

Expert-Informed Hierarchical Diagnostics of Multiple Fault Modes of a Spacecraft Propulsion System

Osarenren Kennedy Aimiyekagbon¹, Alexander Löwen², Amelie Bender³, Lars Muth⁴, Walter Sextro⁵

^{1,2,3,4,5} *Paderborn University, Faculty of Mechanical Engineering,
Chair of Dynamics and Mechatronics, Warburger Str. 100, 33098 Paderborn, Germany
{firstname.lastname}@uni-paderborn.de*

ABSTRACT

This paper presents a comprehensive study on diagnosing a spacecraft propulsion system utilizing data provided by the Prognostics and Health Management (PHM) society, specifically obtained as part of the Asia-Pacific PHM conference's data challenge 2023. The objective of the challenge is to identify and diagnose known faults as well as unknown anomalies in the spacecraft's propulsion system, which is critical for ensuring the spacecraft's proper functionality and safety. To address this challenge, the proposed method follows a systematic approach of feature extraction, feature selection, and model development. The models employed in this study are kMeans clustering and decision trees combined to ensembles, enriched with expert knowledge. With the method presented, our team was capable of reaching high accuracy in identifying anomalies as well as diagnosing faults, resulting in attaining the seventh place with a score of 93.08 % in the data challenge.

1. INTRODUCTION

Technical systems, such as production systems, wind turbines and spacecraft are increasingly complex with multiple components, which can fail unexpectedly due to various faults. Enhanced diagnostics and prognostics are cornerstones of PHM to avoid or mitigate such failures that typically have economical, human and environmental impacts. On one hand, enhanced diagnostics involves early detection of faults before manifestation into failure. Furthermore, it involves isolating the faulty component and identifying the severity of such faults (Katipamula & Brambley, 2005). On the other hand, prognostics involve prediction of the evolution of detected faults and the estimation of the remaining useful life (RUL). This paper focuses on diagnostics as an enabler of prognostics, through transparent and early fault detection, isolation and identification. Although not an exhaustive literature review of all available fault diagnostics methods, an overview of the

general methods is given in the following paragraphs.

Zhang et al. (Zhang, Zhang, Wang, & Habetler, 2020) provided an in-depth literature review on deep learning methods for automatic feature extraction and unsupervised diagnostics. Although deep learning methods yield high accuracy, they require large amount of training data and it is hard to explain the learned relationships.

Rodríguez Ramos et al. (Rodríguez Ramos et al., 2019) employed a fuzzy-logic-based approach to successfully diagnose single, multiple and unknown faults of a simulated continuous-stirred tank heater. However, their approach builds on the residual between the system output and the output of a first principle model. It therefore requires measurements of the system inputs. The proposed method is infeasible, where the system inputs are not synchronously measured with the system output. Furthermore, high-fidelity models can become very costly to evaluate for complex systems.

Gao et al. (Gao, Yang, Xing, & Xu, 2012) utilized principal component analysis for dimensionality reduction and subsequently diagnosed multiple faults on a space satellite via multi-class support vector machine (SVM) with high prediction accuracy. The multi-class SVM involved a one-vs.-all scheme and was trained with sufficient amount of data. A limitation of the proposed method is that mixed fault modes are not accounted for.

A main contribution of this paper is the realization of a comprehensive diagnostics methodology by appropriately incorporating several machine learning techniques and expert knowledge. Specifically, an ensemble of decision trees is utilized to detect and isolate known faults and k-means is employed to cluster known faults from unknown faults. The next section describes the employed feature extraction and selection methods for time series data. Furthermore, a brief overview of decision trees is provided. Afterwards, the diagnostics methodology is evaluated based on an experimental spacecraft propulsion system. Subsequently, the results are laid out and the main findings are summarized.

Osarenren Kennedy Aimiyekagbon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. THEORETICAL FRAMEWORK

In this section, a general approach is described to address the diagnostics tasks comprising fault detection, fault isolation, and fault identification. The framework involves feature extraction, feature selection, and model development. These pivotal steps form the basis of the presented diagnostic methodology to construct a highly accurate and reliable model. This chapter outlines the fundamental principles and tools guiding the adopted diagnostic approach in the spacecraft propulsion system.

Feature extraction

Feature extraction is employed to unveil meaningful and significant underlying information from condition monitoring data. Commonly considered domains for extracting features from data of technical systems are time, frequency and time-frequency domain features (Kimotho & Sextro, 2014; Nie, Zhang, Xu, Cai, & Yang, 2022). Many toolboxes are available to automatically extract features from time series data. A comprehensive feature extraction toolbox is the MATLAB package HCTSA, short for highly comparative time series analysis (B. Fulcher et al., 2020). This package includes more than 7700 features across all domains, which are extracted from univariate time series (B. D. Fulcher, Little, & Jones, 2013). It comprises more features than available alternative packages such as the python package tsfresh, short for time series feature extraction on basis of scalable hypothesis tests (Christ, Braun, Neuffer, & Kempa-Liehr, 2018). Condition monitoring signals typically comprises waveform signals, making it a suitable candidate for applying the HCTSA-toolbox to extract relevant features from the aforementioned domains. Furthermore, it has been applied successfully on several technical systems (Aimiyekagbon, Bender, & Sextro, 2021; Hennig, Grafinger, Gerhard, Dumss, & Rosenberger, 2020).

