POLITECNICO DI TORINO
Repository ISTITUZIONALE

Multimodal Large Language Models for Few-Shot Bearing Fault Diagnosis: Impact of Prompt Design and Comparison with Prototypical Networks

Original
Multimodal Large Language Models for Few-Shot Bearing Fault Diagnosis: Impact of Prompt Design and Comparison with Prototypical Networks / Di Maggio, L.G., Delprete, C., Brusa, E.. - (2026), pp. 1-7. (2026 International Conference on Control, Automation and Diagnosis, ICCAD 2026 Lisbon (PRT) 07-09 July 2026) [10.1109/ICCAD69956.2026.11643196].

Availability:
This version is available at: 11583/3016245 since: 2026-10-02T16:40:52Z

Publisher:
Institute of Electrical and Electronics Engineers

Published
DOI:10.1109/ICCAD69956.2026.11643196

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

22 October 2026

Multimodal Large Language Models for Few-Shot Bearing Fault Diagnosis: Impact of Prompt Design and Comparison with Prototypical Networks

Luigi Gianpio Di Maggio, Cristiana Delprete, Eugenio Brusa
Dept. of Mechanical and Aerospace Engineering
Politecnico di Torino
Torino, Italy
{luigi.dimaggio,cristiana.delprete,eugenio.brusa}@polito.it

***Abstract*—Few-shot learning (FSL) models for bearing fault diagnosis often require meta-training on dedicated datasets and are limited by the scarcity of fault data in real industrial scenarios. Multimodal Large Language Models (MLLMs) offer an alternative, enabling few-shot classification via prompting on time–frequency images of vibration signals. This work studies medium-to-large bearings tested on a dedicated test rig and analyzes (i) when MLLM performance is statistically superior to a Prototypical Network FSL baseline, and (ii) how prompt design affects few-shot classification.**

On a 4-way task with Continuous Wavelet Transform (CWT) images and 1-, 5-, and 10-shot configurations, we perform multiple repetitions and compare models using independent two-sample t-tests and effect sizes defined by Cohen's d. At a fixed rotation speed, advanced MLLMs significantly outperform the FSL baseline, whereas Prototypical Networks remain competitive under variable speeds. The prompt study shows that more verbose templates do not bring benefits. The concise prompt generally matches or outperforms the detailed one, which can even degrade performance. Overall, concise prompts are sufficient and often preferable, and MLLMs emerge as an effective complement to conventional FSL approaches in realistic industrial scenarios.

***Index Terms*—Few-shot learning, multimodal large language models (LLM), bearing fault diagnosis, prototypical networks, prompt engineering, vibration analysis**

I. Introduction

Fault diagnosis in rolling bearings is a key element of condition monitoring for rotating machines, as these components are often the cause of sudden shutdowns and costly plant downtime [1]–[3]. Machine learning (ML) methods can extract complex patterns from vibration signals and time–frequency representations, achieving high performance in various scenarios [4]–[7]. However, in industrial practice labeled failure data are typically scarce, heterogeneous, and collected under different operating conditions.

Few-Shot Learning (FSL) [8], [9] addresses these issues by learning new tasks from a few observations per class, often via episodic meta-learning and similarity metrics. Many recent works extend FSL to fault diagnosis in bearings and rotating machines, with variants based on meta-learning, multi-scale prototypical networks, relation networks, and contrastive schemes [10]–[17]. In parallel, modern architectures such as Vision Transformers, state-space models, and Kolmogorov–Arnold Networks have been explored to improve generalization and interpretability [18]–[22].

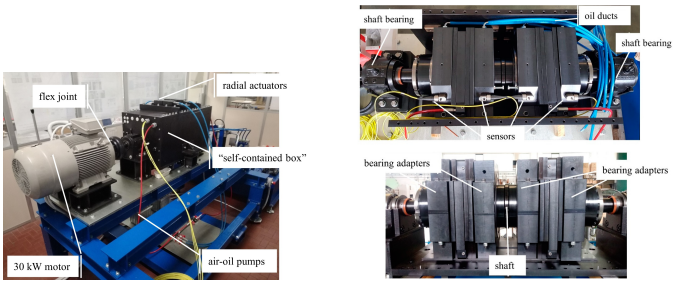
However, most studies, rely on small bearing benchmarks at constant speeds [23], while heavy industrial applications (e.g., rolling or paper mills) require models that handle medium-to-large bearings under varying radial and axial loads [24], [25]. Meta-training-based FSL approaches also depend on domain-specific ground-truth datasets, which limits flexibility when data are fragmented or decentralized, such as in federated settings [26], [27].

Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs) show strong *in-context learning* abilities, solving new tasks from a few examples in the prompt, without further training [28]–[30]. Recent works explore language and vision–language models for mechanical diagnostics, both with textual pipelines and starting from time–frequency images [31]–[36]. Using pre-trained MLLMs as generic backbones for few-shot diagnosis directly on CWT representations is thus appealing, but several questions remain open.

