ORIGINAL ARTICLE

Open Access

Deep Spatiotemporal Convolutional-Neural-Network-Based Remaining Useful Life Estimation of Bearings

Xu Wang[1,2], Tianyang Wang[1*], Anbo Ming[2], Qinkai Han[1], Fulei Chu[1], Wei Zhang[2] and Aihua Li[2]

Abstract

The remaining useful life (RUL) estimation of bearings is critical for ensuring the reliability of mechanical systems. Owing to the rapid development of deep learning methods, a multitude of data-driven RUL estimation approaches have been proposed recently. However, the following problems remain in existing methods: 1) Most network models use raw data or statistical features as input, which renders it difficult to extract complex fault-related information hidden in signals; 2) for current observations, the dependence between current states is emphasized, but their complex dependence on previous states is often disregarded; 3) the output of neural networks is directly used as the estimated RUL in most studies, resulting in extremely volatile prediction results that lack robustness. Hence, a novel prognostics approach is proposed based on a time–frequency representation (TFR) subsequence, three-dimensional convolutional neural network (3DCNN), and Gaussian process regression (GPR). The approach primarily comprises two aspects: construction of a health indicator (HI) using the TFR-subsequence–3DCNN model, and RUL estimation based on the GPR model. The raw signals of the bearings are converted into TFR-subsequences by continuous wavelet transform and a dislocated overlapping strategy. Subsequently, the 3DCNN is applied to extract the hidden spatiotemporal features from the TFR-subsequences and construct HIs. Finally, the RUL of the bearings is estimated using the GPR model, which can also define the probability distribution of the potential function and prediction confidence. Experiments on the PRONOSTIA platform demonstrate the superiority of the proposed TFR-subsequence–3DCNN–GPR approach. The use of degradation-related spatiotemporal features in signals is proposed herein to achieve a highly accurate bearing RUL prediction with uncertainty quantification.

Keywords: Bearing, Remaining useful life, Continuous wavelet transform, Convolution neural network, Gaussian process regression

1 Introduction

Bearings are widely used in rotating machinery, and their prognostic and health management (PHM) is crucial to the precision and reliability of mechanical systems [1–3]. As a significant aspect of the prognostics method, remaining useful life (RUL) estimation contributes significantly to the PHM of bearings [4]. Typical bearing RUL estimation methods primarily include two categories: model-based and data-driven methods [5, 6]. Among them, model-based approaches describe bearing degradation based on physical degradation models, e.g., crack propagation, corrosion, or wear [7]. However, model-based methods are unsuitable for the prognostics of bearing with coupled fault modes or in complex mechanical systems [8, 9]. As a bottom-to-up prognostics approach, data-driven methods can mine the degradation information in the acquired data automatically and have garnered wide attention recently.

*Correspondence: wty19850925@126.com
[1] State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China
Full list of author information is available at the end of the article

Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 2 of 15

In data-driven baring RUL estimation approaches, vibration velocity data are widely used to model the degradation process via statistical or machine learning methods. The primary statistical methods used are the Markov process [10], Kalman filter [11], Wiener process [12], and particle filter [13]. The rapid development of sensing and data-acquiring technologies has resulted in the significant increase in information to be employed for condition monitoring. However, high interference noise and complicated mapping render it difficult to model the statistical distribution of the data obtained [14, 15]. Recently, deep learning methods have provided a good alternative for predicting the RUL of bearings without requiring much prior knowledge regarding their degradation process [16, 17]. Many successful applications have been realized in the RUL estimation of bearings based on the deep belief network (DBN) [18], auto encoder [19], recurrent neural network (RNN) [20], long short-term memory (LSTM) [21], and convolutional neural network (CNN) [22]. However, some problems remain in these methods.

Most previous studies use raw data or statistical features as model inputs only, which does not allow the extraction of complex information hidden in signals. However, signal analysis in a single domain exhibits limited fault-related feature representation capability [23, 24]. An effective prediction depends on the well-informed input for deep-learning-based approaches. However, prediction performance is still limited when powerful data analysis methods are used. Therefore, the time–frequency representation (TFR) method, which can reveal the complex structure of signals with multiple components, has been widely used to analyze nonstationary signals. Many studies have proven the effectiveness of TFR-based bearing-fault diagnosis approaches. Klein [25] suggested a bearing diagnostics solution that applied image-processing techniques to TFRs. Shi [26] demonstrated that the effectiveness of time–frequency analysis depends on its capability in representing the signal energy. Recently, CNN models have demonstrated superiority for the analysis of images, and it can also be used in TFR-based RUL estimation. Hence, Zhu [22] performed the RUL prediction of bearings under one fixed operating condition based on the wavelet transform and the multiscale two-dimensional (2D) CNN (MSCNN) model. However, two limitations exist in TFR–2DCNN-based bearing RUL predictions.

