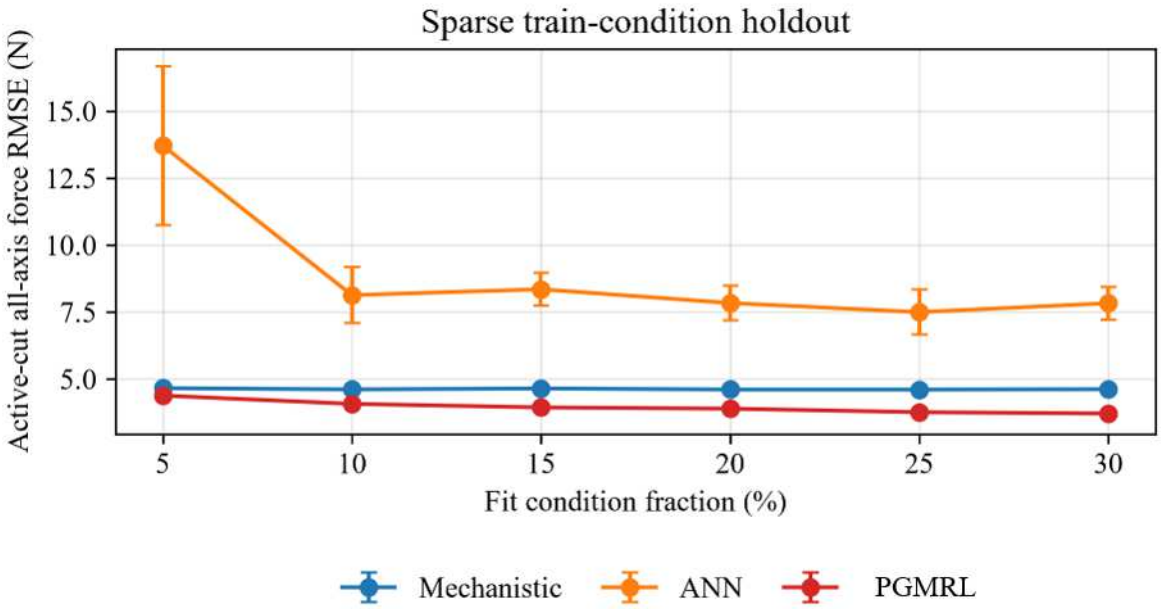


Figure 4. Sparse train-condition holdout performance according to the fitting-condition fraction.

Table 2. Sparse train-condition holdout active-cut three-axis RMSE. Values are reported as mean ± standard deviation over ten repeated condition-level random splits.

Fit fraction (%)	Mechanistic	ANN	PGMRL	PGMRL vs. ANN reduction (%)	PGMRL vs. Mechanistic reduction (%)
5	4.661 ± 0.063	13.713 ± 2.961	4.384 ± 0.097	68.0	5.9
10	4.616 ± 0.041	8.134 ± 1.046	4.076 ± 0.135	49.9	11.7
15	4.650 ± 0.050	8.352 ± 0.613	3.943 ± 0.121	52.8	15.2
20	4.612 ± 0.055	7.838 ± 0.647	3.897 ± 0.095	50.3	15.5
25	4.611 ± 0.061	7.499 ± 0.847	3.766 ± 0.113	49.8	18.3
30	4.623 ± 0.052	7.831 ± 0.617	3.717 ± 0.113	52.5	19.6

4.2 Scatter and Waveform Behavior

Figure 5 shows active-cut measured–predicted scatter plots obtained from a representative 5% sparse train-condition holdout split. This split corresponds to the sparsest setting in the main evaluation, in which only 14 conditions among the total 280 machining conditions were used for fitting. The mechanistic baseline showed

generally stable predictions, but systematic deviations remained in F_z and in some high-load regions. The ANN model exhibited saturation and slope distortion in F_x and F_z, which is consistent with the large RMSE and high split-to-split variability observed in Table 2. In contrast, PGMRL preserved the mechanistic trend while reducing systematic deviations through bounded residual correction, resulting in distributions closer to the 1:1 line for all three force components.

Figure 6 compares the force waveforms of eight representative unseen conditions selected from the same 5% sparse split. The ANN model showed large local overshoots and waveform distortions in several unseen conditions, particularly after engagement and in the force-recovery region. By contrast, the mechanistic baseline and PGMRL followed the main waveform envelopes more stably. In particular, PGMRL did not replace the mechanistic curve with an unconstrained black-box prediction. Instead, it corrected local errors near entry and exit regions and axis-dependent biases through the bounded residual term. This indicates that the performance improvement of PGMRL was not limited to a reduction in average RMSE, but also led to more stable prediction behavior at the waveform level.

