



Generalization of LSTM and CNN autoencoders for anomaly detection across orthogonal and longitudinal turning

Ya-Jing Wu^{1,5} · Justin Kopp² · Jens-Peter M. Zemke¹ · Sebastian Götschel¹ · Pascal Volke³ · Andreas Zabel³ · Sebastian Schibsdatt⁴ · Jan Dege⁴ · Frank Walther² · Daniel Höche⁵

Received: 24 March 2026 / Accepted: 21 June 2026 / Published online: 14 July 2026
© The Author(s) 2026

Abstract

Machine learning models enable automated process monitoring in manufacturing and support adaptive quality assurance and predictive maintenance systems. Autoencoders have emerged as a promising approach to unsupervised anomaly detection in processes. However, their application to turning processes, particularly their generalizability across different process variants such as varying machine kinematics, cutting geometries, and process variability, remains largely unexplored. Existing studies mainly focus on controlled laboratory environments using single machine tools and standardized parameters, limiting industrial scalability. This work addresses this gap by systematically comparing long short-term memory (LSTM) and convolutional (CNN) based autoencoders developed in collaboration with two research projects. The models are trained and evaluated on two harmonized datasets derived from orthogonal and longitudinal turning processes. The datasets differ with respect to machine tools, kinematics, and cutting tools. Each model is trained on both its native process data and on the respective other dataset to evaluate cross-domain performance. The results on normalized reconstruction errors show that LSTM-based autoencoders yield higher temporal sensitivity, whereas CNN-based autoencoders exhibit greater reconstruction stability but reduced sensitivity to transient dynamics. Moreover, relative reconstruction errors are additionally evaluated in the original physical force units to improve process interpretability. This analysis reveals larger relative deviations for LSTM models due to their heightened sensitivity. Furthermore, the variability of the training dataset is identified as a key factor influencing out-of-distribution generalization performance. The findings demonstrate the generalizability of autoencoder based monitoring approaches across different cutting conditions and provide insights for data-driven and grey-box modeling frameworks in manufacturing.

Keywords Anomaly detection · Autoencoder · Turning · AISI 304 (X5CrNi18-10)

1 Introduction

Machining is a complex process influenced by cutting parameters, tool coatings, tool geometry, tool wear, and workpiece properties [1–3]. Reliable process monitoring is therefore essential to increase productivity, manufacturing quality, and prevent machine breakdowns. However, the nonlinear and time-dependent interactions between cutting parameters, tool wear evolution, and dynamic process responses (e.g., cutting forces and vibrations) pose significant challenges for conventional monitoring approaches relying on fixed thresholds or linear models. Consequently, data-driven approaches have been developed to improve

process understanding and enable more accurate condition estimation. Machine learning (ML) techniques have thus become key enablers for automated feature extraction and decision-making in machining processes [4]. Understanding how robustly these models operate across different machine tools and tools provides transferable knowledge for a wide range of intelligent monitoring concepts, including grey-box wear modeling frameworks that combine physics-based insights with data-driven anomaly detection mechanisms.

In manufacturing processes, ML models effectively capture complex relationships from high-dimensional sensor data [4]. Recently, physics-informed approaches have emerged that embed domain knowledge directly into neural

architectures such as the well-informed network by Stemmer et al. [5] or transient wear modeling by Wolf et al. [6]. These advances illustrate current progress in intelligent monitoring yet highlight persisting challenges related to limited labeled failure data. In addition, several studies have specifically addressed the relationship between cutting forces and tool wear progression to estimate wear states during turning and milling operations [7, 8].

However, online temporal sensor data are typically highly imbalanced since tool failure or other abnormal events occur only rarely [9]. This imbalance poses major challenges for supervised learning approaches that rely on labeled failure data and restricts their applicability to continuous industrial monitoring. To overcome these constraints, unsupervised anomaly detection methods, such as autoencoders, have gained increasing attention in recent years [10]. Autoencoders consist of two coupled neural networks, the encoder $e: \mathbb{R}^m \rightarrow \mathbb{R}^n$ followed by the decoder $d: \mathbb{R}^n \rightarrow \mathbb{R}^m$, where usually $n \ll m$. The smaller space $\mathbb{R}^n$ is known as *latent space* that contains a compressed representation of the system state learned during training [11]. During training, the network parameters are optimized by minimizing a reconstruction loss that quantifies the deviation between the input $x \in \mathbb{R}^m$ and its reconstruction $\hat{x} = d(e(x))$. By minimizing this loss on normal operating data, the autoencoder learns a compressed latent representation that captures characteristic patterns of the training data. When applied to new data, elevated reconstruction errors reveal deviations from these learned patterns and thus serve as indicators of anomalous process behavior.

Previous studies on anomaly detection in machining have focused mainly on milling under controlled laboratory conditions. Numerous studies detected tool wear, tool breakage, or material defects using ML [12, 13]. Considerably fewer investigations address anomaly detection using autoencoder architectures [14–16]. Comparable studies for turning operations are still lacking. Most existing works train their models under fixed process parameters on a single machine. This approach ensures high comparability between experimental datasets and enables the targeted analysis of relationships between signal features and process states. However, in real production environments, a model is required that operates reliably regardless of the employed tool types or machines, thus exhibiting strong generalizability.

Previous approaches to improving this generalizability have mainly focused on parameter variations within a single process as a function of process settings [17]. In addition, ML methods have been applied to estimate the operating state of digitally retrofitted machines [18]. Nevertheless, a systematic investigation of cross-domain robustness against domain shifts caused by process setups and machine tools remains largely absent, particularly for turning processes.

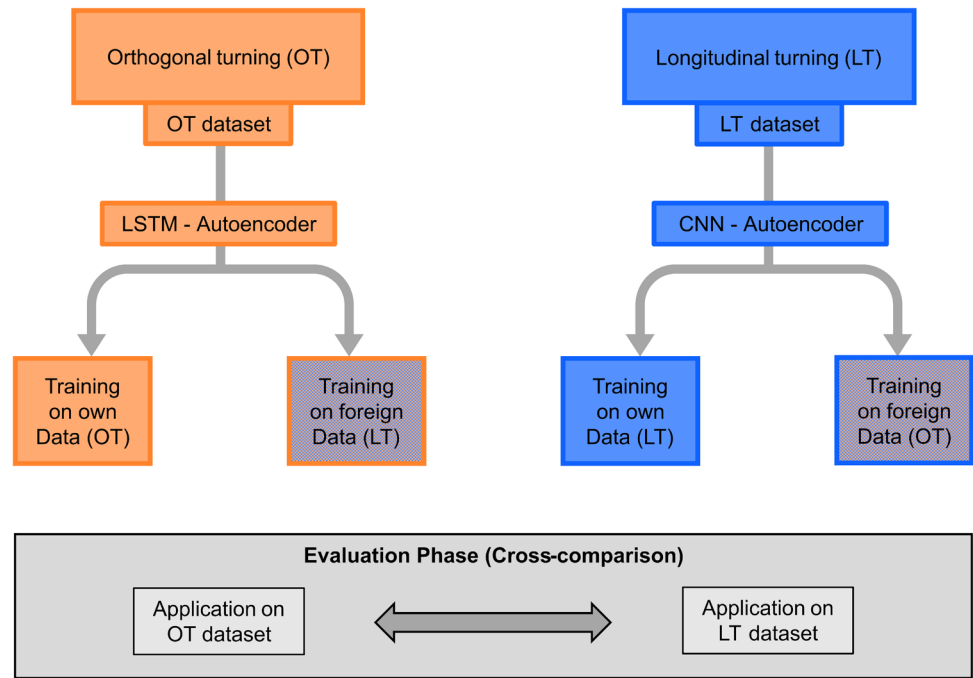
The present study addresses precisely this research gap by conducting two experimental series on two different lathes, each employing distinct indexable inserts. Furthermore, the two experiments differ in their machining process: on the one hand orthogonal turning, and on the other longitudinal turning. This corresponds more to an industrial environment, where different machine tools with different turning processes are used, rather than a laboratory analysis with one machine and only one process.

The main contribution of this work is a systematic cross-domain evaluation of autoencoder-based anomaly detection, analyzing the robustness and generalizability of LSTM- and CNN-based models across heterogeneous datasets. The underlying models and experimental data that originated from two independent research projects [19, 20] are combined to establish a structured framework to assess generalization across machine tools and cutting tool configurations. This enables, for the first time, an experimental quantification of model generalizability under realistic process variability. The findings reveal the relationship between model characteristics and anomaly detection performance under domain shifts, supporting the design of more robust data-driven monitoring methods.

2 Methodology

To systematically evaluate the generalizability of ML models across different machining environments, two experimental series involving orthogonal and longitudinal turning were conducted. Each experiment was carried out on a separate machine tool using distinct indexable inserts. The autoencoder models developed in both projects were retrained according to a cross-domain evaluation scheme illustrated in Fig. 1. Each model was trained twice, once with its own dataset and once with the external dataset provided by the project partner. In total, four differently trained models were obtained and subsequently tested on both datasets. This comparative evaluation makes it possible to analyze how each model performs when exposed to data from previously unseen domains. Differences in machine tool and fixture configurations influence system stiffness and vibration behavior, thereby affecting the measured signals. Variations in cutting tool condition and geometry modify force signatures and frequency content, particularly in the presence of tool wear or tool changes. By incorporating these variations across the datasets, the approach separates the influence of the training data from that of the model architecture itself, enabling a clearer assessment of robustness against variations induced by different tools and machine tools.

