Journal of Machine Engineering, 2025, Vol. 25
ISSN 1895-7595 (Print) ISSN 2391-8071 (Online)

Received*: 19 August 2025* / Accepted: *25 September 2025* / Published online: *17 November 2025*

machine acceptance,
anomaly detection,
anomaly classification

Marvin FRISCH[1*], Robin STRÖBEL[1],
Luca PFLITTNER[1], Samuel DEUCKER[1],
Alexander PUCHTA[1], Jürgen FLEISCHER[1]

ANOMALY DETECTION AND CLASSIFICATION FOR WORKER ASSISTANCE DURING MACHINE TOOL ACCEPTANCE

Machine acceptance is a vital part of the manufacturing process, especially for 5-axis machine tools prevalent in the aerospace industry. It is currently done by skilled workers using their experience and knowledge to iteratively improve the machine tool until it is able to manufacture a test piece that meets the required quality standards. This process is time consuming, requires a lot of expertise, and is not easily transferable to new workers. In this paper, we propose a system that uses machine control signals to detect anomalies during the manufacturing of the test piece and classify them by their cause, like an onset of chatter, positional errors, or others. For this, the machine signals are segmented using a sliding window approach. Multiple strategies to reduce the dimensionality of the segments are evaluated, including autoencoders based on a Convolutional Neural Network or a Long-Short Term Memory Network as well as manually designed features. The reduced segments are then classified using a Random Forest. The results show that the proposed system is able to detect anomalies with high accuracy and classify them correctly.

1. INTRODUCTION

The process of machine acceptance is vital to ensuring that a machine tool is able to manufacture workpieces in the desired quality and quantity. A machine tool that does not pass acceptance cannot be sold, especially in aerospace, where the quality standards are very high. For 5-axis milling tools, the prevalent machine type for this industry, the acceptance standard according to ISO 10791-7:2020 [1] is being established. According to it, a test piece (depicted as a Computer Aided Design (CAD) part in Fig. 1 (left) and as a manufactured part in Fig 1 (right) is to be manufactured using the tool. It is then analysed using a coordinate measuring machine. If the piece does not meet the required quality standards, iterative improvements are necessary.

[1] wbk Institute of Production Science, Karlsruhe Institute of Technology, Germany
[*] E-mail: marvin.frisch@kit.edu
https://doi.org/10.36897/jme/211352

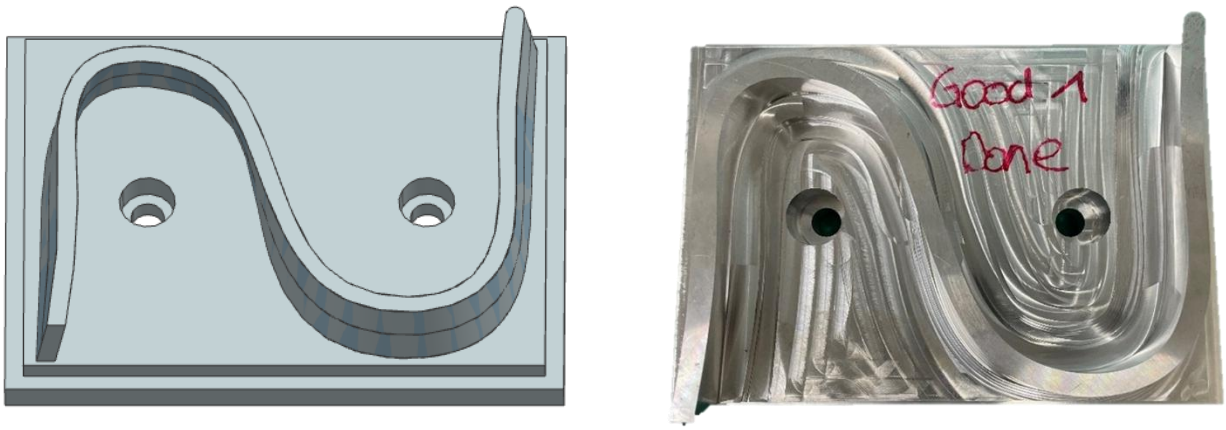

Fig. 1. CAD model of a test piece according to ISO 10791-7:2020 (left). A manufactured test piece (right)

