



HAL
open science

Bearings RUL prediction based on contrastive self-supervised learning

Weikun Deng, Khanh T P Nguyen, Kamal Medjaher, Christian Gogu, Jérôme Morio

► **To cite this version:**

Weikun Deng, Khanh T P Nguyen, Kamal Medjaher, Christian Gogu, Jérôme Morio. Bearings RUL prediction based on contrastive self-supervised learning. IFAC World Congress 2023, Jul 2023, Yokohama, Japan. pp.11906-11911, 10.1016/j.ifacol.2023.10.604 . hal-04632423

HAL Id: hal-04632423

<https://hal.science/hal-04632423>

Submitted on 2 Jul 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Bearings RUL prediction based on contrastive self-supervised learning

WeiKun Deng * Khanh T. P. Nguyen * Kamal Medjaher *
Christian Gogu ** Jérôme Morio ***

* 47 Av. d'Azereix, 65000 Tarbes, (e-mail: weikun.deng@enit.fr).

** Institut Clément Ader, Université de Toulouse, 3 rue Caroline Aigle, 31400
Toulouse, France, (e-mail: christian.gogu@gmail.com)

*** ONERA\DTIS, Université de Toulouse, F-31055 Toulouse, France, (e-mail:
Jerome.Morio@onera.fr)

Abstract: This paper proposes a new contrastive self-supervised learning paradigm for bearing remaining useful life (RUL) prediction based on CNN-LSTM models. It addresses the dilemma of scarce labels and data imbalance in Prognostics and Health Management (PHM) by designing a specific pretext task to mine the potential degradation-related information in unlabelled data. In this paper, we propose a method to build contrastive sample pairs using sequence order information. Then, a Siamese CNN encoder guided by the customized contrastive loss is designed to maximize the differences between encoding features of the contrastive sample pairs. After that, the CNN's parameters are partly frozen, and its encoded features are used as the input of the subsequent LSTM layer to predict the RUL. Finally, on the labeled dataset, LSTM is fine-tuned to optimize the ability of CNN-LSTM for RUL prediction. The proposed method is validated on "PRONOSTIA Bearing Dataset". The obtained results and the analysis of the hidden layer output highlight the performance of the proposed approach, which outperforms the supervised learning paradigm in terms of maintaining the ability to capture sequential discriminatory information for better RUL prediction, especially in the case of a reduced amount of labeled data.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Degenerative trend discovery, Machine learning, Self-supervised learning, Contrastive sample pairs, Remaining useful life.

1. INTRODUCTION

Prediction of bearing remaining useful life (RUL) is crucial for predictive maintenance of rotating machines. Based on automatic feature processing and knowledge mining in monitoring data, machine learning (ML) methods have provided promising prognostic RUL prediction models such as long- and short-term memory neural networks (see Shi and Chehade (2021)), temporal convolutional network (TCN) (see Cao et al. (2021)), and their variants, with a wide range of applications (see Nguyen et al. (2022)). These current efforts, based on fully supervised models, rely heavily on labeled data, i.e., RUL information, which is sparsely available in many safety-critical systems. For instance, in real-world operations, the bearing degradation process is typically truncated through maintenance and refurbishment, resulting in its operation at a significant distance from the failure point. Therefore, monitoring data consists of a large quantity of unlabelled degradation data and infrequent labeled failure data, leading to imbalanced health-failure information distribution and sparse failure time labels in the dataset used for RUL prediction. This creates a dilemma for existing label-dependent and balanced information-hungry supervised machine learning algorithms. The main approaches to solving this dilemma in PHM include both mining degraded information from large amounts of unlabelled degradation data and making more effective use of small amounts of labeled failure data. Pre-training a feature extractor in a pretext task

