



DATA PREPROCESSING IN THE CONTEXT OF PHM -CHALLENGES AND SOLUTIONS

Indrawata Wardhana, Amal Gouiaa-Mtibaa, Pascal Vrignat, Frédéric Kratz

► To cite this version:

Indrawata Wardhana, Amal Gouiaa-Mtibaa, Pascal Vrignat, Frédéric Kratz. DATA PREPROCESSING IN THE CONTEXT OF PHM -CHALLENGES AND SOLUTIONS. CIE 50 Conference Handbook, CIE 50 Conference Handbook, Oct 2023, Sharjah, United Arab Emirates. <hal-04274223>

HAL Id: hal-04274223

<https://hal.science/hal-04274223v1>

Submitted on 7 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

DATA PREPROCESSING IN THE CONTEXT OF PHM - CHALLENGES AND SOLUTIONS

I. WARDHANA ^{1*}, A. GOUIAA-MTIBAA ², P. VRIGNAT ², F. KRATZ ¹

¹INSA Centre Val de Loire, PRISME Laboratory, 18000 Bourges
France (e-mail: {indrawata.wardhana, frederic.kratz} @insa-cvl.fr)

²Orléans University, PRISME Laboratory, 45000 Orléans (e-mail:
{amal.gouiaa-mtibaa, pascal.vrignat}@univ-orleans.fr)

ABSTRACT

This study presents a comprehensive framework for addressing data preprocessing challenges in diverse scenarios. The framework covers various aspects, including feature extraction in the time domain, frequency domain, and wavelet domain, as well as feature selection techniques, noise reduction strategies, and data normalization approaches for time series data. Through experimental analysis, we demonstrate the effectiveness of our proposed strategy in enhancing the Signal-to-Noise Ratio value. In particular, the application of the Savitzky-Golay Filter proves successful in reducing signal noise and achieving normalization across all bearings. These findings highlight the potential of our framework in improving data quality and preprocessing procedures for bearing data set which can be generalized for a wide range of applications.

Keywords: prognosis and health management (PHM), predictive maintenance, preprocessing data, feature extraction

1 INTRODUCTION

SCADA environments are connected to various database management systems. These databases are becoming more and more essential for the efficient management of the company. Industry 4.0 and its technological panel offer today, a very large collection of this information. In fact, this evolution is taking place at all levels, including the maintenance level. This is a crucial activity, like a part of PHM, as it determines the availability and performance of production systems.

However, various perturbations (e.g. power cuts, digital communication network interruptions, wrong choice of technology...) have a significant impact on the quality of the collected data. Data must be checked to see if it incorrect, incomplete, inaccurate, irrelevant, inconsistent, different, repeated, poorly described, or missing. Big data problems are no secret, as it is difficult to compile a large and accurate database.

The main data quality target is accuracy [3]. Previous studies often use three types of data errors: inaccuracy, incompleteness, and mismatch [4]. The method proposed by [5] combines dimensionality reduction with the identification and separation of erroneous data. Due to the measurement data used as input for data acquisition, the presence of uncertainties and unknown measurement inaccuracies would be a concern. Sources of uncertainty in Machine Learning are divided into two groups. Group I is related with data acquisition error, data annotation, imbalance and skewness, incompleteness, and input data sampling. Group II includes feature selection, model selection, and model hyperparameter estimation [6].

* Corresponding Author

Several steps have been carried out to improve for reducing the error. Data preprocessing is an inevitable step before applying Machine Learning methods for energy or load prediction. According to previous experiences, data preprocessing typically accounts for a large portion of the total data mining task, about 80% [7]. As shown in Figure 1, the “observe” step in PHM includes two steps data acquisition and data preprocessing. The common steps of data preprocessing include data imputation, data resolution processing, data normalization, outlier detection, and data smoothing [8]. This latter could reduce the influence of noisy data and improve data quality, which contributes to the improvement of model performance [9].

Signal preprocessing plays a critical role in understanding and analyzing various types of data. In this study, we are interested in bearing data and especially vibration signals obtained from accelerometers. To gain a deeper understanding of these signals, researchers often explore four primary domains: time domain, frequency domain, time-frequency domain, and wavelets [1]. By converting the signal from the amplitude domain to the time domain [10], researchers can examine its behavior over time and identify any patterns or variations. In addition, exploring the frequency domain allows researchers to analyze the signal in terms of its component frequencies and their respective amplitudes. Through extensive research and experimentation, scientists have made significant strides in understanding and interpreting accelerometer signals, enabling advances in fields as diverse as structural health monitoring, vibration analysis, and predictive maintenance. In Figure 1, we present a summary of the data chain with a focus on the step data preprocessing which will be the objective of our study.