Feature selection

Feature selection is necessary for dimensionality reduction and to exclude irrelevant features from the model development. This enables efficient model training using machine learning algorithms and keeps the model size required to generate accurate and reliable results moderate. To generate even more effective models, expert knowledge should also be incorporated. Feature selection can be divided into filter, wrapper, embedded, and hybrid methods, whereby filter methods are generally faster than the other methods (Hoque, Bhattacharyya, & Kalita, 2014). As the tasks are dominated by classification problems, the Chi-Square test is employed to select the most relevant features. The Chi-Square test is a statistics-based filter method used to determine the independency of two statistical variables. It was first used by Karl Pearson in 1900 (Pearson, 1900). The Chi-Square test statistic can be computed as

$$\chi^2 = \sum_{i=0}^{n_v} \sum_{j=0}^{n_c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (1)$$

where $O_{i,j}$ are the observed counts and $E_{i,j}$ are the expected counts of a value of feature i with n_v different occurring feature values. The Chi-Square metric is calculated for each feature, considering n_c distinct classes. A higher Chi-Square value indicates increased feature importance (Magnello, 2005; Liu & Setiono, 1995; Li et al., 2017).

Modeling

Two methods utilized to generate models to accomplish the tasks are described in the following.

- **Decision trees**

Decision trees are simple and interpretable models. They consist of nodes and branches, leading to leaf nodes. The nodes contain conditions or thresholds based on specific features, which determine the data split along the tree. Due to the tree-like structure, the decision rules and the path to the final decision are easy to follow (Loh, 2011). The construction of decision trees is usually based on impurity. The purer a node is, the better the data can be split with respect to the classes. One measure for impurity is the Gini impurity measure

$$G = 1 - \sum p_k^2 \quad (2)$$

where p_k denotes the proportion of data points belonging to class k in one node. Another metric is the entropy impurity measure. Since the choice of metric does not have a significant impact (Géron, 2019; Loh, 2011), the Gini impurity measure is applied. The decision-making process of decision trees can be easily refined by contributing expert knowledge. An option is feature selection, as already described. Another option is to have experts review the models, which for example were automatically generated by machine learning algorithms, and make adjustments according to their preferences by slightly modifying certain paths or decision rules (Tsai, King, Higgins, Pierce, & Patel, 1997).

- **Ensembles**

Ensembles generally consist of several estimators to improve estimation performance. There exist several methods to combine estimators and one important method is bagging, short for bootstrap aggregating. Bagging means the aggregation of multiple similar models trained on a subset of data each. A decision is made usually through voting. A prominent example are Random Forests, which are among today's most powerful machine learning algorithms, consist of individually trained decision trees (Breiman, 1996; Géron, 2019). In general, however, a wide variety of classifiers can be combined to create an ensemble (Géron, 2019). The successful utilization of a combined approach involving expert knowledge-based rules and decision trees by Aimiyekagbon et al. (Aimiyekagbon, Muth, Wohlleben, Bender, & Sextro, 2021) is acknowledged and taken into account in this

study.

3. CASE STUDY

In the following, the case study of an experimental propulsion system is considered. The data set is provided by the PHM society as part of the Asia-Pacific PHM conference's data challenge 2023. The motivation of the data challenge is to improve PHM technology for spacecraft propulsion systems. Such systems are relevant for current and future scientific and maybe even touristic flights in the cosmic space. In the following subsections, the data acquisition and the data challenge tasks are described and subsequently the evaluation of proposed method is presented.

3.1. Experimental Procedure

Water hammer is a phenomenon, which occurs at start-up of a spacecraft during the rapid opening and closing of valves on a liquid-carrying pipeline, and is characterized by sudden pressure peaks and fluctuations. The sudden pressure peak due to water hammer can result in catastrophic failures (Bandyopadhyay & Majumdar, 2014; Bombardieri, Traudt, & Manfletti, 2019; Al-Khayat, Al-Fatlawi, Al-Baghdadi, & Al-Waily, 2022; Tominaga et al., 2023). Water hammer is generated in the experimental propulsion system by opening and closing solenoid valves (PHM Society, 2023). Different opening ratios of the valves are used to simulate valve faults. Furthermore, air bubbles are introduced into the system at different locations. They affect the properties of the hydraulic medium and the characteristics of the pressure fluctuations.

In the experiments, water is considered as the working fluid and is pressurized to 2 MPa. The behavior of thrusters is simulated by solenoid valves SV1 through SV4. To facilitate the development of diagnostics and prognostics algorithms, eight pressure sensors are located near accumulators and valves at the positions P1 through P7, as depicted in Figure 1.

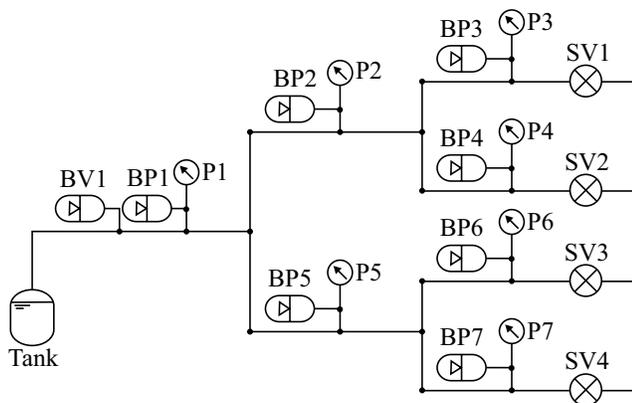


Figure 1. Schematic of the experimental propulsion system, depicting pressure sensor locations (PHM Society, 2023)

Every experiment is 1.2 s long. During this time window the signals of the pressure sensors are recorded at a sampling rate of 1 kHz and three complete phase of solenoid valves opening and closing are observed, as shown in Figure 2.