Fig. 1 Experimental workflow showing training and evaluation of autoencoder models across turning domains



2.1 Experimental data

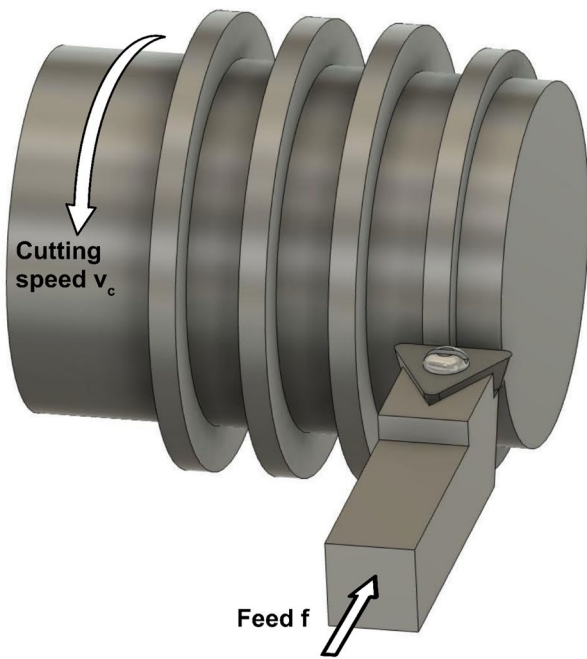
Two distinct experimental datasets form the basis for evaluating the generalizability of the anomaly detection models across different machining domains. The first dataset originates from orthogonal turning experiments, while the second represents external longitudinal turning operations. Both datasets involve the machining of AISI 304 with coated carbide tools but differ with respect to machine tools, tool geometries, process kinematics, and experimental planning. The datasets were generated within two independent research projects addressing different aspects of machining process monitoring and were therefore obtained under different experimental conditions. The orthogonal turning experiments follow a full factorial design, whereas the longitudinal turning experiments were conducted using a partial factorial design based on technologically relevant parameter ranges. Rather than reproducing identical parameter sets, the present study investigates whether characteristic signal patterns of the cutting process can be transferred across such variations. Changes in tool condition, for example due to progressive tool wear, are known to systematically influence cutting forces independent of the exact turning configuration [4, 21]. The investigated cutting parameters of both experiments lie within comparable technological ranges for machining AISI 304 derived from tool manufacturer recommendations and process-specific considerations. Consequently, additional experiments with identical parameter combinations were not required for the intended cross-domain evaluation. The following subsections therefore

describe the two experimental setups representing different but technologically comparable machining domains.

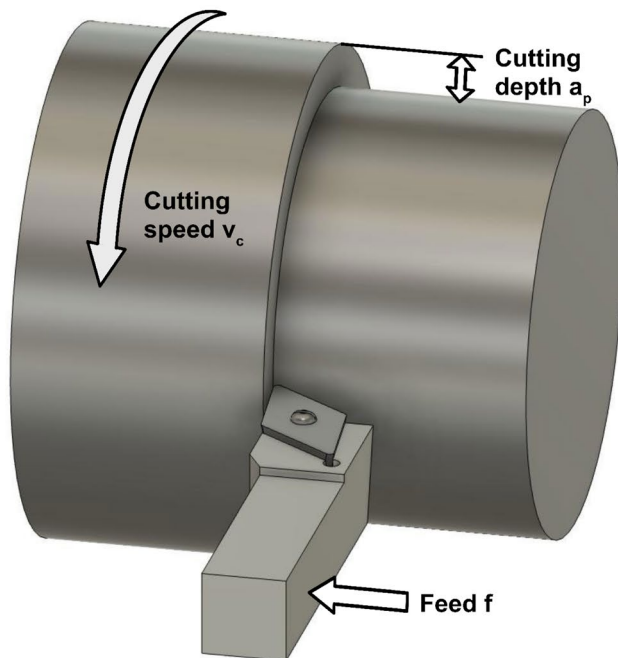
2.1.1 Orthogonal turning process

In orthogonal turning (OT), the cutting edge is oriented perpendicular to the direction of the cutting speed, resulting in a two-dimensional cutting condition along a straight cutting edge. In the present configuration, the tool is fed radially into the rotating workpiece. Thus, the feed motion is realized by the radial infeed of the tool, while the cutting speed v_c is generated by the rotation of the workpiece. To ensure a constant cutting speed, the spindle speed was continuously adjusted according to the actual workpiece diameter. For the investigation of tool wear along the straight cutting edge under defined cutting conditions, grooves were machined into the workpiece prior to the actual experiments. The remaining bars between the grooves were subsequently removed during the turning tests. This preparation ensures a defined width of cut of $b = 2$ mm, which, in the case of external longitudinal turning, corresponds to a cutting depth a_p of 2 mm. The engagement of the tool corner radius was prevented by a defined lateral positioning of the tool relative to the bars, so that only the straight cutting edge participated in the cutting process. The removal of one bar required approximately 1 s of cutting time (Fig. 2a).

All machining operations were carried out on an Index ABC 65 production turning center. Each tool was operated for a maximum cutting time of 240 s. This limit was chosen because the tools exhibited significantly longer tool lives (> 800 s), while 240 s was sufficient to capture the relevant



(a) Orthogonal turning



(b) Longitudinal turning

Fig. 2 Comparison between orthogonal and longitudinal turning process**Table 1** Process parameters and force component definitions for orthogonal and longitudinal turning

Parameter	Unit	Orthogonal turning (OT)	Longitudinal turning (LT)
feed rate f	mm	0.1, 0.2	0.15, 0.25, 0.35
Cutting speed v_c	m/min	155, 205, 255	130, 170, 210, 250, 290
Cutting depth a_p	mm	2	1.2, 1.6
Feed force F_f	N	F_x	F_x
Passive force F_p	N	F_z	F_y
Cutting force F_c	N	F_y	F_z
Sampling rate	kHz	10	51.2
Termination criterion	–	240 s cutting time	VB _{max} = 200 μ m

process dynamics. During the experiments, the force signals were recorded in all three spatial directions using a three-component dynamometer of type 9129AA (Kistler Instrumente AG). The charge signals were conditioned using charge amplifiers (type 5011, Kistler) and digitized via an A/D module (NI-9239, National Instruments) at a sampling rate of 10 kHz. Details of the measurement chain are provided in [22]. A TiAlN-coated cemented carbide tool of type TPUN160304EN manufactured by Ceratizit was used. The coating exhibited a thickness of approximately 4 μ m, determined by cross-sectional analysis using focused ion beam (FIB) preparation. The workpiece material was austenitic stainless steel AISI 304 (X5CrNi18-10).

2.1.2 Longitudinal turning process

External longitudinal turning (LT) is a machining process involving the removal of material from the outer surface of a rotating workpiece using a single-point cutting tool, where the cutting tool moves parallel to the workpiece's axis of rotation. The material is removed at a depth of cut a_p and a cutting speed v_c , while the tool moves parallel to the axis of rotation at a feed rate f (see Fig. 2b). The experiments conducted comprised repeated, time-controlled cutting intervals of 30 s. After each interval, the machined surface and tool wear were measured. During machining, force signals were recorded in three spatial directions.

For each cutting edge, the cutting intervals were repeated under identical cutting parameters until a predefined tool life criterion was reached, defined as a maximum flank wear of $VB_{\max} = 200 \mu$ m. If tool failure or unsuitable chip formation, such as excessive chip length, segmentation, or irregular chip breakage associated with process instabilities and fluctuating cutting forces, occurred before this wear limit was reached, the experiment was terminated prematurely.

The experimental matrix included variations in depth of cut, cutting speed, and feed rate (see Table 1). A cemented carbide insert (CNMG160608-MM5, Walter AG) with a

custom TiAlN coating was used as the cutting tool. The coating, which had a thickness of around 3.5 μm , was applied using physical vapour deposition (PVD). The workpiece material was AISI 304 (X5CrNi18-10) in the form of a shaft with an initial diameter of 150 mm and a usable cutting length of 500 mm fixed on the Max Mueller Gildemeister MD5S CNC lathe. The shaft was flattened to a depth of 4.5 mm on one side to enable the temperature of the rake face to be measured once per rotation. Cutting forces were measured using a three-component piezoelectric dynamometer (Type 9257B, Kistler Instrumente GmbH) mounted between the tool holder and the machine turret. The dynamometer was integrated via a dedicated adapter platform that preserves turret rotation. A detailed description of the mechanical integration, sensor fixation, experimental setup, materials, and data synchronization is provided by Schibsdatt et al. in [20].

2.1.3 Joint data pre-processing

Since the datasets originate from different turning processes, they need to be preprocessed to enable a direct comparison. The raw data from the orthogonal and longitudinal turning experiments differ in several aspects, as summarized in Table 1. In ideal two-dimensional orthogonal cutting, there is no passive force component. However, for consistency and comparability with longitudinal turning, the force component perpendicular to the cutting and feed forces is referred to as the passive force F_p in the present study. The force signals were recorded by dynamometers along the three spatial directions x , y , and z . However, for analysis purposes, the relevant force components are the cutting force, feed force, and passive force. Since the orientation of the measured force components differs between setups, all coordinate systems were aligned so that equivalent physical directions, namely cutting, feed, and passive, could be directly compared across datasets.

The force signals were acquired at different sampling rates (51.2 kHz and 10 kHz). This discrepancy could lead to deviations during cross-domain evaluation if an autoencoder trained on high-rate data expects slightly different temporal dynamics than one trained on lower-rate data. To eliminate this effect, all signals were resampled to a common rate of 10 kHz. Signals with higher sampling frequency were first processed using a low-pass filter removing frequencies above 5 kHz; subsequently, downsampling was performed to 10.24 kHz. As the proposed approach focuses on detecting deviations from learned normal patterns, which are primarily governed by the dominant (lower- to mid-frequency) components of the signal, any residual attenuation of high-frequency content is not expected to significantly affect model performance. The remaining difference of

approximately 2.4% between datasets is negligible and ensures comparable temporal resolution for all subsequent analyses. The nomenclature and definition of the force components are based on DIN 6580 [23].

