Data Descriptor

Not peer-reviewed version

Curated Vibration Features and an Interpretable Gearbox Health Index (GHI) Baseline for Condition Monitoring Bench-Marking

Krisztian Horvath *

Posted Date: 6 March 2026

doi: 10.20944/preprints202603.0506.v1

Keywords: gearbox condition monitoring; vibration features; gear mesh frequency; sideband index; health indicator

Data Descriptor

Curated Vibration Features and an Interpretable Gearbox Health Index (GHI) Baseline for Condition Monitoring Bench-Marking

Krisztian Horvath

Department of Vehicle Development, Audi Hungaria Faculty Of Engineering, Széchenyi István Universi-ty, H-9026, Győr, Egyetem tér 1; horvath.krisztian@ga.sze.com

Abstract

This data descriptor provides a standardized and reproducible subsystem-level represen-tation of the NREL wind turbine gearbox condition monitoring benchmarking dataset. The released records are derived from Healthy (H1–H10) and Damaged (D1–D10) meas-urement files and include subsystem-level standardized indices (KHI_HS, KHI_IMS, KHI_PL) together with a calibrated 0–1 Gearbox Health Index (GHI). The indices are gen-erated using a fully specified and deterministic feature extraction and aggregation work-flow based on established vibration indicators and healthy-referenced normalization. The Zenodo deposit contains machine-readable CSV tables intended to support transparent benchmarking across supervised classification and anomaly detection studies. The pro-posed GHI is introduced as an interpretable and reproducible reference baseline rather than an optimized diagnostic model. Technical validation demonstrates condition-level separability within the analyzed dataset while emphasizing the descriptive nature of the index. By releasing structured derived records and a documented regeneration procedure, this work enables implementation-independent comparison of gearbox condition moni-toring approaches and supports reproducible evaluation of alternative health index for-mulations.

Dataset: 10.5281/zenodo.18832721

Dataset License: Creative Commons Attribution 4.0 International (CC-BY 4.0)

Keywords: gearbox condition monitoring; vibration features; gear mesh frequency; sideband index; health indicator

---

1. Introduction

Vibration-based gearbox condition monitoring has been extensively investigated and is considered one of the most mature approaches in rotating machinery diagnostics. Numerous studies show that vibration signals contain rich information about gear defects and degradation mechanisms, and both classical signal-processing methods and modern machine learning approaches have demonstrated high diagnostic accuracy [1].

At the same time, several reviews emphasize that reproducible benchmarking remains limited due to methodological heterogeneity. Differences in signal preprocessing (e.g., filtering, segmentation), spectral estimation (e.g., windowing, averaging), feature engineering (e.g., statistical indicators, envelope features), and normalization strategies often lead to inconsistent evaluation results across studies, even when similar datasets are used [2]. As a result, reported performance metrics are frequently influenced by implementation-specific design choices rather than reflecting solely the intrinsic information content of the vibration data. This challenge has been repeatedly noted in the context of data-driven prognostics and health management research [3].

To address these issues, the present work introduces a structured, machine-readable, feature-level representation of a gearbox condition monitoring benchmark dataset. Instead of proposing a new diagnostic or classification model, the focus is placed on consistent and fully specified feature extraction procedures, standardized feature representation, and deterministic regeneration of all derived records. Prior research highlights that transparent and well-documented preprocessing pipelines are essential for reproducibility and fair comparison in machine learning-based diagnostics [4].

In addition to curated per-channel features and healthy-referenced standardized scores, a simple hierarchical Gearbox Health Index (GHI) is provided as a transparent baseline. Composite health indicators have been shown to support interpretability and condition trend analysis in rotating machinery monitoring [5]. The proposed index is intentionally simple, interpretable, and reproducible; it is not intended as an optimized diagnostic solution, but rather as a stable reference anchor that enables fair comparison of alternative modeling approaches built on the same standardized feature representation.