(1) The degradation process is typically regarded as a nonlinear Markov process with non-Gaussian characteristics in the 2DCNN-based RUL estimation of bearings, in which the complex temporal dependence of the data under previous states is disregarded [27, 28].

(2) The output of neural networks is used directly as the estimated RUL in 2DCNN-based bearing RUL estimation approaches, which results in extremely volatile predictions that lack robustness. In practice, it is difficult to ensure the accuracy of each discrete predicted health indicator (HI) for the model [29].

Because the defect expansion of bearings is a gradually evolving process, the Markov property is unsuitable for RUL estimation tasks to some extent. Wang [30] constructed a degradation indicator for wear using the Monte Carlo method. For a specified time series, it is assumed that the generation of each observation result depends on an implicit variable that cannot be directly observed. Furthermore, Zhu and Liu [31] proposed a non-Markovian hidden semi-Markov model method to provide prognostics that offered more powerful modeling capabilities for practical problems. These studies indicate the importance of temporal correlation for RUL prediction, which is disregarded in the 2DCNN-based bearing RUL estimation approach. In the gradual process of bearing degradation from normal to failure, each observation must be correlated significantly to several previous observations. Therefore, establishing a correlation between different degradation states is crucial for accurate RUL predictions. In addition, the overall degradation trend of the bearing obtained based on predicted HIs should be considered in the RUL estimation of bearings to compensate for the high volatility and poor robustness of the model.

In this regard, a novel bearing RUL estimation method is proposed based on the continuous wavelet transform (CWT) method, a three-dimensional convolutional neural network (3DCNN) model, and a Gaussian process regression (GPR) model. It primarily comprises two aspects: 1) construction of HIs using the proposed TFR-subsequence–3DCNN model, and 2) bearing RUL estimation based on the GPR algorithm. A flowchart of the proposed TFR–3DCNN–GPR framework is presented in Figure 1. In practice, vibration signals in both the horizontal and vertical directions were monitored. It is meaningful to analyze signals in both directions because they can reflect the bearing status more comprehensively. The CWT method was used to convert the raw bearing signal into a TFR, and a set of TFR subsequences was created using an overlapping strategy. The 3DCNN model was applied to extract the hidden features from the TFR-subsequences and construct HIs. The GPR algorithm was adopted to model the distribution of HIs and calculate the confidence interval of the RUL prediction. The performance of the proposed approach was evaluated using vibration signals acquired from the PRONOSTIA platform.
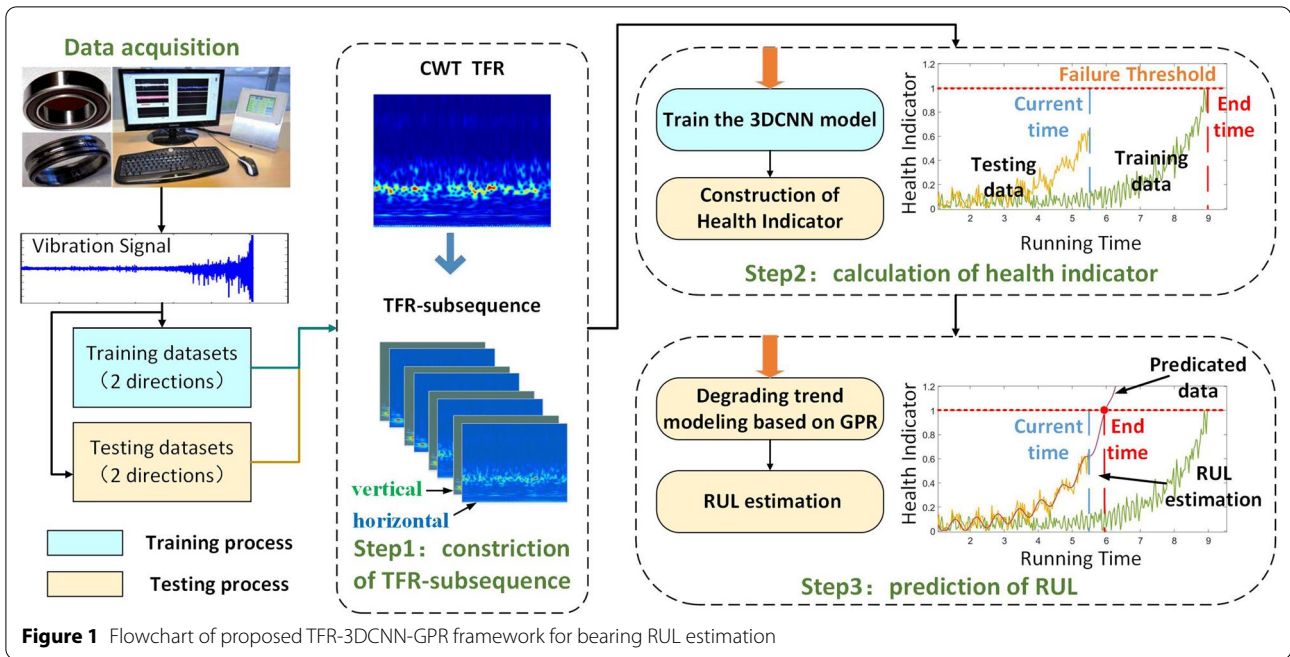
Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 3 of 15

