

Physics-Informed Deep Neural Network for Bearing Prognosis with Multisensory Signals

Xuefeng Chen,¹ Meng Ma,¹ Zhibin Zhao,¹ Zhi Zhai,¹ and Zhu Mao²

¹School of Mechanical Engineering, Xi'an Jiatong University, Xi'an Shaanxi 710049, China

²Department of Mechanical and Materials Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

(Received 01 November 2021; Revised 14 March 2022; Accepted 23 March 2022; Published online 17 August 2022)

Abstract: Prognosis of bearing is critical to improve the safety, reliability, and availability of machinery systems, which provides the health condition assessment and determines how long the machine would work before failure occurs by predicting the remaining useful life (RUL). In order to overcome the drawback of pure data-driven methods and predict RUL accurately, a novel physics-informed deep neural network, named degradation consistency recurrent neural network, is proposed for RUL prediction by integrating the natural degradation knowledge of mechanical components. The degradation is monotonic over the whole life of bearings, which is characterized by temperature signals. To incorporate the knowledge of monotonic degradation, a positive increment recurrence relationship is introduced to keep the monotonicity. Thus, the proposed model is relatively well understood and capable to keep the learning process consistent with physical degradation. The effectiveness and merit of the RUL prediction using the proposed method are demonstrated through vibration signals collected from a set of run-to-failure tests.

Keywords: deep learning; physics-informed neural network (PiNN); Prognostics and Health Management (PHM); remaining useful life

I. INTRODUCTION

Prognostics and Health Management (PHM) of machine systems plays an important role in performing the digital transition of industry by combining the digital, physical, and human dimensions together. It is a computation-based paradigm that leverages physical knowledge, monitoring data, and human experience to achieve the goal of fault detection, degradation assessment, evolution prediction, and remaining useful life (RUL) prediction [1]. Up to now, a lot of efforts have been made to develop the PHM techniques, such as development of hardware (i.e., Internet of Things), smart sensors, and software including data analytics. PHM mainly contains three aspects: fault diagnosis, evolution prognosis, and decisions for management. Nowadays, a large amount of research work focuses on fault diagnosis and prognosis, which are the prerequisites of health management [2].

The initial fault detection of mechanical components has been addressed for many years and made great success in many fields [3]. If an initial fault is detected, it is more challenging to accurately predict how long the machine system will work before failure occurs, namely RUL prediction. The RUL prediction methods are categorized into three classes, namely model-based methods, data-driven methods, and hybrid methods [4]. Model-based approaches usually take advantage of physical knowledge to model the degradation process for RUL prediction. Li et al. proposed an improved exponential model for RUL prediction of rolling element bearings, where an adaptive predicting time was developed based on the three-sigma interval. The simulation and four tests of bearing degradation processes were employed to demonstrate its effectiveness [5].

Singleton et al. used both time and time-frequency domain features to track the degradation process of bearing and predicted the RUL under different operating conditions through extended Kalman filter [6]. The model-based approaches require degradation models, which means one has to master the physical knowledge of bearings' evolution. An alternative solution is to predict the RUL from historical data without physical models.

Data-driven methods are more widely investigated compared to model-based methods, because data-driven approaches only rely on the historical data without fully understanding the degradation models [7]. With accumulation of monitoring data, machine learning models including deep learning architectures are built to predict the RUL without physical models. For examples, Sun et al. proposed a deep transfer learning (DTL) network based on sparse autoencoder for RUL prediction. The RUL prediction of cutting tool using DTL model have higher accuracy compared with other methods [8]. Ma et al. proposed a convolutional neural network for RUL prediction, where time-frequency features were adopted to capture long-term dependencies through convolution operation [9].

Since deep learning has made breakthroughs in many applications such as image recognition, speech recognition, and language translation [10], it is widely investigated in prognosis of mechanical components, such as bearings [11] and gears [12]. However, it is hardly to apply deep learning methods to real mechanical systems. A primary factor is the block-box nature of deep learning framework which is complex to understand the learned features. Even though the deep learning models may achieve somewhat more accurate prediction, they do not provide the ability to understand the underlying processes. Moreover, an interpretable model including physical knowledge will stand a better chance of

Corresponding author: Xuefeng Chen (e-mail: henxf@mail.xjtu.edu.cn).