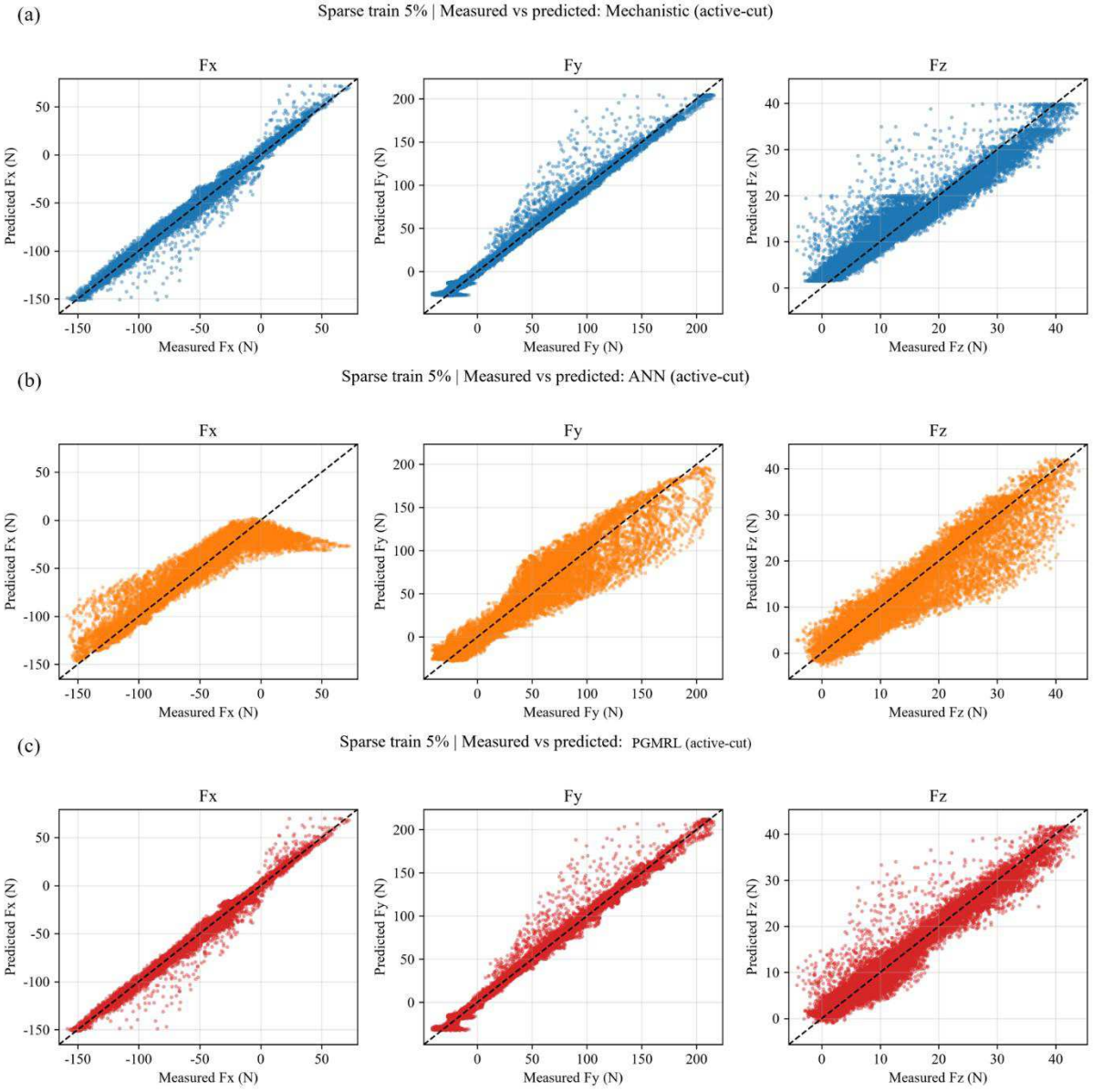