2. Data Description

The source waveform data originate from the publicly available Wind Turbine Gearbox Condition Monitoring Vibration Analysis Benchmarking Dataset released by the National Renewable Energy Laboratory (NREL) [Sheng, 2014, DOI: 10.25984/1844194]. The original dataset contains raw vibration time histories in MATLAB (.mat) format for both Healthy (H1–H10) and Damaged (D1–D10) gearbox states.

The present work does not redistribute raw waveform data. Instead, it releases structured, machine-readable derived records generated from the original dataset through a fully specified and deterministic feature extraction workflow.

The released records include:

- Subsystem-level standardized indices (KHI_HS, KHI_IMS, KHI_PL)
- A calibrated Gearbox Health Index (GHI) mapped to a 0–1 interval
- Supplementary intermediate feature exports and trend-analysis tables used for figure generation and benchmarking transparency

All derived records are provided as CSV tables and are reproducible using the accompanying code and documented procedure.

3. Feature Extraction Workflow

3.1. Preprocessing

Each vibration channel is processed independently. The selected signal segment is detrended to remove constant offsets. Spectral features are computed using Welch power spectral density estimation. Where computational efficiency is required, downsampling may be applied while preserving relevant frequency content.

The rotational speed signal is used to compute the fundamental rotational frequency, defined as

$$f1_x = rpm/60, \tag{1}$$

The gear mesh frequency is defined as

$$f_{GMF} = Z \times f_{1x}, \tag{2}$$

,where Z is the reference tooth count of the high-speed pinion.

3.2. Curated Feature Set

A compact and interpretable feature set is computed for each file and each measurement channel, following established practices in vibration-based gearbox diagnostics. The selected indicators capture both global vibration energy and fault-sensitive modulation characteristics that are widely reported in the literature [7,8].

- RMS (Root Mean Square) — Represents the overall vibration energy and is commonly used as a baseline health indicator, as increases in mechanical damage often manifest as elevated global vibration levels [7,8].
- Kurtosis (excess) — Measures deviation from a Gaussian amplitude distribution and is particularly sensitive to impulsive events caused by localized gear tooth or bearing defects [9].
- GMF_Energy — Quantifies spectral energy in a narrow band centered around the gear mesh frequency (GMF). Faults such as tooth wear or breakage typically produce amplitude growth at the GMF and its harmonics, making this component fundamental in gearbox diagnostics [10].
- SidebandIndex — Expresses the ratio of modulation sideband energy relative to the GMF-centered component. Modulation sidebands around the GMF are well-known indicators of gear defects and transmission errors, and sideband energy ratios are widely applied in gear fault detection [11].
- EnvelopePeak — Captures the peak amplitude of the envelope spectrum within a defined band. Envelope analysis enhances fault-induced impulsive components and is especially effective for early detection of localized gear and bearing damage [1].

These features represent commonly used indicators in gearbox diagnostics and enable direct comparison across analytical approaches.

Table 2. Feature Definitions and Parameters.

Feature	Domain	Definition	Parameters	Diagnostic Relevance
RMS	Time	√(mean(x^2))	20 s segment	Overall vibration energy
Kurtosis	Time	4th standardized moment (excess)	–	Impulsiveness indicator
GMF_Energy	Frequency	Energy in ±10% band around f_GMF	Welch PSD, nperseg = 4096, Hann window, 50% overlap, fs = 40 kHz	Gear mesh excitation
SidebandIndex	Frequency	(E_GMF−1× + E_GMF+1×) / E_GMF	±1× spacing	Modulation / defect growth
EnvelopePeak	Envelope spectrum	Maximum PSD of Hilbert envelope	200–5000 Hz band	Bearing/impact sensitivity

The individual channel-level feature tables are used internally for subsystem aggregation. The Zenodo deposit releases subsystem-level standardized outputs and the aggregated GHI baseline rather than per-channel raw feature tables.

3.3. Healthy-Based Standardization

For cross-channel comparability, feature values are standardized using statistics derived exclusively from the Healthy subset. Baseline normalization using healthy-condition statistics is a common strategy in rotating machinery diagnostics, as it reduces variability caused by operational differences and improves sensitivity to fault-induced deviations [12,13]. For each channel and feature, the Healthy mean and standard deviation are computed, and standardized z-scores are obtained accordingly.