safeguarding against the building of spurious models from the historical data that may cause nongeneralizable performance. This is especially critical when dealing with predictions of complex systems that the failure would cause significant accidents. As a first step for moving beyond the black-box models of deep learning, the physical knowledge is integrated with deep learning models to improve the interpretability of the models. Motivated by embedding physical knowledge into deep neural models, in this study, a degradation-knowledge-based deep learning models are proposed for RUL prediction. Instead of using purely data-driven methods, we principally embed well-known physical principles into the recurrent neural networks. As stated in [13], the degradation process of mechanical systems is monotonic, which means that components cannot heal without repairing. Thus, an ideal degradation indicator should be monotonic over time. In this study, temperature signals collected during bearing run-to-failure tests are used to describe the degradation process because it has better monotonic characteristic compared to vibration signals. To ensure that the learned features of deep models are consistent with the physical knowledge, i.e., the monotonic characteristic of degradation process, a degradation consistency deep neural network is proposed which preserves the monotonicity of degradation. The proposed degradation consistent recurrent neural network (DcRNN) is informed by the physical knowledge in the training stage, which will make it more interpretable. The main contributions of this work are summarized as follows:

- (1) A novel physics-informed deep neural network, named DcRNN, is proposed for RUL prediction by integrating the natural degradation knowledge of mechanical components, which makes the model more interpretable and generalizable.
- (2) To incorporate this knowledge of monotonic degradation, which is generated by temperature signals over the whole life of bearings, is introduced to keep the monotonicity of the degradation process.
- (3) RUL prediction is performed utilizing the proposed DcRNN model on a set of run-to-failure bearing tests. By comparing with other methods, the results demonstrate the priority of the proposed model.

The rest of this article is organized as follows. Section II presents the literature review about physics-informed deep neural network. Section III introduces the proposed degradation consistent deep neural network. Experimental tests of bearing's run-to-failure tests are presented in Section IV. The RUL prediction and results discussion are presented in Sections V and VI, respectively. Finally, Section VII concludes the study.

II. LITERATURE REVIEW

With the breakthroughs of deep learning models in many fields, there is growing interest in the scientific community to take advantage of the benefits of deep models for prognosis of mechanical components [14], and this is because one can directly build the mapping functions with datasets of the whole degradation trajectories, but it neglects the knowledge information. To overcome the drawback of purely data-driven methods, knowledge-guided data science is investigated which aims to leverage the wealth of physical information to increase the generalization of the data-driven models [15]. Karpatne et al. proposed a physics-guided neural network (PGNN) to combine scientific knowledge of physics-based models with neural network for lake temperature modeling. By leveraging the scientific knowledge to guide the modeling of neural network, it demonstrates that PGNN has better generalizability and scientific consistency [16]. Raissi et al. introduced a physics-informed neural network (PiNN) to solve supervised learning problems while keeping any given principles that are governed by nonlinear partial differential equations [17]. The effectiveness is illustrated through some cases in the fields of fluids, quantum mechanics, etc. Furthermore, the models incorporating physical knowledge may produce scientifically interpretable models. There are various ways of embedding the physical knowledge in deep neural networks. Daw et al. developed a physics-guided framework of neural network to integrate the models with uncertainty quantification. The results show that the Monte Carlo estimates match the distribution of actual measurements correctly [18]. Karniadakis et al. summarized some prevailing trends in embedding physics into machine learning for forward and inverse problems, such as discovering hidden physics [19].

In the area of bearing's RUL prediction, data-driven methods commonly neglect the degradation knowledge. During the whole life of bearings, the degradation process is monotonic, which is usually ignored when predicting RUL with vibration signals. Thus, it is necessary to consider the degradation properties when constructing the deep learning models. In this study, the degradation process is embedded into a deep neural network, which is expected to produce more interpretable models.

III. THE PROPOSED DcRNN for RUL PREDICTION

In this study, a novel method, named degradation consistency RNN, is proposed for prognosis of mechanical components. The framework of the RUL prediction procedure with DcRNN is shown in Fig. 1.

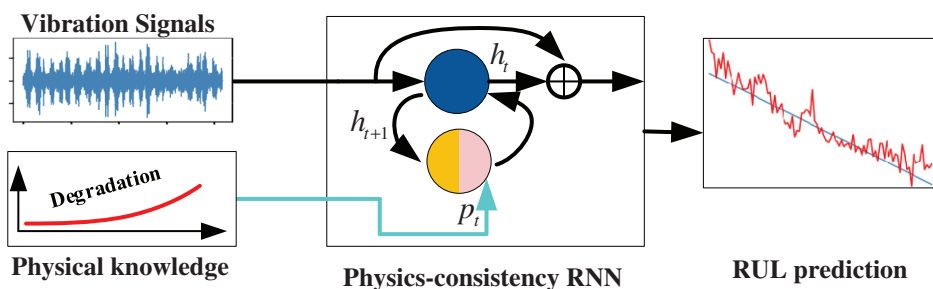


Fig. 1. DcRNN paradigm aims to infuse degradation knowledge for RUL prediction.