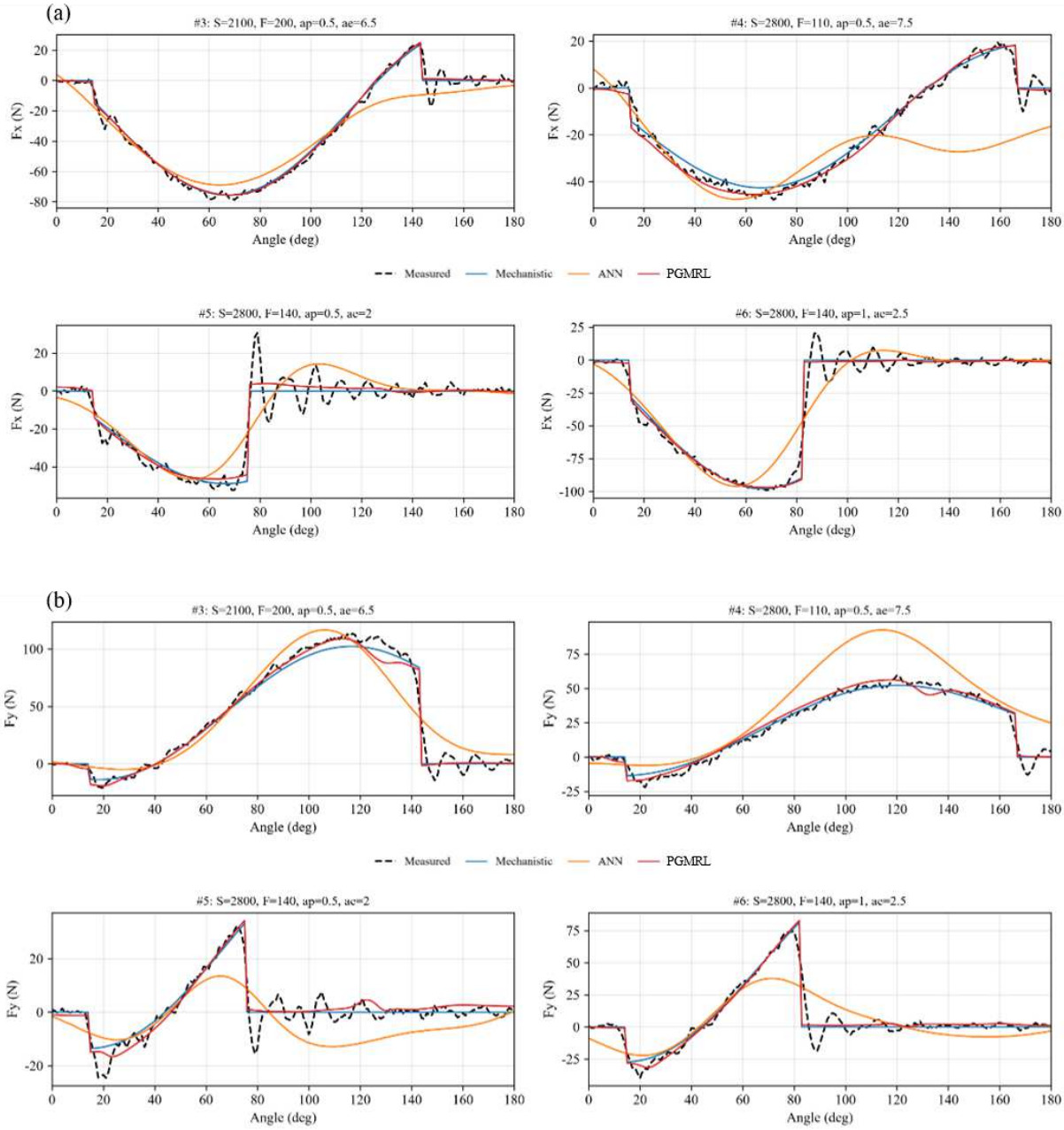