For this, a machine expert uses his acquired knowledge and years of experience to derive parameter adjustments, sensor calibrations or changes in the machine control [1, 2]. There are multiple issues within this approach. Iteratively manufacturing and measuring pieces and especially the reclamping from the machine tool to the measuring machine requires a significant amount of time during which the machine cannot be used productively [3]. Due to the amount of possible causes for problems, even skilled workers with a lot of experience can struggle to find the correct solution [4]. Further, the amount of machine experts able to perform this task is on the decline due to the demographic change [5]. Since the knowledge required to deduce the correct changes from a manufactured piece is not possible to be formulated as a set of rules and the transfer of it from an experienced expert to a new worker cannot be relied upon [6]. In order to facilitate the integration of new workers, to retain knowledge from aging experts, and to minimise reliance on their input, an assistance system for machine tool acceptance is needed.

To reduce the number of measurements on the coordinate measuring machine (and therefore the amount of reclamping), the assistance system needs to detect and classify anomalies using just the information available during manufacture of the piece. The control signals of the feed axes of the machine tool (i.e. current, velocity, position, control difference, etc.) show promise for this purpose. These signals demonstrate significant potential in the areas of process monitoring [7, 8], condition monitoring [9, 10] and the creation and synchronising of digital twins [11–13]. This is due to them being sensitive to changes (both from the machine and the process) and methods based on them being potentially transferable between different machines [14]. They also provide additional insight into the workings of the machine that pure analysis of the manufactured parts cannot. While the part itself cannot be used to derive definitive conclusions about the source of anomalies [1], this additional information can bridge that gap. Furthermore, the machine signals are available for all modern machine tools and are therefore not limited to a specific machine type or manufacturer.

The aim of this paper is to harness the machine control signals in the machine acceptance process, an area they are currently underutilized in. We propose a system that based on anomaly detection and classification algorithms can assist workers of all experience levels to find the problem faster or at all, respectively. It predicts whether a manufactured test piece is satisfactory and fit to be reclamped and formally measured on a coordinate measuring machine. If not, a classification of the likely source of the error is done to shorten the

adjustment process. To best utilize and retain the knowledge of experienced experts, it includes a feedback loop and implements iterative improvements. The structure of the intended system is shown in Fig. 2.

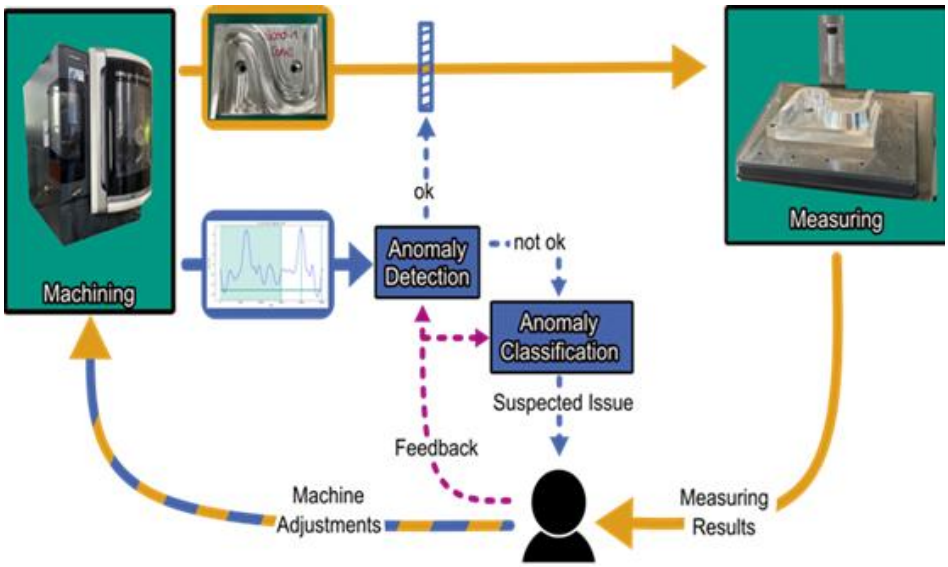

Fig. 2. Structure of a machine acceptance process. The established process in orange, the elements added in this work in blue and purple

2. MATERIALS AND METHODS

2.1. USED DATASET

All data used for the development of the proposed system was recorded on a DMG MORI DMU 65 monoBLOCK (shown in Fig. 3 (left)). The S-shaped test piece proposed by ISO 10791-7:2020 was manufactured multiple times using different machine settings and modifications to generate both "good" and "faulty" labelled datapoints. Five different types of anomalies were induced within the machine:

Chatter: The feedrate of the machine was manipulated in order to induce chatter
PosX: The position value in x was manually changed by +0.05 mm after probing
PosY: The position value in y was manually changed by +0.05 mm after probing
Thermal: Thermal elongations were imitated by using a different tool length
Unbalance: Additional mass was added to the main spindle to induce unbalance (shown in Fig. 3 (right)).