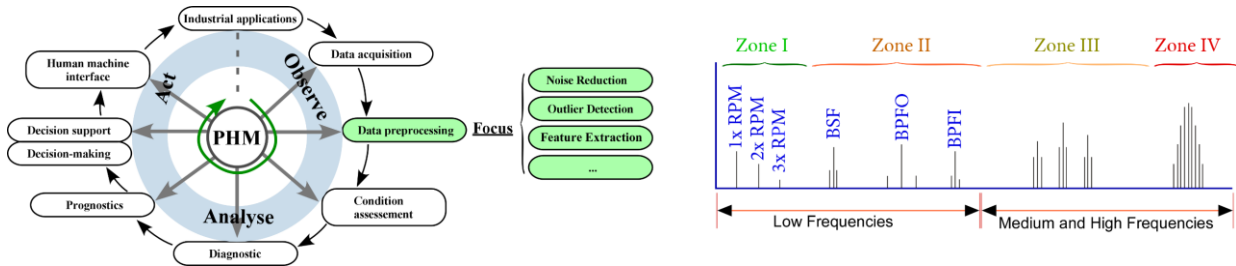


Figure 1: (a) PHM cycle (adopted from [1]) and (b) Bearing Damage Stage III and Multiple Zone (adopted from [2])

The organization of this paper is as follows: Section 2 provides a comprehensive review of preprocessing steps in signal bearing, highlighting their relevance to the state of the art. In Section 3, a case study involving the Bearing FEMTO dataset is presented, along with the framework, method algorithm, or equation employed. Section 4 presents and discusses the results derived from the analysis. Finally, the paper concludes by summarizing the findings and suggesting potential future research directions.

2 STATE OF THE ART

The quality of input data has a significant impact on the quality of the model and the accuracy of prediction results. Many problems have been detected in the bearing data and a preprocessing step is compulsory. One common issue encountered is the presence of excessive unwanted frequencies in the data. A high Signal-to-Noise Ratio (SNR) indicates a clear and easily detectable or interpretable signal, whereas a low SNR suggests that the signal is corrupted or obscured by noise, making it challenging to distinguish or recover meaningful information. To achieve a high SNR [11], the separation of low-frequency signals containing information related to bearing rotation and bearing faults from the middle and

high-frequency components has been the focus of exploration. The operational status of rotating machines (embedded bearing) and the presence of four specific faults are restricted to zones 1 and 2 [2], as shown in Figure 1 (b). This means that the occurrence of these faults is associated with low frequencies, while medium and high frequencies are ignored, potentially leading to degradation of signal quality due to the presence of noise. Several techniques have been investigated for this purpose, including Variational Mode Decomposition (VMD), Empirical Wavelet Transform (EWT), Empirical Mode Decomposition (EMD), Vold-Kalman Filter Order Tracking (VKF_OT) [12], Maximal Overlap Discrete Wavelet Transform (MODWT) [13], and others. These methods aim to effectively extract the desired low frequency signals while filtering out the unwanted medium and high frequency components. By using such techniques, researchers aim to improve the SNR and increase the accuracy of fault detection and prediction in bearing analysis. These approaches have been investigated under different SNR ranges, and the research results have demonstrated prediction accuracy levels of up to 99 percent [14]. Researchers often use denoised signals for fault prediction to improve the noise immunity of diagnostic models and minimize the influence of noise on the overall performance.

Before determining the appropriate feature extraction method, stationarity tests such as Augmented Dickey-Fuller (ADF) are commonly used to assess the stationarity of the signal. Once the time domain features have been generated, it is essential to perform an effective feature selection process in order to identify the most optimal feature for the data input. The RMS time domain feature has been widely used by many researchers as a health indicator [15], [16], [17]. The RMS has been found to have a positive correlation with changes in vibration intensity, providing valuable insights into the progressive degradation path of bearings.

