IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 April 2025, accepted 15 May 2025, date of publication 26 May 2025, date of current version 2 June 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3573400

RESEARCH ARTICLE

Hybrid Machine Learning for CNC Process Monitoring

ROBIN STRÖBEL[1], SAMUEL DEUCKER[1], HANLIN ZHOU[1], HAFEZ KADER[2], (Graduate Student Member, IEEE), ALEXANDER PUCHTA[1], BENJAMIN NOACK[2], (Senior Member, IEEE), AND JÜRGEN FLEISCHER[1]

[1]wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany
[2]AMS–Autonomous Multisensor Systems, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany

Corresponding author: Robin Ströbel (robin.stroebel@kit.edu)

This work was supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK), based on a decision by German Bundestag via Gesellschaft zur Förderung angewandter Informatik e.V.—GFaI under Grant 22849 BG/2.

ABSTRACT The transition to highly customized, one-off production in modern manufacturing necessitates sophisticated process monitoring to reduce waste, minimize downtime, and alleviate operator burden. Computer Numerically Controlled (CNC) axes represent a fundamental component of automated manufacturing and offer a universal and accessible monitoring option through power supply data. By accurately predicting reference signals and comparing them with real-time measurements, deviations can be used for effective model-based process monitoring and anomaly detection. This study explores the efficacy of hybrid machine learning (ML) models in predicting reference signals for CNC axes using features derived from a physical model. Furthermore, relevant but difficult-to-measure features such as process forces and material removal rate (MRR) were made accessible through soft sensors. Various ML models were evaluated, including tree-based models (e.g. random forest (RF) and gradient boosting (GB)) and deep learning (DL) models (e.g. feed-forward neural networks (FNN), long short-term memory (LSTM) and transformers-based models (TF)). A feature importance analysis was performed to gain a better understanding of the influencing factors, which revealed that velocity, acceleration, process forces, spindle torque, and MRR are relevant. Tree-based models, particularly RF and GB, have been shown to be more accurate and robust than DL approaches, particularly when data is sparse and processes are complex. Although DL models improved with larger data sets, their performance remained inferior to that of tree-based methods. This study emphasizes the advantages of incorporating physical knowledge into hybrid ML models to improve model-based process monitoring.

INDEX TERMS Machine tool, CNC, process monitoring, signal prediction, machine learning.

I. INTRODUCTION

The miniaturisation of sensors and enhanced data-processing capabilities enabled companies to collect and analyse production data with higher efficiency. At the same time, the growing demand for customised products is driving the need for greater flexibility in production [1], resulting in a shift towards one-off production. This trend also affects individual aspects of production, such as process monitoring. It must become more flexible and adaptable to produce customised products with high quality and low cost.

Computer numerical control (CNC) milling machines are essential for modern manufacturing, playing a pivotal role in the production of complex components with high precision. The functionality of machine tools is contingent upon the accurate and precise control of their CNC axes. Consequently, the power supply is a valuable source of data for the monitoring of machine tool operations providing real-time insight into the system's behaviour. This provides an accessible and cost-effective option for process monitoring.

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits.

Conventional process monitoring methods are often based on statistical or physical approaches [2]. Statistical methods rely on historical data [3], while physical methods are closely tied to fundamental scientific principles [4], [5]. They exploit physical relationships and reliably describe machine behaviour within a defined range. Outside this range, the underlying equations are no longer able to provide reliable predictions. In highly dynamic environments where system states can quickly move outside this range due to changing conditions such as variations in workpieces, materials, or tools, these methods reach their limits [6].

Due to promising progress in the field of ML, more and more data-driven approaches are applied in production engineering [7]. These have the advantage of learning from existing data and identifying patterns. In addition, ML is capable of describing highly complex relationships, particularly within high-dimensional data structures. However, ML models are unable to represent states that are not present in the training data. The integration of physical knowledge into ML has shown promising results in real-world applications [8], particularly in anomaly detection [9], [10], [11].

As both physical and ML-based approaches have been beneficial, there is a need to investigate the performance of hybrid model structures. To address this, a physical model for CNC axes and main spindle power supply is derived and coupled via perturbation theory. The features have been integrated into hybrid ML models, including RF, GB, LSTM, TF, and FNN. The evaluation is based on their ability to predict reference signals, with a particular focus on a comparison of their strengths and weaknesses. The paper is structured as follows: Section II presents a review of the state-of-the-art in CNC milling process monitoring, with a focus on both physical and data-driven approaches. Section III outlines the methodology, detailing the physical and hybrid ML models used in the study. Section IV presents the experimental setup, followed by the results in Section V. Section VI discusses the key findings, and Section VII concludes with a summary of insights and directions for future research.

II. STATE OF THE ART

Machining processes are influenced by various physical parameters, which, during the production, reflect a defined machine state. By observing this state, the machining process can be monitored; however, the question arises regarding how this machine state can be represented based on existing knowledge. While sensors can record the physical parameters, the relationships between these parameters and the actual process state remain initially unknown. The present study explores various approaches to bridge this gap between recorded signals and the actual process state [2], [10]. The following section presents and evaluates methods from recent research aimed at addressing this challenge.

A. PHYSICS-BASED APPROACHES

A direct approach to process monitoring is given by observing and evaluating key influencing process factors and their subsequent evaluation. Physics-based methods employ well-established physical laws to model the ideal process state via these factors. One of the critical process parameters is represented by process force. During machining, the force vector changes continuously [12]. Changes in process force have proven to be a key parameter for describing the production process [13], [14]. It has already been demonstrated that these approaches are effective for tool wear monitoring. Therefore, [15] is able to develop a monitoring system based on internal machine signals. A force model was developed using spindle current due to correlations, which is used to define a process quality indicator. This physics-based indicator can be continuously monitored, thereby enabling the tracking of process quality during ongoing machining operations. Another force-based methodology is proposed by [4], where the calculated cutting force vector serves as a reference for monitoring. By comparing the actual force vector with the reference value calculated, it is possible to identify the tool wear. However, it is imperative to conduct predefined calibration tests to ascertain the coefficients of the physical equations.

A model-based approach for milling process monitoring, developed by [16], uses CAD data and the NC program to define the ideal process state. Process forces are determined through a force model, forming a reference for the machine's power values. In practice, these reference values are compared with the machine's current performance data for online tool wear monitoring.

Another way to exploit the physical relationships between process parameters and process states is by evaluating specific correlations between measured values and key state variables. For example, process forces can be indirectly monitored by observing the spindle motor current [17]. Such correlations can also be used to monitor tool conditions [18] introduced a method for calculating the "*Current Rise Index*". Similar to [15], current signals are monitored, and their changes are carefully analysed.

In summary, physics-based methods play a crucial role in process monitoring, whereby an ideal process state can be determined using physical models or by evaluating correlations between measured variables and target variables. However, these methods exhibit certain weaknesses in practical applications. Firstly, physical models must always be calibrated through experimental trials before use, and similarly, methods based on correlations between variables require calibration through reference measurements. Additionally, the complexity of the mapping is limited by the assumptions made during the modelling. A major advantage of such methods is their straightforward interpretability and their data efficiency. Abnormal process states often have clearly defined causes that can be directly recognised by experts during evaluation. The relationship between process

parameters and process state allows causes to be reliably inferred through experience or intuition.

B. DATA-DRIVEN APPROACHES

In contrast to physical-based approaches, data-driven methods do not rely on physical equations as their foundational principles. Rather, relationships between measured parameters are determined by analysing historical data. Previous publications have demonstrated the utilisation of statistical methodologies for process monitoring [3]. Data analysis can be performed either through simple statistical evaluations [19] or using ML-based techniques [20], [21], [22].

Jennings and Drake present a statistically based method for continuous process monitoring. During the machining process, various process-dependent and measurable parameters are recorded on a control chart [19]. The trend of these parameters directly reflects the state of the process. Sudden changes that statistically deviate significantly from the ideal process value indicate a potential error. The increasing use of artificial intelligence and the ongoing miniaturization of sensors have led to a growing adoption of ML-based methods in production environments. Compared to traditional approaches, ML models can process significantly larger amounts of data and identify critical patterns. This trend can also be observed in the field of process monitoring [10]. There are already numerous applications where ML models are successfully employed.

Various studies also utilize vibration signals [20] and acoustic signals [21] for tool condition monitoring. The recorded data are analysed using DL or ML-based methods. Surface anomalies can also be successfully identified through signal analysis [22].

One example [23] is based on spindle current signals and (DL) models. During the training phase, the network is trained on data from error-free production. This allows the model to learn the normal behaviour patterns of the machine. Conversely, the model is sensitized to detect abnormal machine states. A significant deviation of the measured data from the training data indicates potential faults.

Another approach [17], [24] demonstrates the potential of ML methods for the generation of reference values. Therefore, the production process is divided into several elemental subprocesses. A simulation provides the corresponding material removal and process forces for each subprocess, modelling the ideal state of the subprocess. Based on the correlation between material removal and spindle current, this study successfully used a linear regression model to estimate spindle current as a reference for process monitoring.

Data-driven approaches have been shown to offer certain advantages over physics-based methods, particularly in situations where the system under investigation cannot yet be fully modelled physically. ML methods also excel in handling complex relationships. Furthermore, ML models are capable of efficiently identifying patterns in high-dimensional data without relying on physical fundamental equations [25]. However, it is imperative that a data set that is both sufficient and representative is available.

C. RESEARCH QUESTIONS

The power consumption of a machine tool and its axis has been shown to correlate with both long-term effects, such as tool wear, and short-term process anomalies. Consequently, it can be considered a pivotal indicator of the process condition. The use of both physical and ML models is imperative for the implementation of robust process monitoring systems [8]. To develop a flexible process monitoring system based on power signals of individual axes requires investigation of the optimal prediction model design. Therefore, the following research questions are addressed in this study:

1) How to model axis and main spindle power supply using ML and physical models, and which features are essential for a hybrid ML model?
2) How well can different ML model architectures learn and generate meaningful predictions from the derived physical features?
3) How do the models perform under realistic operating conditions?