A. THE BASIC RNN ARCHITECTURE

Recurrent neural networks have been widely used for time series data prediction, speech recognition, language translation, and many other applications by incorporating the sequential information of time series signals [20]. RNNs model the sequential context among the signals by transforming a vector of hidden state h_t from the last step to current time step t :

$$h_t = f(x_t, h_{t-1}) \quad (1)$$

where $t \in [0, 1, \dots, T]$ means the time discretization, h is the hidden states, which represents the latent features learned from sequential data, and $f(\cdot)$ is the activated function. The hidden states are repeatedly used as inputs to update the next time states. However, it is difficult to train such a standard RNN model due to the accumulation of computed errors in backpropagation algorithm that results in gradient exploding or vanishing problem. To deal with this problem, long short-term memory (LSTM) network was developed and has been used in many application areas successfully. The architecture of LSTM is constructed through purpose-built memory cells, which show advantages in extracting and exploiting long sequential context. A LSTM is implemented by the following functions:

$$\begin{aligned} i_t &= \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \\ f_t &= \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \\ o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_t + b_o) \\ h_t &= o_t \circ \tanh(c_t) \end{aligned} \quad (2)$$

where $\sigma(\cdot)$ is an active function, W terms represent weight matrices of a LSTM model (e.g., W_{ih} is the input-hidden weight matrix), and b terms are bias vectors. $i_t, f_t,$ and o_t are the input gate, forget gate, and output gate, respectively.

B. DEGRADATION CONSISTENCY RNN

To improve the generalizability and scientific interpretability of machine learning models, the physical knowledge should be considered, which will ensure the models that are consistent with known principles. In this study, not only predicted loss in the target space y but also the violations of physical knowledge in the model outputs \hat{p} are leveraged. Both of them are used to compute the final loss function:

$$L = \arg \min \{L_{data}(y, \hat{y}) + \lambda L_{physics}(p, \hat{p})\} \quad (3)$$

where λ is a trade-off hyperparameter and controls the weight between physical consistency and empirical loss. In this way, the weights of deep neural model will be searched in the restrictions which keep consistency with physical knowledge.

In this study, the degradation information of bearings is considered when building the deep neural model, which tries to keep the learning process of model consistent with the physical degradation process. Since the degradation is monotonic, it is assumed to be increased over time, and the degradation trajectory over time is expressed as:

$$d_t = d_{t-1} + \Delta d_t \quad (4)$$

where $\Delta d_t > 0$ is the degradation increment over time due to working under loads. This means that the degradation process is irreversible; thus, the learned features of deep neural network should be consistent with the irreversible evolution of bearing's health condition.

The degradation consistency RNN is constructed based on the basic LSTM architecture by embedding the degradation knowledge. The monotonic characteristic is modeled in the proposed DcRNN through building the relationship of monotonic trend. To ensure that the deep learning model is consistent with physical knowledge, the monotonic is preserved by introducing degradation change. Instead of using vibration signals for RUL prediction in an end-to-end way, the learned features are informed by degradation monotonicity, which is represented as the physical intermediate variables that increase over time. However, it is hard to obtain the monotonic index directly because it cannot be measured through sensors. As an alternative way, the temperature signals collected in the bearing's run-to-failure tests are used as degradation indicator, because the temperature signals have the better monotonic characteristic. Vibration signals are commonly used for RUL prediction because they contain more information and are sensitive to initial fault.