It deals with high-dimensional data, where the number of features greatly exceeds the number of samples. One popular method for feature selection is the Laplacian Score, which effectively measures the relevance of features based on the underlying manifold structure of the data. The advantage of using the Laplacian Score is its ability to efficiently handle high-dimensional data in unsupervised learning [18], resulting in a reduced feature set without sacrificing predictive performance. This leads to improved computational efficiency and enhanced interpretability of the selected features, making it a valuable tool in various domains. Noise reduction and normalization are also two important steps before moving on to training and testing. Several methods are used depending on the type of signal [19], [20]. The Savitzky-Golay Filter (SGF) is a powerful tool for denoising and smoothing fault-generated signals [21].

3 METHOD AND CASE STUDY

Rolling bearings are critical to the operation of rotating machinery, ensuring efficient performance and minimizing friction. Consisting of components such as the outer ring, inner ring, cage, and rollers. They play a key role in various industrial equipment. The condition of rolling bearings has a direct impact on the overall safety and reliability of the system. Research shows that rolling element bearings are responsible for a significant portion, approximately 45-55% [22], of equipment failures.

3.1 Dataset

The challenge data was obtained from a well-designed experimental platform called PRONOSTIA [23]. The dataset consists of vibration and temperature data collected during run-to-failure tests. The sampling frequency used in the experiment is 25.6 kHz, with a

recording interval of 10 seconds and a data length of 0.1 seconds per recording. In other words, each data recording consists of 2560 data points. For example, the total duration of the bearing degradation data in Bearing 1-1 is 28030 seconds. There are a total of 17 run-to-failure bearing records representing three different operating conditions. These different operating conditions are summarized in Table 1. Of these, six training datasets contain complete run-to-failure data, while the remaining dataset provides only censored bearing life data.

Table 1: Operation Conditions of FEMTO Bearings

	Operation Condition 1	Operation Condition 2	Operation Condition 3
Load (N)	4000	4200	5000
Speed (rpm)	1800	1650	1500
Freq (Hz)	30	27.5	25
Training dataset	Bearing1_1 Bearing1_2	Bearing2_1 Bearing2_2	Bearing3_1 Bearing3_2
Testing dataset	Bearing1_3 Bearing1_4 Bearing1_5 Bearing1_6 Bearing1_7	Bearing2_3 Bearing2_4 Bearing2_5 Bearing2_6 Bearing2_7	Bearing3_3

3.2 Framework

We present a novel approach, illustrated in Figure 2, for the initial processing of rolling bearing components. The scheme is derived from a previous study [24], which employed various techniques such as time-frequency, and time-domain feature extraction, wavelet transform with different decomposition levels depending on SNR values, and feature selection for optimal feature identification. In addition, this technique was further enhanced with noise reduction and normalization.

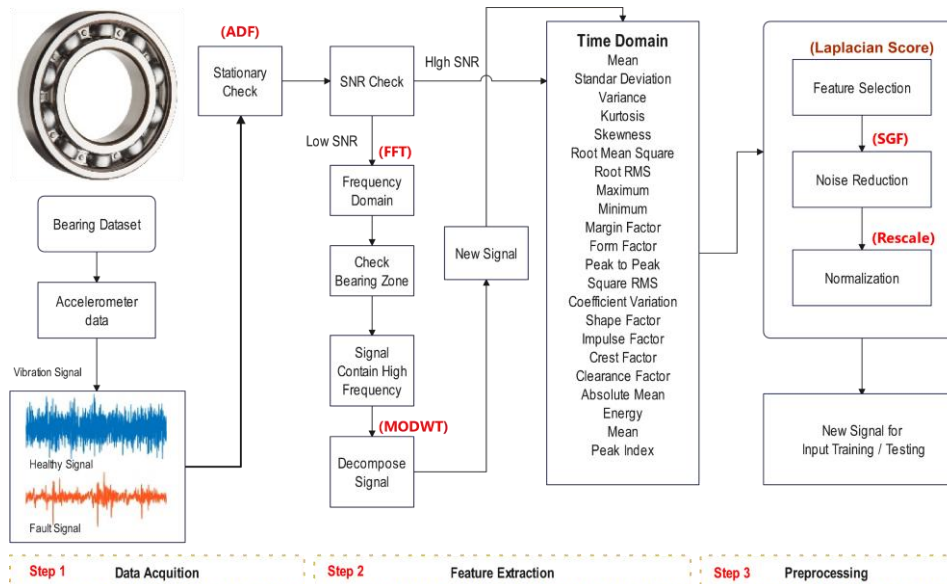


Figure 2: Framework for Bearing Preprocessing (adopted from [24])