Missing raw feature values are preserved as NaN in the primary feature table. During standardized aggregation, missing contributions are excluded from summation to maintain numerical stability. Power spectral density estimates were computed using Welch's method with a

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 March 2026 doi:10.20944/preprints202603.0506.v1

segment length of 4096 samples ($n_{perseg} = 4096$), a Hann window, and 50% overlap between adjacent segments at a sampling frequency of 40 kHz.

3.4. Example Hierarchical Health Index

To provide a transparent benchmark reference, a hierarchical aggregation scheme is included. Standardized channel-level scores are combined into Channel Health Indices (CHI). Channels are then grouped into subsystem-level indices (KHI). A Gearbox Health Index (GHI) is computed as a weighted combination of subsystem indices and mapped to a 0–1 interval using logistic calibration. The GHI defined in this study is a deterministic reference metric introduced for benchmarking purposes. It is not intended to represent an industry-standard diagnostic indicator.

The GHI is intentionally not optimized for maximal discrimination; it is provided as a simple, interpretable, and reproducible reference baseline. It is included solely as a reproducible and interpretable baseline against which alternative models may be compared. The hierarchical structure is provided for transparency and interpretability, not for claiming diagnostic optimality. The complete deterministic processing pipeline from raw waveform data to subsystem indices and the calibrated GHI baseline is shown in Figure 1.

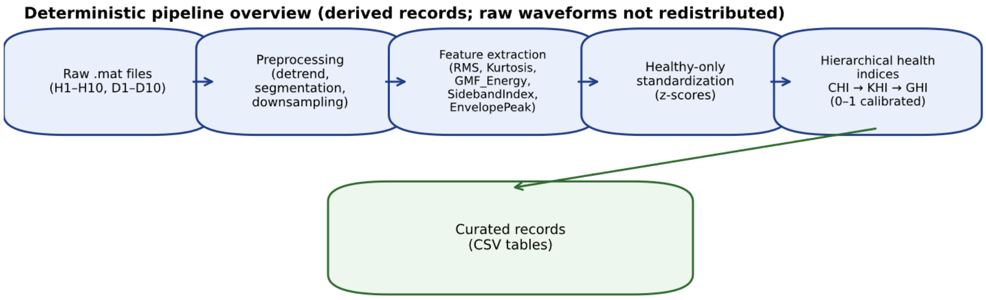

Figure 1. Pipeline overview from raw (.mat) files to feature tables and health indices.

4. Data Records

The Zenodo deposit contains structured derived records generated deterministically from the original NREL waveform dataset.

4.1. GHI_per_file.csv

Granularity: file-level (one row per measurement file).
Columns:

- file – file identifier (H1–H10, D1–D10)
- group – condition label (healthy or damaged)
- KHI_HS – subsystem index for high-speed stage
- KHI_IMS – subsystem index for intermediate stage
- KHI_PL – subsystem index for planetary stage
- GHI_raw – uncalibrated aggregate score
- GHI – calibrated 0–1 Gearbox Health Index
- This table constitutes the primary benchmarking baseline product of the deposit.

4.2. features_with_zscores.csv

Granularity: file-level.
Columns:

- file
- group
- KHI_HS

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 March 2026 doi:10.20944/preprints202603.0506.v1

- KHI_IMS
- KHI_PL

This table provides subsystem-level standardized outputs intended for direct use in supervised classification, anomaly detection, and comparative benchmarking studies.

4.3. Supplementary Analytical Tables

Additional CSV files included in the deposit provide intermediate feature exports, D1–D2 comparisons, D1–D10 trend summaries, and figure-support tables used in the manuscript. These records are included for transparency and reproducibility of reported figures but are not required for baseline benchmarking usage.

5. Technical Validation

Technical validation confirms that the curated records retain condition-relevant information and support benchmarking.