The basic LSTM architecture explicitly obtains recurrence relationships through hidden units, which is not informed by physics. The proposed DcRNN embeds the degradation knowledge into the LSTM structure by a monotonic intermediate variable, as shown in Fig. 2. This intermediate unit is employed as the physical knowledge that represents the degradation process. The monotonicity is achieved by adding a positive constant to the unit d_{t-1} . In this way, the degradation process is modeled and will only increase over time. Forward propagation equations that describes the DcRNN are shown as follows:

$$i_t = \sigma(W_i x_t + W_i h_{t-1} + W_i d_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{fx} x_t + W_{fh} h_{t-1} + W_{fd} d_{t-1} + b_f) \quad (6)$$

$$\begin{aligned} c_t &= f_t \circ c_{t-1} + i_t \circ \\ &\tanh(W_{cx} x_t + W_{ch} h_{t-1} + W_{cd} d_{t-1} + b_c) \end{aligned} \quad (7)$$

$$o_t = \sigma(W_{ox} x_t + W_{oh} h_t + W_{od} d_t + b_o) \quad (8)$$

$$h_t = o_t \circ \tanh(c_t) \quad (9)$$

$$\Delta_t = \text{ReLU}(W_{\Delta} \cdot \sigma(W_h h_t + b_h) + b_{\Delta}) \quad (10)$$

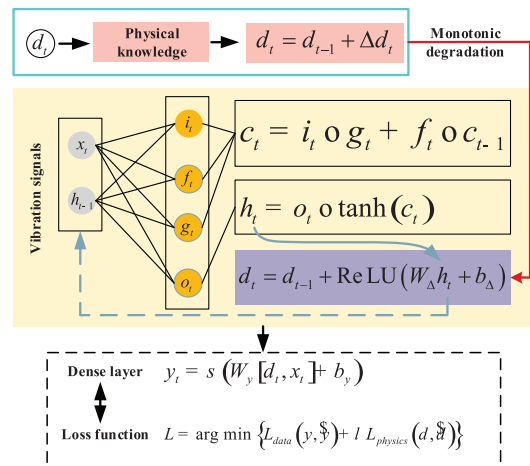


Fig. 2. The proposed DcRNN framework.

$$d_t = d_{t-1} + \Delta_t \tag{11}$$

$$y_t = \sigma(W_y[d_t, x_t] + b_y) \tag{12}$$

Equations (10)–(12) represent the embedding of physical knowledge that describes degradation process, which is used to inform feature learning and preserve monotonic degradation. This idea is motivated by degradation physics of a bearing’s evolution. To keep the intermediate unit d_t consistent with physical knowledge, a loss function between d_t and actual degradation is defined. However, it is difficult to obtain the actual degradation index, temperature signals are used instead to describe the degradation state. The RUL prediction is achieved by a dense layer of the concatenation of intermediate variable and vibration signals. The final loss function with physical consistency is defined as follows:

$$L = \arg \min \{L_{data}(y, \hat{y}) + \lambda L_{physics}(d, \hat{d})\} \tag{13}$$

IV. EXPERIMENT SETUP

Bearings’ run-to-failure tests are carried out on a special design test beds to observe the natural degradation process. The test rig is specially designed for bearing run-to-failure experiments, which includes a power and drive system, a hydraulic loading system, a lubrication system, a control system, and an independent data recording system. The main part of test rig is designed consisting of a support beam structure, where two test bearings are installed on both ends of the shaft, as shown in Fig. 3. During the experiments, both the radial load and axial load were applied to testing bearings. The parameters of test and steady bearings are given in Table I.

To monitor the health condition of bearings, both vibration and temperature signals are collected. If the temperature amplitude exceeds a certain value, it means a failure occurs; thus, the experiment should be ended to avoid an accident that may cause damage to the test rig. The vibration signals were recorded every 5 minutes with 32,768 data points, while the sampling frequency of temperature signals is 1 Hz. This is because vibration signals contain frequency characteristics compared with temperature signals. The rotating speed and radial load were set to be 2500 r/min and 12 kN, respectively.

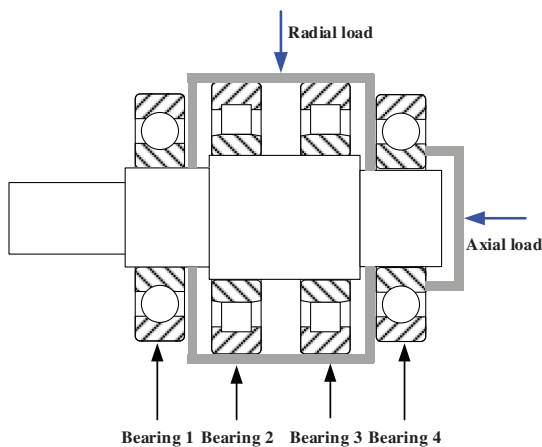


Fig. 3. Structure of test rig.