In particular, there is a lack of studies that quantify:

1) when pre-trained MLLMs are statistically superior to a robust FSL baseline;
2) how prompt design (concise instructions vs. detailed defect descriptions) affects few-shot classification in realistic industrial scenarios.

We address these points on a medium-to-large spherical bearing test rig with independent radial and axial loads and a systematic vibration dataset with four conditions and ten speeds [24], [37]–[39]. We consider a 4-way K-shot protocol on CWT images, comparing proprietary MLLMs with a Prototypical Network with Swin Transformer backbone, in both heterogeneous (all speeds mixed) and homogeneous (single speed) regimes. Two-sample t-tests and effect sizes quantify differences and the impact of *concise* vs. *detailed* prompts.



(a) (b)

Fig. 1: Test rig for medium-to-large bearings. (a) External view of the structure with motor and shaft line. (b) Detail of the bearing housing and accelerometer positions.

II. Test Rig, Dataset, and Few-Shot Protocol

A. Test Rig and Data Acquisition

The experiments were conducted on a test rig for medium-to-large bearings at Politecnico di Torino [39], [40]. The rig consists of a self-supporting frame with a shaft driven by a three-phase asynchronous motor and a hydraulic system that applies independently controlled radial and axial loads (Fig. 1).

The bearing under investigation is an SKF® 22240 CCK/W33 spherical roller bearing, instrumented via piezoelectric accelerometers mounted on the housing. Acceleration signals are acquired at 20.48 kHz. Four operating conditions are considered:

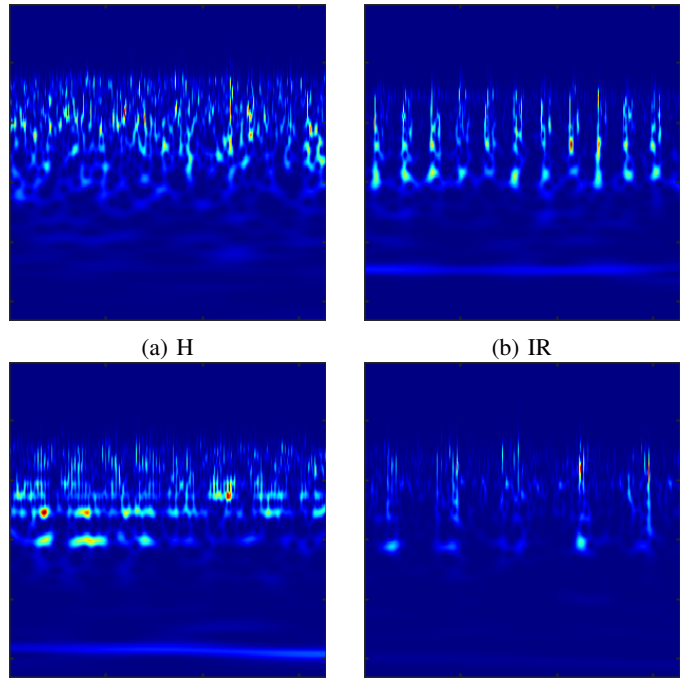
- H: healthy bearing;
- IR: defect on the inner race;
- OR: defect on the outer race;
- B: defect on a rolling element.

Defects are introduced by material removal, with geometry consistent with standard bearing fault datasets. A single load case is analyzed (124.8 kN radial load, zero axial load), while the rotational speed takes ten nominal values: 127, 227, 353, 457, 523, 607, 727, 877, 937, and 997 rpm. For each speed and condition, 30 s of data are recorded.

B. Dataset Construction and Time–Frequency Representation

Each 30 s recording is divided into 50 overlapping segments of duration $T_{\text{seg}} = 0.8$ s (50% overlap). A standard preprocessing chain is applied to each segment. Namely, mean removal and normalization, bandpass filtering in the 1.4–2.8 kHz band, envelope extraction, and computation of a time–frequency representation via a Morse wavelet with 24 voices per octave. The magnitude of the CWT is mapped to 300×300 pixel grayscale images, compatible with the considered vision backbones.

Combining all speeds and conditions (H, IR, OR, B) yields 8000 CWT images, 2000 per class. Example scalograms at 607 rpm are shown in Fig. 2; differences in impulsive and modulated high-frequency components provide discriminative information.



(a) H (b) IR

(c) OR (d) B

Fig. 2: Examples of CWT scalograms for conditions H, IR, OR, and B (607 rpm, 124.8 kN radial load).

C. Few-Shot Task Definition

The problem is formulated as a 4-class few-shot task with label set $\mathcal{C} = \{\text{H}, \text{IR}, \text{OR}, \text{B}\}$. For a given $K \in \{1, 5, 10\}$, each N-way K-shot episode ($N = 4$) consists of:

- a *Support Set* with K labeled examples per class;
- a *Query Set* with one sample per class to be classified.