Figure 5. Active-cut measured–predicted force scatter plots obtained from a representative split of the 5% sparse train-condition holdout case. (a) Mechanistic baseline, (b) ANN baseline, and (c) PGMRL.



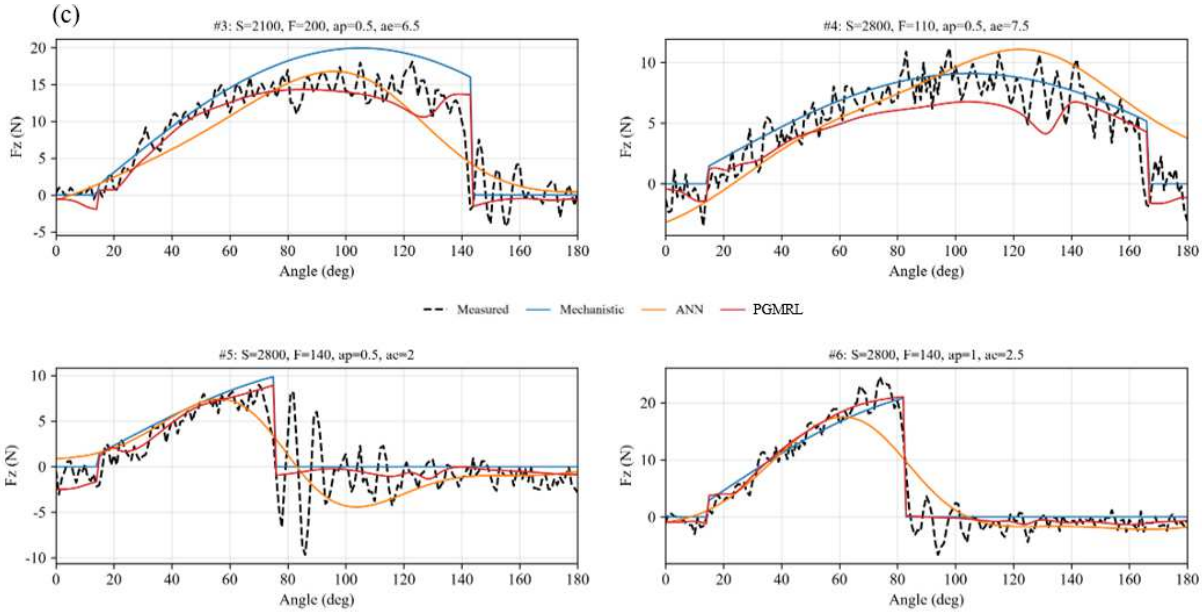


Figure 6. Comparison of force waveforms for eight representative unseen conditions selected from the 5% sparse train-condition holdout case: (a) F_x, (b) F_y, and (c) F_z.

4.3 Auxiliary Generalization and Robustness Tests

Table 3 summarizes the active-cut three-axis RMSE obtained from the auxiliary evaluation protocols. The auxiliary tests were conducted to examine whether the advantages of PGMRL observed in the main sparse train-condition holdout evaluation were maintained under other generalization scenarios. Specifically, interpolation-gap, boundary-extrapolation, increasing holdout-ratio, limited training-condition coverage, and training-force noise conditions were evaluated.

In the interpolation-gap tests, PGMRL achieved an RMSE of 3.520 ± 0.534N, which was lower than those of both the ANN model and the mechanistic baseline. In the boundary-extrapolation tests, PGMRL showed an RMSE of 3.898 ± 0.768N, whereas the ANN model produced a larger error of 8.775 ± 1.928N. At the protocol level, the mean RMSE reductions of PGMRL relative to the ANN model were 51.4% for the interpolation-gap tests and 55.6% for the boundary-extrapolation tests. Compared with the mechanistic baseline, PGMRL also reduced the RMSE by 23.1% and 17.5%, respectively. These results indicate that the ANN model was more sensitive to the absence of boundary condition levels, whereas PGMRL maintained the stabilizing effect of the mechanistic basis.

In the held-out ratio sweep, limited condition coverage, and train-noise robustness tests, the ANN model was also more sensitive to reduced condition coverage or increased training noise. PGMRL maintained a lower protocol-level mean RMSE than the mechanistic baseline in all auxiliary protocol summaries listed in Table 3. Specifically, PGMRL showed RMSE reductions of 51.8%, 57.1%, and 50.4% compared with the ANN model in the held-out ratio sweep, limited condition coverage, and noise robustness tests, respectively. Compared with the mechanistic baseline, the corresponding reductions were 19.2%, 18.9%, and 21.6%.

Figure 8 shows the robustness results obtained when Gaussian noise was added to the fitting force samples. In this test, noise was injected only into the fitting data, while the final-test force samples were kept clean.

Therefore, this result evaluates how stably a model trained under noisy fitting conditions can generalize to clean unseen conditions.

Because the number and characteristics of evaluation cases differ among the auxiliary evaluation protocols, n in Table 3 denotes protocol-level evaluation instances rather than repeated random splits of the same form. Therefore, Table 3 should be interpreted as a descriptive summary of generalization behavior under different stress-test conditions, rather than as a single pooled inferential statistic. This distinction is particularly important to avoid directly treating the boundary-extrapolation test with $n = 4$ and the holdout-ratio sweep with $n = 90$ as comparable statistical-test targets.