coupled with downstream parameter fine-tuning is the main manner in the application of this idea. When the pretext task is unsupervised, the methods are semi-supervised learning. When the pretext task is supervised, the methods are self-supervised learning (SSL) (see Chen et al. (2020)). Compared to unsupervised learning, which builds objective functions from the perspective of the data distribution or patterns, self-supervised learning (SSL) methods aim to construct supervised information or pseudo-labels from the data itself. SSL is more appropriate for solving the above problem because the degradation data passed to the algorithm is rich in internal structure, and the targets and rewards used for purely unsupervised training are sparse. The SSL method in RUL prediction is not only useful for understanding the internal structure patterns but also for applying them to determine time variation trends. Therefore, it is delightful to see and review the SSL ideas in RUL prediction here. Although only a few SSL methods have been developed in current advances, one development is based on pre-training a mask autoencoder (MAE) by reconstructing the RUL data (see Guo et al. (2022)). The randomly masked signal is recovered by pre-training coding. Without MAE parameter freezing, the downstream decoded structures are then connected with a regression head for RUL prediction in the training on the labeled dataset. This method is viewed as a generative framework and is commonly applied in diagnostics and anomaly detection (see Hahn and Mechefske (2021)). This reconstruction task is easy to implement but cannot be fully consistent with the RUL predictions. It only guarantees a relatively better initialization of the model. It facilitates the

* This work is supported by the CSTE Funding, and achieved within the "Laboratoire Génie de Production de l'École Nationale d'Ingénieurs de Tarbes".

adaptive extraction of representative features from the original data, but these features do not directly and explicitly represent the degradation process.

Another approach for SSL is based on contrastive learning, which trains a CNN extractor on the regression prediction of next-moment vibration timing slices and contrastive time-series pairs' divergence. This information is embedded as a feature set into the hard negative contrast CNN-GRU model (see Xu et al. (2022)). Another similar approach focuses on the differences in sequential prediction results driven by metric learning, using Siamese network structures and a pre-trained feature extractor designed to learn a potential feature space structure (see Krokotsch et al. (2021)). Besides, auto-regressive pretext tasks have also been proposed, where the pretext objective is to predict one or multi-step ahead raw sensor data, followed by fine-tuning a smaller network on limited labeled RUL data (see Akrim et al. (2022)). The frozen part is equivalent to fixed knowledge for downstream tasks.

To ensure the accuracy of RUL prediction based on the above studies, the core challenges of SSL lie in:

- whether representative features are extracted.
- whether sequential information is captured and considered and the learning of the feature extractor reflects the basic logic of the degradation process.

This paper aims to provide an explicit method and theory to guide the design of pretext tasks for feature extractors to consider information related to the degradation sequential trend. The principles and research inspirations on this issue are detailed in Section 2. The proposed method is described in Section 3, and the performance of the method is analyzed and validated in Section 5. Finally, the conclusion and perspectives are presented in Section 6.

2. BACKGROUND AND ASSUMPTIONS

In this section, the intuitive motivation behind the proposed methods will be laid out, and the pre-text task will be described in detail, highlighting how it has been meticulously designed to achieve our goals.

2.1 Inspired by the medical process

This research is inspired by the medical process. Determining the patient's status by reading the medical imaging collected from different stages of the disease is an important aspect of a doctor's advanced training. During this process, the patient's medical records must be compared and analyzed over time. While the shape and size of the cell may be similar at certain points in the progression and recovery process, the doctor can provide valuable insights by capturing positive or negative sequential order information, which is summarized from the disease status comparison. The stages of treatment for malignant cell proliferation are depicted in Fig. 1, providing a visual representation of the process. It should be noted that the doctor's knowledge is utilized throughout this process. By analogy, an SSL-based prognostic feature extractor is expected to be effective in processing equipment health monitoring information from various states of degradation using the same hidden knowledge and in maximizing the discrimination related to degradation across these states.

2.2 Pretext task design

The above paragraph discusses the implementation of SSL-based RUL prediction. A feature extractor is designed to sim-

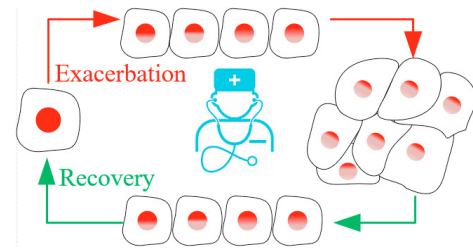


Fig. 1. Physiologically degraded monitoring processes.

ulate "fixed knowledge," and a Siamese CNN is used to capture degradation-related discrimination that is hidden in different sequential directions of the series. The CNN produces maximum differentiated outputs when processing various contrastive input pairs.