In the first step, data is collected from sensors using only the vibration accelerometer signal and excluding the temperature signal. In the second step, the raw signals are examined using the stationarity method to define the signal. The SNR is used to evaluate the signal

quality. Signals with low SNR are extracted using Fast Fourier Transform (FFT) to identify the frequency distribution. The high frequencies based on the bearing zone are not used and can be removed. To eliminate the high frequencies, a signal decomposition method is also applied for all bearings. This step generates new high SNR signals that can be further processed for time-domain features. The time-domain features contain 21 different features, including Mean, Crest Factor, Standard Deviation, and Margin Factor. In the third step, the 21 time-domain features must be evaluated to determine which one has the highest correlation with the health indicator. Unsupervised learning techniques are devoted in bearing analysis to evaluate the correlation between features and variations. To account for the presence of noise in the signal, noise reduction methods are implemented. In addition, to facilitate accurate comparison and analysis, the signals are normalized to bring them into a consistent range. All of the methods used in this framework are described in [21], [25], [26], and [27]. These methods are shown in Figure 2 in bold red text.

4 RESULTS AND DISCUSSIONS

In order to investigate data preprocessing in FEMTO Bearing with different SNR, we conducted experiments using the new proposed framework. The main focus on this step consists to extract the decomposed signal based on the SNR value.

The ADF test is employed to analyze the signal under three different operating conditions. Table 1 shows that almost all bearings have a stationary signal. The SNR is in the range of -12 to -9 dB for normal bearings and -13 to -11 dB for faulty bearings. Given the high level of noise present, it is necessary to process the signal by removing the high frequencies using a decomposition method.

Table 1: Stationary check and SNR of Three Operation Condition

Name of Bearing	Stationary Check		SNR (dB)	
	Normal	Fault	Normal	Fault
Bearing 1_1	Stationary	Stationary	-12.84	-13.25
Bearing 2_1	Stationary	Stationary	-9.93	-11.78
Bearing 3_1	Stationary	Stationary	-12.37	-11.13

Contrary to the expectation, the normal bearing contains predominantly low frequencies. Figure 3 (1)(b) shows the raw signal in blue and the decomposed signal in orange. The low frequency signal has only low amplitudes. Figure 3 (2)(a) shows the presence of low, medium, and high frequencies in the normal condition. However, in this particular case, only the rotational frequency (30 Hz for operating condition 1) and the harmonics, should be present in the normal bearing signal. The remaining frequencies, especially the medium and high frequencies, are likely to be considered as noise. Then, in Figure 3 (2)(b), the elimination of the medium and high frequencies is observed. In addition, the power spectral density decreases from a maximum of 150 to 60. Consequently, the SNR increases from -12.71 dB to a new value of -0.21 dB. These results were obtained by using the MODWT with the sym4 wavelet and performing the transform down to level 4.

In Figure 3 (3) (a), observed as Bearing 1-2, it is evident that the highest power spectrum occurs toward the end of the bearing rotation. This power spectrum contains frequencies exceeding 1000 Hz, which are associated with bearing degradation and potential faults. To reduce noise, it is essential to reduce the frequencies above 1000 Hz in Bearing 1-2. On the other hand, Figure 3 (3) (b) shows samples obtained using a moving window of 20, where only the lower frequencies are present.

The analysis involves the use of 21 time-domain features, including mean, standard deviation (Std), variance, Kurtosis, skewness, RMS, maximum, minimum, margin factor, form factor, peak to peak, square RMS, coefficient variation, shape factor, impulse factor, crest factor, clearance factor, absolute mean, energy, mean, and peak index (Figure 2). These features are used in both the training of six bearings and the testing of eleven bearings. In Figure 3 (5 -a), (1, 2, 3)), the standard deviation of the raw signal is depicted for Bearing 1-2, 2-1, and 3-1. Figure 3 (5-a), (1)) illustrates that Bearing 1-2 exhibits noise values between time 100 to 700. However, the use of the decomposed signal as an input permit to obtain the standard deviation feature values. The noise reduction is evident in Figure 3 (5-b), (4), where the noise has been effectively eliminated. In Figures 3 (5-a), (2, 3), fluctuations in the signal are observed. However, after signal decomposition, Figures 3 (5-b), (5, 6) show a relatively stable and consistent signal without pronounced peaks. This indicates a consistent behavior of the signal over time.