III. METHODOLOGY

The objective of the present article is to develop a hybrid ML model that predicts the power absorbed by the axes of a milling machine. The high mapping complexity of ML models should be supported by known physical relationships. Relevant features are derived from a physical model. In instances where these features cannot be measured by conventional internal sensors and are thus unobservable, the integration of the underlying physical knowledge is facilitated by soft sensors. Therefore, physics-based knowledge is integrated via features selection [26], [27] and simulation of unobservable but relevant data [8], [28]. The proposed approach is structured around two main components: The development of a physical model and the integration of the derived features into ML model structures.

A. PHYSICAL MODEL DEVELOPMENT

As presented in Section II, the power and current supplied to a CNC axis are frequently used for process monitoring purposes. Due to the significant academic interest in machine tool power consumption [29], [30], [31], numerous physical models have been proposed. Similar to process monitoring, recent research has focused on ML-based prediction approaches [32], [33]. However, due to the predominant emphasis on total energy consumption, there is a paucity of prediction models for energy consumed on an axis level. As demonstrated in [34], the current supply of a CNC axis can be derived from a dynamic system description. In order to further enhance the quality of the physical description, analogous to [35], the milling machine and its axis are

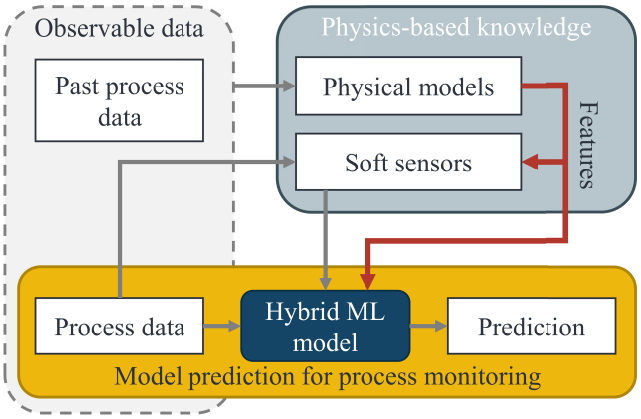


FIGURE 1. Interaction between physics-based knowledge and the hybrid ML model (style analogous to [8]).

considered as a thermodynamic system. Mass flows, such as those of removed chips or adhering coolant, are assumed to be much smaller than the mass of the entire system. Consequently, these effects are neglected, and an axis system can be described using the first law of thermodynamics. The change in internal energy can be expressed by the following equation:

$$\frac{d}{dt}U = \sum \frac{d}{dt}W + \sum \frac{d}{dt}Q \tag{1}$$

where $\frac{d}{dt}U$ represents the change of internal energy of an axis system, $\frac{d}{dt}W$ represents the work, and $\frac{d}{dt}Q$ the heat flows interacting with the system.

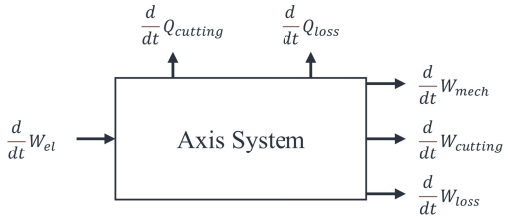


FIGURE 2. Heat and energy flows interacting with an isolated axis system.

As shown in Fig. 2, heat and energy flows result from:

- $\frac{d}{dt}W_{\text{el}}$: Temporal change in electrical energy. It is introduced to the system in the form of electrical power

$$P = \frac{d}{dt}W_{\text{elect}} = u(t) \cdot i(t).$$

- $\frac{d}{dt}W_{\text{cutting}}$: Corresponds to the cutting work. Empirical studies [36], [37], [38] show a dependence of the cutting work on the material removal rate *MRR* and the cutting force F. For simplification,

$$\frac{d}{dt}W_{\text{cutting}} \approx c \cdot F \cdot \text{MRR}$$

is assumed.

- $\frac{d}{dt}W_{\text{mech}}$: Mechanical work resulting from axis movement.
- $\frac{d}{dt}W_{\text{loss}}$: Losses resulting from axis movement.
- $\frac{d}{dt}Q_{\text{cutting}}$: Heat leaving the axis system as a result of the milling process. According to [39], the cutting heat depends on the processing parameters such as the feed rate, the spindle rotation, the material and the tool, which are assumed to be constant.
- $\frac{d}{dt}Q_{\text{loss}}$: Represents heat losses in motors, electrical losses, and other heat losses. These are also assumed to be constant.

Therefore, (1) results in

$$\frac{d}{dt}U = -\frac{d}{dt}Q_{\text{loss}} - \frac{d}{dt}Q_{\text{cutting}} - \frac{d}{dt}W_{\text{cutting}} - \frac{d}{dt}W_{\text{mech}} - \frac{d}{dt}W_{\text{loss}} + \frac{d}{dt}W_{\text{el}},$$

for heat and energy flow of an isolated axis system. If the internal energy is set to constant,

$$\frac{d}{dt}U = 0$$

follows. Assuming linear motion, the mechanical work

$$W_{\text{mech}} = \int F\, ds$$

can be simplified to

$$\frac{d}{dt}W_{\text{mech}} = F \cdot \dot{x}.$$

Due to the equation of motion of structural dynamics, force and torque can be defined as

$$F = m\ddot{x} + k\dot{x} + cx,$$

and

$$M = J\ddot{\varphi} + k_d\dot{\varphi} + c_d\varphi.$$

For a linear motion, using the constants c_1 and c_2, we obtain

$$\frac{d}{dt}W_{\text{mech}} = c_1 \cdot \dot{x} \cdot \ddot{x} + c_2 \cdot \dot{x}^2.$$

The same applies to rotation with the use of

$$\frac{d}{dt}W_{\text{mech}} = M \cdot \dot{\varphi},$$

resulting in

$$\frac{d}{dt}W_{\text{mech}} = c_1 \cdot \dot{\varphi} \cdot \ddot{\varphi} + c_2 \cdot \dot{\varphi}^2.$$

If voltage is assumed to be constant and the sign function is used to model the direction of current flow

$$I_n \approx c_1 \cdot \dot{x}_n \cdot \ddot{x}_n + c_2 \cdot \dot{x}_n^2 \cdot \text{sgn}(\dot{x}_n) + c_3 \cdot F_n \cdot MRR + c_4 \cdot \text{sgn}(\dot{x}_n) + c_5, \quad \textit{where } n \in \{x, y, z\} \textit{ and } c_1, c_2, c_3, c_4, c_5 \in \mathbb{R}, \tag{2}$$

follows. Thus, I_n represents the current flow of axis n, with $\dot{x}_n$ as the velocity and $\ddot{x}_n$ as the acceleration. F_n is the force in the direction of axis n, and MRR represents the material

removal rate. Similarly, the main spindle current supply can be described by

$$\begin{aligned} I_{\text{sp}} \approx\ & c_1 \cdot \dot{\varphi} \cdot \ddot{\varphi} + c_2 \cdot \dot{\varphi}^2 \cdot \text{sgn}(\dot{\varphi}) \\ & + c_3 \cdot M_{\text{sp}} \cdot MRR + c_4 \cdot \text{sgn}(\dot{\varphi}) + c_5, \\ & where \quad c_1, c_2, c_3, c_4, c_5 \in \mathbb{R}. \end{aligned} \tag{3}$$

B. COUPLING VIA PERTURBATION THEORY

In section III-A, the axes of the milling machine were modelled as decoupled systems. However, in reality, coupling through the milling tool, suboptimal axis alignment, and other effects cause the axes to interact. To represent this mutual influence of the different axes and the spindle, the axis systems are coupled via perturbation theory to find an approximate solution.

In the example of the X-axis, according to

$$I_x \approx I_{x,0} + \varepsilon_y^1 \cdot I_{y,0} + \varepsilon_z^1 \cdot I_{z,0} + \varepsilon_{sp}^1 \cdot I_{sp,0} + \mathcal{O}(\varepsilon^2)$$

where $0 < \varepsilon \ll 1$, the approximate solution is given by

$$\begin{aligned} I_x \approx\ & c_1 \cdot \ddot{x}_x \cdot \dot{x}_x + c_2 \cdot \dot{x}_x^2 \cdot sgn(\dot{x}_x) + c_3 \cdot F_x \cdot MRR \\ & + c_4 \cdot sgn(\dot{x}_x) + c_5 \\ & + c_6 \cdot \ddot{x}_y \cdot \dot{x}_y + c_7 \cdot \dot{x}_y^2 \cdot sgn(\dot{x}_y) + c_8 \cdot F_y \cdot MRR \\ & + c_9 \cdot sgn(\dot{x}_y) \\ & + c_{10} \cdot \ddot{x}_z \cdot \dot{x}_z + c_{11} \cdot \dot{x}_z^2 \cdot sgn(\dot{x}_z) + c_{12} \cdot F_z \cdot MRR \\ & + c_{13} \cdot sgn(\dot{x}_z) \\ & + c_{14} \cdot \ddot{\varphi} \cdot \dot{\varphi} + c_{15} \cdot \dot{\varphi}^2 \cdot sgn(\dot{\varphi}) + c_{16} \cdot M_{sp} \cdot MRR \\ & + c_{17} \cdot sgn(\dot{\varphi}). \end{aligned} \tag{4}$$

The same can be done for the other axes (y, z) and the main spindle (sp).

C. FEATURE SELECTION AND MODEL STRUCTURE

In the case of 3-axis milling, the hybrid ML model structure can be described by axis- and spindle-specific ML models

$$Out_n = M_n(In) = I_n, \ \text{where } n \in \{x, y, z, sp\} \tag{5}$$

which map the input vector

$$\text{In} = \left[v_x, v_y, v_z, a_x, a_y, a_z, \dot{\varphi}, \ddot{\varphi}, F_x, F_y, F_z, M_{sp}, MRR\right]^T \tag{6}$$

to the respective target I_x, I_y, I_z, and I_{sp}. The structure shown in figure (3) allows mapping of complex relationships.

