

A DATASPACE-DRIVEN EDGE COMPUTING AND FEDERATED LEARNING FRAMEWORK FOR SENSORY TOOLING SYSTEMS

T. Trautner^{1*}, F. Zink¹, E. Slimane¹, L. Graff², W. Wunschel², M. Breurather¹, M. Weigold², F. Bleicher¹

¹TU Wien, Institute of Production Engineering and Photonic Technologies, Vienna, Austria

²TU Darmstadt, PTW Institut für Produktionsmanagement, Technologie und Werkzeugmaschinen, Darmstadt, Germany

*Corresponding author; e-mail: trautner@ift.at

Abstract

Although manufacturing becomes increasingly data-driven, sensor data is typically available only locally, with limited integration across systems or organizations. This lack of available, heterogeneous data prevents the development of robust AI models to predict key machining outcomes such as tool wear and surface roughness. To overcome this challenge, a cooperative approach that leverages Dataspaces and Compute-to-Data is proposed. This enables access to heterogeneous data and the subsequent development of robust algorithms without sharing the actual data, thereby keeping know-how and intellectual property secure. Therefore, an edge computing architecture is proposed, integrating sensory tool holders with machine tools and enabling high-frequency measurements of acceleration during machining. These measurements are transmitted via Bluetooth from the tool to a stationary transceiver unit, labeled and subsequently stored on a local data storage. The data is then made accessible through a Dataspace. To prevent stakeholders from accessing raw data of each other, the AI model training is conducted in Compute-to-Data environments. The approach is illustrated through an exemplary case study, focusing on the prediction of surface roughness during machining. Thus, this work contributes to the cooperative creation of intelligent machine tools. Future research can build on this approach and explore more industrialized implementations.

Keywords:

Dataspace, Pontus-X, Gaia-X, Ecosystem, Surface Roughness Prediction, AI, XGBoost

1 INTRODUCTION

Advances in manufacturing demand machine tools that continuously improve their performance through data-driven insights. In this context, self-optimizing machine tools, capable of autonomously adjusting cutting parameters to maximize productivity and surface quality, have emerged [Möhring 2020]. Sensory tooling systems, which embed sensors directly into cutting tools to monitor process variables such as vibrations and temperature in real time are a promising solution. Using an integrated acceleration sensor, sensory tools can detect chatter and adapt cutting conditions during machining [Bleicher 2021]. More recently, in-situ surface roughness measurement systems have been explored to provide data for AI training directly from within the machine tool [Sulz 2023], and thereby enable self-learning capabilities of machine tools.

Despite these advances, current self-learning approaches typically require massive volumes of machine-specific data to train robust models for applications such as tool condition monitoring, thermal compensation, and dynamic stability control. Each of these applications, whether predicting tool wear, compensating for tool-holder thermal drift, or

stabilizing cutting forces, relies on distinct data sources (sensor streams, machine metadata, tool geometry, etc.) and bespoke algorithms. Gathering sufficient data in a single machine or shop can be prohibitively time-consuming and risks overfitting to idiosyncratic operating conditions. Federated Learning, where models are trained collaboratively across multiple machines or facilities without sharing raw data, offers a way to overcome these limitations.

To address concerns around data privacy, ownership, and business models, there is a clear need for sovereign collaboration frameworks that allow companies to share insights without relinquishing control over their raw data. Data ecosystems based on the Dataspace concept and coupled with Compute-to-Data (CtD) techniques, with which algorithms are dispatched to execute within a data holder's secure environment, provide such a framework [Gehrer 2024]. In this setup, participants retain full control of their data, while contributing to a global federated model. Ensuring data quality, establishing ground truth measurements (e.g., calibrated roughness or flank wear benchmarks), and achieving semantic interoperability across heterogeneous sensor and metadata formats are

essential prerequisites for such a federated intelligence network.

This work investigates a suitable edge-computing architecture for Federated Learning within a Dataspace that preserves data sovereignty, specifically tailored to sensory tooling systems. To answer this, an edge-centric data acquisition and processing pipeline for the shopfloor, integrated with a Dataspace-driven Federated Learning framework, is proposed. Therefore the paper outlines related work on application fields and collaborative data acquisition and analysis, presents a Dataspace-driven Federated Learning framework, and demonstrates it in an example application, which will be scaled in future work.