Table 2 displays the SNR values for Bearing 1_1 in normal condition, which is -12.84 dB. However, after applying the decomposed signal using the MODWT method, the SNR increases to 1.83 dB. In the fault condition, Bearing 1_3 initially has an SNR of 2.08 dB, which then spectacularly increases to 293.38 dB. The SNR values of this particular raw signal differ from those of other signals, indicating a higher signal level compared to the noise. This can be observed in Figure 3 (4), (a), where the signal predominantly exists in low frequencies. The signal decomposition shown in Figure 3 (4), (b). The medium and high frequencies are eliminated, resulting in a high impact on the SNR values. These results suggest that the MODWT method is highly effective in significantly reducing noise.

To order the 21 features time domain, unsupervised feature selection Laplacian-score is used. Figure 4 (1) show that the Standard Deviation, Root Mean Square get the same value for high scores. In this context, several authors propose results with RMS ([28], [29]). The use of the standard deviation (Std) shows that signal processing is more efficient compared with RMS (Figure 3 (6-a), (1, 2, 3)). Nevertheless, upon decomposition as depicted in Figure 3 (6-b), (4, 5, 6), the noise reduction in the RMS values is not as prominent as observed in the Std. This contrasts with the findings reported in various articles, which could be attributed to the effective elimination of noise in the Std.

