

DATA NOTE

Open Access



HUST bearing: a practical dataset for ball bearing fault diagnosis

Nguyen Duc Thuan¹ and Hoang Si Hong^{1*}

Abstract

Objectives The rapid growth of machine learning methods has led to an increase in the demand for data. For bearing fault diagnosis, the data acquisition is time-consuming with complicated processes. Existing datasets are only focused on only one type of bearing, which limits real-world applications. Therefore, the objective of this work is to propose a diverse dataset for ball bearing fault diagnosis based on vibration.

Data description In this work, we introduce a practical dataset named *HUST bearing*, which provides a large set of vibration data on different ball bearings. This dataset contains 99 raw vibration signals of 6 types of defects (inner crack, outer crack, ball crack, and their 2-combinations) on 5 types of bearing (6204, 6205, 6206, 6207, and 6208) at 3 working conditions (0 W, 200 W, and 400 W). Each vibration signal is sampled at a rate of 51,200 samples per second for 10 s. The data acquisition system is elaborately designed with high reliability.

Keywords Dataset, Bearing fault, Fault diagnosis

Objective

Electric motors are indispensable equipment in manufacturing plants. They often have to work continuously and are subjected to a variety of heavy loads, and operate in harsh environments. Therefore, bearings in electric motors become vulnerable components. Statistics show that electric motor faults related to bearings account for nearly 50% of the total number of common faults [1]. The consequences when a bearing fault occurs are severe, including the destruction of mechanical structures, production stoppage, and costly repair [2]. For these reasons, the early detection and diagnosis of bearing faults is an urgent task and plays a crucial role in electrical machine health monitoring.

When a bearing defect appears, it generates abnormal signals such as vibrations, sound waves, and temperature [3]. In particular, the vibration signal often contains useful information, so it is widely used to detect and diagnose bearing faults [4]. Vibration signal-based methods are usually approached in three approaches, i.e., physical model-based, signal processing, and machine learning-based methods [5]. Machine learning methods have shown superiority over conventional methods when giving results with high accuracy, especially thanks to deep learning models [6]. The rapid growth of machine learning methods has led to an increase in the demand for data. These data are collected from defective bearings to support the learning process.

In practice, the data acquisition process takes time with difficult and complicated processes [7]. Data acquisition tasks include defect generation on bearings, measurement system setup, and data recording. Fortunately, many open-access datasets have been published for scientists and engineers to research and develop. Some of the most popular bearing fault data sets are CWRU [8],

*Correspondence:

Hoang Si Hong
hong.hoangsy@hust.edu.vn

¹School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, No. 1 Dai Co Viet Road, Hanoi, Vietnam



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Paderborn [9], IMS [10], and Pronostia [11]. However, the common point of the mentioned above data sets is to focus on only one type of bearing and this is the limitation for the in-real applications. This is the motivation for us to propose a novel dataset for ball bearing fault diagnosis with a high diversity compare to previous datasets.

Data description

Defects generation

In general, there are two ways to create bearing defects: artificial and natural. However, the natural fault generation process is relatively complex, so the bearing defects in our dataset are artificially generated. Our object is electric motors operating at low loads, so the bearings used to create defects consist of ID 6204, 6205, 6206, 6207, and 6208 ball bearings from KG Bearing India. As for the fault types, from our observation, when a bearing appears with one defect, it easily leads to another defect on another component due to mechanical interaction. Therefore, our data set includes common single faults such as inner race fault, outer race fault, and ball fault. In addition, double faults are also incorporated as inner and outer faults, inner and ball faults, and outer and ball faults (see Data file 1 in Table 1). We use the wire-cutting method to create a 0.2 mm micro-crack width, simulating the initial state of the fault. We only created one early-stage fault size because we believe that early diagnosis of faults is much more important than diagnosing the extent of the fault so that 0.2 mm fault is enough. Descriptions for the bearing dimensions (inner diameter in mm, outer diameter in mm, ball diameter in mm, and ball number) are: (20, 47, 7.6, 8) for 6204, (25, 52, 7.8, 9) for 6205, (30, 62, 9.0, 9) for 6206, (35, 72, 11.0, 9) for 6207, and (40, 80, 12.0, 9) for 6208.

Test bench setup

The basic layout of the test bench is demonstrated in Data file 2 in Table 1. The test bench consists of a 750 W (1 HP) induction motor driving a multi-step shaft and a powder brake of Leroy Somer. The multi-step shaft means the shaft with multiple step changes in diameter. The motor

is controlled by an inverter and the powder brake plays as a simulated load. Furthermore, a torque transducer and a dynamometer are also mounted to the shaft to monitor the load and velocity of the motor. Defective bearings are mounted to different types of housings and these housings can be flexibly replaced on the multi-step shaft. On the bearing, an accelerometer of PCB 325C33 is installed in the vertical direction to measure vibration.