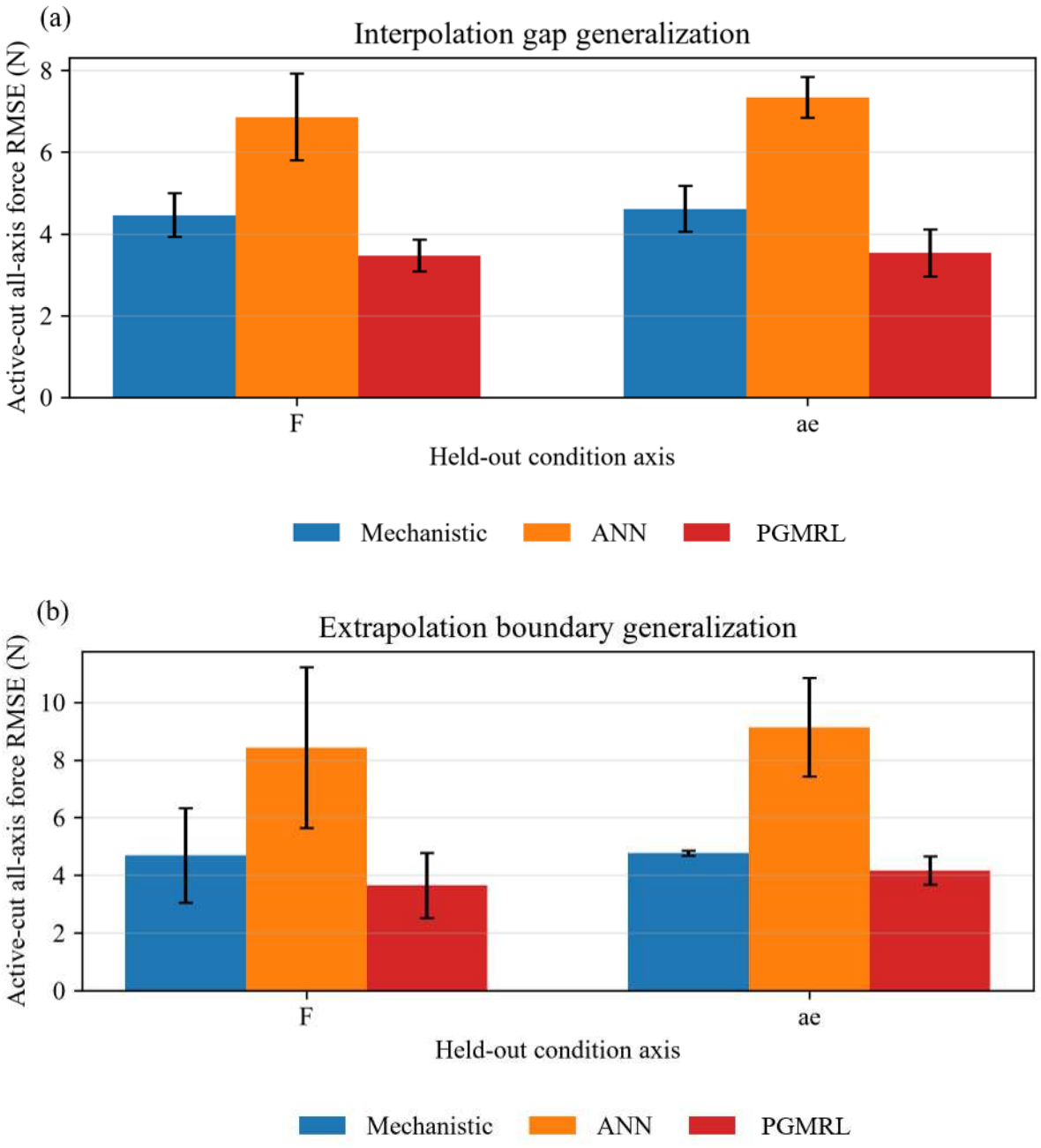


Figure 7. Auxiliary gap-generalization evaluation results. Active-cut three-axis RMSE obtained from (a) the

interpolation-gap test and (b) the boundary-extrapolation test.

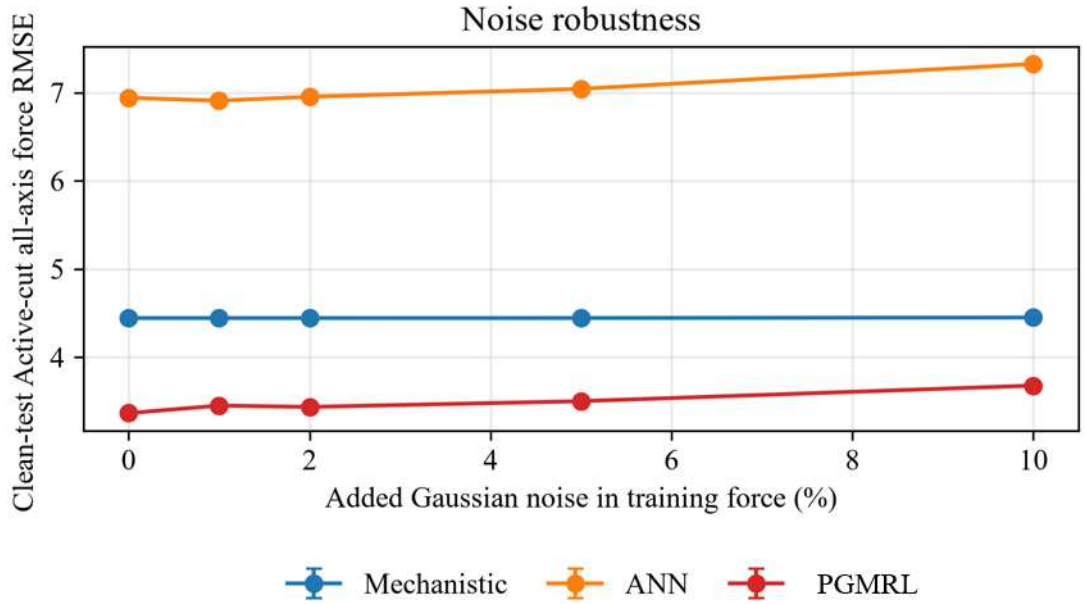


Figure 8. Train-force noise robustness evaluation results.