For each of these, one test piece was done. Starting from a rough pre-machined part, the piece was finished in multiple tool passes, each reducing the radial depth of cut a_e (starting at 6 mm and ending at 0.5 mm). The passes are clearly visible in Figure 1 (right). During four of these passes (5 mm, 3.5 mm, 2 mm, and 0.5 mm) the respective anomaly was induced and the machine signals captured. The other passes (6 mm, 4.5 mm, 3 mm, and 1.5 mm) were done normally in order to clear the part. This results in 20 different "faulty" runs.

Additionally, two “good” parts were manufactured in the same way, resulting in 8 “good” runs. The captured signals include the Current i_q (Axes X, Y, Z, A, C, Spindle), the ContourDeviation (Axes X, Y, Z), the ControlDeviation (Axes X, Y, Z, A, C), and the EncoderDifference (Axes X, Y, Z, A, C), each with a sampling frequency of 500 Hz.

Fig. 3. DMG MORI DMU 65 monoBLOCK (left). „Unbalance” anomaly induced by adding a mass on one side of the spindle, causing an unbalanced spindle (right). Even though the mass added is small, the deviations are noticeable in the signals

2.2. DATA PROCESSING AND SEGMENTATION

The preprocessing starts with data scaling. In order to be able to compare the different value ranges of the signals, Min-Max-Scaling [15] was chosen. To segment the time series, the sliding window method [16] was used. It separates a time series X of length T with the format

$$X = (x_1, x_2, \dots, x_T) \tag{1}$$

into N segments of length L and spacing S

$$F_i = (x_{i*S+1}, x_{i*S+2}, \dots, x_{i*S+L});\ i = 0, \dots, N-1; N = \left\lfloor \frac{T-L}{S} + 1 \right\rfloor \tag{2}$$

Since these segments overlap for values of S smaller than L, redundant information is introduced into the dataset. This is a form of data augmentation [17]. The principle is shown in Figure 4. The sliding window method generalizes well and enables an in-process reaction. The best value for the segment length L depends on the use case and the interaction with the other used methods and is therefore determined by experiments. The results for this are shown in Section 3.1.