Since the model should represent the ideal process state, it must be represented by the process force and the MRR. Since these known relationships can only be learned from the data with a great deal of effort, physical knowledge should be infused. Therefore, a soft sensor is defined as a software-based method that calculates the process forces in all three spatial directions (x, y and z) as well as the torque in the spindle's rotational direction. The calculation is based on the Kienzle force model, which determines the active forces during the machining process in the cutting and normal cutting directions using known material and process

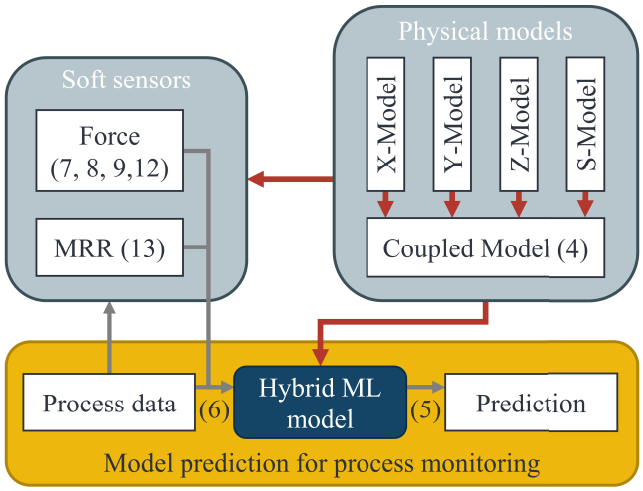


FIGURE 3. Hybrid ML-model structure with formulas for the infusion via physical models and soft sensors for a given axis.

parameters [40], [41]. Using a coordinate transformation, the calculated forces in the cutting direction can then be converted into the tool's movement direction [42].

If tool movement remains within the x, y plane during the machining process, the active forces also lie within this plane. Following [40], [41], forces per tooth in the cutting and normal cutting directions are calculated using

$$F_{cz} = b \cdot h_m^{1-z} \cdot k_{c1.1} \cdot K \tag{7}$$

and

$$F_{cnz} = b \cdot h_m^{1-x} \cdot k_{f1.1} \cdot K. \tag{8}$$

In x-y plane machining processes F_z corresponds the passive force F_{pz}. It can be calculated with the use of formula [40], [41]

$$F_{pz} = b \cdot h_m^{1-y} \cdot k_{p1.1} \cdot K. \tag{9}$$

The parameters b and h_m describe the chip thickness and mean width, respectively. They depend on the cutting depth a_p and width a_e cutting arc angle φ_s, setting angle κ, cutter diameter D, and feed per tooth f_z.

The parameters $k_{p1.1}$, $k_{c1.1}$, $k_{f1.1}$, x, y and z are martial specific tabulated values. b and h_m can be calculated using [40], [41]

$$b = \frac{a_p}{\sin \kappa} \tag{10}$$

and

$$h_m = \frac{114,6}{\varphi_s} \cdot f_z \cdot \frac{a_e}{D} \cdot \sin \kappa. \tag{11}$$

The process forces in the cutting direction can be determined using Equations (7) to (11). However, for the input parameters of the hybrid ML models (6), the process forces in the x, y and z directions of the machine coordinate

system are required. Therefore, a coordinate transformation around the z-axis

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = A \cdot \begin{bmatrix} F_c \\ F_{cn} \\ F_p \end{bmatrix} \tag{12}$$

with

$$A = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

must be performed. θ describes the directional angle of the cutter's velocity vector within the machine coordinate system.

In addition, the MRR is determined using a voxel-based method. Therefore, the workpiece is divided into individual cubes with edge length l. The relative position of the tool and workpiece is determined using control system data. When the tool with its given geometry enters the workpiece, the overlapping regions between are considered as removed volume [43], [44]. By tracking the tool's position relative to the workpiece, the removed volume can be determined throughout the production process. From the voxel-based material removal simulation, the number n of removed voxels between two measurement time points can be determined. By multiplying this with the volume of a single voxel, the total removed volume per second can be calculated by formula (13). For this, the volumes at the measurement time points t_1 and t_2 as well as the sampling rate f must be defined.

$$MRR = (V_{t_2} - V_{t_1}) \cdot f \tag{13}$$

IV. EXPERIMENTAL SETUP

A. MACHINES AND DATA SETS

The data sets were recorded with a Siemens Industrial Edge on a DMG MORI CMX 600 V milling machine (C, [45]) and a retrofitted DECKEL MAHO DMC 60 H milling machine (D, [46] and [47]). All relevant parameters are described below.

References [45] and [46] are equal data sets for general model evaluation. The first component (C_1) represents a complex part milled with a 20 mm, 10 mm, and 5 mm milling cutter. The cutting depth a_p varies from 10 to 5 mm, and cutting width a_e varies from 20 to 2.5 mm. For S235 JR the feedrate F varies from 256 to 635mm/min, and the spindle rotation speed S varies from 800 to 6350rpm. For AL 2007 T4 F varies from 140 to 304mm/min, and the S varies from 900 to 7600rpm. The second component (C_2) was milled with a 10 mm tool with an a_p from 5 to 3mm and a_e from 10 to 5mm. In case of S235 JR F varies is from 360 to 576mm/min, and S from 2000 to 3200rpm. In case of AL 2007 T4 F varies is from 212 to 350mm/min, and S from 2300 to 3800rpm. All components were produced without a workpiece as an aircut ($A1$, $A2$).

Reference [47] represents three parts (Gear G, Notch N, Plate P) milled with a 10mm end mill and a default a_p of 6mm and a_e of 10mm. In case of S235 JR, F is 576mm/min and S is 3200rpm, whereas in case of AL 2007 T4, F is 350mm/min and S is 3800rpm. Furthermore, in $DAL/DST_{Ga_p/Pa_p}$ a_p ranges from 3 to 12mm. In $DAL_{PSF/GSF}$ F varies from 267 to 442mm/min and S from 2900 to 4800rpm. For S235 JR F varies from 432 to 720mm/min and S from 2400 to 4000rpm. All data sets extracted for use in this publication are listed in table (1).

TABLE 1. Data sets extracted from [45], [46] and [47].

Machine	Material	Name of the data sets
CMX 600V [45]	AL 2007 T4	CAL_{C_1}, CAL_{C_2}, CAL_{A_1}, CAL_{A_2}
	S235 JR	CST_{C_1}, CST_{C_2}, CST_{A_1}, CST_{A_2}
DMC 60H [46]	AL 2007 T4	DAL_{C_1}, DAL_{C_2}, DAL_{A_1}, DAL_{A_2}
	S235 JR	DST_{C_1}, DST_{C_2}, DST_{A_1}, DST_{A_2}
DMC 60H [47]	AL 2007 T4	DAL_G, DAL_N, DAL_P, DAL_{GSF}, DAL_{PSF}, DAL_{Ga_p}, DAL_{Pa_p}
	S235 JR	DST_G, DST_N, DST_P, DST_{GSF}, DST_{PSF}, DST_{Ga_p}, DST_{Pa_p}

Data processing was done in Python, where all JSON files were converted into the `Pandas.DataFrame` format. The sampling rate was reduced from 500 Hz to 50 Hz using `scipy.decimate` to facilitate more efficient processing and to minimize noise. In absence of missing values, the necessity for interpolation was negated. Subsequently, the input channels of $v_x, v_y, v_z, a_x, a_y, a_z, \dot{\varphi}, \ddot{\varphi}$ were given. F_x, F_y, F_z, M_{sp} and MRR are calculated via formula (7) - (13) with $k_{c1.1} = 1990N/mm^2$, $k_{f1.1} = 351N/mm^2$, $k_{p1.1} = 247N/mm^2$ for S235 JR and $k_{c1.1} = 472N/mm^2$, $k_{f1.1} = 20N/mm^2$), $k_{p1.1} = 32N/mm^2$ for AL 2007 T4.

A force measurement plate was installed in [47], allowing a comparison between the values calculated by soft sensors and the actual force values. To ensure that the calculated values closely match real conditions, a material- and machine-specific scaling factor

$$F_{adjusted} = \frac{F_{soft}}{\alpha} \tag{14}$$

is introduced. It is determined through a least-squares optimization applied to the measured and calculated time series. As a result, $\alpha_{Alu} = 5,42$ and $\alpha_{Steel} = 2,78$ are given for the DMC data sets (see [47]).

B. EXPERIMENTAL SERIES 1

The objective of experimental series 1 is to examine the predictive capabilities of various machine learning (ML) model architectures. To this end, experiments are conducted in a consistent manner with data sets on CMX 600V (Experiment 1.1, 1.2 and 1.3) and DMC 60H (Experiment 1.4, 1.5 and 1.6). In order to reduce complexity and understand general behaviour, air cut data is used in experiment 1.1 and 1.4, as the process forces and the material removal rate can be neglected. Consequently, the fundamental suitability of the hybrid approach is analysed by training on component one and tested on component two. The same components are utilised in experiments 1.2 and 1.4 for machining aluminium. It is

TABLE 2. Hyper/Params for the ML model architectures evaluated. A grid search was performed to identify the optimal configurations.

Architecture	Hyper/Params
RF / ET	Estimators: 10, 30, 60, 100 Max Depth: 10, 30, 60, 100 Leaves: 5, 10, 20, 40 Nodes: 1000, 3000, 6000, 10000
GB	Estimators: 10, 30, 60, 100 Max Leaves: 1000, 3000, 6000, 10000
FNN / LSTM	Epochs: 50 Optimizer: Adam Learning Rate: 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2} Layers: 1, 2, 3, 4 Batch Size: 64, 128, 256 Dropout: 0.1, 0.2, 0.3, 0.4 Hidden Units: 32, 64, 128, 256
Temporal Fusion TF	Learning Rate: 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2} Hidden Size: 32, 64, 128, 256 Attention Heads: 2, 3, 4, 5 Dropout: 0.1, 0.2, 0.3, 0.4 Continuous Size: 2, 4, 8, 16

TABLE 3. Setup and data sets of experimental series 1.