Support and Query samples are drawn without overlap. For each configuration (value of K, model, prompt strategy), multiple independent episodes are generated to estimate accuracy and its variability. Two operating regimes are considered:

- *all-speeds*: Support and Query sets are sampled over the full speed range (127–997 rpm), inducing strong intra-class heterogeneity;
- *single-speed*: Support and Query sets are restricted to 607 rpm, yielding a more homogeneous scenario.

III. Models and Prompting Strategies

A. Prototypical Baseline with Swin Transformer

As a conventional FSL baseline, we use a Prototypical Network classifier [41] with a Swin Transformer visual backbone pre-trained on ImageNet. The backbone is used as a frozen feature extractor. Then, given a CWT image, the model computes a fixed-dimensional embedding. For each class, a prototype is obtained as the mean embedding of its K support examples, and query samples are classified by nearest prototype in Euclidean distance.

No meta-training is performed on the fault domain. The Swin backbone is fixed and the classifier is fully defined by

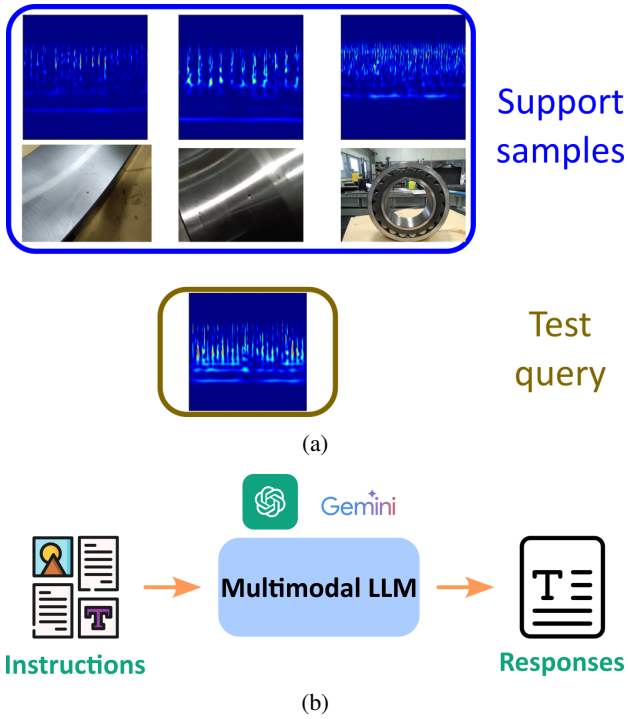


Fig. 3: Few-shot framework. (a) 4-way K-shot episode based on CWT images for conditions H, IR, OR, and B. (b) Interaction with the MLLM via a structured prompt containing Support Set and query image.

the episode-specific Support sets. This mirrors the constraint imposed on MLLMs, which are not fine-tuned on the diagnostic dataset.

B. Multimodal Large Language Models

The MLLM counterpart consists of four generative multimodal models that process text and images jointly:

- GPT-4o;
- GPT-5.1;
- Claude 4.5 Haiku;
- Claude 4.5 Sonnet.

CWT images are provided as visual input via the multimodal APIs, and few-shot classification is handled through in-context learning. To ensure a controlled comparison with the prototypical baseline:

- all MLLMs are used with temperature set to zero;
- CWT images are resized to the API-required resolution (aspect ratio preserved), without additional augmentation;
- for each episode, the same Support Set and Query sample are used for both the Prototypical Network and all MLLMs.

The task is always a 4-way problem (H, IR, OR, B) in 1-, 5-, and 10-shot settings. The overall few-shot framework is illustrated in Fig. 3.

C. Prompting Strategies: Concise vs. Detailed

Each episode is repeated under two prompting conditions:

- **Concise prompt** provides a minimal instruction and defines the model as a machine diagnosis assistant, lists the four classes (H, IR, OR, B), and requires that only the class code is returned.
- **Detailed prompt** provides a more verbose instruction and frames the model as an advanced AI specialized in interpreting CWT images, with richer task description and redundant reminders of the required output format.

In both cases, the context is constructed identically. Namely, a sequence of support examples per class (short sentence plus CWT image), a message introducing the Query image, and a final instruction to respond with only the class code in {H, IR, OR, B}. The only difference between the two prompts is the verbosity of the initial instruction; episode structure, class set, and output constraints remain identical. For reproducibility and to clarify the exact difference between the two prompting conditions, the verbatim initial instruction prompts are reported in Appendix A.