2.2 Autoencoder models

Two different autoencoder architectures were implemented to capture the complex relationships within the sensor data and to evaluate their respective robustness under varying machining conditions. Both models aim to accurately reconstruct normal cutting conditions but differ in their mechanisms for handling temporal dependencies. Their architectures, training procedures, and anomaly-detection approaches are described in the following subsections to ensure comparability between the models.

2.2.1 Long short-term memory autoencoder

The sensor data consist of temporally recorded cutting force signals obtained during orthogonal turning operations. Since anomaly assessment requires consideration of temporal dependencies, a recurrent neural network architecture based on long short-term memory (LSTM) units was employed [19]. The recurrent structure enables the model to retain information from past time steps and thus capture characteristic temporal patterns related to tool wear or process instability. The objective was to design a lightweight and efficiently trainable model that provides stable reconstruction performance while maintaining low complexity. The final architecture is presented in Table 2. The hyperparameters were manually tuned based on empirical testing summarized in Table 2. The LSTM autoencoders were implemented using the PyTorch library [24].

2.2.2 Convolutional autoencoder

In industrial applications, one-dimensional convolutional neural networks (1D CNNs) essentially enable the direct use of time-series sensor data from manufacturing processes without requiring transformation. 1D fully CNN autoencoders not only still keep the temporal and channel dependencies without frequency transformation but also have lower computation complexity [25]. The final configuration of the CNN autoencoders, obtained from force data with an input dimension of [1024, 3] based on all the experimental tests, consisted of two convolutional layers with kernel sizes of [35, 27] and filter numbers of [56, 52], along with one average pooling layer of window size 2. This architecture demonstrated the most effective reconstruction performance, indicating a balanced trade-off between model complexity and signal resolution. The model structure in Table 2 was

Table 2 Architecture and training configuration of LSTM and CNN autoencoders

Component	LSTM autoencoder	CNN autoencoder
Encoder architecture		
Input	shape (x, 3)	shape (1024, 3)
Layer 1	LSTM, output (x, 64)	Conv1d (56), output (990, 56)
Layer 2	–	Conv1d (52), output (964, 52)
Pooling	–	AvgPool1d (2), output (482, 52)
Latent space	output (x, 32)	Conv1d (2), output (482, 2)
Decoder architecture		
Layer 1	LSTM, output (x, 64)	ConvTranspose1d (52), output (482, 52)
Upsampling	–	output (964, 52)
Layer 2	–	ConvTranspose1d (56), output (990, 56)
Output	output (x, 3)	ConvTranspose1d (3), output (1024, 3)
Training configuration		
Optimizer	Adam	AdamW
Learning rate	0.001	0.001
Batch size	1	64
Loss function	MAE	MSE
Activation function	tanh	Leaky ReLU
Weight initialization	Xavier uniform	Xavier uniform
Normalization	Z-score	Z-score
Maximum epochs	500	500
Early stopping	20 epochs	20 epochs

used to develop fully convolutional autoencoder models under different experiments [20]. The hyperparameters and training setup of the CNN autoencoders are shown in Table 2. The resulting models form the basis for the subsequent cross-domain evaluation to assess their robustness under domain shift conditions. The CNN autoencoders were implemented using the TensorFlow library [26].

2.2.3 Model training

Each autoencoder architecture was trained twice according to the cross-domain evaluation scheme. One training run used its own dataset, while the second used the external dataset provided by the project partner. This procedure ensured identical training conditions across both domains and enabled a direct assessment of model robustness independent of architectural differences. As a result, four models were obtained that were later evaluated under domain shift conditions. Different training configurations were selected for the LSTM- and CNN-based autoencoders to ensure stable and effective model convergence. The reconstruction loss function differs between both architectures (Table 2).

The LSTM-based autoencoders are trained using the mean absolute error (MAE) loss, computed over all time steps and all three input force channels. For an input sequence $x \in \mathbb{R}^{n \times 3}$ and its reconstruction $\hat{x} \in \mathbb{R}^{n \times 3}$, the training objective is defined as:

$$\text{MAE}(x, \hat{x}) = \frac{1}{3n} \sum_{c=1}^3 \sum_{i=1}^n |x_{i,c} - \hat{x}_{i,c}|, \quad (1)$$

where n denotes the number of time samples and $c = 1, 2, 3$ indexes the three force-signal channels. It provides stable convergence and is less sensitive to local outliers in time-series data. In contrast, the CNN-based autoencoder was trained using mean squared error (MSE)

$$\text{MSE}(x, \hat{x}) = \frac{1}{3n} \sum_{c=1}^3 \sum_{i=1}^n (x_{i,c} - \hat{x}_{i,c})^2, \quad (2)$$

which emphasizes larger deviations and therefore increases the penalization of pronounced reconstruction mismatches. These differences originate from the independent development of both autoencoder models [19, 20].

For the OT dataset, 2,059,911 samples were used for training, 285,738 for validation, and 286,885 for testing. For the LT dataset, 3,932,158 samples were used for training, 491,522 for validation, and 548,022 for testing. The cutting parameters for this data setup are described in Table 1. The data for the cross-domain evaluation were not part of the training data setup.

The training of the models followed identical training procedures. A maximum number of 500 epochs with early stopping with a patience of 20 epochs and a learning rate reduction strategy (reduction factor = 0.7, patience = 5, $\epsilon = 10^{-4}$, min. learning rate = 10^{-4}) was employed. The best model in the training process was saved for the test dataset.

During machining, the cutting forces were recorded in all three spatial directions for both OT and LT. Due to the process kinematics, the passive force F_p component in orthogonal turning is close to zero, and the measured signal in this direction mainly reflects structural vibrations of the machine–tool system rather than the cutting process itself. In contrast, longitudinal turning exhibits a considerably higher passive force level with comparatively smaller relative fluctuations. Including the passive force component in the autoencoder input would therefore introduce a systematic bias between both datasets after normalization. For OT, the model would learn comparatively strong relative fluctuations around a force level close to zero, whereas for LT it would learn a higher absolute force level with smaller

relative variations. This difference could distort the reconstruction behavior of the autoencoder and affect the cross-domain evaluation. To avoid such bias related to the passive force component, the anomaly detection analysis was performed using only the active force components, namely the cutting force F_c and the feed force F_f , which directly represent the chip formation process and provide the most relevant information about the cutting dynamics.

2.2.4 Cross-domain evaluation of anomaly detection models

The cross-domain evaluation methodology provides a systematic assessment of the responses of autoencoder architectures to data originating from different machining domains. The anomaly detection is based on the reconstruction error between the input and reconstructed signals, which quantifies deviations from the learned normal cutting conditions. Elevated reconstruction errors therefore indicate potential anomalies in the process behavior. For each model, reconstruction errors are computed for both its own dataset and the external dataset. This procedure enables a direct comparison of model responses under domain-shift conditions without retraining. The analysis focuses on identifying how training data origin, model architecture, and the inherent characteristics of the applied datasets influence anomaly sensitivity and robustness.

Although the experimental parameter sets for OT and LT differ, the investigated cutting speeds, feed rates, and depths of cut lie within a comparable technological parameter space. The number of investigated parameter combinations also differs between the datasets (six for OT and nine for LT) due to the independent experimental designs. Consequently, no major differences in the model response are expected solely due to the selected cutting parameters, as the fundamental relationships between tool condition and cutting forces remain similar for both turning processes. Furthermore, the input signals are normalized prior to model training and evaluation, which reduces the influence of absolute force magnitudes and emphasizes characteristic temporal signal patterns. This enables a meaningful comparison of model behavior across the two machining domains.

The datasets exhibit different temporal characteristics in their force signals. For the investigation, the trained models are applied to both domains and their reconstruction errors are compared to assess sensitivity under varying process conditions. Anomaly detection is performed by defining a threshold above a specified quantile of the reconstruction error distribution. This quantile-based approach enables consistent evaluation across models and provides a comparison independent of absolute error magnitudes or model-specific scaling effects. Such threshold definitions

Table 3 Relative composition of cutting forces for both turning processes

Process	F_c [%]	F_f [%]	F_p [%]
OT	65.96 ± 4.27	32.70 ± 4.06	1.33 ± 0.26
LT	54.64 ± 2.61	26.98 ± 2.06	18.38 ± 1.76

Table 4 Average standard deviation of cutting forces across all experiments

Process	Average standard deviation [N]		
	F_c [N]	F_f [N]	F_p [N]
OT	8.91 ± 2.85	6.99 ± 2.18	2.27 ± 0.69
LT	49.66 ± 28.84	50.73 ± 23.57	36.68 ± 31.74

are commonly applied in unsupervised anomaly detection when no labeled failure data are available and the distribution of anomalous events is unknown [27]. In the present study, a high quantile was selected in order to isolate only rare deviations from the learned normal behavior.

To further reduce the influence of isolated noise peaks or short transient fluctuations, consecutive samples exceeding the threshold are grouped into anomaly events. Values separated by gaps of less than 50 samples (corresponding to 0.05 s) are treated as belonging to the same anomaly. This procedure prevents individual outliers from being interpreted as independent anomalies and improves the robustness of the detection results.

3 Results and discussion

This section presents the results of the cross-domain evaluation and discusses the performance of both autoencoder architectures regarding reconstruction accuracy and anomaly-detection capability. The analyses include a comparison of raw process data, training behavior, and detection results to assess model robustness and generalization across machining domains. Independent of the training loss function (MAE or MSE), the reconstruction error is evaluated using MSE.

3.1 Raw-data comparison

Machine learning models capture specific correlations present within the training data. Correlations that do not exist in the data cannot be learned by the model. Due to the differences between the two datasets analyzed in this study, the underlying process behavior is examined to assess its potential influence on model performance.