Table I Bearings’ parameters

Bearing type	Inner diameter (mm)	Outer diameter (mm)	Roller diameter (mm)	Roller number
61911	55	80	7.1	16
N312	60	130	19.1	16

Two sets of bearing run-to-failure tests were analyzed in this study. In the experiments, the failure mode of bearing in Test I is inner race, outer race, and rolling element faults, while that in Test II is inner race fault. There are 4071 and 763 datasets for Test I and Test II, respectively. The vibration and temperature signals over the whole life of bearings in Test I are presented in Figs. 4 and 5. The bearing works under the health state for a long time, then an initial defect occurs leading it to enter into a degradation stage. With the fault development and damage accumulation, the bearing’s performance deteriorates over time. Vibration and temperature signals in Test II are shown in Figs. 6 and 7.

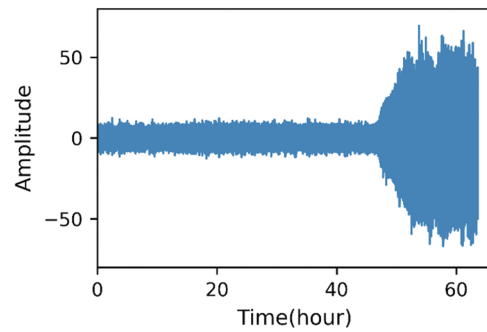


Fig. 4. Bearing’s vibration signals of Test 2.

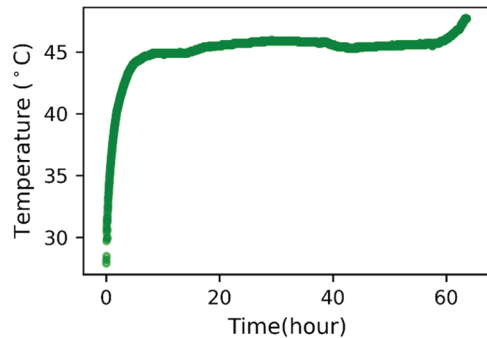


Fig. 5. Bearing’s temperature signals of Test 2.

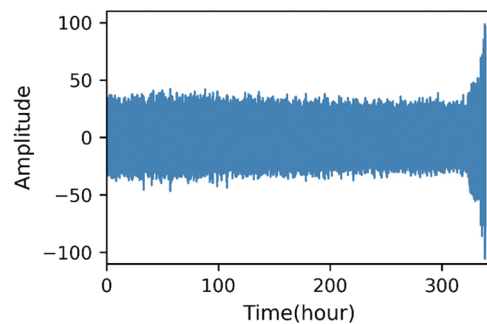


Fig. 6. Bearing’s vibration signals of Test 1.

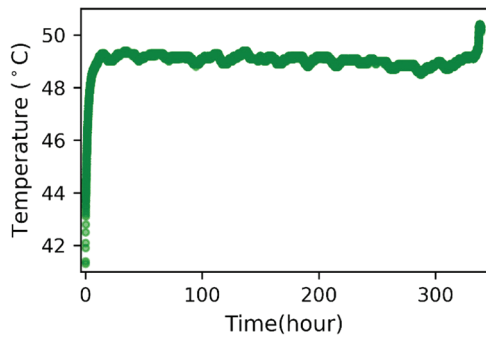


Fig. 7. Bearing’s temperature signals of Test 1.

A similar degradation process is observed from the figures. The amplitude of vibration signals decreased at 343.3 h in Test I and 57.1 h in Test II, but it does not mean that the bearing’s health condition becomes better. This is because a bearing’s degradation is an irreversible process without maintenance. Thus, the degradation should be monotonic over time. When we predict RUL with vibration signals, the physical knowledge of monotonic degradation should be embedded into the neural network to improve the performance.

V. RUL PREDICTION

The proposed model is employed for RUL prediction with vibration signals and temperature signals. The designed deep learning model has two layers with each layer of 128 hidden units. Both the weights and biases are initialized randomly. To compare the predicted accuracy of different inputs, time-domain signals, frequency domain features, and statistical features of time-domain signals are used as inputs of the DcRNN model. Both time-domain signals and frequency features are viewed as raw signals, because frequency features are obtained by fast Fourier transform (FFT) without information loss. The reason for selecting frequency features is that the fault frequency characteristics are reflected in frequency domain. The statistical features, defined in Table II, are also used as inputs of the model. To some extent, they are capable of reflecting the degradation process, as shown in Figs. 8 and 9.