Data records

The data acquisition system includes an acceleration sensor, a measurement module, a chassis, and Labview software. All listed components built a fully automated measurement system for the data acquisition step. Vibration analog signals from the sensor are sent to the NI-9234 sound and vibration input module which provides an analog-to-digital conversion. Then converted digital signals from NI-9234 are synchronized and conditioned by NI-CompactDAQ chassis before transferring to the software using USB connectivity. In the software, vibration signals are presented, analyzed, and acquired. The configuration of the measurement module and chassis is set in the software.

The description of all 99 data files is shown in Data set 1 in Table 1. It includes vibration data of 27 prototype defective bearings and 3 healthy bearings at three load conditions of 0 W, 200 W, and 400 W. In addition, the introduced dataset provides 30 vibration data of the defective bearings at a run-up time of 5 s. The vibration during run-up may early alert the adverse condition of the bearings, especially for high-power electric motors which consume starting time. Each data file consists of one or two letters representing the bearing condition, followed by the bearing type and loading condition. For example, for the file I402.mat, the letter I indicates Inner race fault, the digit 4 indicates bearing type 6204, and digit 2 indicates 200 W loading condition (see also abbreviations and [12]). In the data files, “data” is the vibration signal, “fs” is the shaft frequency in Hz, “rpm” is the shaft angular velocity in RPM during run-up time, “ru” and “ru_raw” are the vibration signals during run-up time. It is worth noting that all data are collected with a high sample rate of 51.2 kHz in which these data can record signal changes in detail.

Limitations

- The proposed dataset is missing 2 cases which are ball fault (B) and inner and ball fault (IB) for bearing ID 6204.
- The intensity of defect in different bearings is different because the depth of the crack is not the same. Therefore, the mean/peak amplitude of the vibration signal is markedly different in different cases.

Table 1 Overview of data files/data sets

| Label | Name of data file/data set | File types (file extension) | Data repository and identifier (DOI or accession number) |
|-------------|----------------------------|-----------------------------|--|
| Data set 1 | HUST bearing dataset | MAT files (.mat) | Mendeley Data (https://doi.org/10.17632/cb-v7jyx4p9) [13] |
| Data file 1 | defects | PNG file (.png) | Mendeley Data (https://doi.org/10.17632/cb-v7jyx4p9) [13] |
| Data file 2 | testbench | PNG file (.png) | Mendeley Data (https://doi.org/10.17632/cb-v7jyx4p9) [13] |

- This dataset does not take into account environmental noise as well as mechanical imperfections.

Abbreviations

| | |
|----|-----------------|
| ID | Identification |
| I | Inner |
| O | Outer |
| B | Ball |
| IO | Inner and outer |
| IB | Inner and ball |
| OB | Outer and ball |

Acknowledgements

We thank Assoc. Prof. Dang Quoc Vuong and Dr. Pham Hong Hai (School of Electrical and Electronic Engineering, Hanoi University of Science and Technology) for their support.

Author contributions

NDT designed the experiments. HSH wrote the manuscript. All authors reviewed the manuscript.

Funding

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2022-PC-006.

Data availability

The data described in this Data note can be freely and openly accessed on the Mendeley Data repository under <https://doi.org/10.17632/cbv7jyx4p9>. Please see Table 1 and references [13] for details and links to the data.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Received: 14 March 2023 / Accepted: 18 June 2023

Published online: 06 July 2023

References

1. Khan SA, Kim J-M. Automated bearing Fault diagnosis using 2D analysis of vibration acceleration signals under variable speed conditions. *Shock Vib.* 2016;2016:1–11.
2. Zhang X, Liu Z, Miao Q, Wang L. Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time–frequency atom dictionary. *Mech Syst Signal Process.* 2018;107:29–42.
3. Sohaib M, Kim J-M. Fault diagnosis of Rotary Machine Bearings under Inconsistent Working Conditions. *IEEE Trans Instrum Meas.* 2020;69:3334–47.
4. Yang D-M. The detection of Motor Bearing Fault with maximal overlap Discrete Wavelet Packet transform and Teager Energy Adaptive Spectral Kurtosis. *Sensors.* 2021;21:6895.
5. Jiao J, Zhao M, Lin J, Liang K. A comprehensive review on convolutional neural network in machine fault diagnosis. *Neurocomputing.* 2020;417:36–63.
6. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science (80-).* 2006;313:504–7.
7. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: a systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process.* 2018;104:799–834.
8. Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mech Syst Signal Process.* 2015;64–65:100–31.
9. Kimotho JK, Lessmeier C, Sextro W, Zimmer D. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification. *Third Eur Conf Progn Heal Manag Soc* 2016. 2016; Cm:152–6.
10. Qiu H, Lee J, Lin J, Yu G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J Sound Vib.* 2006;289:1066–90.
11. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests. *IEEE Int Conf Prognostics Health Manage PHM'12.* 2012;Denver, Colorado, United States: IEEE Catalog Number : CPF12PHM-CDR:1–8.
12. Thuan ND, Hong HS. HUST bearing: a practical dataset for ball bearing fault diagnosis. 2023. <https://doi.org/10.48550/arXiv.2302.12533>.
13. Hong HS, Thuan N. HUST bearing. Mendeley Data. 2023. <https://doi.org/10.17632/cbv7jyx4p9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.