Tables 3 and 4 summarize the statistical characteristics derived from the training datasets for each variation of the process parameters. For every experiment, the values were calculated over a reference segment of 0.5 seconds

corresponding to approximately 5000 samples at a sampling rate of 10 kHz. This interval represents stable cutting conditions and ensures comparability across all parameter settings. Table 3 presents the average percentage contribution of each force component to the total resultant cutting force $F_{\text{total}} = F_c + F_f + F_p$ across all process parameters. The most pronounced difference between both processes is observed in the passive force component, which is considerably higher for longitudinal turning due to the inherent characteristics of this machining operation.

The standard deviations of the measured force components listed in Table 4 show significantly higher variability in the force signals for LT than for OT. The increased variability indicates differences in the structural–dynamic behavior of the respective machine tool systems. In particular, deviations in structural stiffness and damping characteristics of the tool holders may influence the dynamic excitability of the system and thereby affect cutting-force stability [28].

As illustrated in Fig. 3, the temporal evolution of the cutting force signals is analyzed for the cross-domain evaluation process. In the context of LT, pronounced fluctuations in force are observed throughout the duration of the process, aligning with the higher standard deviations documented in Table 4. Over the machining period of approximately 140 seconds, a gradual escalation in cutting force amplitude is evident, indicative of progressive tool wear. There are hardly any obvious anomalies in the force curve. At the end of the experimental phase, a maximum flank wear VB_{max} of 257 μm was measured.

In contrast, in the context of OT, several localized anomalies emerge along the process sequence, signifying transient disturbances. The detected anomalies correspond to distinct local peaks in the cutting force signals visible in the raw data. Tool wear can largely be excluded as the primary cause of these events, since the overall flank wear progression remained low during the experiment and no abrupt changes in wear state were observed in the corresponding time intervals. The final measured flank wear reached only VB_{max} of 91 μm , indicating comparatively mild wear conditions.

Austenitic stainless steels such as AISI 304 are known to exhibit unstable chip formation and the occurrence of irregular chip breakage or entangled chips, which can temporarily alter the contact conditions between tool, chip and workpiece [29]. In the present experiments, continuous chip formation without regular chip breakage was observed. Such long chips can promote chip entanglement or irregular chip flow, which may temporarily disturb tool–chip contact conditions. These transient disturbances can lead to short-term force peaks which appear as anomalies.

3.2 Training results

The training of all models shows stable convergence resulting in low reconstruction errors. Table 5 displays the average final loss values for the training, validation and test datasets. The LSTM autoencoder achieves a lower loss than the CNN architecture. Such architecture-dependent performance differences have also been observed in recent large-scale benchmarking studies on reconstruction-based

Fig. 3 Cutting forces F_c in orthogonal turning (OT) and longitudinal turning (LT). The upper graph shows the global force signal and its averaged representation, while the lower graph highlights a magnified segment to illustrate the local signal

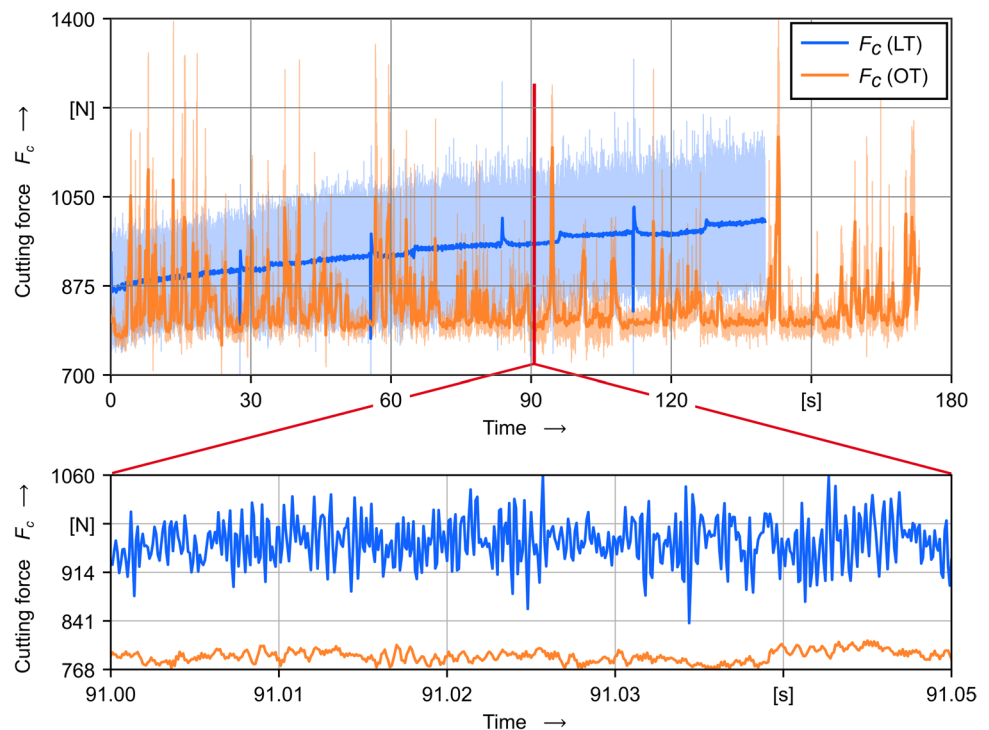


Table 5 Average final loss values on training, validation and test datasets

Model	Dataset	Training		Validation		Test	
		MAE	MSE	MAE	MSE	MAE	MSE
LSTM	OT	0.0035	0.0015	0.0087	0.0035	0.0058	0.0009
LSTM	LT	0.0141	0.0013	0.0125	0.0007	0.0124	0.0006
CNN	OT	0.0623	0.0075	0.0819	0.0156	0.0822	0.0140
CNN	LT	0.2189	0.0846	0.2155	0.0808	0.2196	0.0829

Table 6 Comparison of the reconstruction error of all models on the orthogonal turning (OT) dataset, with the parameters $v_c = 205$ m/min and $f = 0.2$ mm

Model	Mean	Standard deviation	Maximum error
LSTM OT	0.0053	0.2035	41.21
LSTM LT	0.0816	1.0276	75.55
CNN OT	0.0217	0.0696	12.65
CNN LT	0.1694	0.7349	43.35

time-series anomaly detection [30]. This can be attributed to its ability to capture temporal dependencies within the force signals more effectively through recurrent structures than convolutional filters [30, 31]. In contrast, the higher architectural complexity of the CNN may have led to a slower convergence rate. This observation is supported by the fact that the LSTM-based autoencoder was trained using MAE as loss function and still achieved more stable convergence than the CNN-based model trained with MSE. Typically, the use of MSE tends to accelerate convergence due to its stronger penalization of large deviations. Therefore, the superior stability observed for the LSTM model trained with MAE is particularly noteworthy [32].

Nevertheless, both models demonstrate successful training and accurate data reconstruction. In relation to the process data, both architectures yield reduced losses when trained on OT in comparison to LT. Although LT shows no obvious anomalies, it exhibits a wider natural spread in force magnitudes. Normalization reduces the influence of global variations in the signal amplitude but does not remove local temporal dynamics. This suggests that the local fluctuations in LT sensor data are more difficult to predict and reconstruct, whereas OT data contain distinct anomalies but otherwise present lower variability in normal operation, which likely facilitates the learning of consistent patterns and results in smaller loss values.

3.3 Orthogonal turning data analysis and anomaly detection

The reconstruction error is fundamental to the identification of anomalies. Consequently, its statistical behavior offers insight into the model's sensitivity and robustness prior to detailed detection analysis. For this purpose, an independent OT dataset was used, obtained from measurements at a cutting speed of $v_c = 205$ m/min and a feed rate

of $f = 0.2$ mm. This dataset is independent of the training, validation and test datasets.

As demonstrated in Table 6, the CNN architecture demonstrates larger mean reconstruction errors with respect to the training data in comparison to the LSTM models. The elevated error levels can be related to the higher training loss observed for CNNs, suggesting less effective learning of stable signal patterns during model training.

With respect to the influence of the training dataset, models trained on LT data show increased sensitivity, reflected by higher maximum reconstruction errors and larger standard deviations compared with those trained on OT data (see Table 6). The elevated error levels indicate that the greater variability and dynamic behavior present in the LT training signals make it more difficult for these models to generalize to the comparatively stable OT process. Although both LSTM models achieve smaller mean values, but the higher standard deviation and maximum error indicate that the LSTM-based autoencoder reconstructs most samples accurately but suffer from occasional large reconstruction errors on anomalies. In contrast, the CNN OT model provides more stable reconstruction performance with fewer extreme outliers, despite a slightly higher average error.

The anomaly detection analysis is predicated on the active force components, F_c and F_f . These components dominantly represent the process dynamics and tool condition. For multivariate signals, the per-sample MSE at time index t is defined as

$$\text{MSE}(t) = \frac{1}{K} \sum_{k=1}^K (x_k(t) - \hat{x}_k(t))^2, \quad (3)$$

where $K = 2$ is the number of the input channels of cutting and feed forces. Anomalies are identified through the implementation of a threshold at the 99.5th percentile of the reconstruction error distribution shown in Fig. 4. This quantile-based approach is adaptable to individual model characteristics and reliably separates abnormal from normal process behavior. Consecutive values that exceed this threshold, with a maximum gap of 50 values (0.05 seconds) between them, are grouped into single anomaly events for the purpose of comparison across datasets.