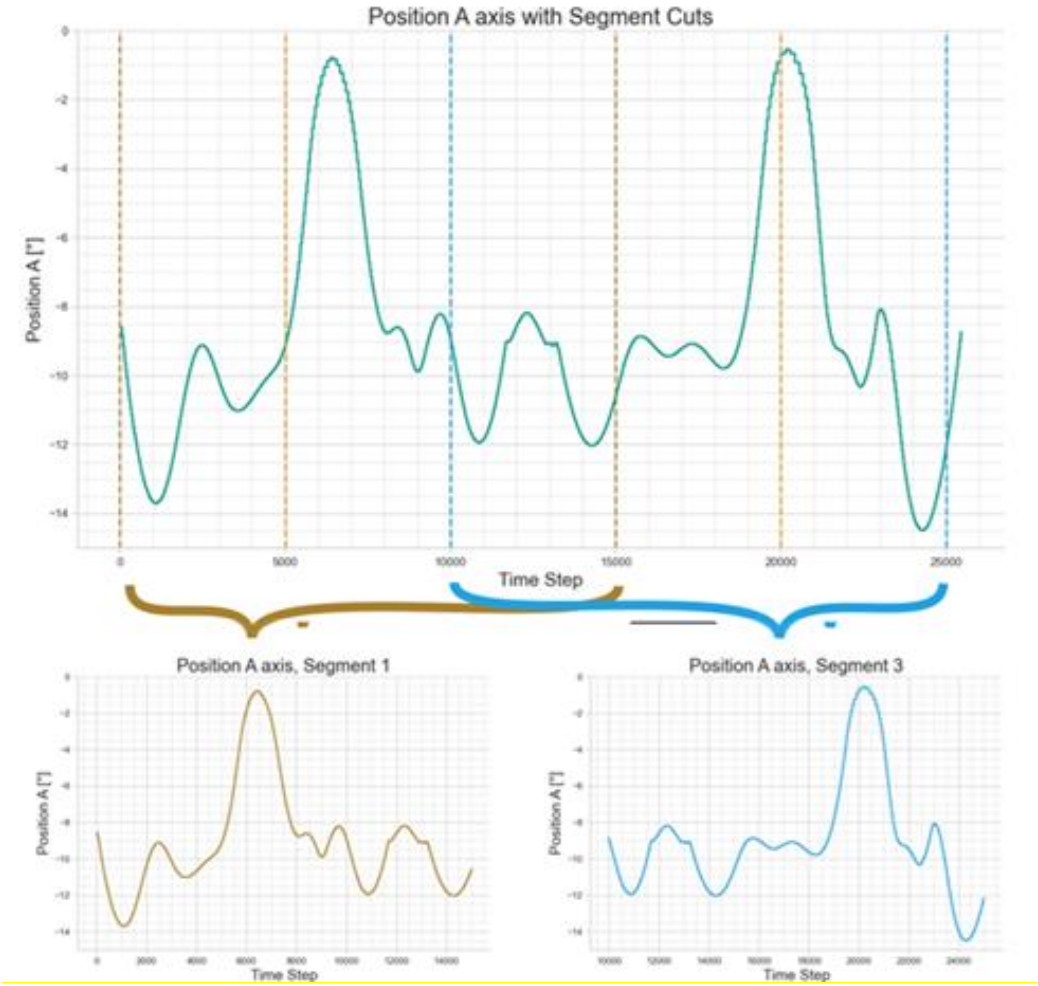

Fig. 4. Segmentation of a time series using the sliding window method for a segment length L = 15000 and spacing S = 5000

2.3. ANOMALY DETECTION

In order to separate anomalous from good segments, multiple different approaches were tested. They include methods belonging to either classical data and time series analysis or Deep Learning. The classical method includes expert features, that is the extraction of the time series features max, mean, root mean square, standard deviation, energy, skewness, kurtosis, peak-to-peak-distance, and the first four coefficients of the Fast Fourier Transform. The binary classification is then done using a Random Forest (RF) approach. The Deep Learning approaches all use autoencoders, that is the compression and subsequent decompression of the raw time series segments. If the autoencoder is trained using only "good" data, the reconstruction error (RE), that is the difference between the compressed and decompressed time series and the original time series, can be used to detect anomalies. Another approach is the clustering and classification of the compressed time series using a separate classification structure. If the encoder is well trained, the features in the embedded space contain all relevant information about the segment. While costlier to design and train,

this approach offers deeper insights into both the workings of the encoder and the nature of the time series itself. For the design of the autoencoder, both a Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) structure were tested. LSTMs are the more "natural" fit for time series data, but CNNs are generally faster in both training and inference and are robust against changes in the trajectory of the segment, since they work by analysing local structures within a time series. Additionally, a self-attention-mechanism was introduced into the embedded space for both structures. This improved the robustness and prediction quality of the models. The results for all the different methods are shown in Section 3.2.

2.4. ANOMALY CLASSIFICATION

To classify the detected anomalies by their cause, the methods described in Section 2.3, that are a RF in combination with an autoencoder as well as in combination with manually chosen expert features, are adapted to fit the multi-class requirement. The RE of the autoencoder methods is used as an additional input for the classification model, where applicable. The RF is used for classification due to its high robustness, low training and inference time, and good performance in general.

2.5. ASSISTANCE SYSTEM AND FEEDBACK LOOP

The assistance system is a central element of the proposed system. It provides the interface between the user and the underlying functionality. This interface works both ways, as the worker is able to label new data points. This gives rise to a new problem: Determining the confidence in this new label. According to [18], the assumption that a human judgement can in general be regarded as correct is only valid for a large sample size. Therefore, the training weight of the newly labelled datapoint needs to be defined. For this, the user needs to provide a judgement of their own experience E and their certainty C on a scale of 1–10. From there, the resulting weight W of the datapoint can be calculated by

$$W = (\alpha * E + \beta * C)^{\gamma} \tag{3}$$

with α, β, and γ as tuneable parameters. The resulting curve values high experience and certainty much stronger than middling values and does not overly punish extremes in either direction. For the "good" class, only datapoints with a maximum certainty value are used, since all "good" parts are verified by the coordinate measuring machine. It is therefore expected that the user enters the maximum certainty value for these verified points.

