

# TSFM in-context learning for time-series classification of bearing-health status

Michel Tokic<sup>1,2</sup>, Slobodan Djukanović<sup>1,3</sup>, Anja von Beuningen<sup>1</sup>, Cheng Feng<sup>1</sup>

1- Siemens AG, Data & Artificial Intelligence, Otto-Hahn-Ring 6,  
81739 Munich, Germany

2- Faculty of Mathematics, Informatics and Statistics,  
Ludwig-Maximilians-University Munich, Munich, Germany

3- Faculty of Electrical Engineering, University of Montenegro,  
Podgorica, Montenegro

**Abstract.** This paper introduces a classification method using in-context learning in time-series foundation models (TSFM). We show how data, which was not part of the TSFM training data corpus, can be classified without the need of finetuning the model. Examples are represented in the form of targets (class id) and covariates (data matrix) within the prompt of the model, which enables to classify an unknown covariate data pattern alongside the forecast axis through in-context learning. We apply this method to vibration data for assessing the health state of a bearing within a servo-press motor. The method transforms frequency domain reference signals into pseudo time-series patterns, generates aligned covariate and target signals, and uses the TSFM to predict probabilities how classified data corresponds to predefined labels. Leveraging the scalability of pre-trained models this method demonstrates efficacy across varied operational conditions. This marks significant progress beyond custom narrow AI solutions towards broader, AI-driven maintenance systems.

## 1 Introduction

In recent years, industries such as manufacturing, energy, transportation, and healthcare have increasingly recognized the importance of anomaly detection to prevent the failure of critical machines and systems. The traditional approach to maintenance involves reactive strategies, which can lead to costly downtime, production losses, and safety risks. As a result, many companies are shifting towards predictive and preventive maintenance strategies, aiming to optimize operational efficiency by preemptively addressing signs of equipment deterioration.

Despite advancements in predictive maintenance, current solutions are largely tailored for specific applications and assets, relying heavily on domain-specific knowledge and handcrafted rules [1]. While effective in controlled contexts, these systems often lack flexibility and require substantial effort to adapt to new environments or equipment.

The rise of artificial intelligence (AI) has introduced AI-driven maintenance strategies that rely on data-driven approaches. However, existing implementations primarily depend on narrow AI solutions, trained on asset-specific datasets,

and typically struggle to scale across different machines, industries, or failure patterns [1]. This limitation is due to their inability to generalize beyond predefined conditions, forcing companies to rely on custom-built solutions that hinder the full potential of AI-driven maintenance.

This work addresses these challenges by introducing a novel approach for classification of data, specifically vibration data, using a time-series foundation model (TSFM). Unlike traditional narrow AI models, TSFMs offer the generalization capabilities necessary for scalable and flexible classification solutions across different assets and operational conditions. Notably, they enable reliable classification of data without necessitating extensive finetuning of large reference datasets, representing a significant move towards more adaptive and generalizable AI models [2].

## 2 Method

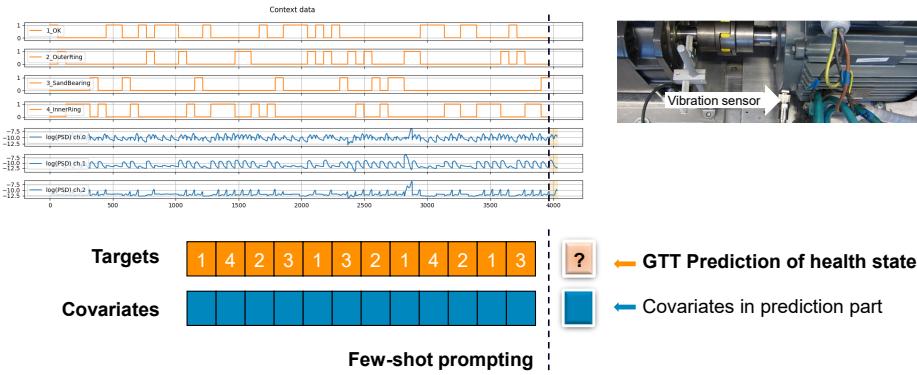
This chapter details the proposed methodology for classifying the health state of a motor bearing using spectral data.

### 2.1 Time Series Foundation Models

TSFMs have emerged as a transformative paradigm for time-series forecasting, delivering strong zero-shot and few-shot performance across a broad spectrum of datasets and application domains [3–5]. Recent studies demonstrate that TSFMs can match or even surpass specialized statistical models and earlier task-specific models on unseen benchmarks. Our method is centered on the *General Time Transformer* (GTT) architecture [6] that captures complex temporal patterns and multivariate dependencies through alternating temporal and channel attentions in Transformer encoder blocks. By adding a learnable sink token to the end of channels for target variates similar to [3], and replacing the original point forecast head with a Gaussian mixture probabilistic forecast head, we adapt the GTT architecture for multivariate probabilistic forecast with covariates that are available in the forecast horizon. Pretrained on a large-scale cross-domain dataset, GTT is a general-purpose zero-shot forecaster that takes a look-back window of  $L$  time points for one or more target variables (optionally with covariates) and predicts their next  $H$  values. Let  $\mathbf{x}_{1:L}$  and  $\mathbf{c}_{1:L+H}$  be the target variates and corresponding covariates. Targets are available only in the context window, whereas the covariates can be available also in the forecast horizon. The model is a function that predicts the distribution of  $\mathbf{x}_{L+1:L+H}$ , such that  $P(\mathbf{x}_{L+1:L+H}) = f(\mathbf{x}_{1:L}, \mathbf{c}_{1:L+H})$ .