IV. Statistical Analysis Methodology

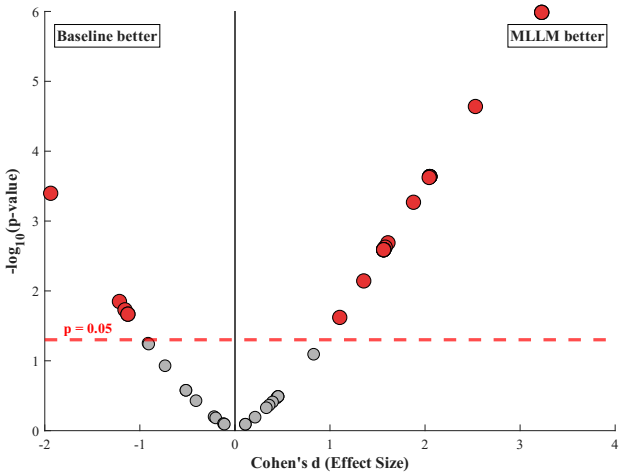
For each configuration (model, K-shot setting, operating scenario), we run $R = 10$ independent few-shot episodes with new random Support and Query sets, and report mean accuracy with 95% confidence intervals (CIs). Precision, recall, and F1-score show trends consistent with accuracy and are omitted for brevity, given the use of a balanced dataset. While $R = 10$ is sufficient to provide an initial estimate of central tendency and variability across randomly sampled episodes, we acknowledge that this sample size limits the statistical power of pairwise significance tests and the stability of effect-size estimates. Therefore, inferential comparisons should be interpreted as exploratory, and future analyses will benefit from a larger number of episodes to increase statistical sensitivity and robustness.

Statistical significance is assessed via independent two-sample t-tests at significance level $\alpha = 0.05$, applied to the accuracy distributions when comparing (i) each MLLM against the Prototypical–Swin baseline and (ii) *concise* versus *detailed* prompts. Effect sizes are quantified by Cohen's d.

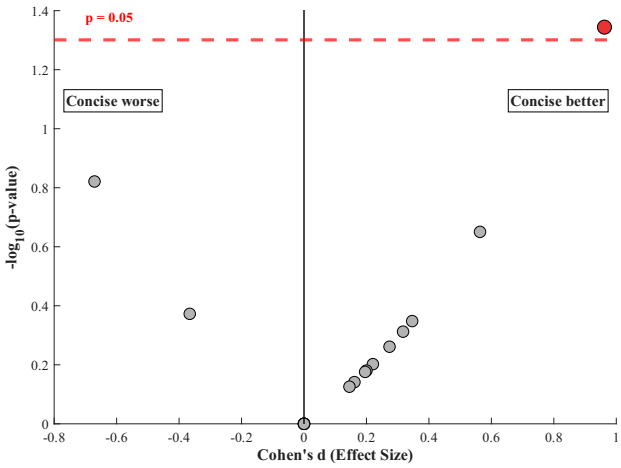
Cohen's d is a standardized effect size measure that quantifies the magnitude of the difference between two groups, expressing the distance between the means in units of standard deviation. For two independent samples, it is defined as $d = (\bar{x}_1 - \bar{x}_2)/s_p$, where $\bar{x}_1$ and $\bar{x}_2$ are the sample means and s_p is the combined standard deviation of the two groups. Unlike the p-value, which provides information on statistical significance, Cohen's d provides an estimate of the practical relevance of the effect. Typically considered reference values are $d \approx 0.2$ (small), $d \approx 0.5$ (medium), and $d \approx 0.8$ (large), although this depends on the application context.

A. Volcano Plots

To jointly visualize statistical significance and effect size over all configurations, we use volcano plots (Fig. 4). Each point corresponds to a specific combination of model, K-shot setting, and operating scenario; the horizontal axis reports



(a) MLLM vs. Prototypical–Swin baseline.



(b) Detailed vs. concise prompts.

Fig. 4: Volcano plots of Cohen's d versus $-\log_{10}(p)$ for all configurations. Points above the $p = 0.05$ threshold correspond to statistically significant differences.

Cohen's d, and the vertical axis reports $-\log_{10}(p)$, where p is the t-test p-value.

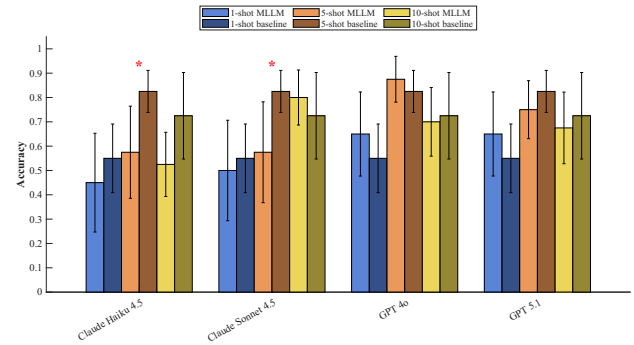
In Fig. 4a, $d > 0$ indicates that the MLLM outperforms the Prototypical–Swin baseline on average, whereas $d < 0$ indicates the opposite. In Fig. 4b, $d > 0$ corresponds to better performance with the detailed prompt and $d < 0$ to better performance with the concise prompt. The horizontal line marks the threshold $p = 0.05$; points above this line denote statistically significant differences.