The paper proposes an assumption that the time series of the bearings correspond to a monotonically decreasing RUL in the absence of maintenance. Generally, the RUL decreases as the operation duration increases. Features for prognostic can be built by capturing the trends of positive and negative order in each monitoring data slice. A positive order sequential indicates a reduction trend in RUL, while a negative order indicates the opposite. This information is represented by the difference in encoding features of the CNN extractor in pre-training. The direction with the largest encoding difference directly indicates the potential degradation trend. This hypothesis is verified in Section 5.2.

To summarize, different sequential order inputs in the feature extractor correspond to different trend information. Therefore, the supervised learning objective of the pretext task is to maximize the gap between the extractor outputs when processing different sequential direction data. This objective can be achieved by mining the contrastive loss.

3. DESCRIPTION OF THE PROPOSED METHOD

This paper proposes and validates a self-supervised learning paradigm based on a CNN-LSTM framework, as shown in Fig. 2. The framework includes a CNN-LSTM encoder-decoder architecture, contrastive sample pairs, and metric-based loss function learning. The input to the CNN is the spectrum of the bearing monitoring data, which is then interpreted by the LSTM.

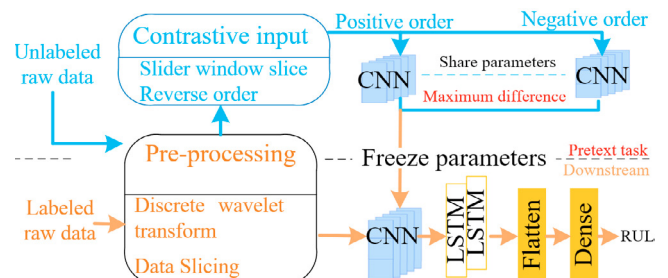


Fig. 2. Contrastive SSL-based CNN-LSTM.

3.1 Siamese CNN in pre-training

The CNN depicted in Fig. 2 follows a Siamese structure that is accomplished by reusing a single CNN, as illustrated in Fig. 3. The objective of the Siamese CNN is to minimize the loss function L , which is defined in Equation (1) and comprises two components: $loss_1$ and $loss_2$.

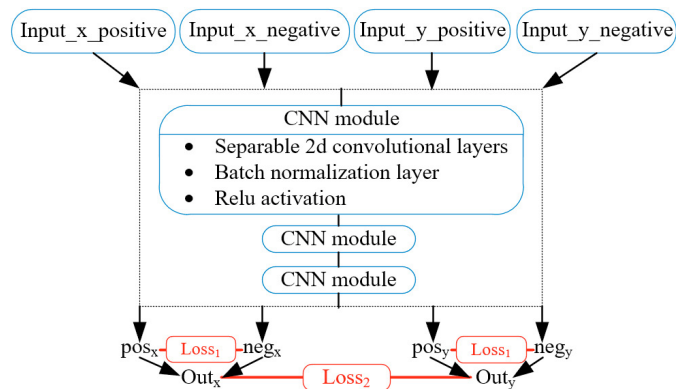


Fig. 3. Siamese CNN implementation diagram.

$$L = \underbrace{\sum_{i=x,y} \frac{1}{\sqrt{(pos_i - neg_i)^2}}}_{loss_1} + \underbrace{out_x - out_y}_{loss_2}. \quad (1)$$

On this basis, it is important to consider that monitoring data are collected from multiple channels, each from a different direction (e.g., horizontal direction x and vertical direction y), at the same time. It is crucial that the same metric results are not interfered by these channels. Therefore, $loss_2$ is designed to ensure that there is agreement among the processing results of different channels. The purpose of $loss_1$ is to maximize the ability of the CNN to capture sequence information in different directions. Specifically, anc_i represents the feature encoding result of positive sequential order, while neg_i represents the encoding result of negative sequential order.

3.2 Contrastive sample pairs

The contrastive pairs' positive and negative samples are generated based on the wavelet transformation results of the vibration data. To ensure consistency, we slice the data using the method presented in Deng et al. (2022). In this paper, we aim to predict the RUL for the next moment using a segment of several time slices. If a segment does not have RUL labels, it is used as a positive sample in sequential order for the feature extractor in the pretext task. Negative samples are obtained by inverting the time sequence of positive sequential order samples, as shown in Fig. 4.