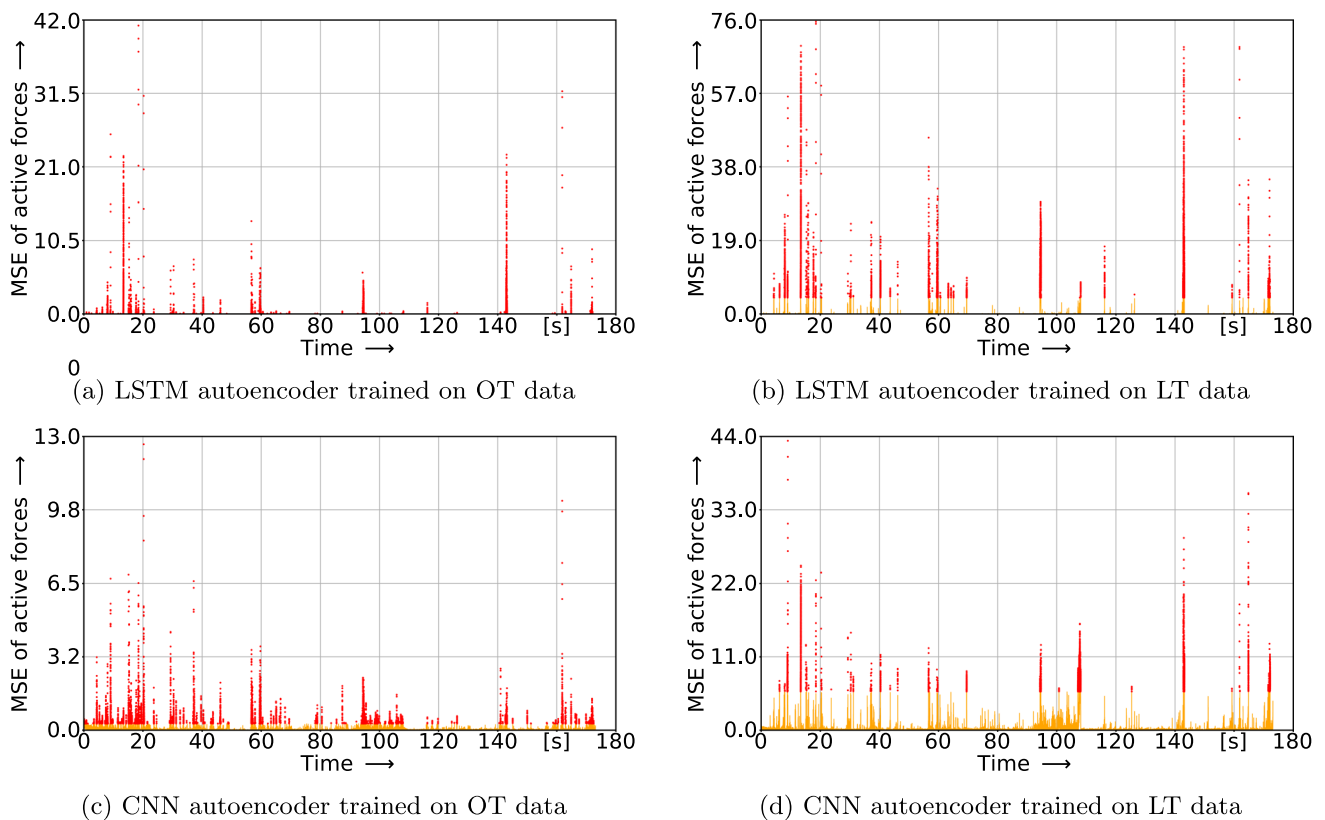


Fig. 4 Mean squared reconstruction errors (MSE) of autoencoders on normalized orthogonal turning data (orange); anomalies during the process (red)

Table 7 Comparison of detected anomaly events between all models on the orthogonal turning (OT) dataset, with the parameters $v_c = 205$ m/min and $f = 0.2$ mm

Model	Model	Groups	Groups	Matching	Matching	Average
A	B	A	B	A	B	[%]
CNN OT	CNN LT	479	86	62 (13 %)	56 (65 %)	39
CNN OT	LSTM OT	479	189	19 (4 %)	19 (10 %)	7
CNN OT	LSTM LT	479	103	13 (3 %)	14 (14 %)	8
CNN LT	LSTM OT	86	189	4 (5 %)	2 (1 %)	3
CNN LT	LSTM LT	86	103	1 (1 %)	1 (1 %)	1
LSTM LT	LSTM OT	103	189	103 (100 %)	101 (53 %)	77

The comparison is summarized in Table 7. A high degree of agreement is observed within each architecture type. This consistency is particularly pronounced for the LSTM models, where all anomalies detected by the LT model were also found by the OT model. The CNN architecture shows still a considerable overlap of approximately 65%. In contrast, cross-domain comparisons between LSTM and CNN architectures reveal only few common anomalies, indicating that network architecture influences what is classified as anomalous behavior. Regarding training data, both models trained on the LT data respond with fewer distinct anomaly events overall. This may be related to the higher dynamic variability within the LT dataset, although it contains no clearly defined failures or breakages. Consequently, smaller

variations are likely interpreted as normal process behavior rather than as anomalies.

Overall, the analysis demonstrates that model architecture and training data domain both significantly affect reconstruction accuracy and anomaly sensitivity on OT data [30, 33].

The detected anomalies in the reconstruction error occur at the same time instances where pronounced increases in the cutting forces can be observed in the raw signal (see Fig. 3). This indicates that the anomalies correspond to real fluctuations in the measured force signals rather than being purely artifacts of the reconstruction model. However, the models differ in their sensitivity with respect to which force increases are classified as anomalous. A direct assignment

Table 8 Comparison of the reconstruction error of all models on the longitudinal turning (LT) dataset, with the parameters $a_p = 1.6$ mm, $v_c = 210$ m/min and $f = 0.25$ mm

Model	Mean	Standard deviation	Maximum error
LSTM OT	0.0154	0.0701	9.28
LSTM LT	0.0007	0.0159	7.54
CNN OT	0.4165	0.4863	14.33
CNN LT	0.0929	0.1027	2.58

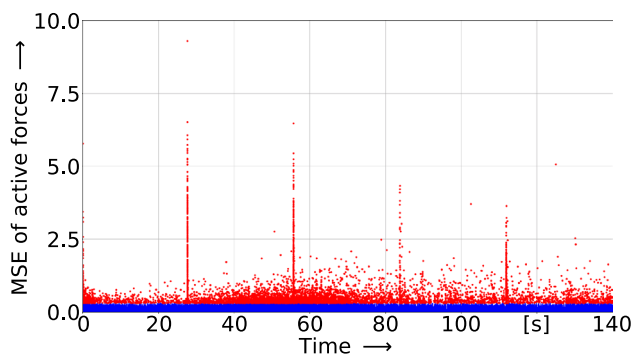
of the detected anomalies to specific physical events in the cutting process is not possible within the scope of the present experiments, since no additional process monitoring such as high-speed imaging or vibration measurements was performed. Consequently, the anomaly detection approach should primarily be interpreted as an indicator of unusual process behavior rather than as a direct identification of specific physical mechanisms. Future studies could extend this approach by combining anomaly detection with additional process observations in order to distinguish between normal process fluctuations and critical events, such as unstable chip behavior or tool failure.

3.4 Longitudinal turning data analysis and anomaly detection

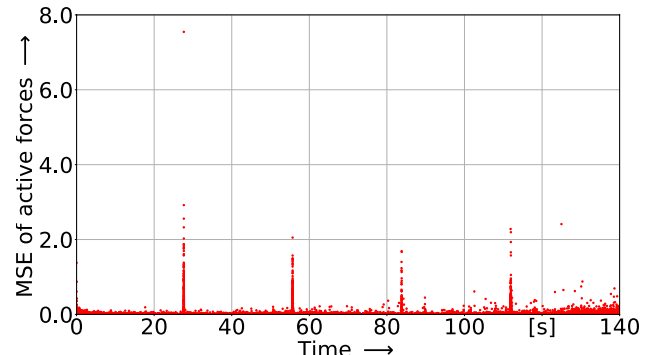
As summarized in Table 8, the CNN architectures exhibit larger mean reconstruction errors and higher standard deviations compared with the LSTM models. This trend is consistent with the results obtained for orthogonal turning. This suggests that, for the investigated force signals, the implemented convolutional autoencoder architecture provides lower predictive capability than the LSTM-based approach. In contrast, both LSTM models achieve smaller mean values and lower variability of reconstruction errors, demonstrating greater robustness under these conditions.

Regarding the influence of the training dataset, models trained on the OT data show increased sensitivity, reflected by higher maximum reconstruction errors and larger standard deviations compared with those trained on the LT data. The CNN model trained on the OT data yields the highest overall reconstruction error, suggesting that the stronger dynamic variations present in the LT signals can be better reconstructed once they have been learned during training.

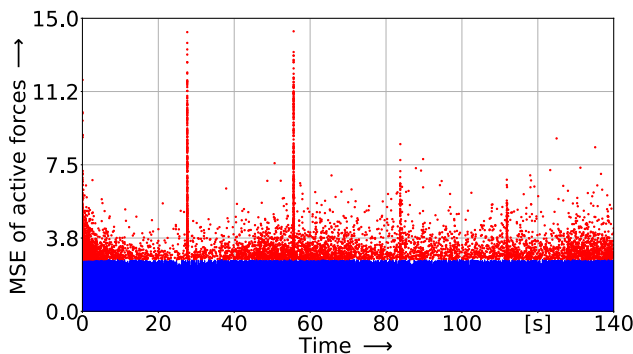
Figure 5 illustrates the distribution of reconstruction errors for all autoencoder models applied to the LT dataset across both active force components (F_c and F_f) together



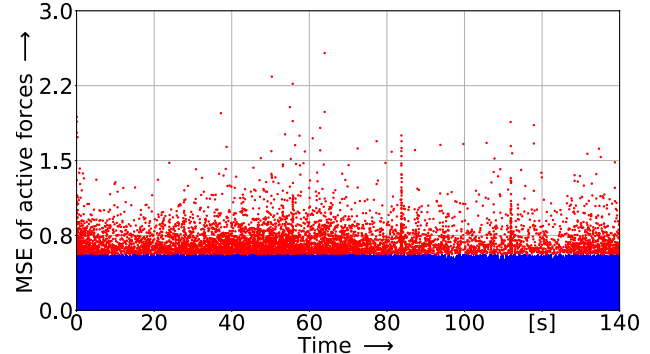
(a) LSTM autoencoder trained on OT data



(b) LSTM autoencoder trained on LT data



(c) CNN autoencoder trained on OT data



(d) CNN autoencoder trained on LT data

Fig. 5 Mean squared reconstruction errors (MSE) of autoencoders on normalized longitudinal turning data (blue); anomalies during the process (red)

with the corresponding 99.5th percentile thresholds for each model.