V. Experimental Results and Discussion

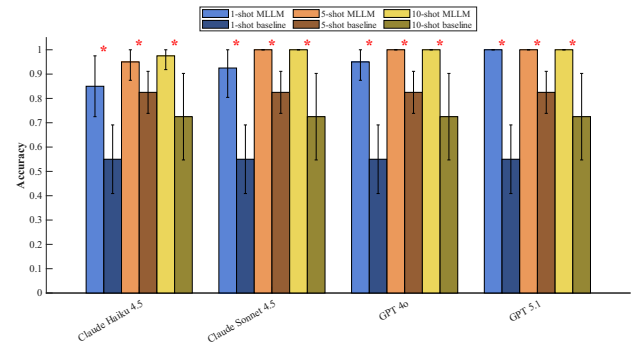
We now discuss the results obtained with the protocol in Section II and the analysis in Section IV, focusing on (i) when MLLMs outperform the Prototypical–Swin baseline and (ii) how the choice of prompt affects few-shot performance.

A. MLLM vs. Prototypical–Swin Comparison

Fig. 5 reports the average accuracy (with 95% CIs) for all models and $K \in \{1, 5, 10\}$ in the *all-speeds* and 607 rpm



(a) All-speeds scenario.



(b) Single-speed scenario (607 rpm).

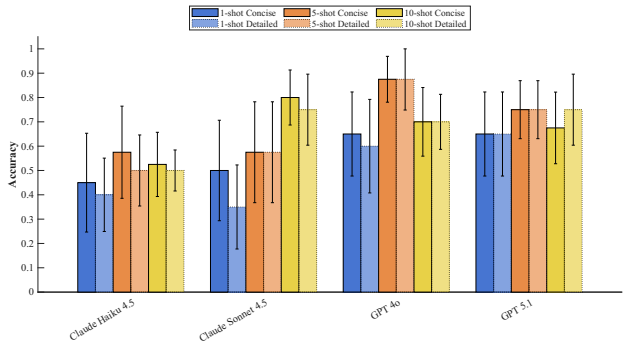
Fig. 5: Comparison between MLLMs and the Prototypical–Swin baseline in the 4-way K-shot configuration. Bars show mean accuracy over R repetitions with 95% CIs; asterisks denote statistically significant differences from the baseline ($p < 0.05$).

scenarios. In the *all-speeds* case (Fig. 5a), where episodes mix all ten spin speeds, the Prototypical–Swin baseline remains competitive. For 1-shot it is among the best models, and confidence intervals overlap with most MLLMs. Asterisks mark configurations where the t-tests detect a significant difference from the baseline. In these cases, the corresponding effect sizes are generally small or moderate, indicating no systematic, large advantage for MLLMs under strong operating variability.

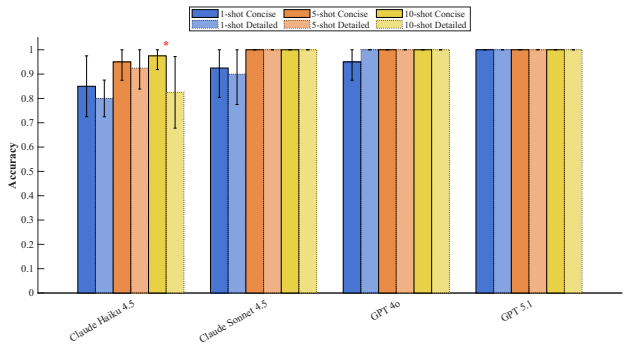
In the single-speed case at 607 rpm (Fig. 5b), intra-class variability due to speed is reduced and the picture changes. The best MLLMs achieve accuracies close to 100% for all K, with narrow CIs, while the Prototypical–Swin baseline reaches lower and more variable accuracy. Here, several configurations show statistically significant improvements over the baseline with medium-to-large effect sizes. Under relatively homogeneous operating conditions, MLLMs thus exploit the CWT information more effectively than the Prototypical–Swin reference in the few-shot regime.

B. Impact of Prompt Design on MLLMs

Fig. 6 compares concise and detailed prompts for the same MLLMs, in both *all-speeds* and 607 rpm scenarios and for $K \in \{1, 5, 10\}$. In many configurations, the bars and CIs overlap and the t-tests do not indicate significant differences,



(a) All-speeds scenario.



(b) Single-speed scenario (607 rpm).

Fig. 6: Comparison between concise and detailed prompts for the considered MLLMs. Bars show mean accuracy with 95% CIs; asterisks indicate statistically significant differences between prompts ($p < 0.05$).

suggesting that prompt choice often has limited impact on accuracy.

When the difference is significant ($p < 0.05$), the trend generally favors the concise prompt. The detailed variant tends to yield similar or slightly lower accuracy, and only rarely provides a benefit. In this setup, increasing verbosity and textual complexity in the initial instructions does not improve few-shot performance and can even slightly reduce robustness.