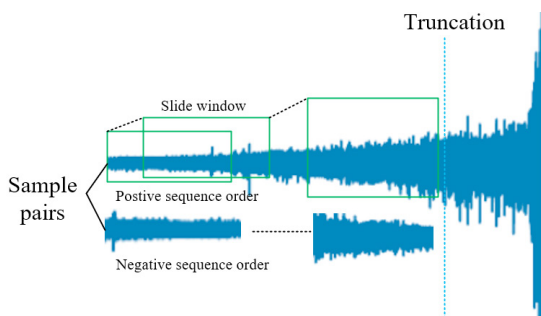


Fig. 4. Construction of contrastive sample pairs.

3.3 SSL-based RUL prediction

After slicing the data into sliding windows and constructing sample pairs, the remaining steps for predicting the remaining useful life (RUL) can be completed by the following stages:

- (1) Firstly, in the pretext stage, a vector of weights is used to convolve on both positive and negative order input sequences. This can be achieved in Keras by concatenating models that handle different inputs.
- (2) Next, the loss function from Equation 1 is minimized. This involves maximizing neg_i , which is the difference in encoding features between the positive and negative sequential order series.
- (3) After that, the hidden layer of the feature extractor is frozen, and a CNN-LSTM model is trained on the dataset with RUL labels. The objective is to minimize the prediction mean squared error. The best-performing model is saved and used for testing.

4. CASE STUDY

This section validates the model performance on “PRONOSTIA Bearing Dataset”. It is generated from the running to-failure experiment on the platform shown in the research (see Nectoux et al. (2012)).

4.1 Data preprocessing

An example of the original vibration data is shown in Fig. 5. It can be seen that the fluctuation of the vibration is not obvious in the early stage of the fault, and its amplitude only changes significantly near the failure point. In this paper, the “db8”

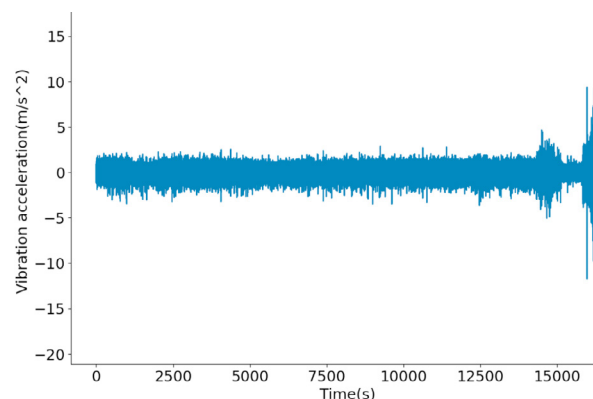


Fig. 5. Raw vibration signals of bearing 2_1.

wavelet transformation is applied to the original data. The obtained approximate and detail coefficient truncated spectrum is shown in Fig. 6. The truncation length is half of the whole spectrum, and it is used as the input for the CNN with a computation size of (2, 1280). Following the method described in Section 3.2, this paper builds contrastive sample pairs of “PRONOSTIA Bearing Dataset” in the following steps:

- (1) In the first two-thirds of each degradation trajectory, random cuts are made to build the unlabeled degradation part. The randomly cut length is half of the whole degradation trajectory.
- (2) Every twelve historical samples are sliced as a segment that has a positive sequential order in every randomly cut part to make the best use of the data. The step size of the sliding window is 1.
- (3) The order of the segment is inverted to form a negative sequential order segment. Both segments are spliced to obtain the contrastive pairs.

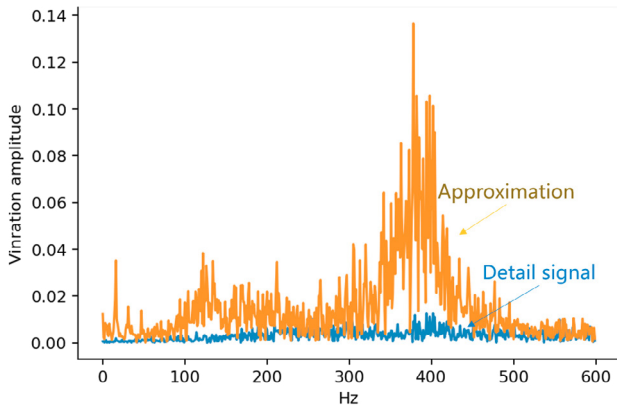


Fig. 6. The result of the discrete wavelet transformation.

