Explainable Artificial Intelligence in Mechanical Engineering: A Synthetic Dataset for Comprehensive Failure Mode Analysis

Raphael Wallsberger, Ricardo Knauer, Stephan Matzka
University of Applied Sciences Berlin
School of Engineering II – Technology and Life
12459 Berlin, Germany
{raphael.wallsberger, ricardo.knauer, stephan.matzka}@htw-berlin.de

***Abstract*—Transparency and comprehensibility of artificial intelligence (AI) algorithms is of central importance across vertical industries. However, comprehensive datasets for evaluating explainable AI (XAI) methods are lacking in mechanical engineering. This paper introduces a new synthetic dataset for failure mode analysis in drilling that is generated with real-world applicability in mind. It is optimized for XAI purposes and can be used to evaluate explanations of such methods. We show that local explanations and overall feature importance scores derived from SHapley Additive exPlanations (SHAP) values match expert-defined failure modes on our data, and thus lay the groundwork for future XAI research in the field.**

***Keywords*—explainable artificial intelligence, SHAP, synthetic dataset generation, drilling**

I. Introduction

The integration of artificial intelligence (AI) across industries has been a transformative force, reshaping traditional methodologies and introducing innovative solutions. Mechanical engineering, a discipline deeply rooted in precision, analysis, and optimization, stands to benefit immensely from AI. While AI's predictive capabilities have been utilized in numerous applications such as predictive maintenance and failure mode detection [1], its potential in providing transparent, understandable, and explainable solutions, known as explainable AI (XAI) [2], is underrepresented in mechanical engineering research, mainly due to the lack of domain-specific datasets that allow for systematic evaluations of XAI methods.

Drilling processes are a fundamental aspect of mechanical engineering that have evolved over the years through the incorporation of advanced technologies such as using carbide material or drill bit coating to increase efficiency and accuracy [3]. However, a significant gap exists: the absence of a comprehensive dataset optimized for XAI in this field. Such a dataset can be the foundation for trustworthy explanations, which can used to identify and resolve problem causes, even for non-subject matter experts. This is crucial for real-world applications where errors can be costly.

This paper aims to bridge the current gap by presenting a synthetic dataset meticulously designed for XAI in drilling processes [4]. This dataset serves as a template for future research, providing a baseline for evaluating AI explanations. Therefore, the dataset does not only allow for predicting significant drill bit failure, but also allows for understanding the decision-making of the underlying (X)AI algorithms. Due to the expert-driven data synthesis, the algorithmic explanations can be checked against the true factors contributing to a drill bit failure.

This paper is structured as follows: In section II, each feature of the synthetic dataset is described in detail and the generation methodology with the underlying equations is shown. Section III provides an overview of the selection, training, and evaluation of an exemplary baseline classifier. Furthermore, a first XAI analysis is reported. The paper concludes in section IV with a discussion on the potential applications, challenges, and future usage of the dataset.

II. XAI Dataset

A. Description of the dataset

The heart of this research lies in the careful design of a synthetic dataset for XAI methods in drilling processes, specifically failure detection on drill bits. It is intended to be a comprehensive representation of real-world drilling scenarios, designed to facilitate in-depth analysis and explanations. The dataset comprises 20,000 data points, i.e., drilling operations, stored as rows; 10 features, one binary main failure label, and 4 binary subgroup failure modes, stored in columns. The features that constitute this dataset are as follows:

- ID: Every data point in the dataset is uniquely identifiable, thanks to the ID feature. This ensures traceability and easy referencing, especially when analyzing specific drilling scenarios or anomalies.
- Cutting speed vc (m/min): The cutting speed is a pivotal parameter in drilling, influencing the efficiency and quality of the drilling process. It represents the speed at which the drill bit's cutting edge moves through the material.
- Spindle speed n (1/min): This feature captures the rotational speed of the spindle or drill bit, respectively.
- Feed f (mm/rev): Feed denotes the depth the drill bit penetrates into the material with each revolution. There is a balance between speed and precision, with higher feeds leading to faster drilling but potentially compromising hole quality.
- Feed rate vf (mm/min): The feed rate is a measure of how quickly the material is fed to the drill bit. It is a determinant of the overall drilling time and influences the heat generated during the process.
- Power Pc (kW): The power consumption during drilling can be indicative of the efficiency of the process and the wear state of the drill bit.