Table 3. Auxiliary active-cut three-axis RMSE summary. Values are reported as mean ± standard deviation over protocol-level evaluation instances.

Protocol	Mechanistic	ANN	PGMRL mechanistic part	PGMRL	n
Interpolation gap	4.577 ± 0.542	7.237 ± 0.629	4.617 ± 0.539	3.520 ± 0.534	15
Boundary extrapolation	4.724 ± 0.949	8.775 ± 1.928	4.765 ± 0.891	3.898 ± 0.768	4
Holdout-ratio sweep	4.621 ± 0.141	7.747 ± 1.048	4.663 ± 0.144	3.733 ± 0.262	90
Limited condition coverage	4.446 ± 0.010	8.396 ± 2.517	4.494 ± 0.013	3.604 ± 0.232	5
Noise robustness	4.444 ± 0.003	7.033 ± 0.170	4.488 ± 0.003	3.486 ± 0.118	5

4.4 Effective Cutting Coefficients and Residual Interpretation

Both the mechanistic baseline and PGMRL retain a cutting-coefficient-based mechanistic force term. Table 4 and Figure 9 summarize the effective coefficients identified from a total of 179 split-level evaluations, including the main sparse holdout splits and auxiliary evaluation cases. Here, K_{tc}, K_{rc}, and K_{ac} denote the chip-thickness-dependent coefficients in the tangential, radial, and axial directions, respectively, whereas K_{te}, K_{re}, and K_{ae} denote the edge-related coefficients.

The values of K_{tc} and K_{rc} remained highly similar between the mechanistic baseline and PGMRL. This suggests that the dominant tangential and radial chip-thickness-dependent terms were largely preserved in PGMRL. In contrast, larger differences were observed in K_{ac} and K_{ae}. This may indicate that the axial-direction force component was more strongly affected by the interaction among the active-edge model structure, helix-based axial slicing, and the residual network.

Across all 60 main sparse-train splits, the final PGMRL output reduced the active-cut RMSE by an average of 0.715 N compared with the PGMRL mechanistic component alone. In the residual diagnostics, the residual-induced RMSE improvement was 0.694 ± 0.267N, the RMS of the residual correction was 1.727 ± 0.213N, and the residual-to-mechanistic force ratio was 0.063 ± 0.008. Therefore, the residual term remained limited in magnitude relative to the mechanistic force. This supports the interpretation that PGMRL does not replace the mechanistic model with an unconstrained ANN-like predictor, but instead corrects systematic mechanistic deviations through a bounded residual term.