3.2. Data Description

The experimental time series data set comprises raw pressure sensor data collected under four operational settings, which simulates different spacecraft. Under every operational setting, it encompasses normal working conditions and different fault modes of the experimental propulsion system. The fault modes are bubble anomalies and solenoid valve faults, as previously described. Although the data set is relatively small, it is balanced with approximately equal number of instances of each class. The training data includes normal and abnormal data from spacecraft 1 through 3, with a total of 177 instances. However, the test data comprises 46 instances of the spacecraft 1 through 4. Furthermore, undisclosed and mixed fault modes are included in the test data set to account for real-world scenarios.

3.3. Task Description

Five tasks were defined in the data challenge:

1. Classify normal and abnormal data.
2. If the data is classified as faulty, then declare the fault type.
3. If a bubble anomaly is detected, then determine the location of the bubbles (BV1, BP1 through BP7).
4. If a solenoid valve fault is detected, then determine which valve failed (SV1 through SV4).
5. For each failed valve, predict its opening ratio between 0 % and 100 %.

In terms of fault diagnostics, the first task can be described as fault detection, which requires a binary decision.

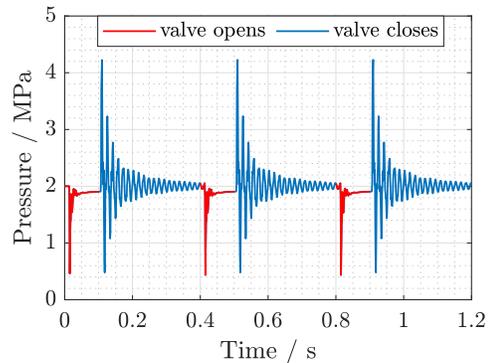


Figure 2. Exemplary pressure sensor signal (PHM Society, 2023)

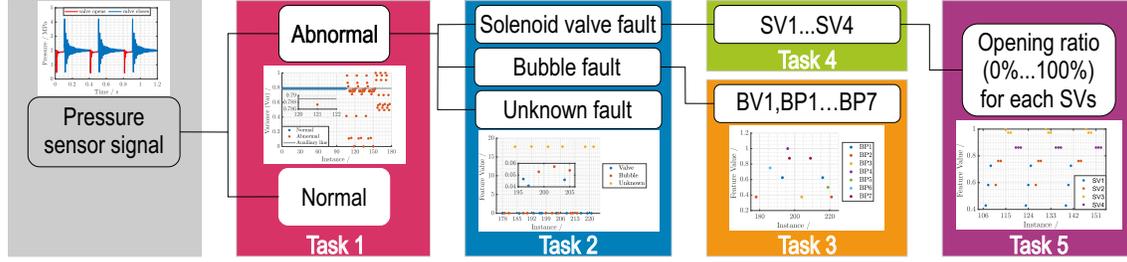


Figure 3. Comprehensive diagnosis based on hierarchical modeling

The second, third, and fourth tasks involve classifying the fault type and isolating the faulty component. The fifth task can be coined as fault identification, where the extent of the fault is to be predicted.

3.4. Method Evaluation

As proposed in section 2 and illustrated in Figure 3, each task is modeled individually and the respective solutions are combined to obtain a comprehensive diagnosis of the propulsion system. The subset of the data available for training becomes less the further right in the hierarchy the model is located. Hence, more expert-knowledge has to be incorporated and data-hungry algorithms become less relevant.

Task 1

For the first task, only the pressure sensor signals up to approximately 100 ms, which represents the solenoid valve opening phase, were utilized to distinguish between normal and abnormal instances. During this time window, the amplitudes at distinct frequencies in the frequency domain remain constant for normal instances. The amplitudes at various frequencies change significantly for the bubble anomaly and slightly for the solenoid valve faults, as exemplarily shown for $P1'$ in Figure 4(a). Here, $P1'$ represents signal $P1$ truncated to approximately 100 ms length. Although sophisticated features from the afordescribed time-series toolboxes provide better

distinction, classical time-domain features, such as variance of the truncated pressure signals, were sufficient to successfully model a binary decision tree, as exemplarily shown for $P1'$ in Figures 4(b) and 4(c), respectively. The decision rule R1 for a healthy instance can be derived from the decision tree in Figure 4(c) as in equation 3a. Furthermore, since faults can occur and be registered at the sensor locations between each accumulator (BV1, BP1 through BP7) and each valve (SV1 through SV4), decision trees are also modeled based on the sensor signals at the sensor locations P2 through P7. The further resulting rules for an healthy instance are in equations 3b through 3g, where $F(x)$ denote the top selected feature per sensor signal x , while considering only classical features. The feature $F(x)$ for signals $P1'$, $P2'$, $P4'$ and $P5'$ is Variance. The feature $F(x)$ for $P3'$, $P6'$ and $P7'$ are Line_Integral, Kurtosis, and Clearance_Factor, respectively.