2 RELATED WORK

2.1 Application Fields

The success of AI applications depends heavily on the quality and quantity of the available data [Zha 2023]. In the machining context, data quality is determined by the sensors and the overall machine tool setup, while the data quantity is influenced by factors such as the number of data sources, in this case the machine tools. Thus, increasing the number of machine tools that collect and provide data available promises to improve AI model performance. The following section presents three frequent application fields of AI in machining - tool condition monitoring, thermal error compensation and surface roughness prediction - along with selected approaches from the literature.

[Hassan 2024] implemented a machine learning based tool condition monitoring system, achieving a high detection accuracy of tool wear. Another article investigates the usage of Long Short-Term Memory (LSTM) networks for real-time tool life predictions in milling operations. Their results show that variations between real-world sensor data and the training in controlled environments lead to discrepancies in model performance [Dominguez-Caballero 2025]. [Huang 2023] propose a deep adversarial domain confusion network to address domain shift in tool wear condition monitoring, enabling accurate cross-domain predictions using vibration signal spectrums.

A thermal displacement compensation method based on deep-learning Bayesian dropout has been introduced by [Fujishima 2019]. By dynamically adjusting compensation weights based on prediction reliability, the method achieves high-performance results on a turning center. [Liu 2021] addressed the problem of thermal errors through LSTM based neural network training of an error prediction model for robust compensation.

Surface roughness has been investigated by [Tonejca 2022]. They describe a machine learning based surface roughness prediction model, integrated into Computer Aided Manufacturing planning software, aiding machine operators in their job. Using data on drive power and tool wear as input, [Pimenov 2018] implemented different machine learning methods for surface roughness prediction in CNC machines. [Lin 2019] investigates the use of vibration signals and deep learning models for in-process surface roughness prediction in milling.

The presented cases show, that machine learning based models can be applied for their specifically trained use cases in machining. Nevertheless, the ability to extrapolate to new processes and scenarios remains a big challenge due to the lack of diverse real-world training data.

2.2 Collaborative data acquisition and analysis

To overcome the challenge of scarce training data, a new approach to collaborative data acquisition and model

training has emerged. Federated Learning is an approach to use decentralized data sources for model training, by training sub-models on local data and only transferring the model updates for a centralized aggregation. Raw data is kept on the source system to ensure data sovereignty and minimize network traffic [Abdulrahman 2021].

[Stoop 2023] have developed a cloud-based thermal error compensation method using Federated Learning at the edge, enabling knowledge transfer between two geographically separated machine tools. Despite differences in machine conditions, the approach achieved over 80 % error reduction under critical conditions. A study conducted by [Kaleli 2024] presents a novel tool wear estimation for CNC machines using Federated Learning combined with LSTM neural networks. [Becker 2022] propose an autoencoder-based Federated Learning approach for condition monitoring of rotating machines using on-premise vibration sensor data. By enabling distributed training at the edge, the method preserves data privacy, reduces network load, and achieves competitive performance, facilitating secure knowledge transfer between industrial sites. A new method for the prediction of the remaining useful life of machines, based on Federated Learning between edge clients, is presented by [Guo 2023]. Experimental evaluations show the potential for decentralized learning for RUL predictions.

[Gehrer 2024] have developed a Federated Learning approach for the manufacturing industry, that focuses on data sovereignty, using Compute-to-Data and Gaia-X Dataspaces for tool condition monitoring using CNC data. Another article presents a Federated Learning architecture based on the Gaia-X infrastructure. [Friedrich 2024] harmonize the data exchange through Asset Administration Shell and OPC UA. The paper describes a usage for predicting the machine tool lifetime, but lacks a real-world example of their proposed architecture.

The presented approaches show the potential of Federated Learning while highlighting that the industrial usage of collaborative approaches struggles due to fragmented data acquisition systems and a lack of best practices.

2.3 Challenges

Based on the described articles, three major challenges in advancing the use of AI applications in machining are identified. The first challenge is the absence of a standardized edge infrastructure and the lack of a consistent, semantically interoperable data description, for example through the use of Asset Administration Shells. Second, the need to incorporate a large number of machine tools into AI model training, while ensuring data security and sovereignty. This requires new approaches such as Dataspaces. The third challenge is about ensuring the normalization of machine tool specific dynamics across different machine tools, as variations in frequency domains can limit the ability to combine different datasets.