The number of samples obtained for pretext training is 25,152, and the number of samples with labeled data for downstream training is 7,360. The data are labeled with the degree of degradation expressed as a percentage, which is derived from normalizing the true value of the RUL in seconds for each degradation trajectory.

4.2 Different models' prediction performance

The training is performed using a batch size of 64, and the early stop mechanism is activated after 50 epochs of patience. The

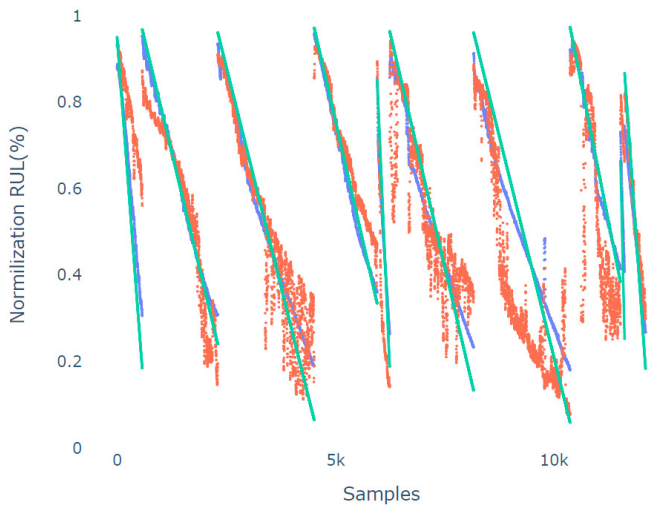


Fig. 7. Prediction results on bearing 1_3, 1_5, 1_6, 1_7, 2_3 2_7, 3_3 degraded trajectories are shown from left to right, with green representing the real trajectory and red representing the benchmark model performance. SSL performance are shown in blue line.

prediction results of the benchmark model are displayed in the top section of Fig. 7. This figure presents a qualitative comparison of the SSL and benchmark methods. The horizontal axis shows the number of segments after the sliding window, which varies based on the length of each bearing's degradation trajectory. The vertical axis represents the normalized RUL in percentage. The results indicate that SSL performs better and is closer to the actual RUL values. In contrast, the benchmark model has a high rate of prediction errors in RUL. On the whole test set, compared to the model in state of the art presented in Xu et al. (2022), the SSL model in this paper has a better prediction performance whereas benchmark (CNN+LSTM) has a similar performance to SSL as shown in Table. 1. "Quantitative

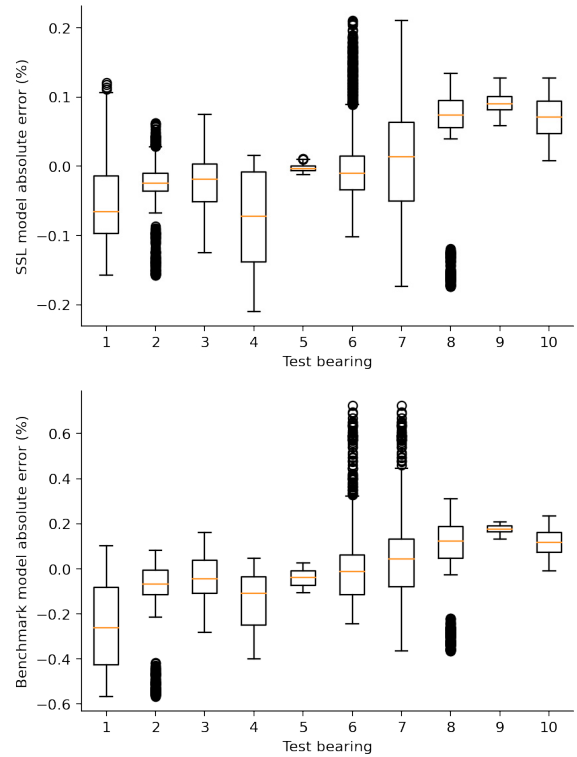


Fig. 8. Prediction errors of different models.

comparisons are made between the benchmark model and the proposed SSL models, and the results are shown in Fig. 8 when 100% of the label data in the training dataset is available. The RMSE error bounds of the proposed models are superior to those of the benchmark model.