- Cooling (%): Effective cooling is paramount in drilling, preventing overheating and reducing wear. This ordinal feature captures the cooling level applied, with four distinct states representing no cooling (0%), partial cooling (25% and 50%), and full cooling (100%).

- Material: The type of material being drilled can significantly influence the drilling parameters and outcomes. This dataset encompasses three primary materials: C45K hot-rolled heat-treatable steel (EN 1.0503), cast iron GJL (EN GJL-250), and aluminum-silicon (AlSi) alloy (EN AC-42000), each presenting its unique challenges and considerations. The three materials are represented as "P (Steel)" for C45K, "K (Cast Iron)" for cast iron GJL and "N (Non-ferrous metal)" for AlSi alloy.

- Drill bit type: Different materials often require specialized drill bits. This feature categorizes the type of drill bit used, ensuring compatibility with the material and optimizing the drilling process. It consists of three categories, which are based on the DIN 1836: "N" for C45K, "H" for cast iron and "W" for AlSi alloy [5].

- Process time t (s): This feature captures the full duration of each drilling operation, providing insights into efficiency and potential bottlenecks.

- Main failure: This binary feature indicates if any significant failure on the drill bit occurred during the drilling process. A value of 1 flags a drilling process that encountered issues, which in this case is true when any of the subgroup failure modes are 1, while 0 indicates a successful drilling operation without any major failures. The main failure rate is about 5.0 % for the whole dataset.

Subgroup failures:

- Build-up edge failure (215x): Represented as a binary feature, a build-up edge failure indicates the occurrence of material accumulation on the cutting edge of the drill bit due to a combination of low cutting speeds and insufficient cooling. A value of 1 signifies the presence of this failure mode, while 0 denotes its absence.

- Compression chips failure (344x): This binary feature captures the formation of compressed chips during drilling, resulting from the factors high feed rate, inadequate cooling and using an incompatible drill bit. A value of 1 indicates the occurrence of at least two of the three factors above, while 0 suggests a smooth drilling operation without compression chips.

- Flank wear failure (278x): A binary feature representing the wear of the drill bit's flank due to a combination of high feed rates and low cutting speeds. A value of 1 indicates significant flank wear, affecting the drilling operation's accuracy and efficiency, while 0 denotes a wear-free operation.

- Wrong drill bit failure (300x): As a binary feature, it indicates the use of an inappropriate drill bit for the material being drilled. A value of 1 signifies a mismatch, leading to potential drilling issues, while 0 indicates the correct drill bit usage.

B. *Generation of the dataset*

The synthetic dataset is based on a simple drilling process. The same type of hole is drilled for three different materials (C45K, cast iron, AlSi alloy). The hole is on solid material, a blind hole, the diameter D is 12 mm, the depth is 60 mm. According to the material, a fitting drill bit type (N, H, W) is chosen, each drill bit has two flutes, and the cooling is externally with an emulsion of 5 %. This process data was put into the Walter GPS online tool [6], and three appropriate drill bits (all high-speed steel - HSS material) were selected. Walter GPS is a machining navigation system that offers an optimum machining solution for specific components and processes. The resulting parameters (Table I) were used as a starting point to ensure a realistic representation of drilling operations.

TABLE I. EXPORT DATA FROM THE WALTER GPS ONLINE TOOL [6]

Material	vc (m/min)	f (mm/rev)	t (sec)
C45K	22.60	0.180	0:35
GJL-250	17.40	0.252	0:33
AlSi	29.99	0.243	0:19,7