3 A DATASPACE-DRIVEN EDGE COMPUTING AND FEDERATED LEARNING FRAMEWORK FOR SENSORY TOOLING SYSTEMS

This section details the proposed framework designed to address the challenges of limited data availability and the need for privacy-preserving collaboration in manufacturing.

The framework facilitates a cooperative approach by integrating edge computing infrastructure for sensory tooling with Dataspaces and enabling CtD paradigms like Federated Learning. The main objective is to create a secure and standardized pipeline, from data acquisition

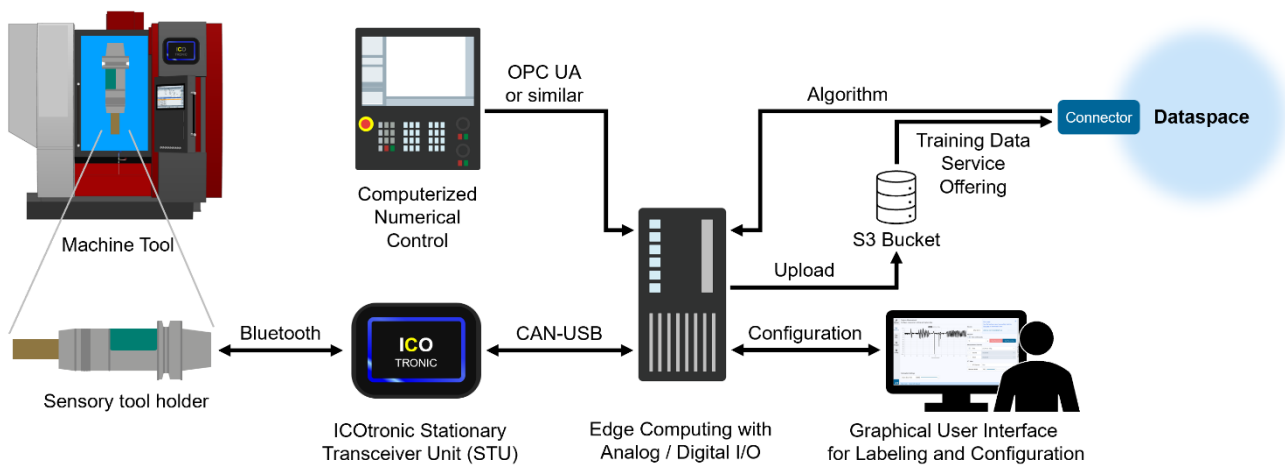


Fig. 1: Edge Computing Framework for Sensory Tooling Systems

during production to making service offerings available within a Dataspace for collaborative algorithm training without exposing sensitive data.

3.1 Edge Infrastructure

The framework is presented in Fig. 1 and makes use of a Sensory Tool Holder (STH) offering functionality for wireless activation, configuration and streaming of acceleration data with up to 10kHz sampling rate via Bluetooth Low Energy (BLE), using a Stationary Transceiver Unit (STU) connected through CAN (Controller Area Network) fieldbus [Bleicher 2021]. The edge device serves as a central hub for local data processing and communication. It combines the STH data with additional data from the Computerized Numerical Control (CNC) of the Machine Tool, using available protocols like OPC UA or FOCAS. For persistent storage at the edge, data is saved locally in one compressed file per dataset. The format chosen was HDF5 due to its ability to compress multiple data tables into a single file. For the purpose of providing a dataset for Federated Learning, it can be uploaded to a private Simple Storage Service (S3) compatible object storage, from where a Service Offering can be created for a single or multiple datasets.

To achieve a standardized and reusable pipeline, a core challenge is data labeling for supervised learning. For this, a Graphical User Interface (GUI), shown in Fig. 2, runs on the edge device. It allows operators or engineers to configure the data acquisition process, monitor the system, and manage the upload process. For labeling, required and optional metadata can be configured via a configuration file.

3.2 Dataspace

The Dataspace acts as the central marketplace and governance layer for secure interorganizational collaboration, built upon the principle of data sovereignty, so participants retain control over their data while gaining the ability to leverage collective knowledge from service offerings such as data provided in a private S3 bucket, as depicted in Fig. 1. To create such a service offering for making the data available in a Dataspace, a connector, such as the Eclipse Dataspace Connector (EDC), or customized solutions based on Pontus-X Nautilus Toolkit, are necessary.