Nr.	Training	Test
1.1	CAL_{A_1}	CAL_{A_2}
1.2	CAL_{C_1}	CAL_{C_2}
1.3	CAL_{C_1}, CST_{C_1}	CAL_{C_2}
1.4	DAL_{A_1}	DAL_{A_2}
1.5	DAL_{C_1}	DAL_{C_2}
1.6	DAL_{C_1}, DST_{C_1}	DAL_{C_2}

acknowledged that the complexity of the process increases significantly due to the presence of process forces and the removal of material. In experiment 1.3 and 1.6, the models are trained on component one, which is a combination of steel and aluminium, and tested on component two, which is solely aluminium. The objective is to assess the performance of the machine learning (ML) model in discriminating between the two materials. This step is necessary in order to consider the approach under approximate real-life conditions.

The following ML models are investigated: RF, extra tree (ET), GB, FNN, LSTMs and TF architectures. To optimize the models on the underlying data sets, a grid search was carried out for each architecture, whereby the relevant hyper-parameters were varied. The hyperparameter spaces are shown in table (2).

C. EXPERIMENTAL SERIES 2

The focus of the second series is the investigation of specific scenarios with increasing complexity. Therefore, the components gear, notch and plate introduce more complex tool paths. Additionally, by $DAL_{Ga_p/GSF/Pa_p/PSF}$ and $DST_{Ga_p/GSF/Pa_p/PSF}$, the models can be confronted with yet unknown technological parameters and thus process states. This enlarged data basis allows a validation of the knowledge gained in the first series with increased complexity over the trials. The experiments are divided into five categories that highlight different aspects of material discrimination, extrapolation and learning from unknown process states. Experiment 2.1 and 2.2 are based on the previous experiments with the goal of material differentiation. The aim is to confirm the material differentiation based on a more diversified training foundation. Therefore, the notch components are not included in the training and used for testing. This is followed by the investigation of increasing complexity by validating the components plate and gear in experiments 2.3 and 2.4, which are therefore not included in the training. The next category deals with extrapolation. Trials 2.5 to 2.8 test whether the model is capable of reliably unknown process states. Both aluminium and steel components are included to enable a comprehensive analysis. In experiments 2.9 and 2.10, the ability of the ML models to extrapolate to unknown components and process conditions at the same time. For this purpose, a validation component that was not included in the training data is used to assess the robustness of the models in unknown scenarios. Finally, the last two experiments focus on learning from unknown states. Therefore, $DAL_{Ga_p,GSF}$ are included in the training data set. The aim is to use the findings from the first series of experiments to achieve more precise and realistic modelling. This structured approach allows for an analysis of the effects of the expanded data base and the integration of unknown states and components into the training process.

TABLE 4. Setup and data sets of experimental series 2.

Nr.	Training	Test
2.1	DAL_{C_1}, DAL_{C_2}, DAL_G, DAL_P	DAL_N
2.2	DST_{C_1}, DST_{C_2}, DST_G, DST_P	DST_N
2.3	DAL_{C_1}, DAL_{C_2}, DAL_G, DAL_N	DAL_P
2.4	DAL_{C_1}, DAL_{C_2}, DAL_P, DAL_N	DAL_G
2.5	DAL_{C_1}, DAL_{C_2}, DAL_G, DAL_P, DAL_N	DAL_{Pa_p}
2.6	DAL_{C_1}, DAL_{C_2}, DAL_G, DAL_P, DAL_N	DAL_{PSF}
2.7	DST_{C_1}, DST_{C_2}, DST_G, DST_P, DST_N	DST_{Pa_p}
2.8	DST_{C_1}, DST_{C_2}, DST_G, DST_P, DST_N	DST_{PSF}
2.9	DAL_{C_1}, DAL_{C_2}, DAL_G, DST_N	DAL_{PA_p}
2.10	DAL_{C_1}, DAL_{C_2}, DAL_G, DST_N	DAL_{SF}
2.11	DAL_{C_1}, DAL_{C_2}, DST_P, DST_N, DAL_G, DAL_{Ga_p}, DAL_{GSF}	DAL_{Pa_p}
2.12	DAL_{C_1}, DAL_{C_2}, DST_P, DST_N, DAL_G, DAL_{Ga_p}, DAL_{GSF}	DAL_{PSF}

Based on the knowledge gained in experiment one, RF, GB and the NN architectures are investigated based on the respective hyper-parameters from table (2). In addition to the main spindle, the X- and Z-axes are representatively examined because of the performance correlation between the translatory axes.

D. EVALUATION METRICS

Mean squared error (MSE) is a common metric used to assess the performance of regression models. Calculated by the average of the squared differences between predicted

values ($\hat{y}_i$) and measured values (y_i)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{15}$$

it penalises larger errors more heavily due to the squaring of residuals, making it sensitive to outliers. A lower MSE indicates better model performance.

Mean absolute error (MAE) is another commonly used metric that calculates the average of the absolute differences between predicted values ($\hat{y}_i$) and actual values (y_i):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{16}$$

Unlike MSE, MAE does not penalize larger errors as heavily, making it less sensitive to outliers. It provides a more intuitive measure of the average error magnitude. A lower MAE indicates better predictive performance.

R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{17}$$

where $\bar{y}$ is the mean of the actual values. R^2 indicates how well the model explains the variation in the data. An R^2 closer to 1 indicates that the model explains most of the variance, while a value closer to 0 indicates poor performance.

When comparing models with different numbers of predictors, **adjusted** R^2 is a more accurate metric. Unlike R^2, which can increase with more predictors, R^2_{adj} accounts for the number of predictors and adjusts for the model complexity. It is calculated as

$$R^2_{adj} = 1 - \left(\frac{(1 - R^2)(n-1)}{n-p-1}\right) \tag{18}$$

where n is the number of data points and p is the number of predictors. R^2_{adj} helps to avoid overfitting by penalizing models with too many variables. A higher R^2_{adj} value indicates a model that fits the data well while avoiding unnecessary complexity.

This study uses MSE, MAE, R^2, and R^2_{adj} to evaluate physical and hybrid ML model performance in predicting reference signals for CNC milling. While MSE and highlights error magnitude, MAE offers a simpler, more interpretable view of model performance, R^2 measures the proportion of variance explained by the model, and R^2_{adj} provides insight into model efficiency by accounting for complexity.

V. RESULTS

A. EXPERIMENTAL SERIES 1

It is generally noticeable that the GB's average R^2 is in the range of 0.4 to 0.7 on the CMX for all translatory axis, depending on the respective test case (see Table 5-7). There are no negative R^2 values except for the spindle, which indicates stability during training. GB often delivers better results than the other models tested, particularly for the spindle. RF and ET show similar behaviour and deliver robust results. The neural network occasionally achieves results well below the level of the tree-based methods. Negative R^2 values or high errors are frequently observed in more complex scenarios. The LSTM model often shows strongly decreasing performance values, often even with negative R^2 values and very high mean square errors. This could be due to an insufficient amount of pure time series information. TF models are difficult to train successfully within the framework of classical time series regression unless large amounts of data are available.

However, looking at the metrics of the models on the DMC, it is noticeable that the statistical key figures are significantly better or more robust across all R^2 metrics. The tree-based methods improve significantly, especially with regard to the spindle. Here, the R^2 of GB increases from almost zero to 0.73. A similar behaviour can also be observed for the translational axes. Here, the tree-based methods are in a range from 0.8 to 0.97.

Experiment 1, 2 and 6 for each translatory axis and the spindle are analysed in more detail here as representative of the first series of experiments. As can be seen from the results of experiment 1, good metrics are achieved for all translatory axes of the DMC machine. This is reflected in an R^2 of 0.90, which means that the model for the X-axis is able to explain 90 percent of the variability contained in the data, and therefore the model fits the data very well. The situation is even better for the Z-axis, where an R^2 of 0.96 is achieved, which can be explained by the fact that less variability is contained in the data here due to the process. It is, therefore, easy for the model to learn this. The Y-axis looks somewhat worse but still solid. Here R^2 deteriorates to 0.88. Metrics such as MSE and MAE follow this underlying pattern. The spindle performs worst with an R^2 value of 0.63, which can be explained by the fact that the spindle has the highest variability in its data. In experiment 2, the complexity increases due to processing. This leads to an increase in the variability contained in the data, which is ultimately reflected in a slight decrease in R^2 and on the other matrices. The largest delta can be seen in the Y-axis, where the metric slips to 0.81. On the other hand, the spindle improves slightly to 0.66, as adding further explanatory features, such as the moment or the material removal rate, is positive for the model. However, the MSE remains virtually unchanged. The results from experiment 6 lead to the conclusion that the tree-based models are able to distinguish between the materials steel and aluminium without any problems. It can also be seen that the spindle, as well as the Y and Z axes, have improved in all metrics, which can be attributed to the slightly larger data base. However, if you look at the results on the CMX Machine, as already stated in the general impression, it is noticeable that the models perform worse here or are unable to find a reasonable mapping, especially for the spindle.

TABLE 5. X-axis results of experimental series 1. The statistical values mean (MN) and variation (VAR) are given for all evaluation metrics.