To be able to retrain the model, all datapoints (and their respective values for E and C) are recorded in a database. In order to save time, the autoencoder is trained incrementally. For that, batches of "good" data are sampled randomly from all datapoints until the model loss reaches a plateau. By not excluding older samples, Catastrophic Forgetting [19] can be avoided. The classification model is retrained from scratch every time due to its fast training speed.

3. RESULTS

In order to decide on the exact algorithms for both the anomaly detection and classification, all in Section 2.3 and Section 2.4 proposed models were implemented and tested on the dataset presented in Section 2.1. The results of these experiments are shown in the following.

3.1 SEGMENT LENGTH

The segment length L was varied in different ranges for the different models (Expert Features, CNN, LSTM, CNN in combination with Expert Features), with a minimum of 4 and a maximum of 60000. The models on trial (and, if necessary, a RF classification model) were then trained on classifying anomalies. The classification was chosen as a benchmark, because it is expected to be a more complex problem than the anomaly detection. The results are shown in Figure 5. With an increase in segment length, the training time of the models decreases due to the lower number of segments. The performance of the models is increasing, but only up to a certain point. After that, the performance decreases again. The chosen segment lengths for all following experiments were 1024 for the LSTM, 8192 for the CNN and the CNN with Expert Features, and 32768 for the Expert Features.

The spacing S was set differently for each segment length. Smaller segment lengths were set to not overlap, while larger segment lengths were set to overlap by up to 50%. This was done to increase the amount of data available without introducing too much redundancy and therefore overfitting the model.

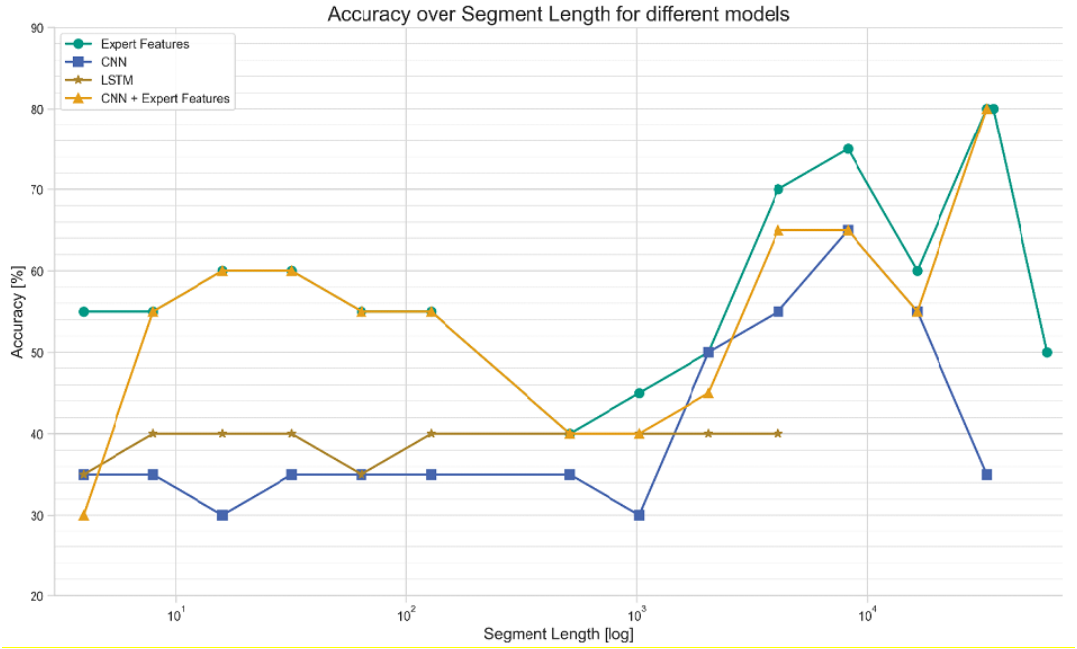

Fig. 5. Performance of the different models with varying segment length. Not shown are the training times, which decrease with increasing segment length

3.2. ANOMALY DETECTION

To assert the anomaly detection performance, the autoencoders were trained on 50 % of the "good" data to ensure a remainder to test on. Because the overall goal is to detect anomalies in a whole manufacturing process, a tool pass is labeled as "good" if all segments within it are "good". The results of the different models for the anomaly detection are shown in Table 1. The models are evaluated using the Balanced Accuracy Score. The autoencoder approaches show a perfect performance, while the Expert Features and CNN with Expert Features are slightly worse.