The loss values over epochs of training data in Test I are shown in Fig. 10. The loss function contains two parts: data loss and physical loss function. Both the loss values

Table II The formula of statistical features

Time-domain features	Formula
ST ₁	$\sqrt{\frac{1}{n} \sum_{i=1}^n v_i^2}$
ST ₂	$\left(\frac{1}{n} \sum_{i=1}^n \sqrt{ v_i }\right)^2$
ST ₃	$\frac{1}{n} \sum_{i=1}^n v_i $
ST ₄	$\frac{1}{n} \sum_{i=1}^n \left(v_i - \frac{1}{n} \sum_{i=1}^n v_i\right)^2$

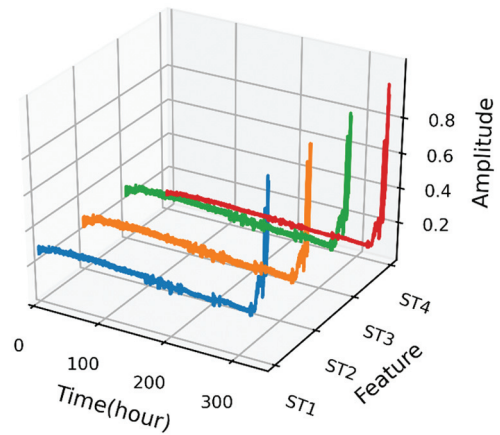


Fig. 8. Statistical features of Test 1.

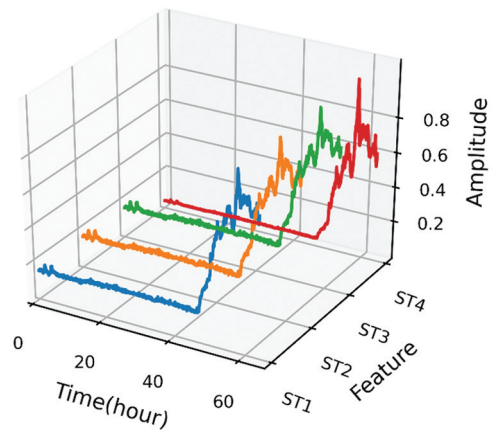


Fig. 9. Statistical features of Test 2.

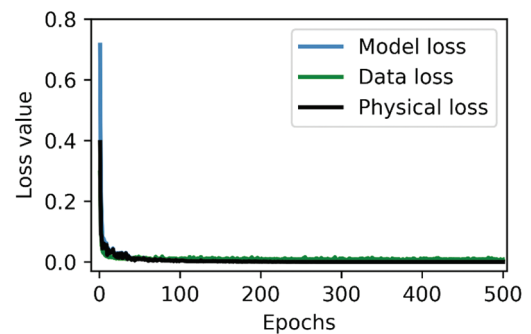


Fig. 10. Loss values of training data in Test 1.

are converged to a small value after 500 epochs, which means the physical information is considered during the training process. To show the physical consistency in the training process, the physical degradation and the learned features are illustrated in Fig. 11. The loss values of training data and physical consistency in Test II are shown in Figs. 13 and 14, respectively. Compared with physical degradation states which are represented by temperature signals, the learned features have the same trend to physical degradation, which means that the proposed model is capable of extracting the features that are consistent with physical knowledge. The predicted results of RUL are shown in Figs. 12 and 15. To evaluate the performance of

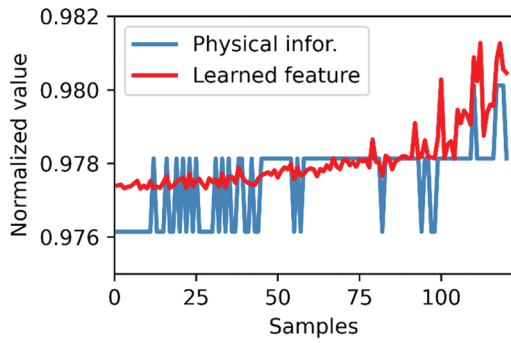


Fig. 11. Physical consistency of Test 1.

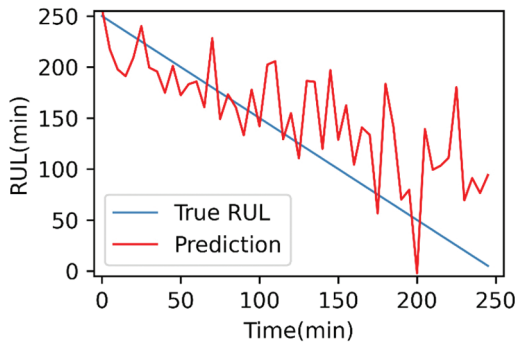


Fig. 12. RUL prediction of Test 1.