5. RESULT DISCUSSION

In this section, firstly, the performance of the SOTA models and SSL models are compared, and secondly, the feature extractor's ability of CNN-LSTM (benchmark) and SSL models on capturing the sequential information is analyzed to verify the conjectures on the relationship between sequential order and degradation trends that are proposed in Section 2.

5.1 Model performance investigation

In this section, the performance of the proposed model is verified at different percentages of labeled degradation data in "PRONOSTIA Bearing Dataset". Then, these results are compared to the ones obtained by different state-of-the-art methods presented in the paper Krokotsch et al. (2021). Additionally, to test the model's generality for incremental data and work conditions, 6 sets of degraded trajectories with labels are used. The training sets are selected in increasing order from 1/6 to 1 ratio, as shown on the horizontal axis of Fig. 9. The vertical coordinates of the figure indicate the RMSE for the entire test set. The results show that the SSL framework is more effective for low labeling percentages, and it can adapt to degraded situations with unknown operating conditions feasibly.

The proposed SSL model shows strong prediction performance with less dependence on labeled data. However, over-fitting can occur when using an inadequately trained model on an unknown dataset, resulting in ineffective feature extraction. In contrast, SSL algorithms acquire a comprehensive feature extractor on unlabeled data, resulting in stronger performance and a fixed knowledge of the feature extractor.

Table 1. Comparison of the prediction MAE of different models

Methods	1_3	1_5	1_6	1_7	2_3	2_4	2_5	2_6	2_7	3_3
SAE+GRU	0.4036	0.2942	0.3489	0.3457	0.3459	0.3336	0.3528	0.2659	0.3368	0.2862
CNN+LSTM(Benchmark model)	0.2074	0.0794	0.0991	0.0884	0.0355	0.1479	0.1864	0.1689	0.0826	0.0459
CNN+BiLstm	0.3511	0.2373	0.2940	0.2912	0.3264	0.3170	0.3356	0.2575	0.3194	0.2293
SSL+GRU(SOTA)	0.3406	0.2288	0.2854	0.2839	0.2732	0.2718	0.2874	0.1987	0.2721	0.2176
BiLstm+Attention	0.5947	0.2317	0.2903	0.3130	0.2815	0.2735	0.2768	0.2109	0.3547	0.2894
Proposed model	0.1214	0.0422	0.0573	0.0844	0.0263	0.0608	0.1063	0.1584	0.0926	0.0634

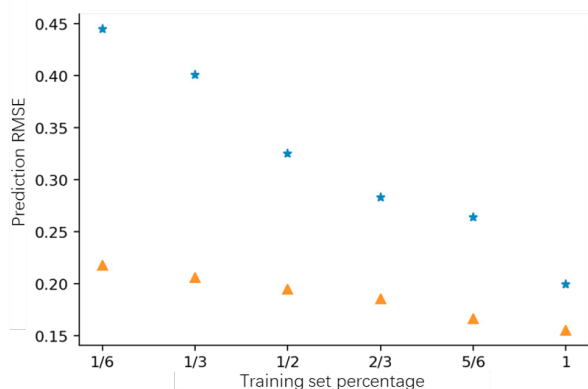


Fig. 9. Performance of SSL and CNN+LSTM (Benchmark) models at different percentages of the labeled degradation trajectories in the training set.

Yellow triangles and blue dots represent the RMSE of the predictions obtained by SSL and Benchmark models respectively.

5.2 Ability to capture degradation information

As the assumption stated in Section 2, the SSL model can capture sequence information that indicates degradation. In this paper, we compare the encoding feature results of the benchmark model and the SSL model at different label percentages. To do this, we train the SSL and benchmark models with varying percentages of labeled data, and then truncate the CNN and part of the stacks Lstm to obtain the sequential encoding features. The layer model takes the input of the CNN-LSTM and projects its output into the input of the LSTM part (see Figure 10). By comparing the Manhattan distance of the encoding features when the layer model processes the contrastive pairs, we can determine the differences in the results. The following results demonstrate these differences.