Machine	Metric	Model	MSE	MAE	R^2	R^2_{adj}
CMX	MW	RF	0.13598	0.22933	0.66772	0.66745
		ET	0.10616	0.21594	0.73358	0.73337
		LSTM	0.45391	0.44896	-0.21024	-
		FNN	0.40361	0.41535	0.03526	0.03453
		GB	0.11657	0.21246	0.71148	0.71125
		TF	0.31128	0.37780	0.16351	-
	VAR	RF	0.01561	0.01731	0.06081	0.06094
		ET	0.00614	0.00848	0.02252	0.02257
		LSTM	0.09945	0.04183	0.70437	-
		FNN	0.17721	0.03648	0.40915	0.41032
		GB	0.00917	0.01031	0.03211	0.03219
		TF	0.02494	0.01647	0.16850	-
DMC	MW	RF	0.06901	0.14185	0.80297	0.80274
		ET	0.04742	0.11422	0.86507	0.86492
		LSTM	0.53548	0.31961	-0.55114	-
		FNN	0.55557	0.27539	-0.70191	-0.70370
		GB	0.04214	0.12827	0.87722	0.87709
		TF	0.40124	0.29160	-0.26001	-
	VAR	RF	0.00293	0.00589	0.01295	0.01299
		ET	0.00132	0.00217	0.00457	0.00459
		LSTM	0.19005	0.01167	1.26611	-
		FNN	0.16365	0.01038	1.65843	1.66243
		GB	0.00061	0.00151	0.00133	0.00134
		TF	0.11359	0.01214	1.24836	-

TABLE 6. Y-axis results of experimental series 1. The statistical values mean (MN) and variation (VAR) are given for all evaluation metrics.

Machine	Metric	Model	MSE	MAE	R^2	R^2_{adj}
CMX	MW	RF	0.19133	0.28395	0.47148	0.47111
		ET	0.15759	0.25888	0.56401	0.56370
		LSTM	0.45703	0.44739	-0.32755	-
		FNN	0.28971	0.32130	0.13091	0.13039
		GB	0.18397	0.27467	0.51355	0.51318
		TF	0.81201	0.51856	-1.16924	-
	VAR	RF	0.01089	0.01156	0.03304	0.03313
		ET	0.00622	0.00762	0.01498	0.01503
		LSTM	0.06765	0.02710	0.55818	-
		FNN	0.00585	0.00717	0.07786	0.07789
		GB	0.01933	0.01699	0.04563	0.04577
		TF	0.68487	0.11142	4.59161	-
DMC	MW	RF	0.05268	0.13031	0.80771	0.80753
		ET	0.09224	0.13865	0.69423	0.69386
		LSTM	0.38312	0.31659	-0.25050	-
		FNN	0.34697	0.36158	-0.26485	-0.26608
		GB	0.04975	0.12545	0.81908	0.81890
		TF	0.27789	0.28249	-0.01276	-
	VAR	RF	0.00040	0.00179	0.00100	0.00101
		ET	0.00724	0.00595	0.04222	0.04237
		LSTM	0.14923	0.05198	0.83953	-
		FNN	0.07740	0.03785	1.12612	1.12767
		GB	0.00036	0.00208	0.00116	0.00117
		TF	0.08922	0.03029	1.30414	-

B. EXPERIMENTAL SERIES 2

Based on the statistical key figures of the second experimental series (see Table 9) in which the X-axis, Z-axis and SP were considered, the following impressions can be gathered. For the X-axis, the GB model shows the best metrics with an average MSE of 0.098 and an R^2 of 0.84. RF follows these results with slightly worse metrics, such as an average MSE

TABLE 7. Z-axis results of experimental series 1. The statistical values mean (MN) and variation (VAR) are given for all evaluation metrics.

Machine	Metric	Model	MSE	MAE	R^2	R^2_{adj}
CMX	MW	RF	0.11016	0.15570	0.34428	0.34372
		ET	0.10305	0.14095	0.35349	0.35294
		LSTM	0.56042	0.42092	-2.58694	-
		FNN	0.15645	0.19394	-0.22030	-0.22101
		GB	0.09034	0.12525	0.40358	0.40309
		TF	0.20575	0.30359	-0.45484	-
	VAR	RF	0.02097	0.00943	0.26068	0.26128
		ET	0.01477	0.00627	0.22962	0.23011
		LSTM	0.51244	0.04768	6.52865	-
		FNN	0.01266	0.00508	0.98827	0.98866
		GB	0.00841	0.00401	0.17079	0.17114
		TF	0.01139	0.00721	0.93701	-
DMC	MW	RF	0.02433	0.07527	0.97137	0.97135
		ET	0.03713	0.08316	0.95950	0.95945
		LSTM	0.41865	0.26163	0.51667	-
		FNN	0.19959	0.17618	0.78021	0.78002
		GB	0.00955	0.05274	0.98870	0.98869
		TF	0.44637	0.26808	0.48298	-
	VAR	RF	0.00007	0.00061	0.00013	0.00013
		ET	0.00061	0.00161	0.00029	0.00029
		LSTM	0.01977	0.00549	0.01585	-
		FNN	0.01057	0.00457	0.00765	0.00764
		GB	0.00002	0.00033	0.00003	0.00003
		TF	0.02105	0.01122	0.03216	-

TABLE 8. SP results of experimental series 1. The statistical values mean (MN) and variation (VAR) are given for all evaluation metrics.

Machine	Metric	Model	MSE	MAE	R^2	R^2_{adj}
CMX	MW	RF	1.62454	0.64610	-1.07363	-1.07495
		ET	1.15721	0.60611	-0.44895	-0.44990
		LSTM	6.18448	1.24611	-7.45333	-
		FNN	1.61233	0.67831	-1.38319	-1.38427
		GB	0.81431	0.48044	-0.00274	-0.00339
		TF	2.31552	0.83498	-1.96531	-
	VAR	RF	0.66732	0.10212	1.14193	1.14365
		ET	0.24333	0.05161	0.34964	0.35022
		LSTM	83.37464	0.73193	141.83464	-
		FNN	1.72008	0.08132	7.01307	7.01562
		GB	0.10564	0.06336	0.06699	0.06711
		TF	2.99619	0.21119	5.10545	-
DMC	MW	RF	2.16597	0.28839	0.68815	0.68786
		ET	4.47272	0.50089	0.35298	0.35237
		LSTM	22.72027	1.50434	-2.36554	-
		FNN	8.92482	0.89852	-0.31187	-0.31301
		GB	1.89961	0.38663	0.72743	0.72719
		TF	7.90538	1.01540	-0.12661	-
	VAR	RF	0.25342	0.01398	0.00347	0.00347
		ET	0.48739	0.03361	0.00416	0.00416
		LSTM	233.67975	1.06529	5.20776	-
		FNN	20.67237	0.24479	0.52922	0.52943
		GB	0.56602	0.02614	0.01121	0.01122
		TF	10.78772	0.27963	0.14364	-

of 0.117 and an R^2 of 0.81. The neural network achieves the worst results here with an average MSE of 0.37 and an R^2 of 0.64. If we look at the Z-axis now, a similar picture is generated. Here, too, GB performs best on average with an MSE of 0.018 and an R^2 of 0.933. In comparison, the neural network and the RF perform only slightly worse, and the NN even performs better with an MSE of 0.02 as opposed to 0.021. The RF achieves the best results for the spindle with an MSE of 0.71, closely followed by the RF. On the other hand, the neural network performs particularly poorly here with an average MSE of 2.589, which is significantly further away from the tree-based methods, suggesting problems with model fitting.

TABLE 9. Results of experimental series 2 for the X- and Y-axis and SP. The statistical values mean (MN) and variation (VAR) are given for all evaluation metrics.

Machine	Metric	Model	MSE	MAE	R^2	R^2_{adj}
X-axis	MW	RF	0.11727	0.19083	0.80623	0.80576
		FNN	0.37051	0.27638	0.64271	0.64159
		GB	0.09833	0.21033	0.83868	0.83829
	VAR	RF	0.00837	0.00754	0.02157	0.02165
		FNN	0.41640	0.00924	0.03895	0.03933
		RB	0.00563	0.00633	0.00928	0.00932
Z-Axis	MW	RF	0.02072	0.06400	0.91925	0.91902
		FNN	0.02009	0.05640	0.93254	0.93233
		GB	0.01843	0.06838	0.93314	0.93300
	VAR	RF	0.00018	0.00096	0.00216	0.00217
		FNN	0.00042	0.00060	0.00102	0.00102
		RB	0.00036	0.00144	0.00410	0.00410
SP	MW	RF	0.70792	0.31381	0.69432	0.69335
		FNN	2.58957	0.36500	-0.28009	-0.86943
		GB	0.83737	0.38944	0.55399	0.55279
	VAR	RF	0.46567	0.06964	0.06150	0.06192
		FNN	1.45497	0.03766	1.20035	9.65400
		RB	0.78708	0.10793	0.37359	0.37571

Experiment 4, 6, 10 and 12 for the X/Z axis and the spindle are described as representative of the second series of trials. These experiments are used to shed more light on the third research question by increasing the underlying data base and increasing the complexity in order to approximate real-life circumstances.

The results for the x-axis show the following differences between the models. In experiment 4, GB dominates with an MSE of 0.027 and an adjusted R^2 of 0.972, closely followed by RF, which has an MSE of 0.0288. The neural network performs significantly worse here with an MSE of 0.144 and an R^2 of 0.85. In experiment 6, on the other hand, the RF achieves the best metrics with an MSE of 0.051. In contrast, GB has an MSE of 0.053 and the neural network 0.069. By a slightly larger margin, GB has the better metrics in experiment 10. It has an MSE of 0.074 ($R^2 = 0.74$), while the RF and the neural network only achieve 0.11 ($R^2 = 0.58$) and 0.13 ($R^2 = 0.53$), respectively. In experiment 12, the whole thing changes again, and the RF west shows the best performance here with an MSE of 0.04.