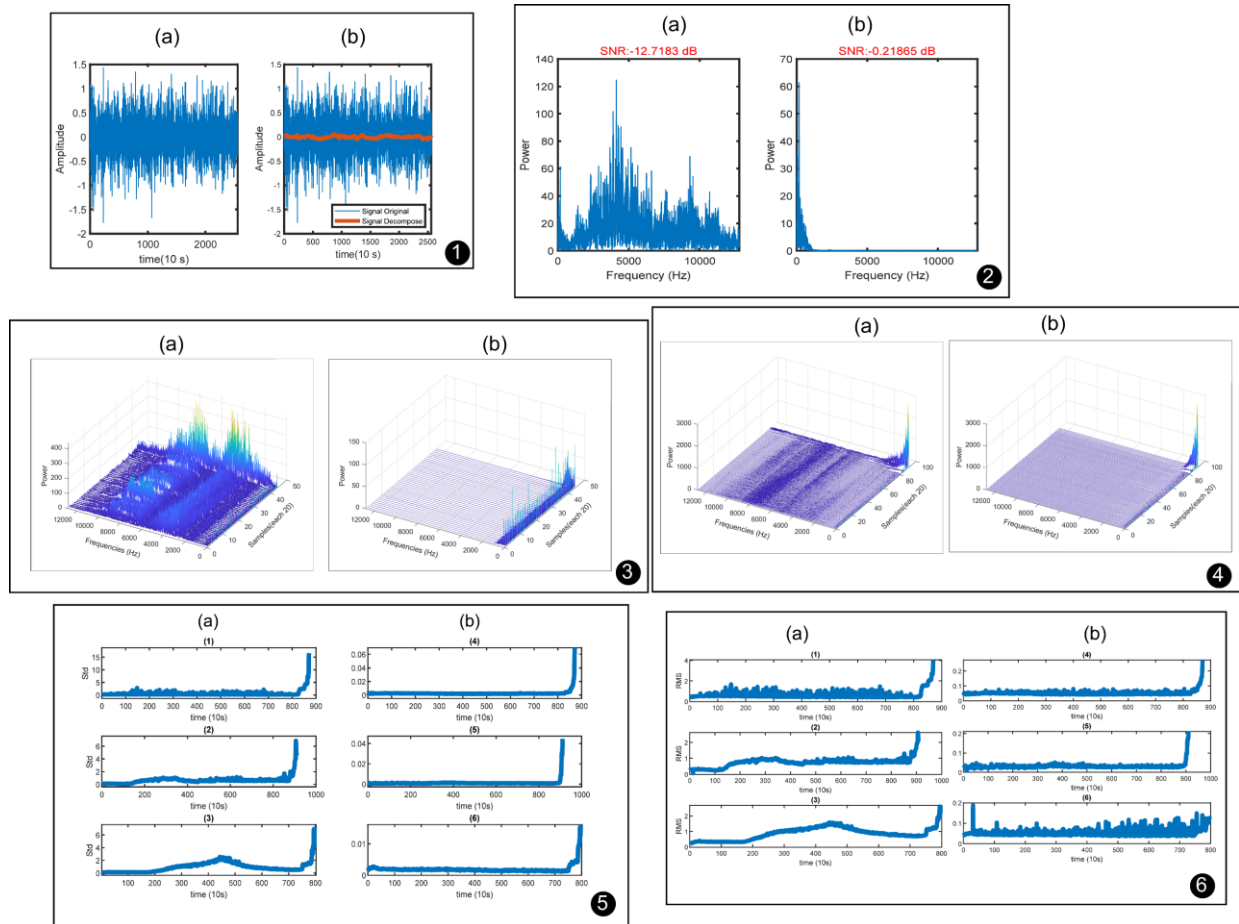


Figure 3: 1 (a) Raw Signal Bearing , 1 (b) Signal After Decompose, 2 (a) FFT Raw, 2 (b) FFT Decompose, 3 (a) FFT-Power-Time Raw Bearing 1-2, 3 (b) FFT-Power-Time Decompose, 4 (a) FFT-Power-Time Bearing 1-3, 4 (b) FFT-Power-Time Decompose 5 (a) (1, 2, 3), 5 (b) (4, 5 ,6) Decompose Std, 6 (a) (1, 2, 3), 6 (b) (4, 5 ,6) Decompose RMS

Table 2: Comparison of Bearing SNR in Normal and Fault Condition

Name of Bearing	SNR (dB) Original		SNR (dB) Decompose	
	Normal	Fault	Normal	Fault
Bearing 1_1	-12.84	-13.25	1.83	-1.62
Bearing 1_2	-12.72	-13.22	-0.22	-9.1
Bearing 1_3	-12.59	2.08	1.82	293.38

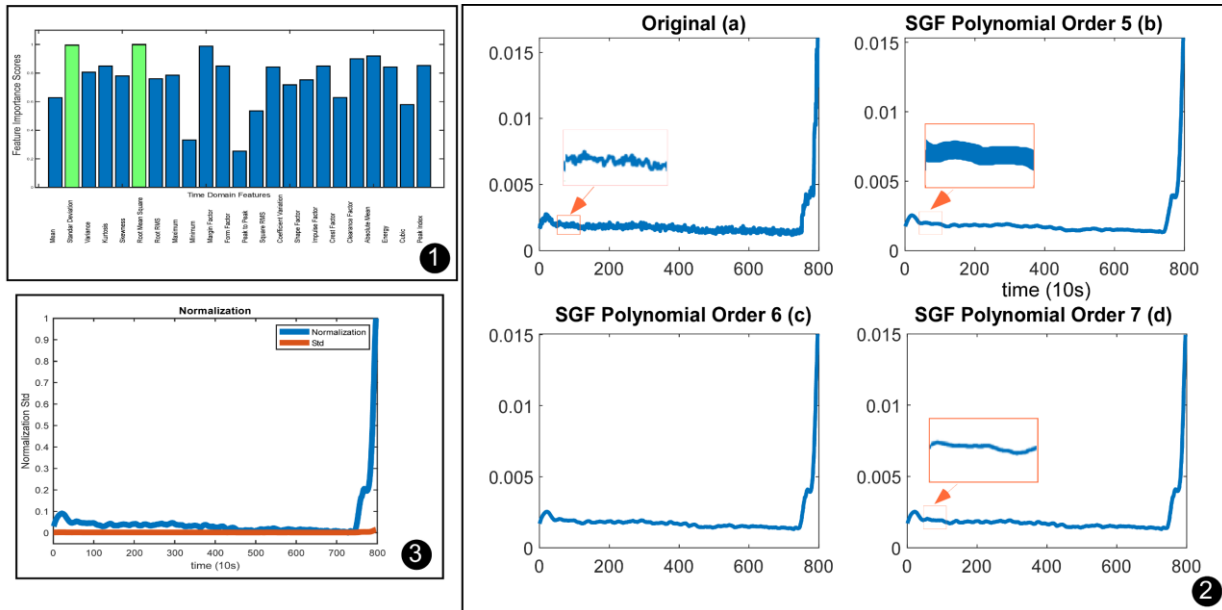


Figure 4: (1) Laplacian Score from 21 Feature Time-Domain, (2) Noise Reduction, (3) Normalization

In Figure 4 (2), (a), it is apparent that the standard deviation of the feature domain in Bearing 2-2 still contains noise. To solve this issue, the SGF is employed with a frame length of 51 samples and a specified Polynomial Order (PolyOd, Polynomial Order=5). Upon closer inspection in the zoomed window of Figure 4 (2), (b), it shows clearly that the noise is still present. After applying a polynomial order of 7, as shown in the zoomed window of Figure 4 (2), (d), a smoother signal is achieved. This suggests that the use of a higher polynomial order enhances the noise reduction capabilities of the SGF in Bearing 2-2. Figure 4 (2), (a, b, c and d) depicts values ranging from 0 to 0.015. These values subsequently undergo a normalization process, resulting in a transformation to the range of 0 to 1. The rescaling of the values to the interval $[0, 1]$ can be observed in Figure 4 (3).