Table 1. Balanced Accuracy Score of the different models for the anomaly detection. The pure Autoencoders could detect all

LSTM (Only RE)	CNN (Only RE)	CNN (Additional RF)	Expert Features (RF)
100	100	97.5	97.5

3.3. ANOMALY CLASSIFICATION

For the training of the anomaly classification, the models were trained on all but one of the "faulty" passes and then tested on the remaining one. This was done for each of the "faulty" passes. This is to simulate the real use case, where the model is trained on every pass beforehand. The results are shown in Table 2. The models were evaluated using the Accuracy Score.

Table 2. Accuracy Score of the different models for the anomaly classification. The Expert Features show the best performance, the encoder strategies are lacking

LSTM (RF)	CNN (RF)	CNN + Expert Features (RF)	Expert Features (RF)
40	65	65	80

A deeper look into the wrongly classified anomalies reveals that *Chatter* and *Thermal* are confused most often. This is due to the fact that both anomaly types result in vibrations of the tool and therefore similar signals. The anomaly *Chatter_50*, the 5 mm pass of the *Chatter* anomaly, was not detected by the classifiers and most often classified incorrectly.

The *PosX* and *PosY* anomalies are also confused. This is due to the nature of the training, as only one pass is left out for testing. The model generalizes better over the depth of cut than over the type of anomaly. For instance, the *PosX_50* pass is more similar to the *PosY_50* pass than to the *PosX_20* pass.

If the *Thermal* and *Chatter* as well as the *PosX* and *PosY* anomalies are each combined, the performance of all classifiers increases up to an accuracy of 95 %. The *Unbalance* anomaly is not confused with any other anomaly, as it is a completely different type of anomaly.

3.4. ASSISTANCE SYSTEM AND USER INTERFACE

A large part of the proposed system is the user interaction. The necessary algorithms and interfaces are designed to be easy to use and intuitive, while being able to run on an edge

device with limited ressources. The response time for the anomaly classification is 3 to 4.5 seconds, depending on the segment length and model chosen.

The User Interface developed for testing the system is shown in Fig. 6.

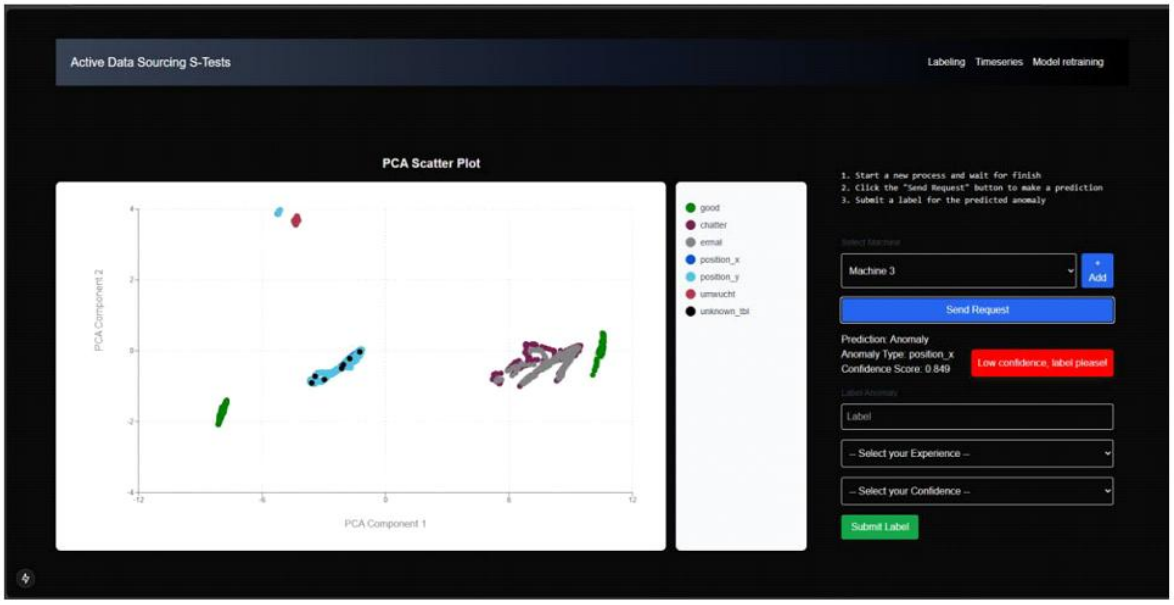

Fig. 6. The proposed User Interface. Due to low model confidence, a label is requested. To better explain the prediction, the PCA of the embedded space is plotted