Due to the absence of clearly separated anomalies in this dataset, an anomaly-detection procedure as conducted in the previous section is not meaningful here. Therefore, a qualitative analysis of the reconstruction-error progression was performed. The results reveal four distinct peaks in the LSTM models' reconstruction error that occur at regular intervals. These peaks correspond to run-in behavior when the tool begins chip formation and cutting forces gradually increase. In this case, the run-in behavior was not fully filtered out during preprocessing, and both autoencoders correctly identified these regions as deviations from their learned steady-state patterns. Although these transients are part of normal process behavior, they appear as anomalies within the learned representation. Comparing the results of both models reveals that the two LSTM autoencoders detect four major peaks in the reconstruction error. When trained on the LT dataset, these peaks show a stronger contrast between background noise and peak amplitude. The model trained on the OT data additionally exhibits two smaller peaks at approximately 83 s and 125 s, whereas the LT-trained model only reproduces the latter peak. The CNN autoencoder presents a distinctly different behavior. When trained on the OT data, it successfully recognizes the first two run-in phases while interpreting the subsequent ones as part of normal operation. Moreover, the smaller peaks observed in the LSTM models do not appear here. When trained directly on the LT data, no distinct peaks are visible at all, indicating that this model has learned to reproduce process behavior very consistently and with minimal sensitivity to transient fluctuations.

To summarize, an analysis of the LT dataset reveals that the CNN autoencoder demonstrates reduced sensitivity to process variations. The model also tends to interpret unusual but recurring behavior from its training data as part of normal patterns. In contrast, the LSTM architecture responds more strongly to such deviations and is therefore more likely to classify them as anomalies. Such architectural trade-offs between sensitivity and stability have also been reported in broader comparative studies of reconstruction-based models [30].

3.5 Relative reconstruction errors

The axes of the datasets have different average values and ranges. Since model training, threshold definition, and cross-domain evaluation were performed on normalized signals, the reconstruction performance was primarily evaluated using the mean squared error (MSE) computed on standardized data to ensure comparability across datasets with different force magnitudes. However, evaluating

the error exclusively in normalized space does not directly reflect the magnitude of deviations in physical units and may therefore limit the interpretability of the detected anomalies with respect to actual process behavior. Therefore, the final model performance for detected anomalies is additionally assessed using the relative reconstruction error, defined as

$$\text{RE}(t) = \frac{\|F_{\text{true}}(t) - F_{\text{recon}}(t)\|}{\|F_{\text{true}}(t)\|} \times 100\% \quad (4)$$

$$F(t) = \begin{bmatrix} F_c(t) \\ F_f(t) \end{bmatrix}$$

on the resultant force of F_c and F_f in the original physical units. This complementary evaluation allows distinguishing genuine process-related deviations from model- or normalization-induced artifacts [34].

The peaks shown in Fig. 6 suggest higher data irregularity of the OT dataset. The consistency of the peak locations with Fig. 4 confirms that the detected deviations correspond to clear local anomalies in the OT data as illustrated in Section 3.1. The relative error of the LSTM models reduces the prominence of peak locations in the two periods of stable and lower dominant cutting force from 40 s to 60 s and from 100 s to 140 s. The CNN models still detect these anomalies as in Fig. 4a with lower overall relative error.

On the other hand, Fig. 7 shows the relative error of the models on the LT data. The four prominent outliers are consistent with the peaks observed in Fig. 5b, confirming that these deviations are not artifacts of normalization. When comparing Figs. 5d and 7d, it becomes evident that the relative error highlights transient deviations across all models, whereas the MSE of the CNN models enables a clearer differentiation between background fluctuations and pronounced anomalies.

In contrast to the MSE-based evaluation, the relative reconstruction error reduces overall magnitude differences between models but does not eliminate domain-specific effects. For the OT dataset, the influence of the training domain remains clearly observable, whereas for the LT dataset, architectural characteristics become more dominant. This indicates that normalization-related scaling effects are reduced, while inherent differences in data structure and model dynamics remain decisive.

3.6 Implications for cross-domain generalization

The comparison between the LT and OT datasets (Tables 6 and 8) reveals a consistent pattern across both evaluation metrics. Regardless of whether the reconstruction error is assessed in the normalized space using MSE or by means of the relative error, each model achieves lower errors when evaluated on the same domain on which it was trained.

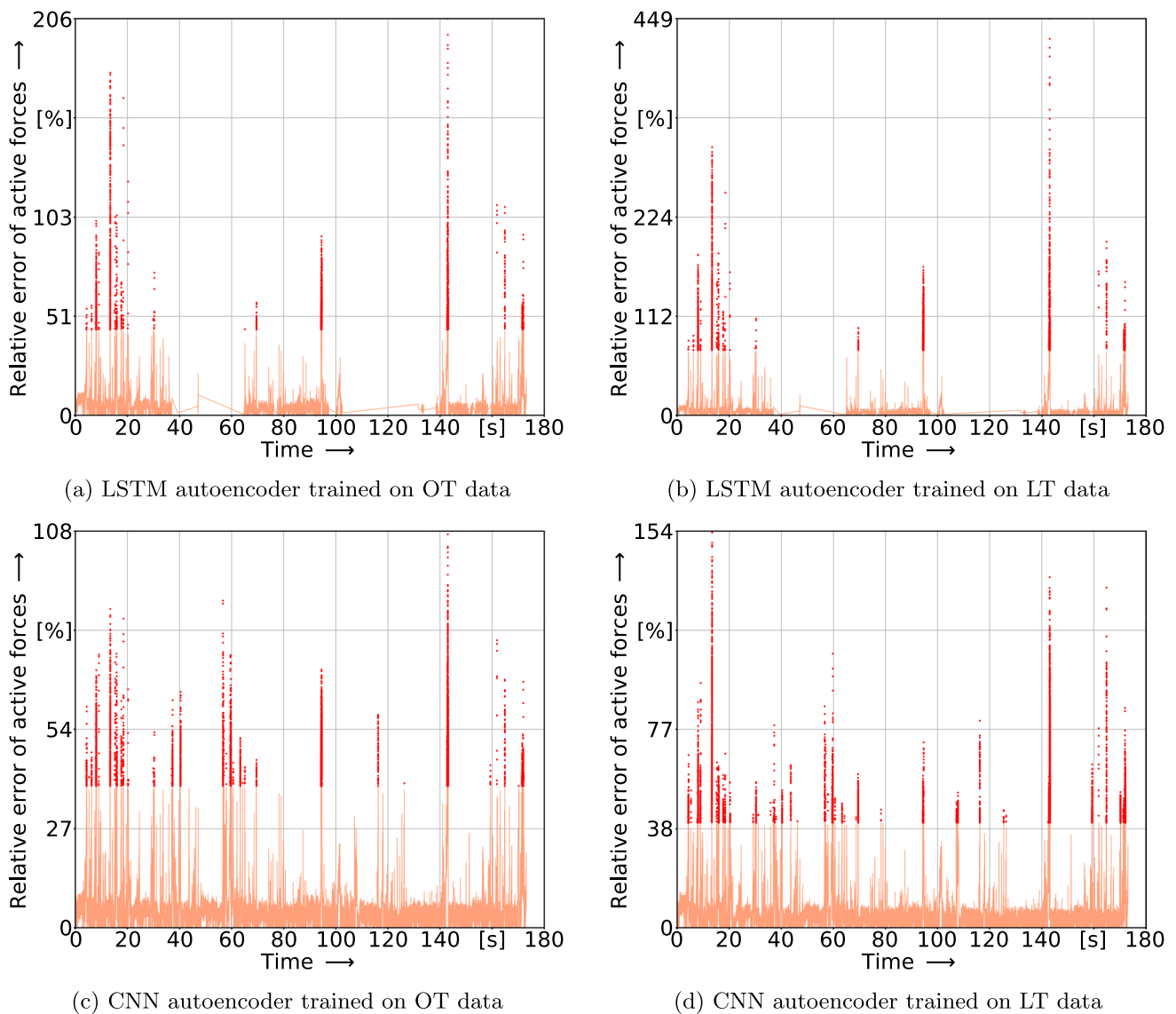


Fig. 6 Relative reconstruction errors of autoencoders on orthogonal turning data (lightsalmon); anomalies during the process (red)

The MSE reflects deviations in the standardized feature space and therefore depends on the statistical properties learned during training. In contrast, the relative reconstruction error normalizes the deviation by the magnitude of the true force vector, reducing the influence of scale differences between components and preventing a single force component from dominating the overall error. This enables a more balanced comparison of reconstruction performance across datasets.

Domain-specific effects remain clearly visible in both metrics. In particular, for the OT dataset, the influence of the training domain is consistently observable, indicating that the reduced cross-domain performance cannot be attributed solely to normalization effects, but rather to structural differences in signal dynamics and variability between OT and LT processes.

These findings highlight a fundamental limitation of reconstruction-based anomaly detection models: they implicitly assume stationarity of the underlying data distribution. When the operational domain changes, the learned representation no longer fully captures the new process characteristics. Consequently, successful industrial deployment requires training data that adequately represent the variability of the intended application domain [35, 36].