Overall, Figs. 4 to 6 indicate that:

- In the *all-speeds* scenario, the Prototypical–Swin baseline remains a strong benchmark and MLLMs provide only limited gains. In the five configurations, where differences are statistically significant, MLLMs are on average 26.0% *less* accurate than the baseline.
- In the single-speed (607 rpm) scenario, advanced MLLMs clearly outperform the baseline, achieving near-perfect accuracy with narrow CIs and large effect sizes across all K-shot settings. In the 22 statistically significant comparisons, MLLMs are on average 27.7% more accurate than Prototypical–Swin.
- Regarding prompt design, the *concise* prompt is at least competitive and often preferable to the *detailed* one; no systematic advantage of longer descriptions is observed. Only one configuration reaches statistical significance, where at fixed speed Claude 4.5 Haiku performs about 15% worse with the detailed prompt than with the concise one.

From an application perspective, MLLMs are particularly attractive for few-shot diagnosis when operation can be segmented into relatively homogeneous regimes, whereas in highly heterogeneous conditions a lightweight model such as Prototypical–Swin remains a robust and easy-to-implement option. Simple, standardized prompts are sufficient to exploit MLLM capabilities in this context, and excessive prompt engineering does not appear necessary.

VI. Conclusions and Future Work

This work investigated Multimodal Large Language Models for few-shot fault diagnosis of medium-to-large bearings, using CWT images acquired on a test rig designed for medium-sized rolling bearings. We compared several MLLMs against a Few-Shot Learning baseline based on Prototypical Networks with a Swin Transformer backbone, and assessed the impact of concise versus detailed prompts.

When few-shot episodes aggregate all rotational speeds, strong intra-class heterogeneity makes the Prototypical–Swin baseline competitive. In a single-speed regime (607 rpm), instead, advanced MLLMs reach very high accuracy for all K and significantly outperform the baseline, showing that multimodal pre-training can be effectively transferred to industrial diagnosis without episodic meta-training. Prompt design has a limited influence. Namely, concise prompts are generally sufficient and sometimes preferable, while more verbose templates do not yield systematic benefits.

Future work will focus on: (i) extending the analysis to varying loads, additional fault types, and other rotating machines; (ii) exploring more complex few-shot protocols and aggregation strategies across operating regimes; (iii) studying hybrid schemes that couple specialized feature extractors with MLLMs, including distributed or federated deployments; and (iv) performing ablation studies on the adopted time–frequency representation, by analyzing the sensitivity to CWT parameter choices and benchmarking against alternatives such as STFT.

Appendix A
Verbatim Prompts Used in the MLLM Experiments

As discussed in Section III, the *concise* and *detailed* prompting strategies differ only in the verbosity of the initial task instruction, while the support-example structure, query-image introduction, and final output constraint remain unchanged. The verbatim initial instructions are reported below.

A. Concise Prompt

```
You are a machine diagnosis
assistant. I will provide you
with example continuous wavelet
transform (CWT) images, each
describing a specific condition.
Then, I'll ask you to analyze a
test image and provide ONLY the
condition code.
```

```
Your response MUST be ONLY one
of these codes: H, IR, OR, or B.
Do NOT provide any explanation or
additional text. ONLY output the
single code.
```

B. Detailed Prompt

```
Imagine you are an advanced
AI specialized in interpreting
Continuous Wavelet Transform (CWT)
images for machine diagnosis.
I will provide you with a set
of CWT images, each accompanied
by a description specifying
its condition, in the format
'[condition description] ([code])',
such as 'bearing damaged on the
inner race (IR)'. After reviewing
these examples, I will present
you with a test image. Your task
will be to classify this test
image based on the similarities
you recognize with the provided
examples.
Your response MUST be ONLY the
code letter in this exact format:
'[code]', for example: 'IR' or
'OR' or 'H' or 'B'. Do NOT provide
any additional text, analysis, or
commentary. ONLY output the single
letter code.
```

For completeness, in both prompting conditions each support example is then introduced with a short textual label followed by the corresponding CWT image, the query image is presented in a separate message, and a final reminder is added in the form: "Respond with ONLY the condition code (H, IR, OR, or B). Nothing else. No explanation. Just the code."

References

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020. [Online]. Available: https://doi.org/10.1016/j.ymssp.2019.106587

[2] M. Jalayer, C. Orsenigo, and C. Vercellis, "Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, Fast Fourier and continuous wavelet transforms," *Computers in Industry*, vol. 125, p. 103378, Feb. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0166361520306126

[3] E. Brusa, C. Delprete, and L. G. Di Maggio, "Eigen-spectrograms: An interpretable feature space for bearing fault diagnosis based on artificial intelligence and image processing," *Mechanics of Advanced Materials and Structures*, pp. 1–13, Jul. 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15376494.2022.2102274

[4] A. R. Mohanty, *Machinery condition monitoring: Principles and practices*. CRC Press, 2014.