Looking at the Z-axis, all models deliver good results. In experiment 4, GB with an MSE of 0.0130 ($R^2 = 0.9123$) and RF with 0.0131 ($R^2 = 0.9115$) achieve similarly good results, while the neural network falls slightly behind with a higher MSE of 0.0161 ($R^2 = 0.8914$). In experiment 6, GB also delivers the best result with an MSE of 0.0074 ($R^2 = 0.9767$), followed by the neural network (MSE = 0.0086; $R^2 = 0.9729$) and RF (MSE = 0.0139; $R^2 = 0.9561$). In experiment 10, GB also achieves the best results (MSE = 0.0081; $R^2 = 0.9744$), but the neural network (MSE = 0.0113; $R^2 = 0.9642$) and RF (MSE = 0.0207; $R^2 = 0.9345$) are not far behind and perform well. Experiment 12 confirms the performance of GB with an MSE of 0.0046 and an R^2 of 0.9854. The neural network with 0.0086 and 0.9730 and the RF with 0.0098 and 0.9691 are also convincing, with very low errors and high degrees of explanation. To summarise, GB on the Z-axis delivers the most stable results overall, closely followed by RF and the neural network, which also achieves solid performances.

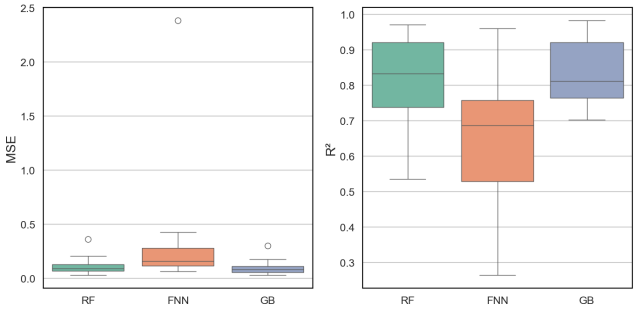


FIGURE 4. Results of experimental series 2 for the X-axis for the metrics MSE (left) and R^2 (right).

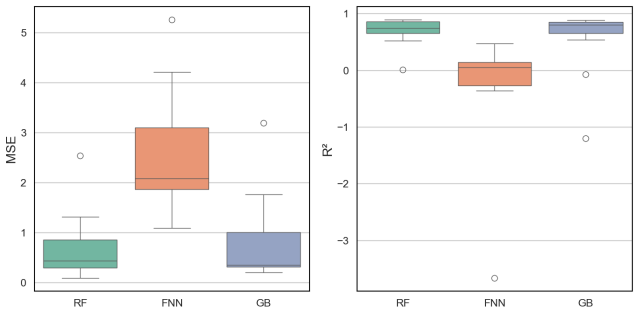


FIGURE 5. Results of experimental series 2 for the main spindle for the metrics MSE (left) and R^2 (right).

The results for the spindle show different performance levels of the models used. In experiment 4, RF achieves the best result with an MSE of 0.092 but with a low adjusted R^2 of only 0.014, which indicates a weak fit to the data. GB (MSE = 0.2049; R^2 = -1.194) and the neural network (MSE = 1.0892; Adj. R^2 = -10.66) show significantly worse results here. In experiment 6, RF with a MSE of 0.3178 ($R^2 = 0.8704$) and GB with 0.3409 ($R^2 = 0.8607$) show good and comparable performance. However, the neural

network clearly lags behind with a MSE of 2.1166 (R^2 = 0.1356), which indicates considerable adaptation difficulties. In experiment 10, the RF again delivers the best results (MSE = 0.2483; R^2 = 0.8986), closely followed by GB (MSE = 0.2704; R^2 = 0.8896). The neural network also shows a significantly poorer performance (MSE = 2.0537; R^2 = 0.1612), which again signals difficulties in model fitting. In experiment 12, RF confirms its performance (MSE = 0.2612; R^2 = 0.8933), while GB lags behind with a slightly higher error (MSE = 0.3595; R^2 = 0.8535). The neural network again achieves a significantly higher error level in this experiment (MSE = 1.5889; R^2 = 0.3511). In summary, RF shows the most robust results on the spindle axis across these experiments, GB follows close behind, while the neural network consistently shows difficulties in fitting the data.

VI. DISCUSSION

In almost all series one experiments, RF, ET or GB achieve more precise results. This is reflected in very good metrics, i.e. low MSE, RMSE and MAE, as well as high values for R^2 and R^2_{adj}, which also only have a low delta. However, GB often comes out as the best model architecture when looking at the spindle. One explanation for the better performance of these architectures could lie in the fact that they are more robust against noise and only require a small data base to reliably model non-linear relationships. Furthermore, there are significantly fewer underlying hyperparameters, which simplifies the handling of these models, and there is less sensitivity with regard to the hyperparameter configuration. The improved performance can also be attributed to built-in data processing techniques such as bootstrapping and subsampling, which increase data diversity within the ensemble and improve generalisation. In addition, tree-based models inherently perform feature selection, making them more robust to limited data. FNN, LSTMs and TF models performed significantly worse in comparison. This can be recognised by the significantly higher metrics for MSE, RMSE and MEA and the lower values for R^2 and adjusted R^2. This could be due to an insufficient data base, i.e. more data is required for training. In addition, there is a greater hyperparameter sensitivity, i.e. FNN are more dependent on the learning rate, batch size and so on. With LSTMs in particular, the chosen sequences may be too short to present enough variation in the data to learn temporal dependencies. However, the sequence length is limited by hardware capacities. In summary, tree-based models perform significantly better and more efficiently on small data sets.

In the case of air-cut, R^2 in the range 0.8-0.9 or higher are achieved. The lack of material removal, process forces and, thus, an uncoupled axis system lead to the reduced complexity. This results in less complex power consumption, which the models can map with higher precision. It is evident that all models demonstrate superior performance on the DMC 60H in comparison to the CMX 600 V (see figure 6).

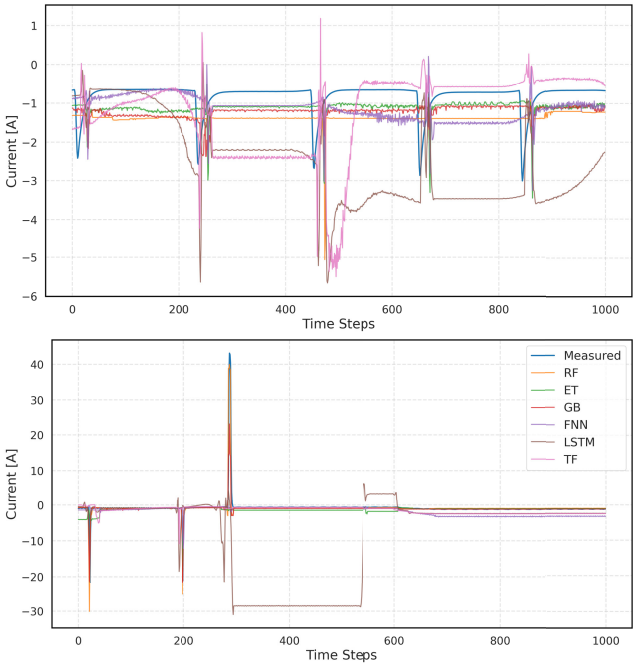


FIGURE 6. Comparison of model output and measurement in experiment 1.6 for the spindle on the CMX (top) DMC (bottom).

Consequently, it can be hypothesised that the calibration of the force soft sensor exerts a substantial influence on the models' performance. This finding is particularly interesting, as it suggests that this factor could also be learned from the data. This finding underscores the value of integrating known physical relationships through soft sensors, demonstrating a notable model performance advantage.

As a result, the first series of experiments can answer the first research question: How can power signals be modelled using ML methods and physical models, and which features are essential for predicting power data? Looking at the results, it can be concluded that tree-based methods are better suited to predicting the power signal than FNN or transformer architectures under the given data. In addition, the results (R^2 and MSE metrics) indicate that the features derived from the physical description, such as speed, acceleration, process forces, spindle torque and MRR, are sufficient to describe the current signal. In particular, it is noticeable that speed and acceleration, as well as the MRR during processing, have a significant influence on the translatory axes, as can be seen in figure 7. For the spindle, its torque is also of importance. Thus, the second research question can be answered: Tree-based approaches tend to outperform DL models, particularly when data is sparse or complex.

However, the functionality of soft sensor-based hybrid ML models is limited by the quality of their formulas. The calculation of physical quantities within the soft sensor system is based on Kienzle's force model (see Section III-C). While the model provides a useful estimate, it has certain limitations that must be considered in practical applications. The key parameters ($k_{p1.1}$, $k_{c1.1}$, $k_{f1.1}$, x, y and z) are

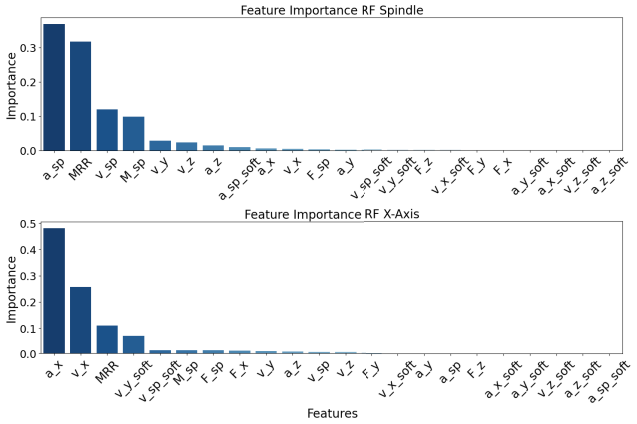


FIGURE 7. Feature importance for the RF architecture for the main spindle (top) and X-axis (bottom). The term *soft* is used to denote the output of the soft sensors.