For the framework presented in this work Pontus-X is chosen because of its built-in support for running CtD algorithms. A custom REST API based on Nautilus is used, to facilitate S3 services, for publishing of assets and algorithms, and to enabling consuming CtD services for model training. Authentication and authorization within this connector are handled via OpenID Connect (OIDC),

ensuring only permitted users can interact with the private data assets.

The pipeline from the Edge Device thereby enables automated publishing of the uploaded and labeled data within the S3 Buckets to Pontus-X, making the data available in the ecosystem for different participants. While data collection could be automated in the future, currently it is done manually to ensure correct labeling of the datasets.

3.3 Federated Learning in Dataspaces using Compute-to-Data (CtD)

A major application of the described Edge Device and Dataspace infrastructure is enabling Federated Learning, as depicted in Fig. 3. Federated Learning shows great potential for enabling secure and privacy-preserving training on distributed data sources, e.g. machine tools or datasets offered by different manufacturers.

On the Dataspace level, other participants do not gain direct read access to the provided data. Instead, Dataspace participants can discover available data assets (based on metadata described in the service offering) and initiate compute tasks (like training in Federated Learning) that run in a predefined environment, where they gain access to the data, adhering to the CtD principle. The CtD environment may either be provided by the data owner or be completely

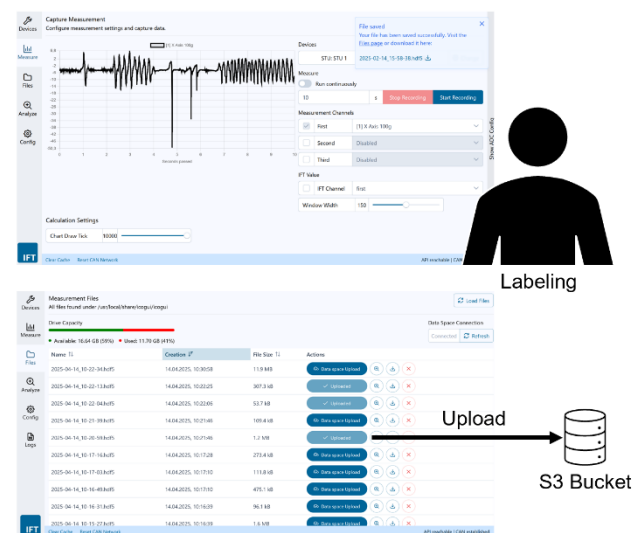


Fig. 2: Graphical User Interface for Labeling and S3 Bucket Upload

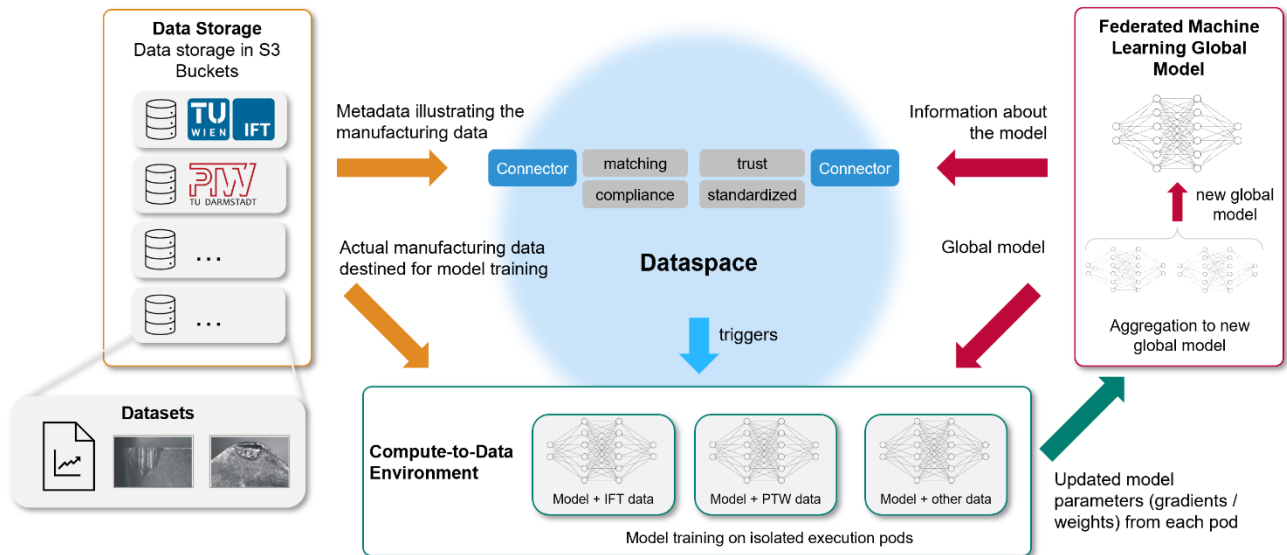


Fig. 3: Architecture for Federated Learning using the Pontus-X Ecosystem

independent, so that neither the owner of the data or the algorithm, nor the consumer can access this environment.

Furthermore, the Dataspace facilitates the secure exchange of algorithms and trained models, allowing participants to deploy externally developed and cooperatively trained models onto their edge devices for local inference or analysis. The concept of Federated Learning is suitable for such a scenario. It describes the training of a machine learning algorithm being performed across multiple clients, allowing the data to remain decentralized and stored locally on each client. Only coordination and