[5] Z. Wang, Q. Luo, H. Chen, J. Zhao, L. Yao, J. Zhang, and F. Chu, "A high-accuracy intelligent fault diagnosis method for aero-engine bearings with limited samples," *Computers in Industry*, vol. 159-160, p. 104099, Aug. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0166361524000277

[6] M. M. Islam and J.-M. Kim, "Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network," *Computers in Industry*, vol. 106, pp. 142–153, Apr. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0166361518302689

[7] L. Giraudo, L. G. Di Maggio, L. Giorio, and C. Delprete, "Dynamic Multibody Modeling of Spherical Roller Bearings with Localized Defects for Large-Scale Rotating Machinery," *Sensors*, vol. 25, no. 8, p. 2419, Apr. 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/8/2419

[8] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, and J. Hu, "Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning," *IEEE Access*, vol. 7, pp. 110 895–110 904, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8793060/

[9] S. Zhang, F. Ye, B. Wang, and T. Habetler, "Few-Shot Bearing Fault Diagnosis Based on Model-Agnostic Meta-Learning," *IEEE Transactions on Industry Applications*, vol. 57, no. 5, pp. 4754–4764, Sep. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9463790/

[10] L. Chang and Y.-H. Lin, "Meta-Learning With Adaptive Learning Rates for Few-Shot Fault Diagnosis," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 5948–5958, Dec. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9844890/

[11] K. Yue, J. Li, J. Chen, R. Huang, and W. Li, "Multiscale Wavelet Prototypical Network for Cross-Component Few-Shot Intelligent Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9999323/

[12] L. Cheng, Z. An, Y. Guo, M. Ren, Z. Yang, and S. McLoone, "MMFSL: A Novel Multimodal Few-Shot Learning Framework for Fault Diagnosis of Industrial Bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10167449/

[13] J. Chen, W. Hu, D. Cao, Z. Zhang, Z. Chen, and F. Blaabjerg, "A Meta-Learning Method for Electric Machine Bearing Fault Diagnosis Under Varying Working Conditions With Limited Data," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2552–2564, Mar. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9749912/

[14] G. Yu, D. Wang, J. Liu, and X. Zhang, "Distribution-Agnostic Few-Shot Industrial Fault Diagnosis via Adaptation-Aware Optimal Feature Transport," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5623–5632, Apr. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9904307/

[15] M. Luo, J. Xu, Y. Fan, and J. Zhang, "TRNet: A Cross-Component Few-Shot Mechanical Fault Diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6883–6894, May 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9878220/

[16] Y. Liu, H. Jiang, R. Yao, and T. Zeng, "Counterfactual-augmented few-shot contrastive learning for machinery intelligent fault diagnosis with limited samples," *Mechanical Systems and Signal Processing*, vol. 216, p. 111507, Jul. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888327024004059

[17] Y. Xu, X. Ma, X. Wang, J. Wang, G. Tang, and Z. Ji, "Unified feature learning network for few-shot fault diagnosis," *Neurocomputing*, vol. 598, p. 128035, Sep. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231224008063

[18] Y. Huang, M. Zhao, X. Liu, X. Xiang, M. Liu, and J. Xie, "Zero-shot bearing fault diagnosis under varying conditions based on prototypical vision transformers," *Engineering Research Express*, vol. 7, no. 4, p. 045554, Dec. 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/2631-8695/ae154c

[19] N.-L. Than, V. Q. Nguyen, G.-B. Truong, V.-T. Pham, and T.-T. Tran, "Mixmamba-fewshot: mamba and attention mixer-based method with few-shot learning for bearing fault diagnosis," *Applied Intelligence*, vol. 55, no. 7, p. 484, May 2025. [Online]. Available: https://link.springer.com/10.1007/s10489-025-06361-0

[20] Z. Chu, S. Xing, B. Han, and J. Wang, "A Novel Cross-Domain Mechanical Fault Diagnosis Method Fusing Acoustic and Vibration Signals by Vision Transformer," *Sensors*, vol. 24, no. 16, p. 5120, Aug. 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/16/5120

[21] Y. Qi, L. Zhao, A. Gu, B. Wu, and B. Suo, "TKAN: a novel method for bearing fault diagnosis in noisy environments*," *Measurement Science*

and Technology, vol. 36, no. 10, p. 106122, Oct. 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6501/ae0e94

[22] Q. Wang, C. Yin, K. She, Q. Tong, G. Lu, H. Zhang, and J. Lu, "Bearing fault diagnosis for variable operating conditions based on KAN convolution and dual branch fusion attention," *Scientific Reports*, vol. 15, no. 1, p. 21442, Jul. 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-04620-1

[23] J. Hendriks, P. Dumond, and D. Knox, "Towards better benchmarking using the CWRU bearing fault dataset," *Mechanical Systems and Signal Processing*, vol. 169, p. 108732, Apr. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888327021010499

[24] E. Brusa, C. Delprete, L. Giorio, L. G. Di Maggio, and V. Zanella, "Design of an Innovative Test Rig for Industrial Bearing Monitoring with Self-Balancing Layout," *Machines*, vol. 10, no. 54, 2022.