The Principal Component Analysis (PCA) of the embedded space is computed to visualize the data and the reasoning behind the classification. Iteratively training the whole model takes around 7 minutes, training just the classification model around 40 seconds. All tests were run on a NVIDIA RTX 3060 with a Ryzen 5 3600 CPU and 48 GB of RAM.

4. CONCLUSION

For the anomaly detection, the autoencoder models based on the RE work flawlessly. The CNN is faster in its prediction time (<0.1 s compared to ~0.13 s for the LSTM) and therefore the better choice for the anomaly detection. The Expert Features and CNN with additional Expert Features are slightly worse, but still show a good performance. This is surprising, as the RE is also included as an input to the RF classifier (without it, the accuracy drops by ~10%). The additional information is therefore not necessary for the anomaly detection, but it is still useful for the classification.

The Expert Features with a RF classifier perform the best for the anomaly classification. This is because the autoencoders are trained to minimize the RE, which is not the same as designing an embedded space that is optimal for classification. Introducing additional supervised training into the autoencoders could improve the performance, which is a future research direction. The CNN with additional Expert Features is slightly worse than the Expert Features alone, which is similar to the anomaly detection case. Additional data that is not useful is therefore harmful for the classifier.

The combination of the different anomaly types increasing the performance of the classifiers shows that a classifier trained on distinguishing all anomaly types is not able to separate anomalies that are similar in nature. A possible future research direction could be to use a cascading classifier, where the first classifier separates the anomalies into groups and the second classifier distinguishes between the anomalies within each group. This could improve the performance of the classification, especially for distinguishing the Chatter and Thermal types.

The anomaly Chatter_50 is not detected by any classifier. This is due to the nature of its induction: By changing the feed rate, it is not guaranteed to actually induce chatter. This anomaly might therefore not actually exist.

For the application of the proposed system, the anomaly detection using the CNN with an additional RF classifier and the anomaly classification using an RF trained on the Expert Features is chosen. Both methods are fast and robust. Around four seconds are needed for the anomaly detection and classification using the aforementioned hardware, which is acceptable for a real-time application. The segmentation takes the longest time, but this is not a problem, as it can be done in the background while the machine is running. The proposed system is therefore able to assist workers during the machine acceptance process.

Future research directions include the integration of the proposed system into a real machine acceptance process, where it can be used to assist workers in finding and classifying anomalies. Additionally, extensions of the system to use Contrastive Learning [20] to actively design the embedded space could improve the performance of the anomaly classification, especially if the number of anomaly types increases.

ACKNOWLEDGEMENTS

The APC was funded by the Open Access Publication Fond of the Karlsruhe Institute of Technology (KIT). The data supporting the conclusions of this article will be made available by the authors on request. Marvin Frisch, Robin Ströbel and Luca Pflittner equally contributed to this work.

REFERENCES

[1] ISO 10791-7:2020, International Standard under systematic review 10791–7, Jan. 2020. Accessed: May 19, 2025. [Online]. Available: https://www.iso.org/standard/73814.html.

[2] WILLOUGHBY P., VERMA M., LONGSTAFF A.P., FLETCHER S., 2010, *A Holistic Approach to Quantifying and Controlling the Accuracy, Performance and Availability of Machine Tools*, Proceedings of the 36th International MATADOR Conference, Hinduja S. and Li L., Eds., London: Springer London, 313–316. doi: 10.1007/978-1-84996-432-6_71.

[3] M. RAHMAN, A.B.M.A. ASAD, T. MASAKI, T. SALEH, Y.S. WONG, A. SENTHIL KUMAR, 2009, *A Multiprocess Machine Tool for Compound Micromachining*, International Journal of Machine Tools and Manufacture, 50/4, 344–356, https://doi.org/10.1016/j.ijmachtools.2009.10.007.