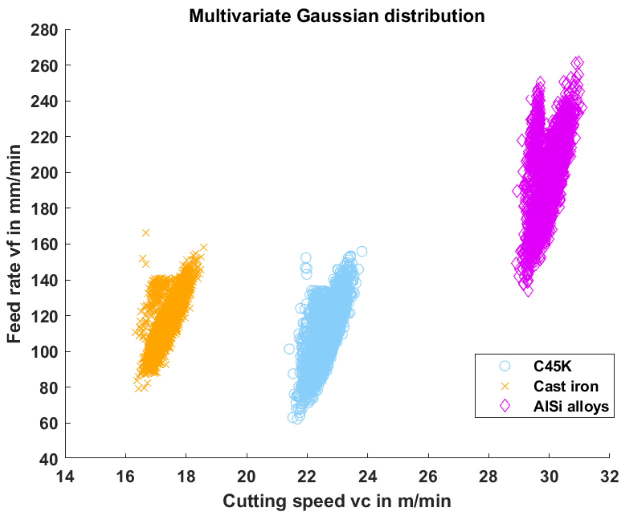

Fig. 1. Multivariate Gaussian distribution of the features cutting speed and feed rate, three clusters represent the different materials.

In the beginning, the two features cutting speed vc and feed f were generated using a multivariate Gaussian distribution. The outputs for vc and f of the Walter GPS online tool (Table I) were used as initial values for the means, for each material separately. Vc and f are correlated through the spindle speed n which is why a multivariate Gaussian distribution was considered reasonable (Fig. 1). The process time t was generated with a univariate Gaussian distribution using the initial values from Table I for the means as well. The rest of the features were calculated according to the following commonly used equations in the manufacturing industry [7]:

$$n = \frac{v_C \cdot 1000}{D \cdot \pi} \quad \text{in 1/min} \tag{1}$$

vc - cutting speed, D - drill bit diameter, n - spindle speed

$$v_f = f \cdot n \quad \text{in m/min} \tag{2}$$

vf - feed rate, f - feed per revolution, n - spindle speed

$$F = f \cdot D \cdot \pi \cdot n \quad \text{in mm/min} \tag{3}$$

F - drilling force, f - feed per revolution, D - drill bit diameter, n - spindle speed

$$P_C = \frac{F \cdot v_C}{1000} \quad \text{in kW} \tag{4}$$

Pc - cutting power, F - drilling force, vc - cutting speed

After generating the features, each of the feature values was rounded and the subgroup failure modes were added according to expert knowledge.

The drilling wear types build-up edge (BUE), compression chips and flank wear are commonly known problems [8,9] when drilling independent of the tool material (HSS or carbide). BUE is generally caused due to low cutting speed and insufficient cooling, compression chips can have up to three reasons: feed rate to high, cooling to low and a wrong drill bit geometry (here wrong drill bit). If the feed rate is too high and the cutting speed too low, flank wear occurs. For each of these three failure categories certain limits were set, according to the specific materials, to define certain data points as true values. For compression chips at least two out of the three causes needed to occur to set it. Wrong drill bit failure is caused by an inadequate drill bit type according to the material.

To enhance the dataset's realism and applicability, several adjustments were made. For instance, cooling classes were modified based on the means of cutting speed and feed rate, simulating the dynamic nature of cooling in drilling processes. Specific rows underwent changes in feed values, introducing anomalies to simulate real-world challenges and deviations. Firstly, for randomly selected 10 % of the rows with low cooling values (<= 25%), the feed rate was increased by 30 %. Secondly, the feed rate was increased by 30 % for 30 % of the rows that had a cutting speed beneath the low limit according to the material. Lastly, 1.5 % of the data points were randomly assigned a different drill bit type. These specific changes were necessary to gain a balanced number of the failure modes.

III. Model Training and XAI Analysis

A. *Model Selection and Training*

Since our dataset is highly imbalanced, only 5 % of the 20,000 data points belong to class 1, an algorithm that can deal with that issue needed to be selected. An effective type of algorithms for imbalanced data are ensemble methods (bagging or boosting) [10]. As our XAI analysis in III B requires a tree-based algorithm for its fast implementation, the XGBoost algorithm (version 1.7.6) was considered a reasonable choice, using the default hyperparameter settings. Furthermore, a combination of over- and undersampling was performed, as suggested in [11]. The combination of Synthetic Minority Oversampling Technique (SMOTE) and Tomek Link during preprocessing lead to a balanced class distribution [12].