$$R1: F(P1') \geq 0.788 \text{ AND } F(P1') < 0.804 \quad (3a)$$

$$R2: F(P2') \geq 0.813 \text{ AND } F(P2') < 0.837 \quad (3b)$$

$$R3: F(P3') \geq 0.913 \text{ AND } F(P3') < 0.917 \quad (3c)$$

$$R4: F(P4') \geq 0.6552 \text{ AND } F(P4') < 0.65536 \quad (3d)$$

$$R5: F(P5') \geq 0.676 \text{ AND } F(P5') < 0.683 \quad (3e)$$

$$R6: F(P6') \geq 0.064 \text{ AND } F(P6') < 0.0692 \quad (3f)$$

$$R7: F(P7') \geq 0.186 \text{ AND } F(P7') < 0.199 \quad (3g)$$

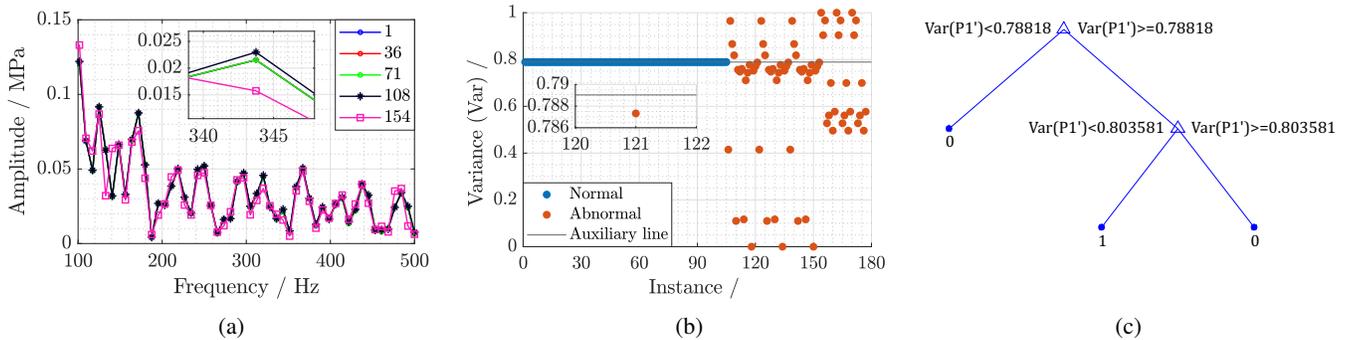


Figure 4: (a) Frequency spectrum analysis in the region of interest for normal instances (1, 36, 71), solenoid valve fault (108) and bubble anomaly (154) (b) Variance of all training instances scaled in the range from 0 to 1 (c) Resulting binary decision tree for $P1'$, class 0 represent abnormal instance and class 1 signifies normal instance

An ensemble of the binary decision trees modeled per sensor in form of a logical conjunction between the rules is the resulting fault detection model for task 1, since the system is deemed healthy if and only if no sensor declares a fault. The resulting fault detection rule for task 1 can thus be formulated as in equation 3h, where Y is the prediction output, 0 denotes abnormal instance and 1 denotes normal instance.

$$Y = \begin{cases} 1, & \text{if AND}(R1,R2,R3,R4,R5,R6,R7). \\ 0, & \text{otherwise.} \end{cases} \quad (3h)$$

Task 2

As illustrated in Figure 3, only the instances classified as *Abnormal* are considered for this task. Since unknown and mixed fault modes are included in the test data and not in the training data, a twofold problem was solved for task 2. Firstly, a binary classification problem was solved for the known solenoid valve faults and bubble anomalies and secondly, a binary clustering problem was solved to identify unknown faults. The binary classification was similarly modeled as in task 1, that is, the classical time-domain features extracted from the approximately 100 ms truncated sensor signals were utilized. Furthermore, a top feature, namely *RMS* for signals $P1'$ and $P3'$, *Peak* for $P2'$, $P4'$, $P5'$ and $P6'$ and *Entropy* for $P7'$, was selected per sensor. Each selected feature per sensor is subsequently employed to train a decision tree model. As opposed to task 1, majority voting ensemble of the decision trees per sensor was adopted here, since some faults might not be registered by sensors farther away from the fault source. For the test data, kMeans clustering algorithm is then employed on the classified fault instances to unveil unknown faults, which should deviate from known faults.

Task 3

This task involves the determination of the location of bubble anomalies, as identified in task 2. In reference to Figure 1, bubble anomalies can occur at eight accumulator positions,

namely BP1 through BP7 and BV1. Thus, the classification problem is multi-class, which involves classifying each identified bubble anomaly instance from the test data to a bubble anomaly location. The one-vs-all scheme was adopted here, which implies a binary classification model was built for each possible bubble anomaly location, resulting in eight models. Considering that there is a sensor near each possible bubble anomaly location, features are extracted from the respective sensor for each model. As opposed to the previous two tasks, features were selected from the HCTSA-time series toolbox, because of the time-frequency-related effects of this anomaly. Figure 5(a) exemplarily depicts the raw pressure signals for bubble anomaly instances 154 (BP1) and 155 (BP2), truncated at approximately 400 ms. Due to the existence of bubbles in the liquid-carrying pipeline, the speed of sound is reduced. Depending on the bubble location, the arrival time of the signal differ, as can be deduced from Figure 5(a). Furthermore, the amplitudes of the signals differ and their shape are distorted, as a result of superimposition of frequencies. Figures 5(b) and 5(c) are spectrograms to exemplarily uncover the time-frequency-related effects. As can be inferred from both figures, low frequency components, and particularly approximately 60 Hz-component are dominant over the time of interest. Furthermore, up to about 200 ms, frequency components in the range of 200 Hz and 300 Hz and between 400 Hz and 500 Hz are also dominant, but at different intensities for both instances. Finally, at about 400 ms, frequency components up to approximately 400 Hz are present, also at different intensities for both instances. Depending on the sensor location, these differences are captured by the features extracted and subsequently selected from the HCTSA-time series toolbox through distribution, Fourier spectrum or information theory analysis (refer to Table 5 in the appendix). As in task 1, the selected features are then employed to train a decision tree model per possible sensor location.