### 2.2 Data Preprocessing

Raw vibration signals sampled at 48 kHz from a servo-press motor are processed to compute the fast-fourier transform (FFT) [7], where each recording has a length of  $\approx 60$  seconds. As illustrated in Fig. 2, the preprocessing divides the whole spectrum into  $N = 20$  *data channels* (covariates). For each channel, the



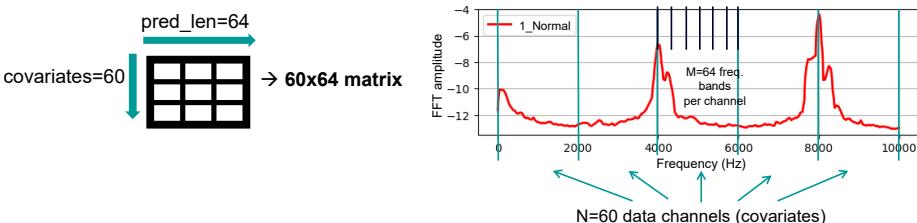
**Fig. 1:** A sequence of random samples from the servo-press dataset with one-hot encoded targets (orange) and corresponding covariates (blue) forms the few-shot prompting context. The model outputs a prediction of the targets w.r.t. covariates in the forecast horizon (blue block to the right of the dashed line).

FFT is segmented into  $M = 64$  frequency sub-bands, with the mean value from each sub-band forming a feature. Overall, one raw vibration signal results in a covariate structure of size  $N \times M$ , used as input to the TSFM.

The system defines target variables for classification, corresponding to four health states (classes): 1. *Normal operation*, 2. *Outer ring fault*, 3. *Sand in bearing\**, and 4. *Inner ring fault*. Targets are one-hot encoded (1 if the class is present, 0 otherwise) over a prediction horizon of  $M = 64$  time steps, enabling sensitivity to subtle variations and emphasizing fault-indicative frequency components (see Figs. 2 and 3).

### 2.3 Few-shot prompting for in-context learning classification

We apply few-shot prompting to adapt the pre-trained GTT for the asset health classification tasks, as shown in Fig. 1.



**Fig. 2:** Preprocessing: Signal spectrum (magnitude of FFT) is transformed into a  $60 \times 64$  matrix ( $N = 60$  data channels and  $M = 64$  frequency sub-bands per channel).

A small sequence of recorded target health states and their corresponding covariates forms the "context" or "prompt". This allows the model to learn

\*Sand has been artificially added in the bearing.

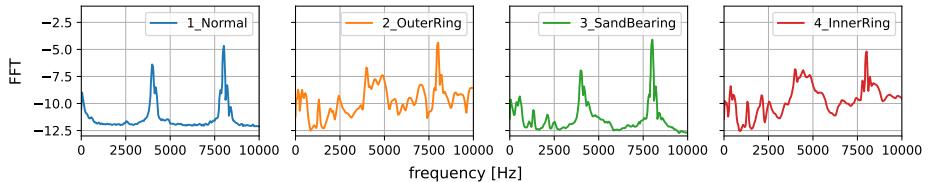
about patterns associated with specific health states from a limited number of examples, which is known as *in-context learning* or *few-shot prompting*.

For a new, unknown covariate input, the model predicts the corresponding class as depicted in Fig. 4. Within 64 time steps, the model predicts the class to which the unknown covariate data belongs. As the classification method, we apply a winner-takes-all rule; alternatively, a discrete probability distribution could be obtained by applying a Softmax function to the predicted intensities.

### 3 Experiments

We conducted experiments with a GTT model with 750 million parameters pre-trained on a large-scale cross domain time series corpus with 124 billion data points. The model can accept at most 64 channels/variates in the input context.

We use 3968 time steps on the input side, and thus are able to draw 62 random samples from our servo-press data, where each sample has a length of  $M = 64$  steps and  $N = 60$  covariate channels for representing the spectrum. The dataset comprises 280 samples almost equally distributed across all four classes depicted in Fig. 3. Including four target variables, we have a total of 64 variables (targets + covariates) per class sample, which corresponds to the current limit of GTT.



**Fig. 3:** Example FFTs from the servo-press dataset. Note similarity between *1\_Normal* and *3\_SandBearing* and between *2\_OuterRing* and *4\_InnerRing*.

A typical classification is depicted in Fig. 4, where the task was to predict class *2\_OuterRing*. In the corresponding prompt, the last example was drawn from class *4\_InnerRing*, explaining why the prediction for this class starts with high intensity (around 0.7) at the beginning, but develops down to almost 0 until step  $t = 64$ . In contrast, the intensity of class *2\_OuterRing* rises towards almost 1. At  $t = 64$ , the applied winner-takes-it all rule classifies the data correctly.