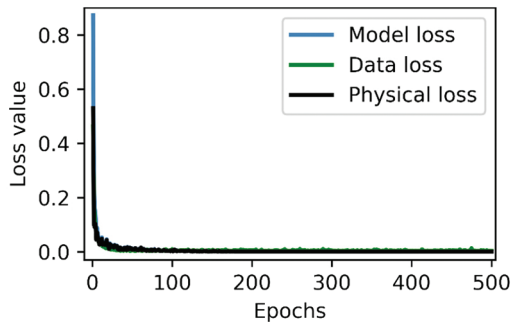


Fig. 13. Loss values of training data in Test 2.

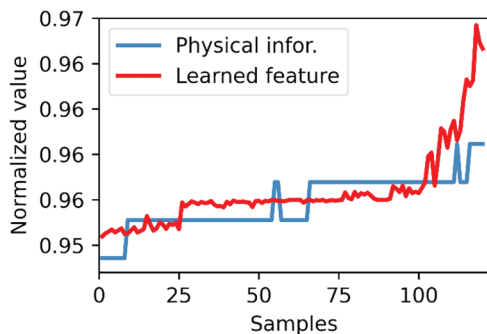


Fig. 14. Physical consistency of Test 2.

the proposed DcRNN for RUL prediction quantitatively, mean absolute percentage error (MAPE), and root mean squared error (RMSE), two commonly used metrics for prediction are shown as follows:

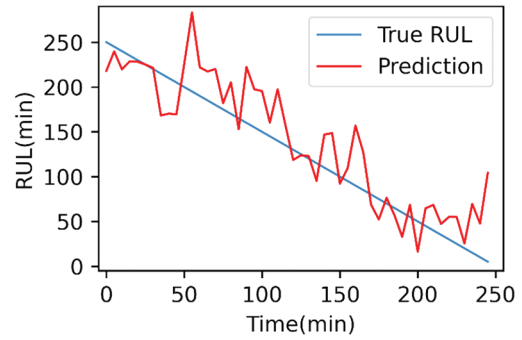


Fig. 15. RUL prediction of Test 2.

Table III RUL prediction with the proposed method

Input features	Predicted errors of Test 1		Predicted errors of Test-2	
	RMSE	MAPE	RMSE	MAPE
Time-domain signals	75.0	82.6	40.1	49.3
Statistical features	63.2	67.6	37.8	48.1
Frequency features	27.7	32.1	24.1	27.7

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - \hat{p}_i)^2} \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{p_i - \hat{p}_i}{p_i} \right| \quad (15)$$

where p and \hat{p} are predicted and actual values, respectively, and n is the number of samples.

The predicted errors are shown in Table III. It is seen that when the frequency features are used as inputs, the predicted errors are the smallest compared with other inputs. If the time-domain signals are used as inputs, the predicted performance is the worst. Because the fault characteristic frequency is able to be reflected in frequency domain, while for time-domain signals, it is hard to recognize the characteristic frequency. The predicted results with statistical features are better compared with raw vibration signals; thus, the manual designed features are useful when used for RUL prediction. It is concluded that transforming time-domain signals into frequency domain will contribute to improve the performance of RUL prediction. As stated in [21], vibration signals are indirect measurements that reflect the degradation process, so when the raw signals are used for RUL prediction, the predicted performance is not as good as that of frequency features.

VI. COMPARISON AND DISCUSSION

To demonstrate the advantages of the proposed model that embeds physical knowledge, the conventional LSTM is adopted for RUL prediction with vibration signals. The conventional LSTM includes two layers, and there are 128 hidden units in each layer. The input features, training and testing datasets are set the same with those of the

Table IV RUL prediction of traditional deep neural network

Input features	Predicted errors of Test 1		Predicted errors of Test 2	
	RMSE	MAPE	RMSE	MAPE
Time-domain signals	105.1	119.1	56.5	60.1
Statistical features	83.2	89.6	45.8	56.1
Frequency features	53.1	56.3	34.8	45.1

proposed LSTM architecture. The RUL predicted results are shown in Table IV.

When raw vibration signals in time domain are used for RUL prediction without any feature extraction, the predicted results are worse than that of statistical features. If the inputs of the model are frequency features, the predicted accuracy is the best, which is similar to the results of the proposed method. This demonstrates that bearing RUL prediction with frequency features will have better performance. The reason is that the frequency features may reflect the fault characteristic frequency compared with time-domain features, though both frequency features and time-domain signals are seen as raw signals for RUL prediction.