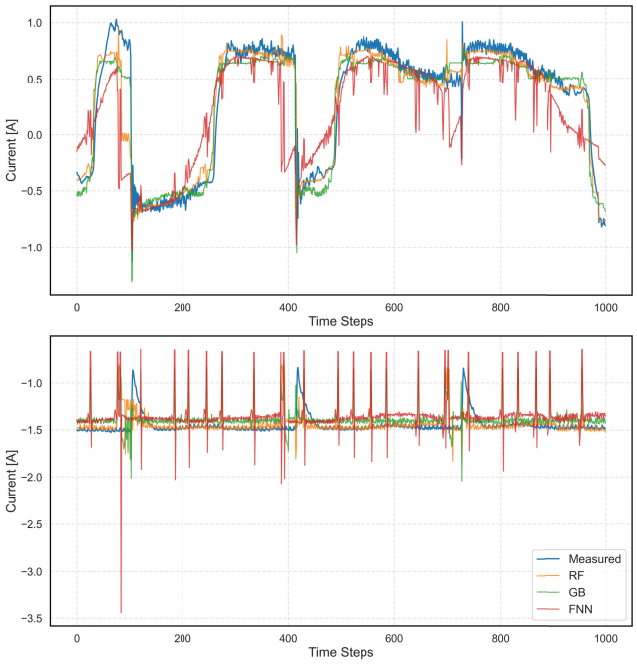


FIGURE 8. Comparison of model output and measurement in experiment 2.4 for the x-axis (top) spindle (bottom).

material-dependent and empirically determined. Significant deviations in force estimation may occur when the model is applied to materials without correct parameters. Additionally, the current state of the force model does not take tool wear into account. While this is a drawback for accurate force prediction, it is advantageous for reference signal prediction in process monitoring. Since the model does not represent tool wear, increasing deviations between the calculated and measured values can be directly interpreted as potential anomalies.

The results of the second experimental series show a significant improvement in FNN performance. This is due to the larger data base and the information it contains. DL methods can therefore unfold their capabilities especially on translatory axes. However, model performance converges fast on the main spindle, and tree-based methods still perform significantly better (see figure 8). Overall, the tree-based methods deliver the most stable results in this series of experiments. There is a trend towards higher accuracy, a development that is particularly attributable to the increased data base. It can also be seen that the performance delta between the two approaches correlates with the complexity of extrapolation. In other words, as extrapolation becomes more complex, the distance between the metrics increases, and FNN performes worse in comparison. This answers the final research question.

VII. CONCLUSION AND OUTLOOK

This study addresses the challenge of monitoring CNC milling processes in agile production scenarios via hybrid ML. Therefore, physical models were derived for the axes and the main spindle. The features have been incorporated into hybrid ML architectures. Additionally, soft sensors were used to incorporate difficult-to-record but relevant features.

It was shown that effective modelling can be achieved through the integration of physical knowledge, particularly velocity, acceleration, process forces, spindle torque and MRR. Different ML architectures demonstrated varying predictive capabilities. Tree-based models, such as RF and GB, were found to consistently outperform DL models such as FNN, LSTMs and TF, particularly when data was limited. Finally, the performance of the models was tested under realistic and complex operational conditions. The robust performance and adaptability of the hybrid ML models were demonstrated. This was particularly the case for GB and RF, even as process complexity and variability increased. DL models improved significantly as data availability increased, but still lagged behind the tree-based approaches.

Future research opportunities to enhance the efficacy and predictive performance of the hybrid ML models include soft sensor calibration and deeper integration of the physical knowledge into architecture and training. To investigate generalisability, the model should be applied to a range of production scenarios, including extreme working conditions. Additionally, the approach should be transferred to other milling machines, such as five-axis machining centres, and other types of machine tools.

In order to address the challenges associated with data availability, it is important to investigate incremental learning approaches. The continuous integration of newly available data could be particularly advantageous for DL. Furthermore, physics-informed ML and active learning can enhance the models' ability to adapt to evolving data conditions. Exploring data augmentation methods is also recommended to overcome the limitations posed by small, incomplete, or imbalanced data sets. Active learning can also be employed to incorporate on-site employees' expertise as an additional knowledge source for model training.

For successful application, it is crucial to develop a comprehensive process monitoring system that integrates

hybrid ML models and incremental learning. The monitoring capabilities must be validated, for example with regard to the detection of tool wear, material quality or poor processing conditions.

REFERENCES

[1] Y. Koren, *The Global Manufacturing Revolution: Product-Process-Business Integration and Reconfigurable Systems*, 1st ed., Hoboken, NJ, USA: Wiley, May 4, 2010, doi: 10.1002/9780470618813.

[2] M. I. Ahmad, Y. Yusof, M. E. Daud, K. Latiff, A. Z. A. Kadir, and Y. Saif, "Machine monitoring system: A decade in review," *Int. J. Adv. Manuf. Technol.*, vol. 108, nos. 11–12, pp. 3645–3659, Jun. 2020, doi: 10.1007/s00170-020-05620-3.

[3] B. Denkena, M.-A. Dittrich, H. Noske, and M. Witt, "Statistical approaches for semi-supervised anomaly detection in machining," *Prod. Eng.*, vol. 14, no. 3, pp. 385–393, Jun. 2020, doi: 10.1007/s11740-020-00958-9.

[4] X. Zhang, Y. Gao, Z. Guo, W. Zhang, J. Yin, and W. Zhao, "Physical model-based tool wear and breakage monitoring in milling process," *Mech. Syst. Signal Process.*, vol. 184, Feb. 2023, Art. no. 109641, doi: 10.1016/j.ymssp.2022.109641.

[5] M. Farhadmanesh and K. Ahmadi, "Online identification of mechanistic milling force models," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107318, doi: 10.1016/j.ymssp.2020.107318.

[6] C. Legaard, T. Schranz, G. Schweiger, J. Drgoňa, B. Falay, C. Gomes, A. Iosifidis, M. Abkar, and P. Larsen, "Constructing neural network based models for simulating dynamical systems," *ACM Comput. Surv.*, vol. 55, no. 11, pp. 1–34, Nov. 2023, doi: 10.1145/3567591.

[7] R. X. Gao, J. Krüger, M. Merklein, H.-C. Möhring, and J. Váncza, "Artificial intelligence in manufacturing: State of the art, perspectives, and future directions," *CIRP Ann.*, vol. 73, no. 2, pp. 723–749, 2024, doi: 10.1016/j.cirp.2024.04.101.

[8] Y. Wu, B. Sicard, and S. A. Gadsden, "Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring," *Expert Syst. Appl.*, vol. 255, Dec. 2024, Art. no. 124678, doi: 10.1016/j.eswa.2024.124678.

[9] J. P. U. Cadavid, S. Lamouri, B. Grabot, R. Pellerin, and A. Fortin, "Machine learning applied in production planning and control: A state-of-the-art in the era of Industry 4.0," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1531–1558, Aug. 2020, doi: 10.1007/s10845-019-01531-7.

[10] J. V. Abellan-Nebot and F. R. Subirón, "A review of machining monitoring systems based on artificial intelligence process models," *Int. J. Adv. Manuf. Technol.*, vol. 47, nos. 1–4, pp. 237–257, Mar. 2010, doi: 10.1007/s00170-009-2191-8.

[11] I. K. Nti, A. F. Adekoya, B. A. Weyori, and O. Nyarko-Boateng, "Applications of artificial intelligence in engineering and manufacturing: A systematic review," *J. Intell. Manuf.*, vol. 33, no. 6, pp. 1581–1601, Aug. 2022, doi: 10.1007/s10845-021-01771-6.

[12] X. P. Li, A. Y. C. Nee, Y. S. Wong, and H. Q. Zheng, "Theoretical modelling and simulation of milling forces," *J. Mater. Process. Technol.*, vols. 89–90, pp. 266–272, May 1999, doi: 10.1016/s0924-0136(99)00076-x.

[13] B. Beri and G. Stepan, "Effect of axial force on the stability of milling: Local bifurcations around stable islands," *J. Vibrat. Control*, vol. 29, nos. 1–2, pp. 440–452, Jan. 2023, doi: 10.1177/10775463211048256.

[14] H. Wang, K. Tao, and T. Jin, "Modeling and estimation of cutting forces in ball helical milling process," *Int. J. Adv. Manuf. Technol.*, vol. 117, nos. 9–10, pp. 2807–2818, Dec. 2021, doi: 10.1007/s00170-021-07817-6.

[15] T. Xi, I. M. Benincá, S. Kehne, M. Fey, and C. Brecher, "Tool wear monitoring in roughing and finishing processes based on machine internal data," *Int. J. Adv. Manuf. Technol.*, vol. 113, nos. 11–12, pp. 3543–3554, Apr. 2021, doi: 10.1007/s00170-021-06748-6.

[16] P. V. Saturley and A. D. Spence, "Integration of milling process simulation with on-line monitoring and control," *Int. J. Adv. Manuf. Technol.*, vol. 16, no. 2, pp. 92–99, Feb. 2000, doi: 10.1007/s001700050013.

[17] B. Denkena, R. Fischer, D. Euhus, and T. Neff, "Simulation based process monitoring for single item production without machine external sensors," *Proc. Technol.*, vol. 15, pp. 341–348, Jan. 2014, doi: 10.1016/j.protcy.2014.09.088.

[18] A. H. Ammouri and R. F. Hamade, "Current rise criterion: A process-independent method for tool-condition monitoring and prognostics," *Int. J. Adv. Manuf. Technol.*, vol. 72, nos. 1–4, pp. 509–519, Apr. 2014, doi: 10.1007/s00170-014-5679-9.

[19] A. D. Jennings and P. R. Drake, "Machine tool condition monitoring using statistical quality control charts," *Int. J. Mach. Tools Manuf.*, vol. 37, no. 9, pp. 1243–1249, Sep. 1997, doi: 10.1016/s0890-6955(97)00016-3.

[20] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool wear monitoring with vibration signals based on short-time Fourier transform and deep convolutional neural network in milling," *Math. Problems Eng.*, vol. 2021, pp. 1–14, Jun. 2021, doi: 10.1155/2021/9976939.

[21] M. C. Gomes, L. C. Brito, M. B. da Silva, and M. A. V. Duarte, "Tool wear monitoring in micromilling using support vector machine with vibration and sound sensors," *Precis. Eng.*, vol. 67, pp. 137–151, Jan. 2021, doi: 10.1016/j.precisioneng.2020.09.025.

[22] D. A. Axinte, N. Gindy, K. Fox, and I. Unanue, "Process monitoring to assist the workpiece surface quality in machining," *Int. J. Mach. Tools Manuf.*, vol. 44, no. 10, pp. 1091–1108, Aug. 2004, doi: 10.1016/j.ijmachtools.2004.02.020.