The overall discriminatory performance for main failure detection was excellent, with an area under the receiver operating characteristic curve of 0.995. This was to be expected considering the synthetic nature of the dataset and failure modes. The confusion matrix can be seen in Table II.

TABLE II. Confusion Matrix for the XGBoost Classifier on our Preprocessed Data

		Predicted	
	Class	***0***	***1***
Actual	***0***	**3759**	**3**
	1	**31**	**3801**

B. *XAI Analysis*

One of the most commonly used XAI algorithms is the SHapley Additive exPlanations (SHAP) value [13,14]. On the one hand, it has a solid theoretical foundation in coalitional game theory; on the other hand, it can produce local explanations as well as overall feature importance scores [15]. The TreeSHAP algorithm is a fast implementation for tree-based algorithms like XGBoost. Therefore, we used TreeSHAP as an XAI algorithm on our trained XGBoost classifier to gain a first insight on the usability of our data as an XAI evaluation dataset.

The overall feature importance scores are shown in Fig. 2. The four primary contributing factors for our subgroup failure modes from section II (feed rate, cutting speed, drill bit type, and cooling) are indeed among the five most important features according to the TreeSHAP algorithm.

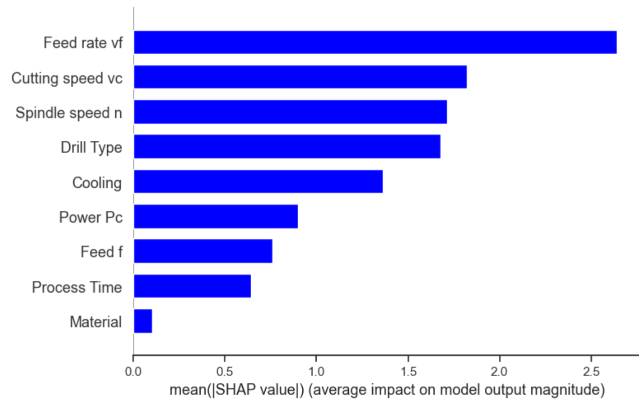

Fig. 2. TreeSHAP feature importance summary plot.

In section II, we also stated that a high feed rate in combination with a low cutting speed defines a flank wear failure. Fig. 3 provides an exemplary local explanation plot for a flank wear failure prediction. The two most important factors for the decision are feed rate and cutting speed, as expected. Thus, the pre-defined rules for that failure mode are in line with the TreeSHAP results.

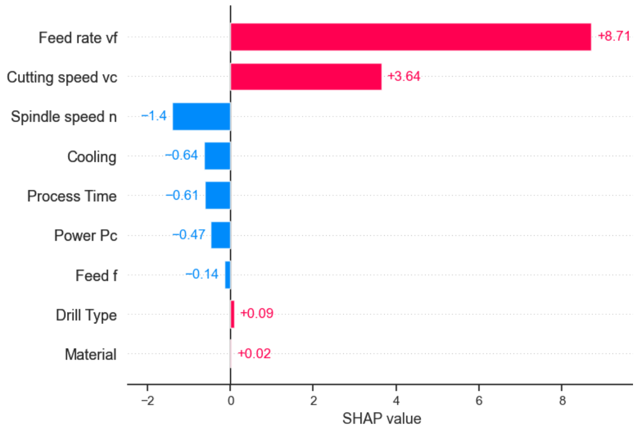

Fig. 3. TreeSHAP local feature importance bar plot for a flank wear failure prediction.

IV. Conclusion and Future Work

The potential of XAI methods in mechanical engineering research has yet to be realized. For drilling, a comprehensive dataset that is optimized for XAI evaluations has been missing so far. This paper presents a synthetic dataset of a drilling process, with a comprehensive feature set and real-world applicability. Since the various failure modes and reference values are based on expert knowledge, the dataset can be used not only for predictive modeling, but also for in-depth explanations. Our XAI analysis shows that local explanations as well as overall feature importance scores derived from SHAP values indeed match pre-defined failure modes. The dataset acts as a blueprint for researchers and domain experts aiming to evaluate XAI algorithm explanations.