By comparing with the proposed method, it is shown that results of proposed DcRNN have higher predicted accuracy, which demonstrates the benefits of the proposed method that incorporates the physical knowledge. In the training process, the learned features are forced to be consistent with degradation process, which will help to improve the predicted results.

VII. CONCLUSION

In this work, a novel physics-informed deep neural network, named DcRNN, is proposed for RUL prediction of bearings. The traditional deep learning models for RUL prediction are purely data-driven methods and ignore the physical information. The proposed DcRNN is able to learn features that are consistent with scientific principles, which moves toward constructing interpretable and generalizable deep neural models. To be more specific, the latent variables are consistent with degradation state, which is monotonic, temperature signals are used to represent the degradation process. Then the latent features and vibration signals are used for RUL prediction. Bearing run-to-failure tests are carried out to obtain the historical data of the whole life. RUL prediction is performed with vibration and temperature signals using the proposed method. The results show that deep neural models which embed physical knowledge have the potential for accurate RUL prediction. As the future work, the models that include more physical knowledge should be constructed, such as the degradation knowledge of dynamic models. With more physical knowledge incorporating, the deep neural networks will be more generalizable and have better performance in prediction.

Acknowledgments

This work was support in part by China Postdoctoral Science Foundation (No. 2021M702634) and National Science Foundation of China (No.52175116).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

References

- [1] E. Zio, "Prognostics and health management (PHM): Where are we and where do we (need to) go in theory and practice," *Reliab. Eng. Syst. Saf.*, vol. 218, p. 108119, 2022.
- [2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Sig. Process.*, vol. 104, pp. 799–834, 2018.
- [3] R. B. Sun et al., "Cyclostationary modeling for local fault diagnosis of planetary gear vibration signals," *J. Sound Vib.*, vol. 471, p. 115175, 2020.
- [4] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Reliab. Eng. Syst. Saf.*, vol. 217, p. 107961, 2022.
- [5] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Trans. Ind. Electron.*, vol. 62, pp. 7762–7773, 2015.
- [6] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, pp. 1781–1790, 2014.
- [7] W. Wang, Y. Lei, T. Yan, N. Li, and A. Nandi, "Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification," *J. Dyn. Monit. Diagn.*, vol. 1, no. 1, pp. 2–8, Dec. 2021. DOI: [10.37965/jdmd.v2i2.43](https://doi.org/10.37965/jdmd.v2i2.43).
- [8] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Inf.*, vol. 15, pp. 2416–2425, 2018.
- [9] M. Ma and Z. Mao, "Deep-convolution-based LSTM network for remaining useful life prediction," *IEEE Trans. Ind. Inf.*, vol. 17, pp. 1658–1667, 2020.
- [10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [11] X. Wang, T. Wang, A. Ming, W. Zhang, A. Li, and F. Chu, "Generalized cross-severity fault diagnosis of bearings via a hierarchical cross-category inference framework," *IEEE Trans. Ind. Inf.*, vol. 18, no. 10, pp. 7240–7251, 2021.
- [12] C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Inf.*, vol. 14, no. 7, pp. 3261–3270, 2018.
- [13] M. Ma, X. Chen, X. Zhang, B. Ding, and S. Wang, "Locally linear embedding on Grassmann manifold for performance degradation assessment of bearings," *IEEE Trans. Reliab.*, vol. 66, pp. 467–477, 2017.
- [14] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, pp. 2727–2736, 2017.
- [15] A. Karpatne et al., "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, pp. 2318–2331, 2017.
- [16] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (pgnn): An application in lake temperature modeling," *arXiv preprint arXiv:171011431*, p. 2, 2017.

- [17] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
- [18] A. Daw, R. Q. Thomas, C. C. Carey, J. S. Read, A. P. Appling, and A. Karpatne, "Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling," in *Proc. 2020 SIAM Int. Conf. Data Mining*, SIAM, 2020, Cincinnati, Ohio, U.S., pp. 532–540.
- [19] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, pp. 422–440, 2021.
- [20] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *2013 IEEE Int. Conf. Acoust. Speech Signal Process.*, IEEE, 2013, Vancouver, BC, Canada, pp. 6645–6649.
- [21] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Trans. Instrum. Meas.*, vol. 66, pp. 3115–3125, 2017.