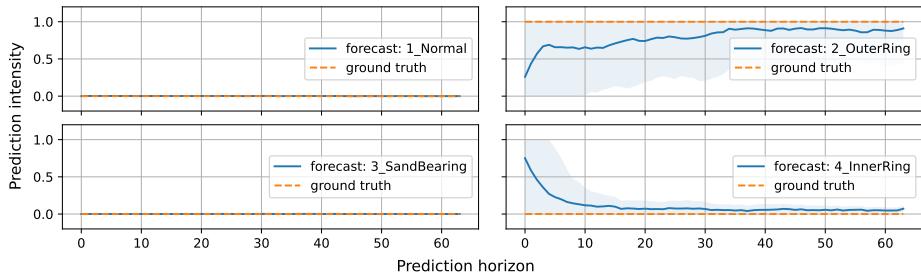
An extensive study has been carried out with 1000 random contexts, each containing 63 uniformly distributed samples from all four classes, where the last sample represents the data to be classified. Due to the random sampling process, we investigate varied starting conditions (one-hot encoded targets within the 62<sup>nd</sup> sample) for the forecast of the 63<sup>rd</sup> sample shown in Fig. 4.

The analysis achieved excellent classification performance with an overall accuracy of 96% across all samples. All four classes were identified with high precision and recall, exhibiting only minor misclassifications. These results demonstrate that the TSFM-based approach is highly reliable and consistent in differ-

entiating between normal and faulty bearing conditions, particularly considering that this dataset was excluded from the model's training data.

True	Predicted				precision	recall	F1	accuracy
	1	2	3	4				
1	245	1	11	8	0.97	0.92	0.95	
2	2	256	1	4	0.98	0.97	0.98	
3	5	2	230	5	0.95	0.95	0.95	<b>0.96</b>
4	0	1	0	229	0.93	1.00	0.96	

**Table 1:** Confusion matrix and classification metrics for TSFM-based servo-press data analysis using FFT representation on 1000 randomly generated contexts. Each context includes 63 uniformly distributed samples from all four classes, with the last sample used for classification. Class labels: 1 = 1\_Normal, 2 = 2\_OuterRing, 3 = 3\_SandBearing, 4 = 4\_InnerRing.



**Fig. 4:** Correct classification of class 2\_OuterRing (upper right), based on corresponding covariate matrix available in the forecast part of the GTT. Note, the prediction intensity stays/develops towards 0 for all other classes.

## 4 Results and Discussion

The presented methodology offers significant advantages over conventional approaches by combining pre-trained TSFMs with few-shot prompting. It enables rapid deployment, eliminates training time, minimizes expert dependency, and ensures scalability across diverse assets and failure modes.

The results demonstrate that aligning covariates and target data as pseudo time-series patterns effectively utilized the TSFM's generalization capacity. The model achieved 96% accuracy across all samples, with high precision and recall ( $F1 \approx 0.95 - 0.98$ ), consistently distinguishing between normal and faulty conditions, although the servo-press dataset was not part of the models training data. The FFT-based approach was crucial in highlighting fault-indicative frequencies, and the presentation of random examples in the prompt context contributed to robust classification, minimizing confusion between classes.

While this method is promising, limitations must be acknowledged. The method was applied to only four target classes. Since the context length in the

examined GTT model is limited in the number of time steps, a higher number of classes would result in less represented examples per class. This can reduce the prediction accuracy of the TSFM and increases the probability of misclassifications. Real-world deployment may also require additional consideration of computational and integration factors, especially if the method is implemented on an edge device with limited memory and computation ability.

This work highlights the broader potential of TSFMs in predictive maintenance across industries. The method's flexibility makes it adaptable to diverse assets and fault types, with future enhancements like domain-specific pre-training or improved interpretability expected to extend its impact further.

## References

- [1] S. Qiu, X. Cui, Z. Ping, N. Shan, Z. Li, X. Bao, and X. Xu, “Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review,” *Sensors*, vol. 23, no. 3, 1305 2023.
- [2] Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, and Q. Wen, “Foundation models for time series analysis: A tutorial and survey,” in *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 2024, pp. 6555–6565.
- [3] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, “Unified training of universal time series forecasting transformers,” in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 53140–53164.
- [4] A. Das, W. Kong, R. Sen, and Y. Zhou, “A decoder-only foundation model for time-series forecasting,” in *41st International Conference on Machine Learning*, 2024.
- [5] A. F. Ansari, O. Shchur, J. Küken, A. Auer, B. Han, P. Mercado, S. S. Rangapuram, H. Shen, L. Stella, X. Zhang, et al., “Chronos-2: From univariate to universal forecasting,” *arXiv preprint arXiv:2510.15821*, 2025.
- [6] C. Feng, L. Huang, and D. Krompass, “General time transformer: An encoder-only foundation model for zero-shot multivariate time series forecasting,” in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 3757–3761.
- [7] M. K. Steven, *Fundamentals of statistical signal processing*. PTR Prentice-Hall, Englewood Cliffs, NJ, 1993.