aggregation of updates (e.g., model weights) are handled by a central orchestrating server. The central server never sees the clients' training data. By avoiding the centralization of data, a higher degree of data security can be ensured. Various training strategies can be employed in such a federated setup. For example, when using gradient-boosted decision trees with libraries like XGBoost, a bagging aggregation technique can be applied by treating each client as a bootstrap sample and subsequently aggregating the resulting trees through the central orchestrator.

4 APPLICATION EXAMPLE

In this section, an example of how the framework can be applied for roughness prediction in milling is presented for a single machine instance, with the purpose of being scaled up in the future.

4.1 Experiment Setup

The experimental setup depicted in Fig. 4, consists of a DMG MORI DMU75 5-axis machine tool and a Jenoptik Waveline W912RC measurement machine. The workpiece features both face milling and end milling operations, with each operation having a constant and a varied operation per workpiece. Varied means that the parameters change during the cut. Depth/width of cut (a_p , a_e), spindle speed (n), and feed rate (v_f) were selected to evaluate their impact on surface roughness and to serve as input features for the prediction model. Data is collected from labeling (workpiece ID, material, tool diameter and number of teeth) and during the milling process, including acceleration data from the sensory tool and NC data using the FANUC FOCAS interface, including position, feed rate, spindle speed, and macro variables to track a_p and a_e during machining. Next to HDF5 capture and S3 upload, the live data is also

streamed to an MQTT Broker to make it accessible for further applications, e.g. when finally deploying the trained prediction model during machining. After machining, the surface roughness profile is obtained using the measurement machine, exported to CSV, converted to HDF5 and labeled with the workpiece ID. So separate datasets can later be matched and finally uploaded manually to the same S3 bucket as the machining data.

In total, 24 workpieces were machined, during which the tool wear was kept nearly constant. Fig. 5 displays the combination of roughness profile and machining data, resulting from data preparation, over the position on the x-axis of the workpiece/machine tool for a single operation. The start and end of the workpiece is indicated by the green