4 Conclusion

Regarding real-world implementation, the proposed method is well-suited for integration into process monitoring systems, as it relies on continuously acquired force sensor signals and performs unsupervised anomaly detection without

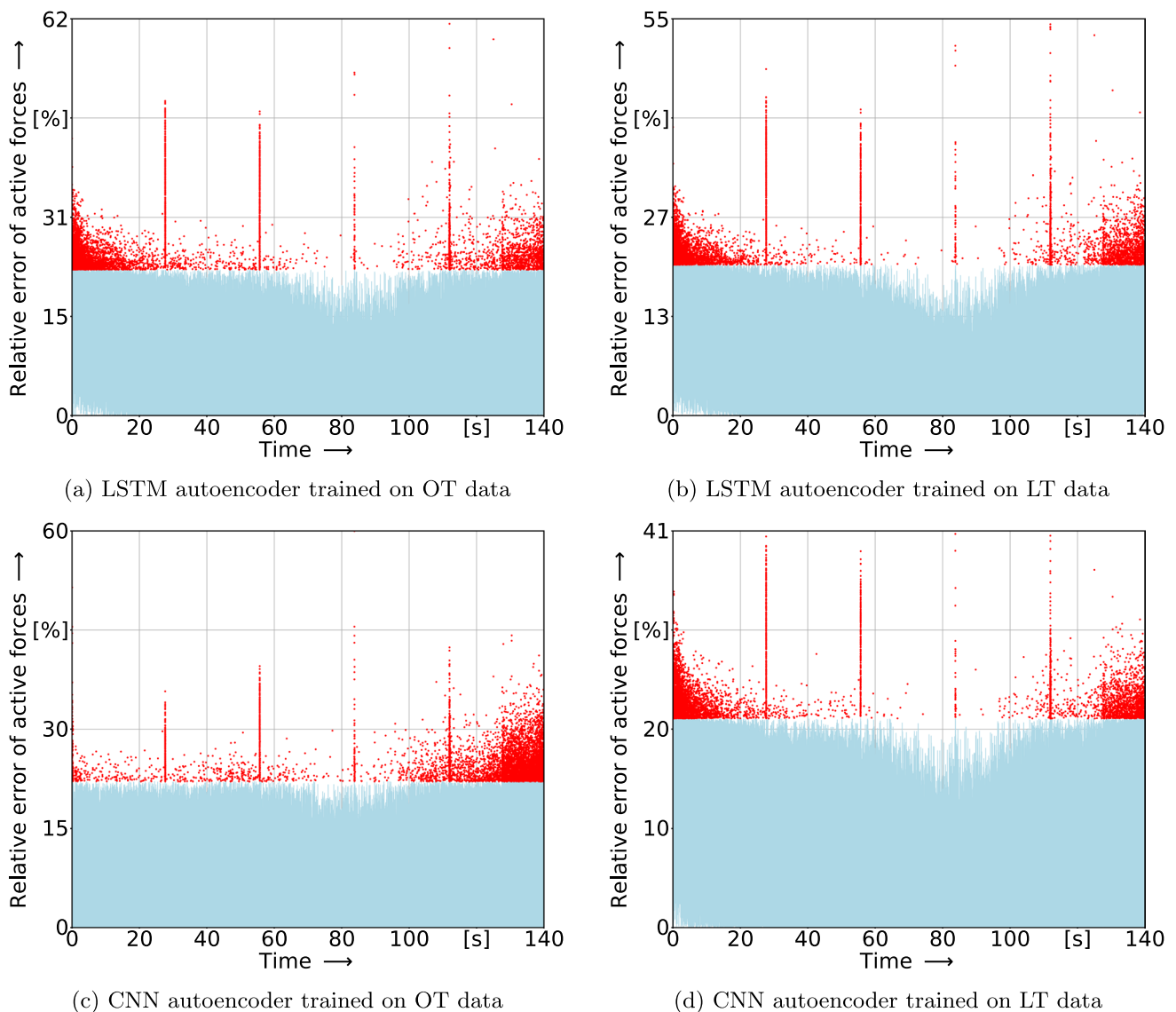


Fig. 7 Relative reconstruction errors of autoencoders on longitudinal turning data (lightblue); anomalies during the process (red)

requiring labeled fault data. The demonstrated cross-domain generalizability across different machine tools and cutting conditions indicates applicability in flexible production environments. This study systematically investigated the generalizability and robustness of LSTM- and CNN-based autoencoders for unsupervised anomaly detection in turning processes. Two fundamentally different machining domains, orthogonal turning and longitudinal turning, were analyzed using harmonized datasets to isolate the influence of machine kinematics, tool geometry, and process variability on model performance. In addition to the conventional MSE-based evaluation on normalized data, the reconstruction performance was further assessed using a relative error metric in original units to ensure a component-balanced and scale-robust comparison across domains. By training each model architecture on the two datasets, cross-domain

evaluation enabled a direct assessment of out-of-distribution generalization. The main findings of this work can be summarized as follows:

1. **Architecture sensitivity:** The LSTM autoencoders show higher sensitivity to transient signal deviations and are therefore better suited to detect subtle anomalies and process instabilities, in the normalized data. This increased sensitivity, however, also makes LSTMs more prone to false alarms in an unsupervised setting, where rare but valid signal variations may be misinterpreted as anomalies. The CNN autoencoders, in contrast, could prevent false alarms in noisy environments by smoothing out short-term fluctuations, and interpret recurring transients as normal behavior. Based on the relative reconstruction error, evaluated in the original physical

force units, the CNN models provide more stable peak reconstruction and lower overall reconstruction errors, indicating superior robustness in signal reconstruction.

2. **Influence of training data domain:** Models trained on dynamic datasets (longitudinal turning) exhibit higher reconstruction errors when applied to more stable processes (orthogonal turning). In contrast, models trained on smoother data struggle to adapt to increased variability. This behavior was consistently observed for both the MSE in normalized space and the relative reconstruction error, indicating that the reduced cross-domain performance cannot be attributed solely to scaling effects but reflects genuine structural differences between the machining domains.
3. **Generalizability:** The recurrent structure of LSTM networks enables improved generalization across machine tools and cutting geometries because temporal dependencies are preserved more effectively than with convolutional filters. In contrast, the CNN autoencoders provide stronger robustness to variability. Training on a sufficiently diverse dataset is more important than focusing on a specific model or architecture to achieve high generalizability, as needed for an industrial environment.

While the experimental setup reflects relevant variations in machining conditions, a full validation under production-scale industrial environments has not yet been performed, and is therefore identified as an important subject for future work. Nevertheless, the considered variations in machine tools, cutting conditions, and process kinematics provide a representative basis for assessing cross-domain robustness and underline the practical relevance of the presented approach. Further studies can extend the cross-domain evaluation to a larger variety of machine tools, tool types, and materials to validate model scalability under realistic production variability. Moreover, long-term stability and robustness against signal drift must be evaluated over extended production periods. Such investigations could provide insights for industrial anomaly detection as well as for other machine learning and grey-box approaches in manufacturing.

Acknowledgements The authors gratefully acknowledge funding by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for the projects “Extrapolative digital grey-box models for describing and predicting the macroscopic system behavior of TiAlN-coated cutting tools” (project no. 521385417) and “Greybox model-based prediction of wear evolution of coated tools through experimental and model-driven identification of relevant loads” (project no. 521377466) within the priority program SPP2402 “Greybox models for the qualification of coated tools for high-performance machining”.

Author Contributions Ya-Jing Wu, Justin Kopp, Jens-Peter M. Zemke, Sebastian Götschel, and Frank Walther contributed to the study con-

ception and design. Investigation, data analysis and methodology were performed by Ya-Jing Wu and Justin Kopp. Material preparation and data collection were performed by Sebastian Schibsdatt and Pascal Volke. The original draft of the manuscript was written by Ya-Jing Wu and Justin Kopp. Jens-Peter M. Zemke, Sebastian Götschel, Andreas Zabel, Jan Dege, Frank Walther, and Daniel Höche reviewed and edited the manuscript and provided supervision. All the authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) projects “Extrapolative digital grey-box models for describing and predicting the macroscopic system behavior of TiAlN-coated cutting tools” (project no. 521385417) and “Greybox model-based prediction of wear evolution of coated tools through experimental and model-driven identification of relevant loads” (project no. 521377466) within the priority program SPP2402 “Greybox models for the qualification of coated tools for high-performance machining”.

Data Availability The data used in this publication are available at <https://doi.org/10.15480/882.16923>.

Declarations

Competing interests 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Tönshoff HK, Winkler J (1997) The influence of tool coatings in machining of magnesium. *Surf Coat Technol* 94–95:610–616. [https://doi.org/10.1016/S0257-8972\(97\)00505-7](https://doi.org/10.1016/S0257-8972(97)00505-7)
2. Capello E (2005) Residual stresses in turning: Part I: Influence of process parameters. *J Mater Process Technol* 160(2):221–228. <https://doi.org/10.1016/j.jmatprotec.2004.06.012>
3. Diniz AE, Machado AR, Corrêa JG (2016) Tool wear mechanisms in the machining of steels and stainless steels. *Int J Adv Manuf Technol* 87(9):3157–3168. <https://doi.org/10.1007/s00170-016-8704-3>
4. Teti R, Jemielniak K, O’Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Ann* 59:717–739. <https://doi.org/10.1016/j.cirp.2010.05.010>
5. Stemmer S, Papenberg B, Langenhorst L, Sölter J, Meyer D, Fischer A, Tracht K, Karpuschewski B (2025) Well-informed neural network: an approach for the prediction of the width of flank wear land in turning processes. *Procedia CIRP* 133:84–89. <https://doi.org/10.1016/j.procir.2025.02.016>