Future work will provide a more in-depth analysis of the different failure modes and an alternative visualization approach as shown in [16]. Another focus will be on the evaluation and performance of various XAI algorithms, such as Shapley Additive Global ImportancE (SAGE) values, Shapley effects, or glass box models [17-19]. While the current dataset is comprehensive, there is some scope for enhancement. Future iterations could incorporate more materials, drill bit specifications, and even simulate more complex drilling scenarios. Furthermore, it is planned to make the dataset available as a function that can be integrated and adapted as needed.

References

[1] T. Zonta, C. André da Costa, R. da Rosa Righi, M. José de Lima, E. Silveira da Trindade, G. Pyng Li, “Predictive maintenance in the Industry 4.0: A systematic literature review” Computers & Industrial Engineering, Volume 150, 2020.

[2] D. Gunning, D. Aha, “DARPA’s Explainable Artificial Intelligence (XAI) Program”, AI Magazine, vol. 40, no. 2, Jun. 2019, pp. 44-58.

[3] B. Awiszus, J. Bast, T. Hänel, M. Kusch, “Grundlagen der Fertigungstechnik”, 7th edition, Carl Hanser Verlag, 2020, pp. 137-140.

[4] R. Wallsberger, R. Knauer, S. Matzka, “XAI drilling dataset”, Dataset, https://www.kaggle.com/datasets/raphaelwallsberger/xai-drilling-dataset, 2023.

[5] DIN 1836, “Groups of tool application for chip removal”, Beuth Verlag, 1984.

[6] Walter AG, Online Tool Walter GPS 6.7.1, https://www.walter-tools.com/en-gb/gps (accessed Aug. 24, 2023)

[7] W. Degner, H. Lutze, E. Smejkal, U. Heisel, J. Rothmund, “Spanende Formung: Theorie - Berechnung – Richtwerte”, 18th edition, Carl Hanser Verlag, 2019

[8] Sandvik Coromant, Online metal cutting knowledge library, https://www.sandvik.coromant.com/en-us/knowledge/drilling/drilling-wear-and-troubleshooting (accessed Aug. 24, 2023)

[9] Hofmann Group - Garant, “Machining handbook”, self published, 2022

[10] He Haibo, Ma Yunqian, “Imbalanced Learning: Foundations, Algorithms, and Applications”, Wiley, 2013

[11] Ruisen, Luo & Dian, Songyi & Wang, Chen & Peng, Cheng & Zuodong, Tang & YanMei, Yu & Shixiong, Wang, “Bagging of Xgboost Classifiers with Random Under-sampling and Tomek Link for Noisy Label-imbalanced Data.”, IOP Conference Series: Materials Science and Engineering, No. 428, 2018

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique”, Journal of Artificial Intelligence Research 16, 2002

[13] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, W. Samek, “Explainable AI Methods - A Brief Overview.”, In: “xxAI - Beyond Explainable AI “ Lecture Notes in Computer Science, vol 13200, 2022, pp. 13-38.

[14] S. Lundberg and S. Lee, “A Unified Approach to Interpreting Model Predictions”, 31st Conference on Neural Information Processing System, 2017

[15] C. Molnar, “Interpretable Machine Learning: A Guide for Making Black Box Models Explainable”, 2nd edition, https://christophm.github.io/interpretable-ml-book/, 2022

[16] R. Wallsberger, T. D. Eberhardt, P.-A. Bartlau, M. L. Dörnte, T. L. Schröter, S. Matzka, “Explainable Artificial Intelligence for a high dimensional condition monitoring application using the SHAP Method“, 5th International Conference on Artificial Intelligence for Industries (AI4I), 2022

[17] I. C. Covert, S. Lundberg, S. Lee, “Understanding Global Feature Contributions With Additive Importance Measures”, 34th Conference on Neural Information Processing Systems, 2020

[18] E. Song, B. L. Barry, J. Staum, “Shapley Effects for Global Sensitivity Analysis: Theory and Computation”, SIAM/ASA Journal on Uncertainty Quantification, 4th edition, 2016, pp. 1060-1083

[19] H. Nori, S. Jenkins, P. Koch, R. Caruana, “InterpretML: A Unified Framework for Machine Learning Interpretability”, https://arxiv.org/abs/1909.09223 , 2019