[25] J. Wang, E. Brusa, Y. Peng, X. Zhao, M. Wang, C. Delprete, and C. Gastaldi, "Dynamics and stability analysis of rolling mill system during variable gauge rolling," *Meccanica*, vol. 58, no. 12, pp. 2439–2454, Dec. 2023. [Online]. Available: https://link.springer.com/10.1007/s11012-023-01735-5

[26] B. Yang, Y. Lei, N. Li, X. Li, X. Si, and C. Chen, "Balance recovery and collaborative adaptation approach for federated fault diagnosis of inconsistent machine groups," *Knowledge-Based Systems*, vol. 317, p. 113480, May 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S095070512500526X

[27] B. Yang, Y. Lei, X. Li, N. Li, X. Si, and C. Chen, "A dynamic barycenter bridging network for federated transfer fault diagnosis in machine groups," *Mechanical Systems and Signal Processing*, vol. 230, p. 112605, May 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888327025003061

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018, version Number: 2. [Online]. Available: https://arxiv.org/abs/1810.04805

[29] T. B. Brown et al., "Language Models are Few-Shot Learners," Jul. 2020, arXiv:2005.14165 [cs]. [Online]. Available: http://arxiv.org/abs/2005.14165

[30] J. Wu, W. Gan, Z. Chen, S. Wan, and P. S. Yu, "Multimodal Large Language Models: A Survey," Nov. 2023, arXiv:2311.13165 [cs]. [Online]. Available: http://arxiv.org/abs/2311.13165

[31] L. G. Di Maggio, "Toward Autonomous LLM-Based AI Agents for Predictive Maintenance: State of the Art, Challenges, and Future Perspectives," *Applied Sciences*, vol. 15, no. 21, p. 11515, Oct. 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/21/11515

[32] D. He, M. He, and J. Yoon, "Full Ceramic Bearing Fault Diagnosis with Few-Shot Learning Using GPT-2," *Computer Modeling in Engineering & Sciences*, vol. 143, no. 2, pp. 1955–1969, 2025. [Online]. Available: https://www.techscience.com/CMES/v143n2/61446

[33] L. Lin, S. Zhang, S. Fu, and Y. Liu, "FD-LLM: Large language model for fault diagnosis of complex equipment," *Advanced Engineering Informatics*, vol. 65, p. 103208, May 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1474034625001016

[34] Z. Pang, H. Zhang, and T. Li, "Hybrid Fine-Tuning in Large Language Model Learning for Machinery Fault Diagnosis," in *2024 IEEE 22nd International Conference on Industrial Informatics (INDIN)*. Beijing, China: IEEE, Aug. 2024, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10774330/

[35] K. M. Alsaif, A. A. Albeshri, M. A. Khemakhem, and F. E. Eassa, "Multimodal Large Language Model-Based Fault Detection and Diagnosis in Context of Industry 4.0," *Electronics*, vol. 13, no. 24, p. 4912, Dec. 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/24/4912

[36] L. G. Di Maggio, "Predictive Maintenance MCP: An Open-Source Framework for Bridging Large Language Models and Industrial Condition Monitoring via the Model Context Protocol," *Applied Sciences*, vol. 16, no. 6, p. 2812, Mar. 2026. [Online]. Available: https://www.mdpi.com/2076-3417/16/6/2812

[37] L. G. Di Maggio, L. Giorio, C. Delprete, and E. Brusa, "Dataset of Vibration, Temperature and Speed Measurements for Multiple Types of Localized Defects on Spherical Roller Bearings across Multiple Operating Conditions," Oct. 2024. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.13913254

[38] L. G. Di Maggio, E. Brusa, and C. Delprete, "Zero-Shot Generative AI for Rotating Machinery Fault Diagnosis: Synthesizing Highly Realistic Training Data via Cycle-Consistent Adversarial Networks," *Applied Sciences*, vol. 13, no. 22, p. 12458, Nov. 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/22/12458

[39] ——, "Novelty Detection in Rotating Machinery: Assessment of Unsupervised Machine Learning Models for Medium-Sized Industrial Bearings," in *2025 International Conference on Control, Automation and Diagnosis (ICCAD)*. Barcelona, Spain: IEEE, Jul. 2025, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/11099203/

[40] L. G. Di Maggio, "Intelligent Fault Diagnosis of Industrial Bearings Using Transfer Learning and CNNs Pre-Trained for Audio Classification," *Sensors*, vol. 23, no. 1, p. 211, Dec. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/23/1/211

[41] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical Networks for Few-shot Learning," Jun. 2017, arXiv:1703.05175 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1703.05175