5 CONCLUSION

In this study, we introduced a framework that addresses the challenges of data preprocessing in rolling scenarios. Our framework encompasses time-domain and frequency-domain feature extraction techniques, feature selection methods, noise reduction strategies and data normalization approaches for time-series data. Experimental results indicate that the proposed strategy effectively reduces the influence of noise on the signal. This reduction in noise will enable better detection of bearing faults. The next stage of our work will enable us to address the challenges associated with fault diagnosis (Figure 1).

6 REFERENCES

- [1] P. Vignat, F. Kratz, M. Avila. 2022. Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review, *Reliab. Eng. Syst. Saf.*, vol. 218, no. October 2020, doi: 10.1016/j.res.2021.108140.
- [2] L. Eren. 2017. Bearing fault detection by one-dimensional convolutional neural networks, *Math. Probl. Eng.*, vol. 2017, doi: 10.1155/2017/8617315.
- [3] S. Li, S. Ren, X. Wang. 2013. HVAC Room temperature prediction control based on

- neural network model, *Proc. - 2013 5th Conf. Meas. Technol. Mechatronics Autom. ICMTMA 2013*, pp. 606-609, doi: 10.1109/ICMTMA.2013.151.
- [4] A. Parssian, S. Sarkar, V. S. Jacob. 2004. Assessing data quality for information products: Impact of selection, projection, and cartesian product, *Manage. Sci.*, vol. 50, no. 7, pp. 967-982, doi: 10.1287/mnsc.1040.0237.
 - [5] L. Ben Amor, I. Lahyani, M. Jmaiel. 2018. Data accuracy aware mobile healthcare applications, *Comput. Ind.*, vol. 97, pp. 54-66, doi: 10.1016/j.compind.2018.01.020.
 - [6] H. Yang, T. Zhang, H. Li, D. Woradechjurnroen, X. Liu. 2014. HVAC Equipment, Unitary: Fault Detection and Diagnosis, *Encycl. Energy Eng. Technol. Second Ed.*, no. November, pp. 854-864, doi: 10.1081/e-eee2-120051345.
 - [7] I. Davidson, G. Tayi. 2009. Data preparation using data quality matrices for classification mining, *Eur. J. Oper. Res.*, vol. 197, no. 2, pp. 764-772, doi: 10.1016/j.ejor.2008.07.019.
 - [8] K. Amasyali, N. M. El-Gohary. 2018. A review of data-driven building energy consumption prediction studies, *Renew. Sustain. Energy Rev.*, vol. 81, no. September 2017, pp. 1192-1205, doi: 10.1016/j.rser.2017.04.095.
 - [9] H. Zhao, H. Liu, Y. Jin, X. Dang and W. Deng. 2021. Feature Extraction for Data-Driven Remaining Useful Life Prediction of Rolling Bearings, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10.
 - [10] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak. 2012. Fault prognostic of bearings by using support vector data description, *PHM 2012 - 2012 IEEE Int. Conf. on Progn. Heal. Manag. Enhancing Safety, Effic. Availability, Eff. Syst. Through PHM Technol. Appl. Conf. Progr.*, doi: 10.1109/ICPHM.2012.6299511.
 - [11] C. Liu, J. Tan, Z. Huang. 2022. Fault Diagnosis of Rolling Element Bearings Based on Adaptive Mode Extraction, *Machines*, vol. 10, no. 4, p. 260, doi: 10.3390/MACHINES10040260.
 - [12] T. Liu, Z. Luo, J. Huang, S. Yan. 2018. A comparative study of four kinds of adaptive decomposition algorithms and their applications, *Sensors*, vol. 18, no. 7, pp. 1-51, doi: 10.3390/s18072120.
 - [13] N. H. M. M. Shrifan, M. F. Akbar, N. A. Mat Isa. 2021. Maximal overlap discrete wavelet-packet transform aided microwave nondestructive testing, *NDT E International*, vol. 119, no. January, p. 102414, doi: 10.1016/j.ndteint.2021.102414.
 - [14] X. Zhang, C. Sheng, W. Ouyang, L. Zheng. 2023. Fault diagnosis of marine electric thruster bearing based on fusing multi-sensor deep learning models, *Measurement*, vol. 214, p. 112727, doi: 10.1016/j.measurement.2023.112727.
 - [15] L. Nie, L. Zhang, S. Xu, W. Cai, H. Yang. 2022. Remaining useful life prediction for rolling bearings based on similarity feature fusion and convolutional neural network, *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 44, no. 8, pp. 1-16, doi: 10.1007/s40430-022-03638-0.
 - [16] J. Liu, C. Pan, F. Lei, D. Hu, H. Zuo. 2021. Fault prediction of bearings based on LSTM and statistical process analysis, *Reliability Engineering & System Safety*, vol. 214, no. March, p. 107646, doi: 10.1016/j.res.2021.107646.
 - [17] M. Yan, L. Xie, I. Muhammad, X. Yang, Y. Liu. 2022. An effective method for remaining useful life estimation of bearings with elbow point detection and adaptive