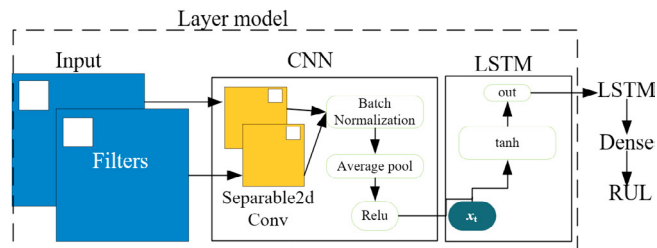


Fig. 10. Layer model generation schematic.

It's noteworthy that the SSL model's section is pre-trained, and some layer parameters are frozen. Meanwhile, the benchmark model adjusts each layer parameter for different data sets completely via error back-propagation.

For instance, the degradation data of bearing 1_3 is used to illustrate how the SSL model and the Benchmark model capture

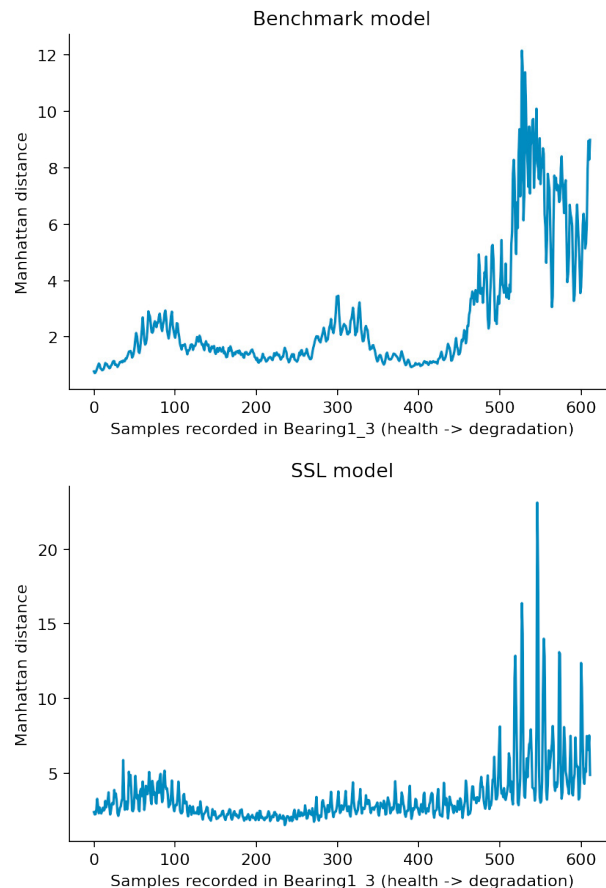


Fig. 11. Recognition ability of different models for sequential data under 100% labeling degradation process.

the difference between the negative and positive degradation segments at different rates of labeled data. The results, shown in Fig. 11, display the Manhattan distances between different positive and negative sequential sequences of the degradation process on the vertical axis, and the number of trajectory samples on the horizontal axis. Larger Manhattan distances on the figure correspond to smaller similarities between positive and negative sequential signals, capturing significant sequential differences. When labeled data is 100% available, both models exhibit similar sequence order recognition performance. They reflect that small differences in encoding features of positive and negative sequences appear early in the degradation process but become more pronounced in later stages. This partially explains the similar performance of a supervised training model and an SSL model, as they can capture degradation trend-related information effectively.

However, when only 50% of the labeled data is used for training, the Benchmark model experiences significant performance degradation due to inadequate training. This is evi-