Figure 1 Flowchart of proposed TFR-3DCNN-GPR framework for bearing RUL estimation

The primary contributions of the proposed approach can be summarized as follows:

(1) A novel TFR-subsequence–3DCNN framework is proposed for bearing degradation modeling. By effectively extracting the spatial features and temporal correlations of the TFR-subsequences generated from the measured vibration signals, the proposed method can support more accurate RUL estimation of bearings.

(2) The GPR algorithm was adopted to model the distribution of HIs and calculate the confidence interval of the RUL prediction. Unlike most deep-learning-based bearing RUL estimation approaches, the proposed method can effectively mitigate prediction volatility and facilitate the quantification of estimation uncertainty.

The remainder of this paper is organized as follows. Section 2 presents the proposed TFR-subsequence–3DCNN model for HI construction comprehensively. In Section 3, the proposed GPR-based bearing RUL estimation approach is described. The validation experiments are presented in Section 4, in which a comparison of related studies is provided as well. The conclusions are presented in Section 5.

2 TFR-Subsequence–3DCNN Model

The 3DCNN model was applied to extract the spatiotemporal features of the TFR-subsequences by performing a three-dimensional (3D) convolution kernel [32] and
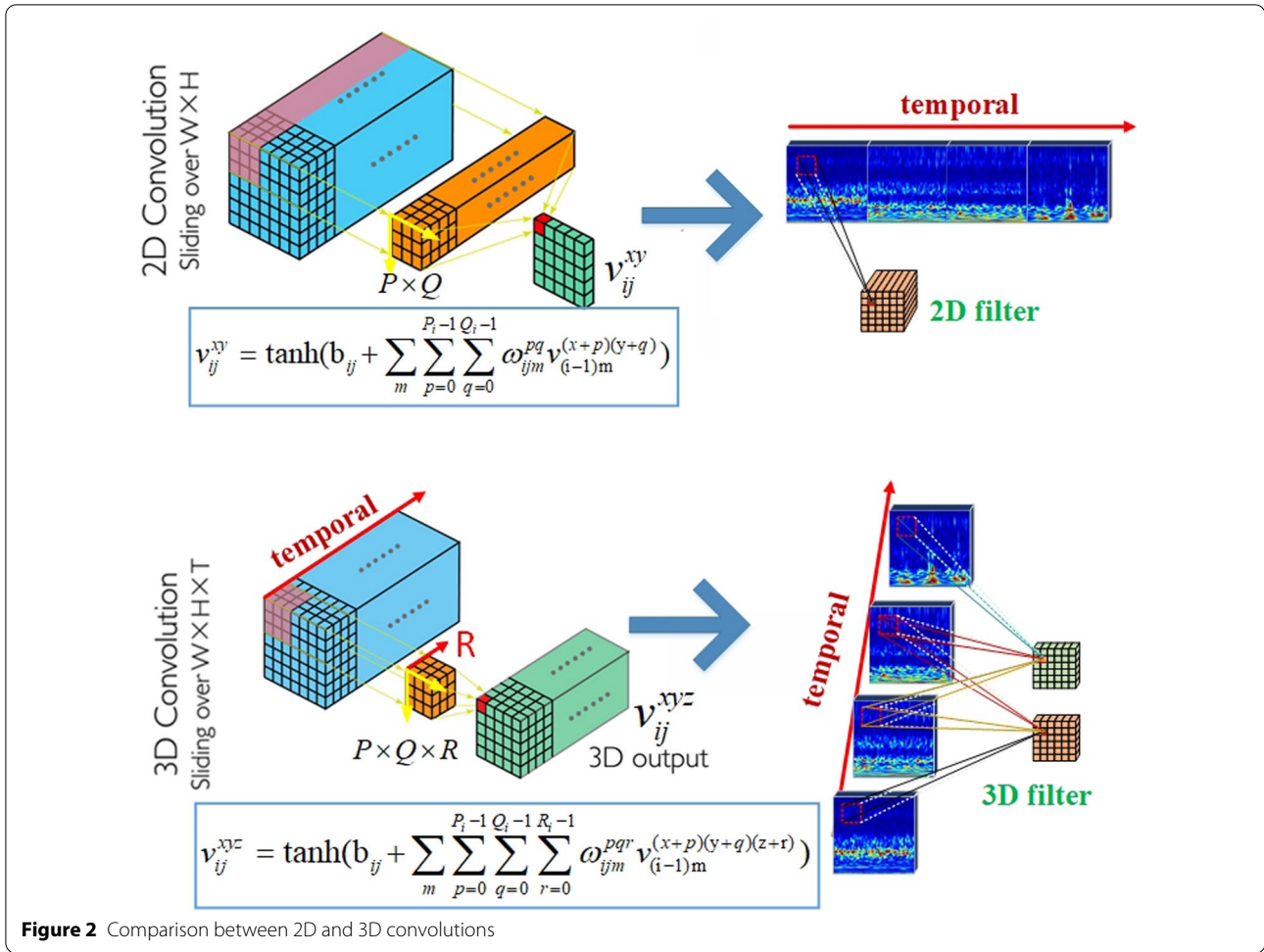
constructing HIs that indicate bearing degradation. Compared with one-dimensional time signal analysis or 2D time–frequency analysis, more hidden characteristic information of the bearing vibration signals can be extracted using the 3DCNN, and a more accurate RUL prediction effect is expected from using the proposed method.

2.1 Three-Dimensional Convolution Kernel for Spatiotemporal Feature Extraction