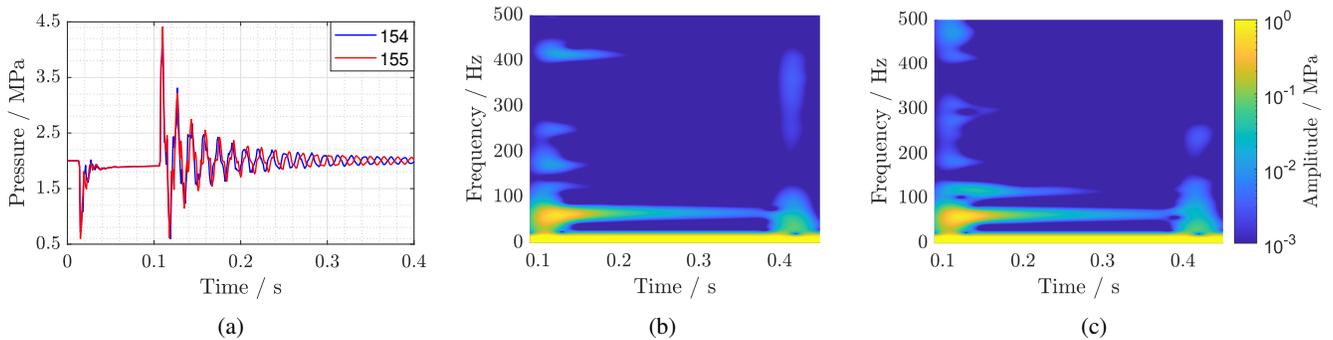


Figure 5: (a) Analysis of bubble anomaly instances 154 (BP1) and 155 (BP2) in the time domain (b) Time-frequency domain analysis of bubble anomaly instance 154 (BP1) (c) Comparative analysis of bubble anomaly instance 155 (BP2) in the time-frequency domain

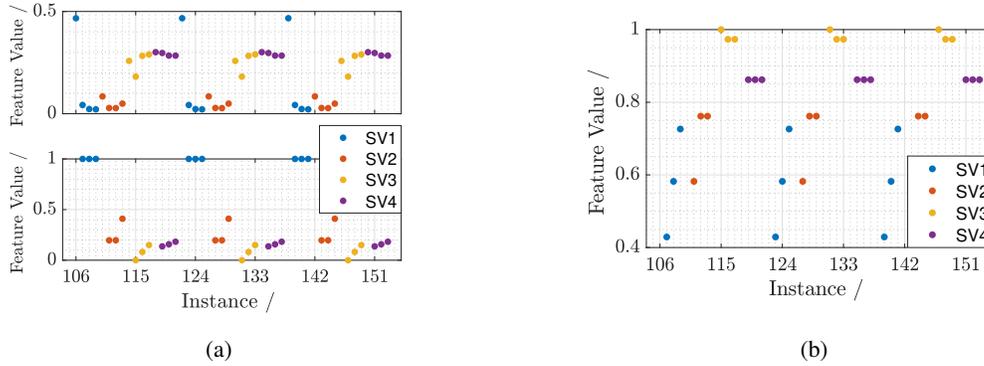


Figure 6: (a) Selected features based on four valve opening ratios (top figure) and excluding the closed valve (bottom figure). (b) Top feature value selected at the truncated signal P1, excluding the closed valve

Task 4

As in the previous task, the current task also involves a multi-class classification problem of identifying the four possible locations (SV1 through SV4) of the solenoid valve faults, as identified in task 2. Typically, the best sensors to isolate the respective fault locations are the sensors closest to the fault source. However, due to the challenge of the fault isolation task and little amount of training data available, further sensors along the direct path from the tank to the valves were also considered. Specifically, sensor P1 near the tank and the respective sensor closest to the fault location were each employed for generating a binary classification model per possible fault location. Furthermore, because faults were induced in the experiments by varying the respective valve's opening ratio, the sensor signals were truncated at approximately 100 ms within the valve opening phase. Subsequently, features were extracted from the truncated sensor signals with the HCTSA-toolbox and selected with the Chi-Square test. The initial selected feature (SP_Summaries_welch_rect.linfiltloglog_lf_sigrat) from sensor P3, while considering the opening ratios 0%, 50%, 75%, and 100% per sensor locations, is exemplarily depicted in Figure 6(a). As can be inferred from Figure 6(a), the feature value for the opening ratio of SV1 at 0% is distant from the other opening ratios of the same valve. Furthermore, since the selected feature barely allow a good separation of the classes, the opening ratio at 0% was consequently excluded during the feature selection process. As seen in Figure 6(b), the resulting top feature allows a clear separation between SV1 and other sensor locations. Furthermore, as can be inferred from Figure 6(b), except for the feature value of instance 111 (SV2) which coincides with the feature of instance 108 (SV1), the top selected feature at the sensor location P1 near the tank allows good separation of the classes. Thus, this feature and the respective selected feature next to the fault location, as in Table 6 in the appendix, were each employed to train a decision tree model per fault location. The resulting trees were then combined to an ensemble of trees to make the prediction for the respective

fault location. A detailed explanation of the corresponding selected features can be found in the documentation of the HCTSA-toolbox (B. D. Fulcher, Little, & Jones, 2022).