5.1. Condition Separability

The released Gearbox Health Index (GHI) was evaluated as a descriptive separability indicator between Healthy and Damaged files. ROC analysis indicates complete separation within the analyzed dataset (AUC = 1.0). Given the limited dataset size (N = 20), this result must be interpreted cautiously. A single misclassification would reduce the AUC to approximately 0.90, illustrating the statistical fragility of perfect separation in small samples. The observed AUC therefore reflects a descriptive property of this homogeneous dataset rather than a claim of generalizable diagnostic performance.

The GHI is intended as a reproducible reference baseline for benchmarking alternative modeling approaches. No train/test split or cross-validation procedure is applied, as the index is presented strictly as a deterministic descriptive baseline.

The file-level distribution of the deterministic GHI values across Healthy and Damaged records is illustrated in Figure 2.

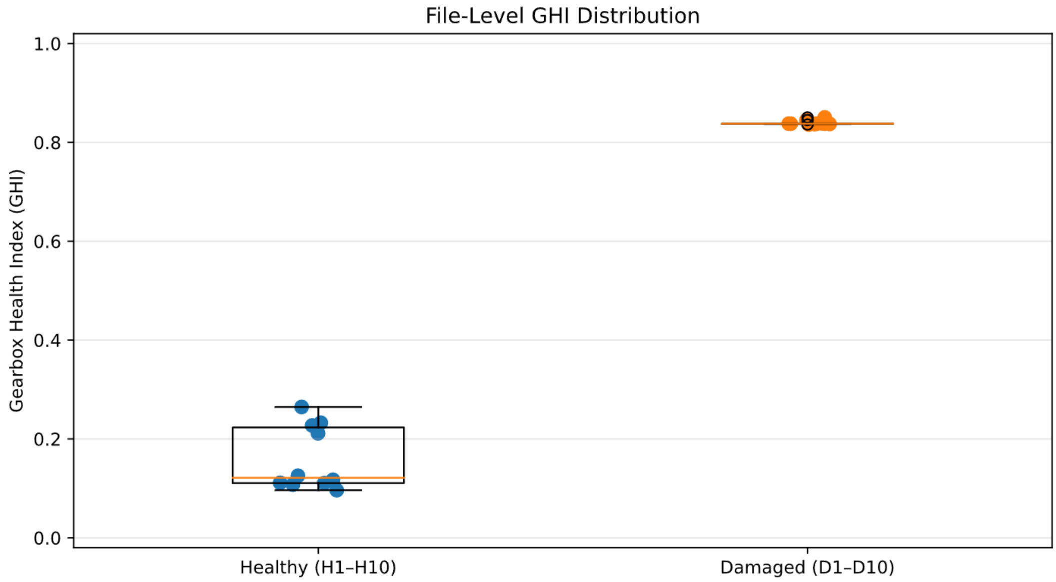

Figure 2. File-level distribution of the deterministic Gearbox Health Index (GHI) across Healthy (H1–H10) and Damaged (D1–D10) records. The clear separation reflects a descriptive property of the curated dataset and should not be interpreted as evidence of generalizable diagnostic performance.

5.2. Progression Behavior

Rank-based correlation analysis across ordered Damaged files indicates limited monotonic progression in the GHI values. This confirms that while the index captures condition-level separation, it is not optimized for degradation trajectory modeling. The baseline is intentionally designed for interpretability and reproducibility rather than predictive optimality.

5.3. Reproducibility

All released tables are deterministic outputs of the documented workflow. Re-running the feature extraction and aggregation scripts with identical inputs produces identical CSV records.

5.4. Limitations

The present dataset comprises 20 files (10 healthy and 10 damaged), which limits statistical robustness. Performance metrics such as ROC-AUC are highly sensitive to individual file-level variations. The dataset represents a controlled experimental campaign conducted under homogeneous operating conditions; therefore, results reflect internal consistency rather than cross-platform generalizability.

The released GHI is a deterministic aggregation of standardized features and does not incorporate cross-validation, hyperparameter tuning, or probabilistic modeling. Consequently, the index should not be interpreted as an optimized diagnostic model.

Future extensions should include larger multi-operating-condition datasets, experimental validation under variable load regimes, and adaptive frequency-band segmentation strategies to enhance robustness and applicability.