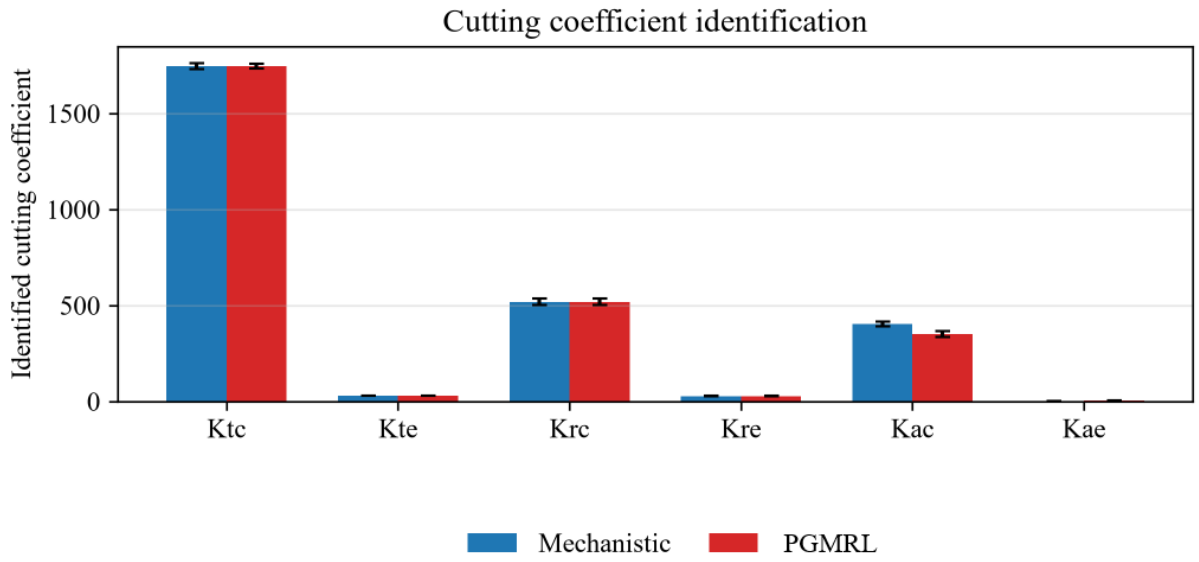


Figure 9. Comparison of the mean identified effective coefficients between the mechanistic baseline and PGMRL.

Table 4. Effective cutting coefficient summary across 179 split-level evaluations. CV denotes the coefficient of variation.

Coefficient	Mechanistic	Mechanistic CV (%)	PGMRL	PGMRL CV (%)
K_{tc}	1746.208 ± 14.766	0.85	1746.144 ± 13.671	0.78
K_{te}	29.668 ± 0.659	2.22	29.612 ± 0.642	2.17
K_{rc}	518.864 ± 16.517	3.18	518.162 ± 17.019	3.28
K_{re}	27.965 ± 0.668	2.39	27.953 ± 0.672	2.40
K_{ac}	403.836 ± 12.850	3.18	351.580 ± 14.799	4.21
K_{ae}	2.571 ± 0.305	11.86	5.158 ± 0.539	10.46

5. Conclusions

This study proposed a physics-guided mechanistic residual learning framework for three-axis milling-force waveform prediction in the spindle-angle domain. The proposed PGMRL model uses an equivalent active-edge mechanistic force map as a physics-based baseline and trains a neural network only to learn a bounded residual correction. Therefore, the neural component does not directly replace the mechanistic model with an unconstrained black-box predictor, but compensates for systematic deviations that remain after mechanistic coefficient fitting. This formulation preserves the physical structure of the milling-force model while improving

its ability to represent measured force waveforms under sparse machining-condition coverage.

The proposed framework was evaluated using dry up-milling force data of S45C steel under 280 machining conditions. Because the measured flute-to-flute tip protrusion difference exceeded the nominal feed-per-tooth range, the mechanistic component was formulated using an equivalent single-active-edge model with helical axial slicing. Model generalization was assessed using a sparse condition-level holdout protocol, in which complete machining conditions, rather than individual angle samples, were excluded from training. Across 60 sparse holdout splits, PGMRL achieved an active-cut three-axis RMSE of 3.964 ± 0.248 N, corresponding to error reductions of 55.4% and 14.4% compared with the ANN baseline and mechanistic baseline, respectively. In the most data-limited 5% fitting-condition case, PGMRL maintained stable accuracy, whereas the ANN baseline showed substantial error growth and split-to-split variability.

Waveform, scatter, and diagnostic analyses confirmed that PGMRL improved prediction accuracy without sacrificing the stability of the mechanistic model. The ANN baseline exhibited saturation, slope distortion, local overshoot, and false-force behavior in non-cutting regions under sparse training conditions. In contrast, PGMRL followed the main mechanistic force envelope while correcting local entry/exit-region errors and axis-dependent bias through the bounded residual term. Additional interpolation-gap, boundary-extrapolation, holdout-ratio, limited-coverage, and force-noise robustness tests further demonstrated that the proposed framework maintained better generalization than both the purely data-driven ANN and the conventional mechanistic baseline under various sparse and imperfect training scenarios.

The coefficient and residual analyses also support the intended role of the proposed residual-learning structure. The dominant tangential and radial chip-thickness-dependent coefficients remained highly consistent between the mechanistic baseline and PGMRL, indicating that the primary mechanistic force structure was preserved. Meanwhile, the residual-to-mechanistic force ratio remained small, approximately 6.3% on average, showing that the neural network acted as a limited correction module rather than as a substitute for the mechanistic model. Overall, the proposed PGMRL framework provides a data-efficient and physically constrained approach for milling-force prediction under unseen machining conditions. Future work should extend the framework to multi-tool, multi-material, and tool-wear-varying cases and investigate adaptive residual bounds or condition-dependent mechanistic coefficients for broader industrial applicability.

Acknowledgments

This work was supported by Korea Planning and Evaluation Institute of Industrial Technology (KEIT) grant funded by the Korea Government(MOTIE) (RS-2024-00416035) and the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2024-00357760).

Conflicts of interest

The authors declare that they have no competing interests.