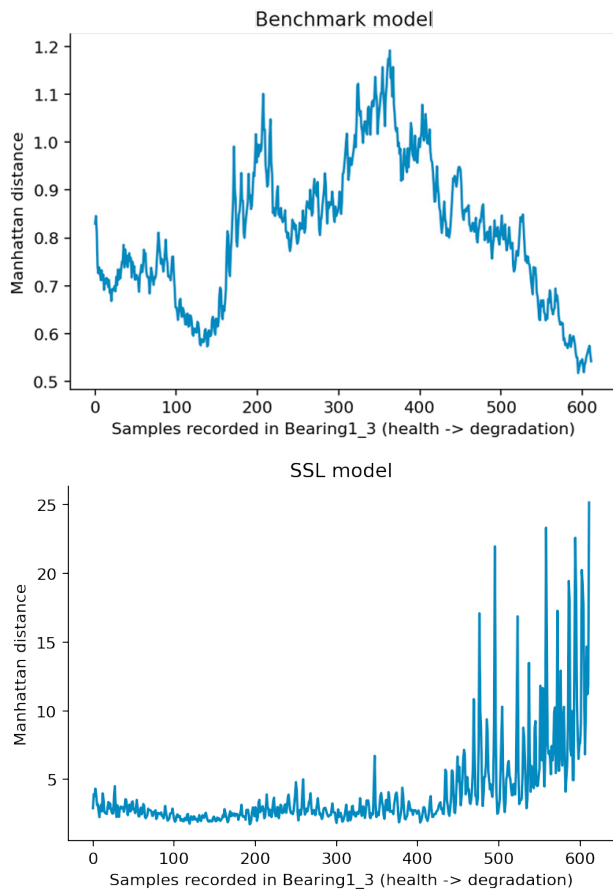


Fig. 12. Recognition ability of different models for sequential data under 50% labeling degradation process.

dent from the results shown in Fig. 12, as the model is unable to identify sequence information, and the difference between early and late degradation is no longer significant. It can be speculated that the model produces over-fitted results since the training data is reduced to 50% of the original length and mostly contains information far from the failure point.

6. CONCLUSIONS

This research proposes a new SSL method based on sequence trend comparison, this paper addresses the problems of sparse labels and data imbalance in RUL prediction. The method is validated in the CNN-LSTM framework on the "PRONOSTIA Bearing Dataset". The SSL model performs similarly to the benchmark model when training labels are available for the entire degradation process, but outperforms the benchmark as the percentage of labeled data decreases. The success of the SSL method lies in a well-designed pretext task that considers sequence information, degradation information, and representative features as the learning target to get an effective feature extractor during pre-training. The SSL model also maintains its ability to discriminate between positive and negative order data with a large amount of unlabeled data and by freezing model parameters when labeled data is reduced. The findings of this study have positive implications for promoting the use of SSL in PHM, particularly in reducing the need for high-quality labeling in machine learning. Additionally, the proposed pretext design, which focuses on extracting knowledge hidden in the sequential order, has the potential to inspire future research in PHM tasks beyond bearings. This could

contribute to the development of more efficient and accurate approaches to prognostics and health management in various industrial settings.

REFERENCES

- Akrim, A., Gogu, C., Vingerhoeds, R., and Salaün, M. (2022). Remaining useful life prediction with a deep self-supervised learning approach. In *Proc. of 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022)* 5-9th June 2022.
- Cao, Y., Ding, Y., Jia, M., and Tian, R. (2021). A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. *Reliability Engineering & System Safety*, 215, 107813.
- Chen, T., Kornblith, S., Swersky, K., Norouzi, M., and Hinton, G.E. (2020). Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33, 22243–22255.
- Deng, W., Nguyen, K.T., Gogu, C., Morio, J., and Medjaher, K. (2022). Physics-informed lightweight temporal convolution networks for fault prognostics associated to bearing stiffness degradation. In *PHM Society European Conference*, volume 7, 118–125.
- Guo, H., Zhu, H., Wang, J., Prahlad, V., Ho, W.K., and Lee, T.H. (2022). Masked self-supervision for remaining useful lifetime prediction in machine tools. *arXiv preprint arXiv:2207.01219*.
- Hahn, T.V. and Mechefske, C.K. (2021). Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder. *International Journal of Hydromechanics*, 4(1), 69–98.
- Krokotsch, T., Knaak, M., and Gühmann, C. (2021). Improving semi-supervised learning for remaining useful lifetime estimation through self-supervision. *arXiv preprint arXiv:2108.08721*.
- Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., and Varnier, C. (2012). Pronostia: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management, PHM'12.*, 1–8. IEEE Catalog Number: CPF12PHM-CDR.
- Nguyen, K.T., Medjaher, K., and Tran, D.T. (2022). A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines. *Artificial Intelligence Review*, 1–51.
- Shi, Z. and Chehade, A. (2021). A dual-lstm framework combining change point detection and remaining useful life prediction. *Reliability Engineering & System Safety*, 205, 107257.
- Xu, J., Qian, L., Chen, W., and Ding, X. (2022). Hard negative samples contrastive learning for remaining useful-life prediction of bearings. *Lubricants*, 10(5), 102.