6. Usage Notes

The released derived records support:

- Baseline benchmarking of classification methods
- Unsupervised anomaly detection experiments
- Comparative evaluation of health index formulations
- Sensitivity analysis across subsystem-level standardized indicators

Users are encouraged to report both subsystem-level (KHI) and global (GHI) performance metrics to ensure comparability with the reference baseline provided in this work.

7. Conclusions

This data descriptor presents a standardized and fully documented feature-level representation of a gearbox vibration condition monitoring benchmark dataset. The primary objective of this work is not algorithmic innovation, but the elimination of methodological ambiguity through a deterministic and reproducible feature extraction and aggregation workflow derived from the original NREL waveform data. The released subsystem-level standardized indices (KHI_HS, KHI_IMS, KHI_PL) and the calibrated Gearbox Health Index (GHI) provide a transparent and interpretable baseline representation of the dataset. All processing steps are explicitly defined, and identical inputs yield identical outputs, ensuring implementation-independent reproducibility. Within the analyzed dataset (N = 20), the example GHI demonstrates clear condition-level separability between Healthy and Damaged files. However, this separability is reported solely as a descriptive characteristic of the curated dataset and does not imply statistical generalization. The index is intentionally simple and not optimized for predictive performance or degradation trajectory modeling. The core contribution of this work lies in structured data curation, standardized healthy-referenced normalization, and the release of machine-readable derived records together with a documented regeneration procedure. By removing variability caused by heterogeneous preprocessing pipelines, the dataset enables fair benchmarking of supervised classifiers, unsupervised anomaly detection methods, and alternative health index formulations. This open and deterministic representation supports methodological transparency and reproducible comparison practices in vibration-based gearbox condition monitoring research. The released records are intended to serve as a stable reference baseline for future methodological developments.

Data Availability Statement: The original waveform dataset analyzed in this study is publicly available from the National Renewable Energy Laboratory (NREL): Sheng, S. (2014). Wind Turbine Gearbox Condition Monitoring Vibration Analysis Benchmarking Datasets. https://doi.org/10.25984/1844194. The derived feature-level and health-index records generated in this study are openly available via Zenodo at: https://doi.org/10.5281/zenodo.18832721.

References

1. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 138, 106587. https://doi.org/10.1016/j.ymssp.2019.106587
2. Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. (2016). Fault Diagnosis for Rotating Machinery Using Vibration Signal Analysis. Sensors, 16(527). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4934321/
3. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016
4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
5. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012
6. Sheng, Shawn. "Wind Turbine Gearbox Condition Monitoring Vibration Analysis Benchmarking Datasets." , Mar. 2014. https://doi.org/10.25984/1844194
7. Tiboni, M. (2022). A review on vibration-based condition monitoring of rotating machinery. Applied Sciences, 12(3), 972. https://www.mdpi.com/2076-3417/12/3/972
8. Caesarendra, W., & Tjahjowidodo, T. (2017). A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. Machines, 5(4), 21. https://doi.org/10.3390/machines5040021
9. Wang, Y., & Randall, R. B. (2016). Review of spectral kurtosis for fault detection, diagnosis and prognosis of rotating machinery. Mechanical Systems and Signal Processing. https://doi.org/10.1016/j.ymssp.2015.09.017
10. Sait, A. S. S., & Sharaf-Eldeen, Y. I. (2016). A review of gearbox condition monitoring based on vibration analysis techniques: Diagnostics and prognostics. ResearchGate.
11. Zhang, M., Li, Z., & Peng, Z. (2021). An improved sideband energy ratio for fault diagnosis of planetary gearboxes. Journal of Sound and Vibration, 500, 116046. https://doi.org/10.1016/j.jsv.2021.116046
12. Antoni, J. (2006). The spectral kurtosis: A useful tool for characterising non-stationary signals. Mechanical Systems and Signal Processing, 20(2), 282–307. https://doi.org/10.1016/j.ymssp.2004.09.001
13. Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing, 96, 1–15. https://doi.org/10.1016/j.sigpro.2013.04.015