[23] G. Li, Y. Fu, D. Chen, L. Shi, and J. Zhou, "Deep anomaly detection for CNC machine cutting tool using spindle current signals," *Sensors*, vol. 20, no. 17, p. 4896, Aug. 2020, doi: 10.3390/s20174896.

[24] B. Denkena, D. Dahlmann, and J. Damm, "Self-tuning of teachless process monitoring systems with multi-criteria monitoring strategy in series production," *Proc. Technol.*, vol. 15, pp. 613–620, Jan. 2014, doi: 10.1016/j.protcy.2014.09.022.

[25] Y. Zhou and W. Xue, "A multisensor fusion method for tool condition monitoring in milling," *Sensors*, vol. 18, no. 11, p. 3866, Nov. 2018, doi: 10.3390/s18113866.

[26] Z. Ma, H. Liao, J. Gao, S. Nie, and Y. Geng, "Physics-informed machine learning for degradation modeling of an electro-hydrostatic actuator system," *Rel. Eng. Syst. Saf.*, vol. 229, Jan. 2023, Art. no. 108898, doi: 10.1016/j.ress.2022.108898.

[27] J. Jakubowski, P. Stanisz, S. Bobek, and G. J. Nalepa, "Roll wear prediction in strip cold rolling with physics-informed autoencoder and counterfactual explanations," in *Proc. IEEE 9th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2022, pp. 1–10, doi: 10.1109/DSAA54385.2022.10032357.

[28] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 107961, doi: 10.1016/j.ress.2021.107961.

[29] R. Ströbel, A. Bott, L. Hutt, S. Groß, and J. Fleischer, "Vom verbrauchsmonitoring zur verbrauchsprognose: Untersuchung Des.umgangs produzierender unternehmen mit energieverbrauchsmonitoring und -vorhersage," *Zeitschrift für Wirtschaftlichen Fabrikbetrieb*, vol. 119, nos. 1–2, pp. 80–84, Feb. 2024, doi: 10.1515/zwf-2024-1009.

[30] N. Sihag and K. S. Sangwan, "A systematic literature review on machine tool energy consumption," *J. Cleaner Prod.*, vol. 275, pp. 123–125, Dec. 2020, doi: 10.1016/j.jclepro.2020.123125.

[31] G. Kant and K. S. Sangwan, "Prediction and optimization of machining parameters for minimizing power consumption and surface roughness in machining," *J. Cleaner Prod.*, vol. 83, pp. 151–164, Nov. 2014, doi: 10.1016/j.jclepro.2014.07.073.

[32] J. Moore, J. Stammers, and J. Dominguez-Caballero, "The application of machine learning to sensor signals for machine tool and process health assessment," *Proc. Inst. Mech. Eng., B, J. Eng. Manuf.*, vol. 235, no. 10, pp. 1543–1557, Aug. 2021, doi: 10.1177/0954405420960892.

[33] V. Nasir and F. Sassani, "A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges," *Int. J. Adv. Manuf. Technol.*, vol. 115, nos. 9–10, pp. 2683–2709, Aug. 2021, doi: 10.1007/s00170-021-07325-7.

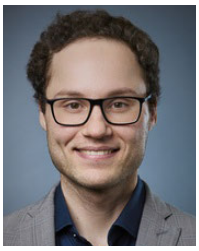
[34] R. Ströbel, Y. Probst, S. Deucker, and J. Fleischer, "Time series prediction for energy consumption of computer numerical control axes using hybrid machine learning models," *Machines*, vol. 11, no. 11, p. 1015, Nov. 2023, doi: 10.3390/machines11111015.

[35] R. I. Asrai, S. T. Newman, and A. Nassehi, "A mechanistic model of energy consumption in milling," *Int. J. Prod. Res.*, vol. 56, nos. 1–2, pp. 642–659, Jan. 2018, doi: 10.1080/00207543.2017.1404160.

[36] T. Gutowski, J. Dahmus, and A. Thiriez, "Electrical energy requirements for manufacturing processes," in *Proc. 13th CIRP Int. Conf. Life Cycle Eng.*, 2006, pp. 623–638.

[37] H.-S. Yoon, J.-Y. Lee, M.-S. Kim, and S.-H. Ahn, "Empirical power-consumption model for material removal in three-axis milling," *J. Cleaner Prod.*, vol. 78, pp. 54–62, Sep. 2014, doi: 10.1016/j.jclepro.2014.03.061.

[38] N. Diaz, E. Redelsheimer, and D. Dornfeld, "Energy consumption characterization and reduction strategies for milling machine tool use," in *Glocalized Solutions for Sustainability in Manufacturing*. Berlin, Germany: Springer, 2011, pp. 263–267, doi: 10.1007/978-3-642-19692-8_46.
[39] N. L. Bhirud and R. R. Gawande, "Measurement and prediction of cutting temperatures during dry milling: Review and discussions," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 39, no. 12, pp. 5135–5158, Dec. 2017, doi: 10.1007/s40430-017-0869-7.
[40] R. Neugebauer, Ed., *Werkzeugmaschinen: Aufbau, Funktion Und Anwendung Von Spanenden Und Abtragenden Werkzeugmaschinen*. Berlin, Germany: Springer, 2012, doi: 10.1007/978-3-642-30078-3.
[41] C. Croitoru, "An empirical estimation of cutting force for face milling using a stationary dynamometer," *Bull. Polytech. Inst. Iaşi. Mach. Constructions Sect.*, vol. 67, no. 4, pp. 53–67, Dec. 2021, doi: 10.2478/bipcm-2021-0023.
[42] B. Denkena and H. K. Tönshoff, *Spanen: Grundlagen*. Berlin, Germany: Springer, 2011, doi: 10.1007/978-3-642-19772-7.
[43] M. Witt, M. Schumann, and P. Klimant, "Real-time machine simulation using cutting force calculation based on a voxel material removal model," *Int. J. Adv. Manuf. Technol.*, vol. 105, nos. 5–6, pp. 2321–2328, Dec. 2019, doi: 10.1007/s00170-019-04418-2.
[44] J. C. Miers, T. Tucker, T. Kurfess, and C. Saldana, "Voxel-based modeling of transient material removal in machining," *Int. J. Adv. Manuf. Technol.*, vol. 116, nos. 5–6, pp. 1575–1589, Sep. 2021, doi: 10.1007/s00170-021-07545-x.
[45] R. Ströbel, *Training and Validation Dataset 2 of Milling Processes for Time Series Prediction*. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT), 2023, doi: 10.35097/1738.
[46] R. Ströbel, Y. Probst, and J. Fleischer, *Training and Validation Dataset of Milling Processes for Time Series Prediction*. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT), 2023, doi: 10.5445/IR/1000157789.
[47] R. Ströbel, *Part Series Dataset of Milling Processes for Time Series Prediction*. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT), 2025, doi: 10.35097/CTVUJ6DGZEPZMK0G.

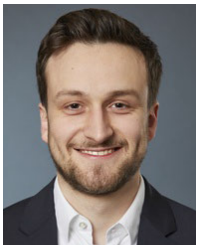


ROBIN STRÖBEL received the M.Sc. degree in mechanical engineering from Karlsruhe Institute of Technology (KIT), in 2022. He is currently pursuing the Ph.D. degree in incremental model training for model-based anomaly detection. Since 2022, he has been a Research Associate with the Intelligent Machines and Components Research Group, wbk Institute of Production Science, where he works on various projects in the field of data science in manufacturing.

SAMUEL DEUCKER received the B.Sc. degree in industrial engineering and management from Karlsruhe Institute of Technology (KIT), in 2023, where he is currently pursuing the master's degree in information systems. Since 2023, he has been a Research Assistant with the wbk Institute of Production Science, KIT, with research interests in the area of data science in the context of production.

HANLIN ZHOU received the M.Sc. degree in mechanical engineering from Karlsruhe Institute of Technology (KIT), in 2025. His research interests include machine tools, process monitoring, and data science.

HAFEZ KADER (Graduate Student Member, IEEE) is currently pursuing the Ph.D. degree with the Autonomous Multisensor Systems (AMS) Research Group, Faculty of Computer Science, Otto von Guericke University Magdeburg (OVGU). His research focuses on analyzing the relevance and significance of different features and detecting anomalies in sensor data to make processes more efficient and optimize the performance of machine learning models.



ALEXANDER PUCHTA received the M.Sc. degree in mechanical engineering from Karlsruhe institute of Technology, in 2020. He is currently pursuing the Ph.D. degree in machine, plant, and process automation, focusing on the topic of autonomous modeling of feed axes for machine tools. He has been the Head of the Intelligent Machines and Components Group, wbk Institute of Production Science, since 2023.

BENJAMIN NOACK (Senior Member, IEEE) received the Ph.D. degree in computer science from the Intelligent Sensor-Actuator Systems (ISAS) Research Group, University of Karlsruhe (TH), in 2013. Since 2021, he has been a Professor with the Faculty of Computer Science, Otto von Guericke University Magdeburg (OVGU), where he heads the Autonomous Multisensor Systems (AMS) Research Group. His research focuses on distributed sensor data fusion and its applications in industrial process monitoring and autonomous mobile robotics.

JÜRGEN FLEISCHER received the degree in mechanical engineering from the Universität Karlsruhe (TH), and the Ph.D. degree from the Institute for Machine Tools and Industrial Engineering (iwb), in 1989. Since 1992, he has been holding several leading positions with industry before being appointed as a Professor and the Head of the wbk Institute for Production Technology, Karlsruhe Institute of Technology (KIT), in 2003. In addition, he has been a Visiting Professor with Tongji University, Shanghai, since 2012. His current scientific research focuses are intelligent components for machine tools and handling systems, the automation of immature processes, and agile production facilities. As a recognized member of the scientific community, he is active in German Academy of Science and Engineering (acatech) and a member of several scientific and industrial advisory boards.