Task 5

This final task entails the identification of the opening ratio for each instance classified as solenoid valve fault. Contrary to the previous tasks, where the prediction target was a discrete value of predetermined classes, except for the anomalous case, the prediction target of the current task is a continuous value, for the test data. Furthermore, the training data comprises discrete values for the opening ratios 0%, 50%, 75%, and 100% per sensor locations. Thus, four values per sensor location are available for solving the resulting regression problem. To this end, features were extracted from approximately 100 ms truncated sensor signals with the HCTSA-toolbox, as in task 4. Subsequently, since the the training data comprises discrete values, features were preselected with the Chi-Square test. Finally, through sensitivity analysis, the top selected feature from sensor location P7 for SV1, SV3 and SV4, and the top selected feature from sensor location P6 for SV2 were utilized to predict the opening ratio of the test data via linear interpolation. The top selected feature for the sensor locations is the Kurtosis.

4. RESULTS

During the preparation of this paper, the ground-truth data was unavailable for a detailed evaluation of the results. However, with the proposed comprehensive diagnostics methodology, our team LDM emerged 7th place, with a score of 93.08%. Furthermore, the predictions for the test data are laid out in the following paragraphs for comparative studies.

Task 1:

Figure 7(a) exemplarily shows the scaled variance value of the truncated sensor signal P2' in the range of interest for the test data, 0 denotes abnormal instance and 1 denotes normal

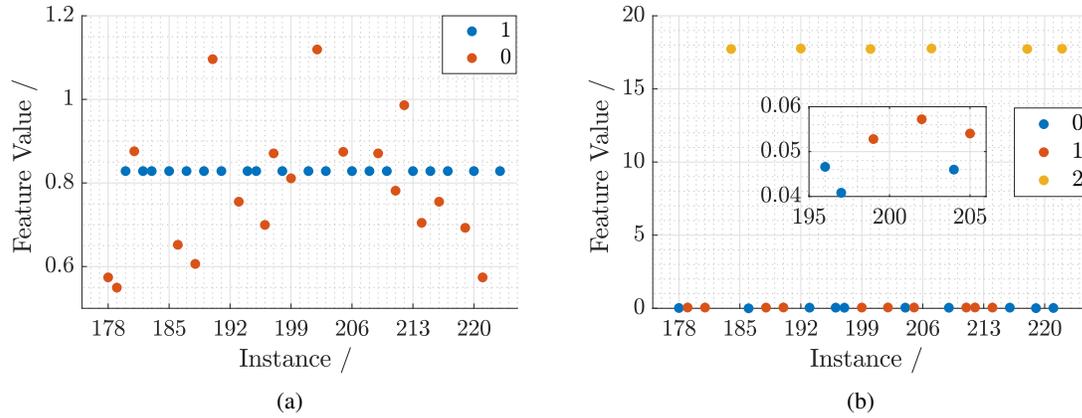


Figure 7: (a) Variance of the test instances, scaled with the training data scaling parameters (b) RMS of the test instances, scaled with the training data scaling parameters

instance.

Based on the ensemble of decision trees as described in Section 3.4 for task 1 and particularly equation 3h, the predictions for task 1 are presented in Table 1 and can be inferred from Figure 7(a).

Table 1. Classes of predicted instances for task 1

Class	Instances
1	180, 182, 183, 185, 187, 189, 191, 194, 195, 198, 201, 203, 206, 208, 210, 213, 215, 217, 220, 223
0	178, 179, 181, 184, 186, 188, 190, 192, 193, 196, 197, 199, 200, 202, 204, 205, 207, 209, 211, 212, 214, 216, 218, 219, 221, 222

Note: 0 denotes abnormal instance and 1 denotes normal instance

Task 2:

Figure 7(b) exemplarily shows the scaled RMS value of the truncated sensor signal $P1'$ for the abnormal test data instances identified in the preceding task.

Valve fault instance is denoted by class 0, bubble anomaly and unknown fault instances are denoted by classes 1 and 2, respectively. As can be deduced from Figure 7(b), the unknown fault instances differ significantly from the training data. Hence, they are easily separable with the kmeans algorithm, as described in Section 3.4 for task 2.

Furthermore, the valve and bubble anomaly instances could be classified with the ensemble of decision trees, as also described in 3.4 for task 2. The predictions for task 2 are laid out in Table 2

Task 3:

Although the DN_Unique feature was selected for classifying BP1, its feature values also allow a distinction of other fault

Table 2. Classes of predicted instances for task 2

Class	Instances
0	178, 186, 193, 196, 197, 204, 209, 216, 219, 221
1	179, 181, 188, 190, 199, 202, 205, 211, 212, 214
2	184, 192, 200, 207, 218, 222

Note: valve fault instance is denoted by class 0, bubble anomaly instance by class 1 and unknown fault instance by class 2

locations, except for BP2 and BP3, as can be seen in Figure 8(a).

The methodology described in Section 3.4 for task 3 is employed to predict the locations of the bubble anomaly instances, as presented in Table 3.