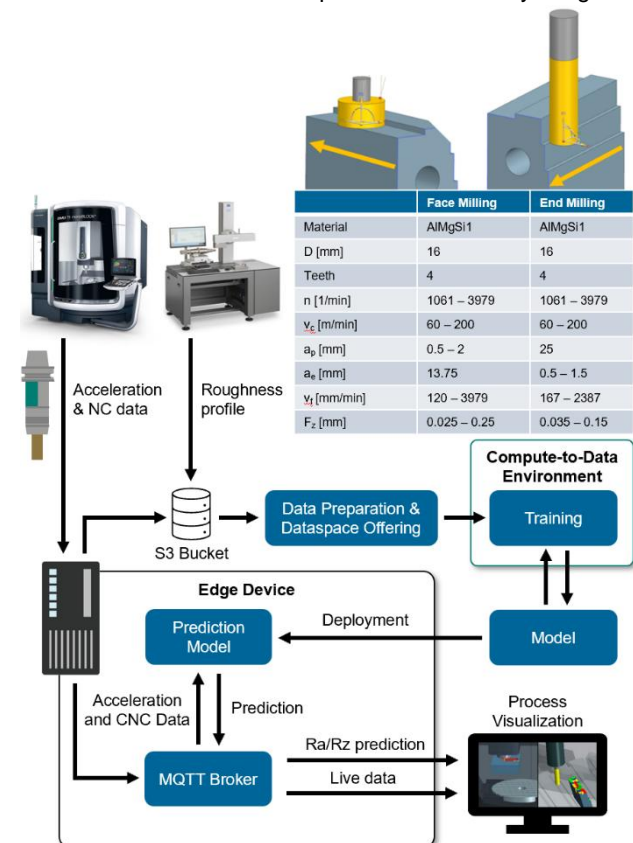


Fig. 4: Experiment setup

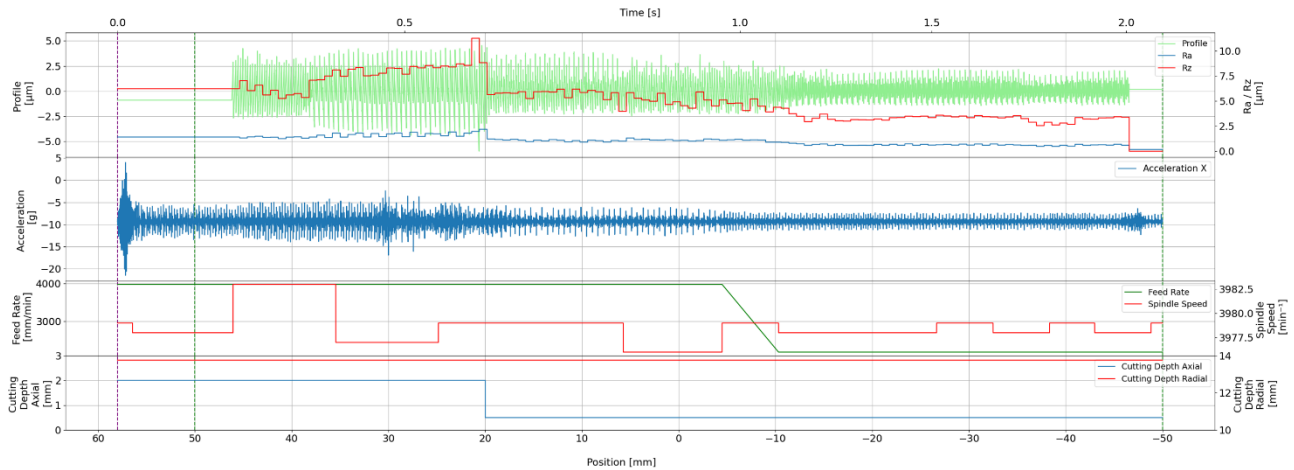


Fig. 5: Experiment data of a face milling operation with varied parameters after data preparation and synchronisation

vertical lines, with the roughness profile starting and ending with a distance of 3 mm to the edges of the workpiece. The figure further shows the measured acceleration from the sensory tool, with the bottom diagrams illustrating the feed rate, and spindle speed and the depth of cut.

4.2 Data Preparation and Modelling

The method used for data preparation is CRISP-DM. For practical application in machining, a feature-based data acquisition approach is applied, enabling fast data cleaning, by using markers included in the CNC code in the form of macro variables, in order to label machining features like face milling and end milling in the timeseries data. Using these markers, it is possible to extract data related operations from the NC data. By relying on the start and end timestamps of each operation available in the NC data, the acceleration data can also be filtered accordingly. For synchronisation of the timeseries acceleration data with the NC data, the point of tool-workpiece engagement is used. This requires the workpiece position, the interpolated tool position including the tool radius together with the spike in acceleration data, as shown by the acceleration signal in Fig. 5 preceding the workpiece by the tool radius.

As the surface roughness data does not directly provide the Ra and Rz values; these must be computed from the roughness profile. The resulting values, the NC parameters, acceleration, and roughness data are merged into a single dataset that serves as input for the prediction model.

After preparing the dataset, it is split into two sets: 80 % for training and 20 % for testing, in order to train the prediction model. The model selected for this task is XGBoost. XGBoost is a machine learning algorithm based on decision trees. It relies on gradient boosting, which aims to correct the errors of previous models and improve overall prediction accuracy [Chen 2016].