6. Wolf J, Reeber T, Bandaru NK, Dienwiebel M, Möhring H (2024) Transient wear modelling of coated cutting tools. *Procedia CIRP* 130:1827–1831. <https://doi.org/10.1016/j.procir.2024.10.323>
7. Dekena B, Bergmann B, Witt M (2020) Feeling machine for process monitoring of turning hybrid solid components. *Metals* 10(7):930–931. <https://doi.org/10.3390/met10070930>
8. Thangarasu SK, Shankar S, Mohanraj T, Devendran K (2020) Tool wear prediction in hard turning of EN8 steel using cutting force and surface roughness with artificial neural network. *J Mech Eng Sci* 234:329–342. <https://doi.org/10.1177/0954406219873932>
9. Hassan M, Sadek A, Attia H (2023) A real-time deep machine learning approach for sudden tool failure prediction and prevention in machining processes. *Sensors* 23:3894. <https://doi.org/10.3390/s23083894>
10. Kim J, Lee H, Jeon JW, Kim JM, Lee HU, Kim S (2020) Stacked auto-encoder based CNC tool diagnosis using discrete wavelet transform feature extraction. *Processes* 8(4):456–457. <https://doi.org/10.3390/pr8040456>
11. Hinton GE, Zemel RS (1993) Autoencoders, minimum description length and Helmholtz free energy. In: *Proceedings of the 7th International Conference on Neural Information Processing Systems, NIPS'93*. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, pp 3–10
12. Wang H, Wang S, Sun W, Xiang J (2024) Multi-sensor signal fusion for tool wear condition monitoring using denoising transformer auto-encoder Resnet. *J Manuf Process* 124:1054–1064. <https://doi.org/10.1016/j.jmapro.2024.07.002>
13. Nam JS, Kwon WT (2022) A study on tool breakage detection during milling process using LSTM-autoencoder and Gaussian mixture model. *Int J Precis Eng Manuf* 23:667–675. <https://doi.org/10.1007/s12541-022-00647-w>
14. Schlagenhauf T, Wolf J, Puchta A (2022) Convolutional-based encoder-decoder network for time series anomaly detection during the milling of 16MnCr5. *Data* 175(12):4–8. <https://doi.org/10.3390/data7120175>
15. Sun S, Liu Y, Hu X, Zhang W (2023) A semisupervised auto-encoder-based method for anomaly detection in cutting tools. *J Manuf Process* 93:315–327. <https://doi.org/10.1016/j.jmapro.2023.03.043>
16. Yan S, Shao H, Xiao Y, Liu B, Wan J (2023) Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noise. *Robotics and Computer-Integrated Manufacturing* 79:102441. <https://doi.org/10.1016/j.rcim.2022.102441>
17. Schueller A, Saldaña C (2022) Generalizability analysis of tool condition monitoring ensemble machine learning models. *J Manuf Process* 84:1064–1075. <https://doi.org/10.1016/j.jmapro.2022.10.064>
18. Wang D, Liu Q, Wu D, Wang L (2022) Meta domain generalization for smart manufacturing: Tool wear prediction with small data. *J Manuf Syst* 62:441–449. <https://doi.org/10.1016/j.jmsy.2021.12.009>
19. Kopp J, Rümenapf F, Donnerbauer K, Saelzer J, Dias NFL, Krumme E, Volke P, Zabel A, Tillmann W, Walther F (2025) Characterization of AlCrVY(O)N coating properties and thermo-mechanical load profiles in machining AISI 304 stainless steel using black box modelling approaches. *Wear* 582–583:206346. <https://doi.org/10.1016/j.wear.2025.206346>
20. Schibsdatt S, Wu Y-J, Keunecke M, Baron S, Höche D, Zemke J-P, Götschel S, Herrmann C, Dege JH (2025) A novel data concept for cutting processes through comprehensive experimental setup enabling grey-box models. *Wear* 582–583:206348. <https://doi.org/10.1016/j.wear.2025.206348>
21. Youn J-W, Yang M-Y (2000) A study on the relationships between static/dynamic cutting force components and tool wear. *J Manuf Sci Eng* 123(2):196–205. <https://doi.org/10.1115/1.1362321>
22. Volke P, Saelzer J, Zabel A, Biermann D (2025) Influence of cutting tool coating on rake and flank face temperatures by operando pyrometric temperature measurement during orthogonal turning of AISI 316L. *Wear* 571:205848. <https://doi.org/10.1016/j.wear.2025.205848>
23. DIN - Deutsches Institut für Normung: DIN 6580:1985-10, Begriffe der Zerspantechnik; Bewegungen und Geometrie des Zerspanvorganges. Norm (1985)
24. Ansel J, Yang E, He H, Gimelshein N, Jain A, Voznesensky M, Bao B, Bell P, Berard D, Burovski E, Chauhan G, Chourdia A, Constable W, Desmaison A, DeVito Z, Ellison E, Feng W, Gong J, Gschwind M, Hirsh B, Huang S, Kalambarkar K, Kirsch L, Lazos M, Lezcano M, Liang Y, Liang J, Lu Y, Luk CK, Maher B, Pan Y, Pührsch C, Reso M, Saroufim M, Siraichi MY, Suk H, Zhang S, Suo M, Tillet P, Zhao X, Wang E, Zhou K, Zou R, Wang X, Mathews A, Wen W, Chanan G, Wu P, Chintala S (2024) Pytorch 2: Faster machine learning through dynamic Python bytecode transformation and graph compilation. In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ASPLOS '24, pp 929–947. Association for Computing Machinery. <https://doi.org/10.1145/3620665.3640366>
25. Jana D, Patil J, Herkal S, Nagarajaiah S, Duenas-Orsorio L (2022) CNN and convolutional autoencoder (CAE) based real-time sensor fault detection, localization, and correction. *Mech Syst Signal Process* 169:108723. <https://doi.org/10.1016/j.ymssp.2021.108723>
26. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jozefowicz R, Jia Y, Kaiser L, Kudlur M, Levenberg J, Mané D, Schuster M, Monga R, Moore S, Murray D, Olah C, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow. Large-scale machine learning on heterogeneous systems. <https://doi.org/10.5281/zenodo.4724125>
27. Ndubaku MU, Anjum A, Liotta A (2019) Unsupervised anomaly thresholding from reconstruction errors. In: *Montella R, Ciaramella A, Fortino G, Guerrieri A, Liotta A (eds) Internet and Distributed Computing Systems*, pp 123–129. Springer, Cham. https://doi.org/10.1007/978-3-030-34914-1_12
28. Gök K, Türkeş E, Karakaş Ö, Gök A (2023) Analyzing the effects of tool holder stiffness on chatter vibration reduction in turning. *J Metal Mater Res* 6:16–24. <https://doi.org/10.30564/jmmr.v6i1.5428>
29. Yilmaz B, Karabulut GA (2020) A review of the chip breaking methods for continuous chips in turning. *J Manuf Process* 49:50–69. <https://doi.org/10.1016/j.jmapro.2019.10.026>
30. Yahya M, Moya A, Ventura S (2025) Deep learning for multivariate time series anomaly detection: an evaluation of reconstruction-based methods. *Artif Intell Rev* 58:1–66. <https://doi.org/10.1007/s10462-025-11401-9>
31. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P-A (2019) Deep learning for time series classification: a review. *Data Min Knowl Disc* 33:917–964. <https://doi.org/10.1007/s10618-019-00619-1>
32. Terven J, Cordova-Esparza D-M, Romero-González J-A, Ramírez-Pedraza A, Chávez-Urbiola EA (2025) A comprehensive survey of loss functions and metrics in deep learning. *Int J Precis Eng Manuf* 58:195. <https://doi.org/10.1007/s10462-025-11198-7>
33. Wu W (2024) Time-series anomaly detection study based on deep learning. *Acad J Sci Technol* 13:165–172. <https://doi.org/10.54097/stbmz463>
34. Luo Z, Wang L, Xu J, Chen M, Yuan J, Tan A (2023) Flow reconstruction from sparse sensors based on reduced-order autoencoder

- state estimation. *Phys Fluids* 35(7). <https://doi.org/10.1063/5.0155039>
35. Sun H, Ammann K, Giannoulakis S, Fink O (2024) Continuous test-time domain adaptation for efficient fault detection under evolving operating conditions. *PHM Society European Conference* 8:254–264. <https://doi.org/10.36001/phme.2024.v8i1.402>
36. Farahani A, Voghoei S, Rasheed K, Arabnia H (2021) A brief review of domain adaptation. *Adv Data Sci Inf Eng*, pp 877–894. https://doi.org/10.1007/978-3-030-71704-9_65

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

Authors and Affiliations

Ya-Jing Wu^{1,5}  · Justin Kopp² · Jens-Peter M. Zemke¹ · Sebastian Götschel¹ · Pascal Volke³ · Andreas Zabel³ · Sebastian Schibsdatt⁴ · Jan Dege⁴ · Frank Walther² · Daniel Höche⁵

✉ Ya-Jing Wu
ya-jing.wu@hereon.de; ya-jing.wu@tuhh.de

Justin Kopp
justin.kopp@tu-dortmund.de

Jens-Peter M. Zemke
zemke@tuhh.de

Sebastian Götschel
sebastian.goetschel@tuhh.de

Pascal Volke
pascal.volke@tu-dortmund.de

Andreas Zabel
andreas.zabel@tu-dortmund.de

Sebastian Schibsdatt
sebastian.schibsdatt@tuhh.de

Jan Dege
jan.dege@tuhh.de

Frank Walther
frank.walther@tu-dortmund.de

Daniel Höche
daniel.hoeche@hereon.de

¹ Institute of Mathematics, Hamburg University of Technology, Am Schwarzenberg-Campus 3, Hamburg 21073, Hamburg, Germany

² Chair of Materials Test Engineering, TU Dortmund University, Baroper Str. 303, Dortmund 44227, Nordrhein-Westfalen, Germany

³ Institute of Machining Technology, TU Dortmund University, Baroper Str. 303, Dortmund 44227, Nordrhein-Westfalen, Germany

⁴ Institute of Production Management and Technology, Hamburg University of Technology, Denickestraße 15, Hamburg 21073, Hamburg, Germany

⁵ Institute of Surface Science, Helmholtz-Zentrum Hereon, Max-Planck-Straße 1, Geesthacht 21502, Schleswig-Holstein, Germany