[4] ZHANG Z., JIANG F., LUO M., WU B., ZHANG D., TANG K., 2024, *Geometric Error Measuring, Modelling, and Compensation for CNC Machine Tools: A Review*, Chinese Journal of Aeronautics, 37/2, 163–198, https://doi.org/10.1016/j.cja.2023.02.035.

[5] ACEMOGLU D., RESTREPO P., 2022, *Demographics and Automation*, The Review of Economic Studies, 89/1, 1–44, https://doi.org/10.1093/restud/rdab031.

[6] ELLWART T., BÜNDGENS S., RACK O., 2013, *Managing Knowledge Exchange and Identification in Age Diverse Teams*, Journal of Managerial Psych, 28/7/8, 950–972, https://doi.org/10.1108/JMP-06-2013-0181.
[7] TETI R., MOURTZIS D., D'ADDONA M.D., CAGGIANO A, 2022, *Process Monitoring of Machining*, CIRP Annals, 71/2, 529–552, 2022, https://doi.org/10.1016/j.cirp.2022.05.009.
[8] BUGDAYCI N.B., WEGENER K., POSTEL M., 2022, *Monitoring of the Average Cutting Forces from Controller Signals Using Artificial Neural Networks*, Journal of Machine Engineering, https://doi.org/10.36897/jme/154801.
[9] SURUCU O., GADSDEN S.A., YAWNEY J., 2023, *Condition Monitoring Using Machine Learning: a Review of Theory, Applications, and Recent Advances*, Expert Systems with Applications, 221, 119738, https://doi.org/10.1016/j.eswa.2023.119738.
[10] ZEGARRA F.C., VARGAS-MACHUCA J., ROMAN-GONZALEZ A., CORONADO A.M., 2023, *Unsupervised and Supervised Machine Learning Methods for Cutting Tool Path Clustering and RUL Estimation in Manufacturing*, Journal of Machine Engineering, https://doi.org/10.36897/jme/171432.
[11] STRÖBEL R., BOTT A., WORTMANN A., FLEISCHER J., 2023, *Monitoring of Tool and Component Wear for Self-Adaptive Digital Twins: A Multi-Stage Approach Through Anomaly Detection and Wear Cycle Analysis*, Machines, 11/11, 1032, https://doi.org/10.3390/machines11111032.
[12] DENKENA B., BERGMANN B., HANDRUP M., WITT M., 2020, *Material Identification During Turning by Neural Network*, Journal of Machine Engineering, 20/2, 65–76, https://doi.org/10.36897/jme/119677.
[13] DEMETGÜL M., GU M., JONAS H., ZHAO Y., GÖNNHEIMER P., FLEISCHER J., 2022, *Misalignment Detection on Linear Feed Axis with FFT and Statistical Analysis Using Motor Current*, Journal of Machine Engineering, https://doi.org/10.36897/jme/147699.
[14] DENKENA B., KLEMME H., STIEHL T.H., 2023, *Failure Sensitivity and Similarity of Process Signals Among Multiple Machine Tools*, Procedia CIRP, 120, 922–927, 2023, https://doi.org/10.1016/j.procir.2023.09.101.
[15] AHSAN M., MAHMUD M., SAHA P., GUPTA K., SIDDIQUE Z., 2021, *Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance*, Technologies, 9/3, 52, https://doi.org/10.3390/technologies9030052.
[16] CHU C.-S.J., 1995, *Time Series Segmentation: a Sliding Window Approach*, Information Sciences, 85/1–3, 147–173, https://doi.org/10.1016/0020-0255(95)00021-G.
[17] DANIEL T., CASENAVE F., AKKARI N., RYCKELYNCK D., 2021, *Data Augmentation and Feature Selection for Automatic Model Recommendation in Computational Physics*, MCA, 26/1, https://doi.org/10.3390/mca26010017.
[18] AMIN H., LU Z., YIN M., 2023, *Give Weight to Human Reactions: Optimizing Complementary AI in Practical Human-AI Teams*, Workshop on Aritificial Intelligence and Human Computer Interaction at the 4th Internatinoal Conference on Machine Learning (ICML), Honolulu, Hawaii.
[19] FRENCH R. M., 1999, *Catastrophic Forgetting in Connectionist Networks*, Trends in Cognitive Sciences, 3/4.
[20] KHOSLA P. et al., 2020, *Supervised Contrastive Learning*, Advances in Neural Information Processing Systems, Curran Associates, 18661–18673. Accessed: Jul. 07, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-bstract.html.