4.3 Results of the Prediction Model

Tab. 1 presents the results of the XGBoost prediction model for estimating the values of Ra and Rz, distinguishing between the two types of milling operations: face milling and end milling. The results obtained for the face milling operation show that the model performs very well in predicting Ra, with a coefficient of determination R^2 of 0.8645, indicating that more than 86% of the variance in the data is explained by the model. The associated errors are very low, with a Mean Absolute Error (MAE) of 0.0634 and Root Mean Square Error (RMSE) of 0.1126, suggesting that the predictions are very close to the actual values. The model is therefore reliable for estimating Ra roughness in this operation.

Regarding the prediction of Rz, the performance is also very satisfactory, with an R^2 of 0.8387. Compared to Ra, the prediction errors for Rz are higher, with MAE of 0.2702 and RMSE of 0.3832. This can be considered acceptable, as Rz is often more variable and more sensitive to surface peaks.

Tab. 1: Performance results of the XGBoost model for surface roughness prediction.

	Face Milling		End Milling	
	Ra	Rz	Ra	Rz
MAE	0.0634	0.2702	1.0493	0.0784
MSE	0.0127	0.1468	14.3876	0.0147
RMSE	0.1126	0.3832	3.7931	0.1213
R^2	0.8645	0.8387	0.0820	0.8946

For the end milling operation, the prediction model for Ra shows a very low R^2 of 0.0820, meaning that it can explain only about 8.20% of the variation in Ra, which indicates a poor level of accuracy. In addition, the model has relatively high errors, with MAE of 1.0493 and RMSE of 3.7931. This may suggest that Ra roughness is influenced by factors not captured by the model, or that there is a degree of variability that is difficult to model accurately with the available data. This result may be explained by only a single tooth being clearly identifiable in the roughness profile during end milling, but requires further investigation.

In contrast, the Rz prediction model in the end milling operation performs well, with an R^2 of 0.8946, indicating that it accurately reproduces the variations in this parameter. Moreover, the errors remain low, with MAE of 0.0784 and RMSE of 0.1213, reflecting high accuracy and reliability in estimating the maximum height of surface irregularities in this type of operation.

4.4 Application of Compute-to-Data (CtD) for Training

For the training of the prediction model through a CtD environment in Pontus-X based on the Ocean Protocol, the algorithm is required to be deployed from a docker image. This can be either created and published in a publicly accessible registry, or provided as a public image containing only the relevant dependencies, and then built by the CtD environment from a resource accessible only via HTTP-header-based authentication. When deployed, this algorithm can only access and return data using local folders without any outside communication, in order to train a model based on input data and return the trained model as a file.

As Fig. 4 shows, that the data preparation process, resulting in a single clean dataset for training, is also responsible for creating a service offering. It is not feasible to integrate the data preparation into the prediction model algorithm, as data from different sources may require distinct preprocessing steps, and Federated Learning requires the same syntax and semantics.

With this as a basis, an offering for an algorithm asset is created, so afterwards the dataset offering can be created, referring to the decentralized identifier of the corresponding algorithm. For running the algorithm on the data, only the identifiers of the dataset and algorithm need to be referenced, resulting in a model trained on the CtD node.

5 CONCLUSIONS AND OUTLOOK

The paper presents a Federated Learning framework for tooling systems, consisting of an Edge Device layer, including interfaces for data acquisition, a GUI for labeling data with the ability for publishing in a Pontus-X Dataspace, and a Dataspace layer, allowing for application of Compute-to-Data (CtD) paradigms for Federated Learning.

A single-machine example for applying the framework is shown for creating an XGBoost roughness prediction model, using NC data together with acceleration data from sensory tools. The model is first developed offline, then applied using CtD in the Dataspace. The model provides good predictions of roughness parameters for the face milling operation. However, for end milling, the model is more effective in predicting Rz, but less accurate in predicting Ra. The results indicate the need to scale up the application in the future to also include the effect of tool wear. Further research will involve further use cases like tool wear prediction, and the standardization of data models and semantics for Federated Learning.

6 ACKNOWLEDGMENTS

The presented work is supported by the Austrian Research Promotion Agency (ID 883413 and 891268) and the German Federal Ministry for Economic Affairs and Energy (FKZ 01MJ21008A). This work is also done in preparation of the the CIRP STC-C Cooperative Work on Sensor-integrated Manufacturing coordinated by Friedrich Bleicher.

7 REFERENCES

[Abdulrahman 2021] Abdulrahman, S. et al. A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond. *IEEE Internet of Things Journal*. 8(7):5476–5497, 2021. doi: 10.1109/JIOT.2020.3030072.