Table 3. Locations of the bubble anomalies of the test instances

Class	BP1	BP2	BP3	BP4	BP5	BP6	BP7
Instances	193, 216	178, 221	204	196	219	186	197, 209

Task 4 and Task 5:

The prediction task for SV3 and SV4 was based on the ensemble method, as described in Section 3.4. However, the ensemble method produced some overlapping predictions for the fault locations of SV1 and SV2, which suggests that these faults might have been mixed in the test data. As can be deduced from Figure 8(b), based on simple rules, the top selected feature of the truncated sensor signal $P1'$ allows to distinguish between the classes SV3 and SV4, but not between SV1 and SV2.

As described in Section 3.4 for task 4 and 5, the resulting models are successively utilized to predict the locations and opening ratio of the valve fault instances, as presented in Table 4.

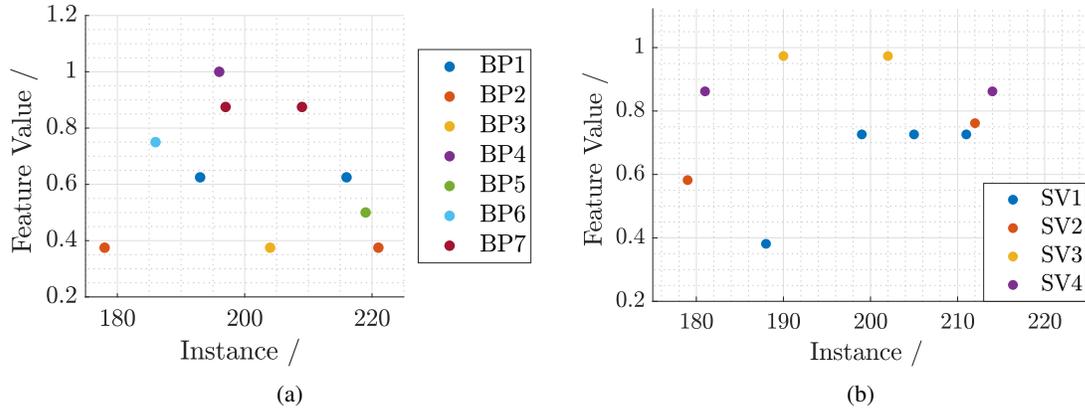


Figure 8: (a) DN_Unique value of the test instances, scaled based on the training data scaling parameters (b) Top feature value of the test instances, scaled based on the training data scaling parameters

Table 4. Classes and respective opening ratios of predicted instances for tasks 4 and 5

Class	Instances	Opening Ratio in %
SV1	188, 199, 205, 211	4.62; 86.17; 85.56; 85.29
SV2	179, 212	24.91; 69.28
SV3	190, 202	44.48; 40.64
SV4	181, 214	73.94; 19.03

5. CONCLUSION

For detecting, isolating and identifying multiple fault modes, a decision tree per fault mode was modeled. The resulting model per fault mode was subsequently combined via different voting ensembles inspired by expert-knowledge. The comprehensive hierarchical diagnostics methodology was evaluated based on an experimental propulsion system and corresponding data sets provided as part of the Asia-Pacific PHM conference's data challenge 2023. Although, only successfully evaluated on the task of the data challenge, the proposed methodology can find applications in other complex systems, such as a wind turbine gearbox to successively diagnose bearing and gear faults. Depending on the application at hand, possible shortcomings of the presented methodology are the supervised feature extraction, selection and modeling approach. Alternatively, several techniques, such as deep learning could be employed to develop models without prior feature engineering. However, at a cost of diminished transparency and explainability.

ACKNOWLEDGMENT

This research was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project number 451737409.

REFERENCES

- Aimiyeqagbon, O. K., Bender, A., & Sextro, W. (2021). On the applicability of time series features as health indicators for technical systems operating under varying conditions. *17. International Conference on Condition Monitoring and Asset Management (CM 2021)*.
- Aimiyeqagbon, O. K., Muth, L., Wohlleben, M., Bender, A., & Sextro, W. (2021). Rule-based diagnostics of a production line. *Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021*, 527–536.
- Al-Khayat, R. H., Al-Fatlawi, A. W. A., Al-Baghdadi, M. A. S., & Al-Waily, M. (2022). Water hammer phenomenon in pumping stations: A stability investigation based on root locus. *Open Engineering*, *12*(1), 254–262.
- Bandyopadhyay, A., & Majumdar, A. (2014). Network flow simulation of fluid transients in rocket propulsion systems. *Journal of Propulsion and Power*, *30*(6), 1646–1653.
- Bombardieri, C., Traudt, T., & Manfletti, C. (2019). Experimental study of water hammer pressure surge. *Progress in Propulsion Physics*, *11*, 555–570.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. doi: 10.1007/BF00058655
- Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, *307*, 72–77. doi: 10.1016/j.neucom.2018.03.067
- Fulcher, B., Cliff, O., Harris, B., Philiphorst, Sethi, S., Lubba, C. H., ... Haohua Li (2020). *benfulcher/hctsa: v1.04*. Zenodo. Retrieved from <https://zenodo.org/record/3955668> doi: 10.5281/ZENODO.3955668
- Fulcher, B. D., Little, M. A., & Jones, N. S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society, Interface*, *10*(83), 20130048. doi: 10.1098/rsif.2013.0048
- Fulcher, B. D., Little, M. A., & Jones, N. S. (2022). *Hctsa documentation*. Retrieved from <https://hctsa-users.gitbook.io/hctsa-manual/> (Last