- regression models, *ISA Trans.*, vol. 128, pp. 290-300, doi: 10.1016/j.isatra.2021.10.031.
- [18] A. J. Ferreira, M. A. T. Figueiredo. 2012. Efficient feature selection filters for high-dimensional data, *Pattern Recognit. Letters*, vol. 33, no. 13, pp. 1794-1804, doi: 10.1016/j.patrec.2012.05.019.
- [19] Z. Liu, M. J. Zuo, Y. Qin. 2016. Remaining useful life prediction of rolling element bearings based on health state assessment, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 230, no 2, p. 314-330. doi: 10.1177/0954406215590167.
- [20] Y. Sun, S. Li, X. Wang. 2021. Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image, *Measurement*, vol. 176, p. 109100, doi: 10.1016/j.measurement.2021.109100.
- [21] M. Biswal, M. Mishra, V. K. Sood, R. C. Bansal, A. Y. Abdelaziz. 2022. Savitzky-Golay Filter integrated matrix pencil method to identify high impedance fault in a renewable penetrated distribution system, *Electric Power Systems Research*, vol. 210, p. 108056, doi: 10.1016/j.epsr.2022.108056.
- [22] T. H. Loutas, D. Roulias, G. Georgoulas. 2013. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression, *IEEE Transactions On Reliability*, vol. 62, no. 4, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/6645455/>
- [23] FEMTO-ST Institute. 2012. IEEE PHM 2012 Prognostic challenge outline, experiments, scoring of results, winners. In: Proceedings of IEEE international conference on prognostics and health management, Denver, CO, 2012, pp.1-11.
- [24] M. S. Rathore, S. P. Harsha. 2022. An Attention-based Stacked BiLSTM Framework for Predicting Remaining Useful Life of Rolling Bearings, *Applied Soft Computing*, vol. 131, p. 109765, doi: 10.1016/j.asoc.2022.109765.
- [25] K. Lotfi, H. Bonakdari, I. Ebtehaj, F. S. Mjalli, M. Zeynoddin, R. Delatolla, B. Gharabaghi. 2019. Predicting wastewater treatment plant quality parameters using a novel hybrid linear-nonlinear methodology, *Journal of environmental management*, vol. 240, p. 463-474., doi: 10.1016/j.jenvman.2019.03.137.
- [26] Y. Liu, Z. Hu, Y. Zhang. 2021. Bearing feature extraction using multi-structure locally linear embedding, *Neurocomputing*, vol. 428, pp. 280-290, doi: 10.1016/j.neucom.2020.11.048.
- [27] C. Barile, C. Casavola, G. Pappaletta, V. Paramsamy Kannan. 2022. Laplacian score and K-means data clustering for damage characterization of adhesively bonded CFRP composites by means of acoustic emission technique, *Applied Acoustics*, 2022, vol. 185, p. 108425, doi: 10.1016/j.apacoust.2021.108425.
- [28] S. Hong, Z. Zhou, E. Zio, K. Hong. 2014. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method, *Digital Signal Processing*, 2014, vol. 27, p. 159-166, doi: 10.1016/j.dsp.2013.12.010.
- [29] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin. 2017. A recurrent neural network based health indicator for remaining useful life prediction of bearings, *Neurocomputing*, vol. 240, pp. 98-109, doi: 10.1016/j.neucom.2017.02.045.