Authors' contributions

Sangjin Maeng conceived the study, designed the methodology, conducted the experiments, developed the mechanistic and machine-learning models, analyzed the data, prepared the figures, and wrote and revised the

manuscript. The author read and approved the final manuscript.

Availability of Data and Materials

The datasets generated and analysed during the current study, including the measured cutting force data and processed data used for model development and validation, are available from the corresponding author on reasonable request.

References

[1] Altintas Y, Kersting P, Biermann D, Budak E, Denkena B, Lazoglu I (2014) Virtual process systems for part machining operations. *CIRP Ann* 63:585–605.

[2] Budak E (2006) Analytical models for high performance milling. Part I: Cutting forces, structural deformations and tolerance integrity. *Int J Mach Tools Manuf* 46:1478–1488.

[3] Mori M, Fujishima M, Inamasu Y, Oda Y (2011) State of the art and future trends in energy efficiency for machine tools. *CIRP Ann* 60:651–673.

[4] Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F (2018) Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol* 94:3563–3576.

[5] Kline WA, DeVor RE, Lindberg JR (1982) The prediction of cutting forces in end milling with application to cornering cuts. *Int J Mach Tool Des Res* 22:7–22.

[6] Armarego EJA, Deshpande NP (1991) Computerized predictive cutting models for forces in end-milling including eccentricity effects. *CIRP Ann* 40:25–29.

[7] Engin S, Altintas Y (2001) Mechanics and dynamics of general milling cutters. Part I: Helical end mills. *Int J Mach Tools Manuf* 41:2195–2212.

[8] Kline WA, DeVor RE (1983) The effect of runout on cutting geometry and forces in end milling. *Int J Mach Tool Des Res* 23:123–140.

[9] Lee P, Altintas Y (1996) Prediction of ball-end milling forces from orthogonal cutting data. *Int J Mach Tools Manuf* 36:1059–1072.

[10] Ozturk E, Budak E (2007) Modeling of 5-axis milling processes. *Mach Sci Technol* 11:287–311.

[11] Wan M, Altintas Y, Perry WH (2007) Calibration of half engagement milling force coefficients with runout. *Int J Mach Tools Manuf* 47:798–806.

[12] Malekian M, Park SS, Jun MB (2009) Modeling of dynamic micro-milling cutting forces. *Int J Mach Tools Manuf* 49:586–598.

[13] Bao WY, Tansel IN (2000) Modeling micro-end-milling operations. Part I: Analytical cutting force model. *Int J Mach Tools Manuf* 40:2155–2173.

[14] Wuest T, Weimer D, Irgens C, Thoben KD (2016) Machine learning in manufacturing: Advantages, challenges, and applications. *Prod Manuf Res* 4:23–45.

[15] Tandon V, El-Mounayri H (2001) A novel artificial neural networks force model for end milling. *Int J Adv Manuf Technol* 18:693–700.

[16] Briceno JF, El-Mounayri H, Mukhopadhyay S (2002) Selecting an artificial neural network for efficient modeling and accurate simulation of the milling process. *Int J Mach Tools Manuf* 42:663–674.

[17] Zuperl U, Cus F (2006) Dynamic neural network approach for tool cutting force modelling of end milling

operations. *Int J Syst Sci* 37:963–973.
[18] Qiao W, Yang S, Hu W, Li Z (2021) Cutting force prediction in milling based on a CNN-LSTM deep learning method. *Int J Adv Manuf Technol* 116:3823–3837.
[19] Liu M, Xie H, Pan W, Ding S, Li G (2025) Prediction of cutting force via machine learning: State of the art, challenges and potentials. *J Intell Manuf* 36:703–764.
[20] Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707.
[21] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L (2021) Physics-informed machine learning. *Nat Rev Phys* 3:422–440.
[22] Willard J, Jia X, Xu S, Steinbach M, Kumar V (2022) Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput Surv* 55:66.
[23] von Rueden L, Mayer S, Beckh K, Georgiev B, Giesselbach S, Heese R, Kirsch B, Pfrommer J, Pick A, Ramamurthy R, Walczak M, Garcke J, Bauckhage C, Schuecker J (2023) Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Trans Knowl Data Eng* 35:614–633.
[24] Zeng S, Pi D (2023) Milling surface roughness prediction based on physics-informed machine learning. *Sensors* 23:4969.
[25] Ren A, Ma M, Wu J, Karandikar J, Tyler C, Shi T, Schmitz T (2025) A cutting mechanics-based machine learning modeling method to discover governing equations of machining dynamics. *Manuf Lett* 44:759–769.
[26] Hathaway J, Stolkin R, Rastegarpanah A (2024) Imitation learning for sim-to-real adaptation of robotic cutting policies based on residual Gaussian process disturbance force model. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp 2899–2906.