- viewed in: 08.2023)
- Gao, Y., Yang, T., Xing, N., & Xu, M. (2012). Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines. In *2012 7th IEEE conference on industrial electronics and applications (iciea)* (pp. 1984–1988). doi: 10.1109/ICIEA.2012.6361054
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (Second edition ed.). Beijing and Boston and Farnham and Sebastopol and Tokyo: O'Reilly.
- Hennig, M., Grafinger, M., Gerhard, D., Dumss, S., & Rosenberger, P. (2020). Comparison of time series clustering algorithms for machine state detection. *Procedia CIRP*, 93, 1352-1357. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2212827120307149> (53rd CIRP Conference on Manufacturing Systems 2020) doi: <https://doi.org/10.1016/j.procir.2020.03.084>
- Hoque, N., Bhattacharyya, D., & Kalita, J. (2014). Mifs-nd: A mutual information-based feature selection method. *Expert Systems with Applications*, 41(14), 6371-6385. doi: <https://doi.org/10.1016/j.eswa.2014.04.019>
- Katipamula, S., & Brambley, M. R. (2005). Methods for fault detection, diagnostics, and prognostics for building systems—a review, part i. *Hvac&R Research*, 11(1), 3–25.
- Kimotho, J. K., & Sextro, W. (2014). An approach for feature extraction and selection from non-trending data for machinery prognosis. In *Proceedings of the second european conference of the prognostics and health management society 2014* (Vol. 5).
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*, 50(6), 1–45.
- Liu, H., & Setiono, R. (1995). Chi2: feature selection and discretization of numeric attributes. In *Proceedings of 7th IEEE international conference on tools with artificial intelligence* (pp. 388–391). doi: 10.1109/TAI.1995.479783
- Loh, W.-Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1), 14–23. doi: 10.1002/widm.8
- Magnello, M. (2005). Chapter 56 - karl pearson, paper on the chi square goodness of fit test (1900). In I. Grattan-Guinness, R. Cooke, L. Corry, P. Crépel, & N. Guicciardini (Eds.), *Landmark writings in western mathematics 1640-1940* (p. 724-731). Amsterdam: Elsevier Science. doi: <https://doi.org/10.1016/B978-044450871-3/50137-6>
- Nie, L., Zhang, L., Xu, S., Cai, W., & Yang, H. (2022). Remaining useful life prediction for rolling bearings based on similarity feature fusion and convolutional neural network. In *Journal of the brazilian society of mechanical sciences and engineering* (Vol. 44). doi: 10.1007/s40430-022-03638-0
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157-175. doi: 10.1080/14786440009463897
- PHM Society. (2023). *Phmap data challenge*. 2023 Asia-Pacific Prognostics and Health Management Society (PHMAP) Data Challenge. Retrieved from <https://data.phmsociety.org/phmap-2023-data-challenge/> (Retrieved on: 22.07.2023)
- Rodríguez Ramos, A., Domínguez Acosta, C., Rivera Torres, P. J., Serrano Mercado, E. I., Beauchamp Baez, G., Rifón, L. A., & Llanes-Santiago, O. (2019). An approach to multiple fault diagnosis using fuzzy logic. *Journal of Intelligent Manufacturing*, 30, 429–439.
- Tominaga, K., Fujii, G., Nagata, T., Wada, D., Hisada, S., Kawatsu, K., & Kasai, T. (2023). Anomaly detection method for spacecraft propulsion system using frequency response functions of multiplexed fbg data. *Acta Astronautica*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0094576523003739> doi: <https://doi.org/10.1016/j.actaastro.2023.07.022>
- Tsai, Y. S., King, P. H., Higgins, M. S., Pierce, D., & Patel, N. P. (1997). An expert-guided decision tree construction strategy: an application in knowledge discovery with medical databases. *Proceedings : a conference of the American Medical Informatics Association. AMIA Fall Symposium*, 208–212.
- Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8, 29857–29881.

APPENDIX

Table 5. Features selected for task 3

Bubble anomaly location	Feature name	Category
BP1	DN_Unique	distribution,raw
BP2	SP_Summaries_fft.numPeaks	FourierSpectrum
BP3	FC_Surprise_T1_20_tau_m2quad_500.min	information,symbolic
BP4	FC_Surprise_dist_5_2_udq_500.uq	information,symbolic
BP5	SP_Summaries_welch_rect.numPromPeaks_1	FourierSpectrum
BP6	SY_KPSStest_1.pValue	stationarity,hypothesistest
BP7	FC_Surprise_dist_5_2_udq_500.uq	information,symbolic
BV1	FC_Surprise_dist_20_2_udq_500.min	information,symbolic

Table 6. Selected features for task 4

Valve fault location	Sensor position	Feature name	Category
SV1	P1	EN_DistributionEntropy_hist_sqrt_0	entropy
	P3	PH_Walker_momentum_2.sw_maxrat	trend
SV2	P1	EN_DistributionEntropy_hist_sqrt_0	entropy
	P4	PH_Walker_biasprop_05_01.res_runstest	trend
SV3	P1	EN_DistributionEntropy_hist_sqrt_0	entropy
	P6	Variance	statistics
SV4	P1	EN_DistributionEntropy_hist_sqrt_0	entropy
	P7	SB_MotifTwo_median.hhhh	symbolic,motifs