[Becker 2022] Becker, S. et al. Federated Learning for Autoencoder-based Condition Monitoring in the Industrial Internet of Things. 2022 IEEE International Conference on Big Data (Big Data) (Dec.-2022), 5424–5433.

[Bleicher 2021] Bleicher, F. et al. Tooling systems with integrated sensors enabling data based process optimization. *Journal of Machine Engineering*. 21(1):5–21, March 2021. doi: 10.36897/jme/134244.

[Chen 2016] Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug.-2016), 785–794.

[Dominguez-Caballero 2025] Dominguez-Caballero, J. et al. Intelligent real-time tool life prediction for a digital twin

framework. *J Intell Manuf*. April 2025. doi: 10.1007/s10845-025-02606-4.

[Friedrich 2024] Friedrich, C. et al. Enabling Federated Learning Services Using OPC UA, Linked Data and GAIA-X in Cognitive Production. *Journal of Machine Engineering*. 24(2):18–33, June 2024. doi: 10.36897/jme/188618.

[Fujishima 2019] Fujishima, M. et al. Adaptive thermal displacement compensation method based on deep learning. *CIRP Journal of Manufacturing Science and Technology*. 25:22–25, 2019. doi: https://doi.org/10.1016/j.cirpj.2019.04.002.

[Gehrer 2024] Gehrer, R. et al. EuProGigant: A decentralized Federated Learning Approach based on Compute-to-Data and Gaia-X. *Procedia CIRP*. 128:710–715, January 2024. doi: 10.1016/j.procir.2024.07.060.

[Guo 2023] Guo, L. et al. FedRUL: A New Federated Learning Method for Edge-Cloud Collaboration Based Remaining Useful Life Prediction of Machines. *IEEE/ASME Transactions on Mechatronics*. 28(1):350–359, February 2023. doi: 10.1109/TMECH.2022.3195524.

[Hassan 2024] Hassan, M. et al. In-process self-configuring approach to develop intelligent tool condition monitoring systems. *CIRP Annals*. 73(1):81–84, January 2024. doi: 10.1016/j.cirp.2024.04.049.

[Huang 2023] Huang, Z. et al. Hybrid learning-based digital twin for manufacturing process: Modeling framework and implementation. *Robotics and Computer-Integrated Manufacturing*. 82:102545, August 2023. doi: 10.1016/j.rcim.2023.102545.

[Kaleli 2024] Kaleli, I.S. et al. A Domain-Aware Federated Learning Study for CNC Tool Wear Estimation. *Mobile Web and Intelligent Information Systems (Cham, 2024)*, 250–265.

[Lin 2019] Lin, W.-J. et al. Evaluation of Deep Learning Neural Networks for Surface Roughness Prediction Using Vibration Signal Analysis. *Applied Sciences*. 9(7), 2019. doi: 10.3390/app9071462.

[Liu 2021] Liu, J. et al. Thermally-induced error compensation of spindle system based on long short term memory neural networks. *Applied Soft Computing*. 102:107094, April 2021. doi: 10.1016/j.asoc.2021.107094.

[Möhring 2020] Möhring, H.-C. et al. Self-optimizing machining systems. *CIRP Annals*. 69(2):740–763, January 2020. doi: 10.1016/j.cirp.2020.05.007.

[Pimenov 2018] Pimenov, D.Yu. et al. Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth. *J Intell Manuf*. 29(5):1045–1061, June 2018. doi: 10.1007/s10845-017-1381-8.

[Stoop 2023] Stoop, F. et al. Cloud-based thermal error compensation with a Federated Learning approach. *Precision Engineering*. 79:135–145, January 2023. doi: 10.1016/j.precisioneng.2022.09.013.

[Sulz 2023] Sulz, C., Bleicher, F. Application Scenarios of a Tactile Surface Roughness Measurement System for In Situ Measurement in Machine Tools. *Metrology*. 3(3):280–291, September 2023. doi: 10.3390/metrology3030016.

[Tonejca 2022] Tonejca Née Plessing, L. et al. AI-Based Surface Roughness Prediction Model for Automated CAM-Planning Optimization. 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA) (2022).

[Zha 2023] Zha, D. et al. Data-centric Artificial Intelligence: A Survey. *arXiv*.