Figure 2 shows a comparison between 2D and 3D convolutions. The feature maps in each convolutional layer are connected with multiple consecutive TFRs in the upper layers in the 3D convolution kernel to capture more temporal information of the bearing TFRs. The 2D convolutional layers were connected to the local receptive field of the feature maps, where the weights were passed through activation functions. If we denote v_{ij}^{xy} as a unit at (x, y) of the jth feature map and the ith layer, then

$$v_{ij}^{xy} = \tanh\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \omega_{ijm}^{pq} v_{(i-1)m}^{(x+p)(y+q)}\right),$$

(1)

where f refers to the activation function, b_{ij} the bias of the corresponding feature map, m the index number, and ω_{ijm}^{pq} the value of the kernel at position (p, q); height P_i and width Q_i correspond to the ith convolution kernel. The Markov property indicates that the state of the model at $t > t_0$ is independent of that before t_0.

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 4 of 15

Figure 2 Comparison between 2D and 3D convolutions

The 3D convolution kernel adds a temporal scale hyperparameter R to the 2D convolution kernel to extract the information at different temporal scales. The value at (x, y, z) corresponding to the jth feature map of the ith layer is expressed as

$$v_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} \omega_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right),$$
(2)

where R_i refers to the temporal scale of the ith layer 3D kernel, and ω_{ijm}^{pqr} is the value at (p, q, r) of the kernel corresponding to the mth feature map. A cube with fixed consecutive TFRs will be formed in the 3D convolution, and it contains both the spatial information of each single TFR and the temporal information along adjacent TFRs.

2.2 Constructed TFR-Subsequence–3DCNN Model for HI Construction

The CWT enables a multiscale refinement analysis of signals through expansion and translation functions,

thereby effectively extracting time–frequency domain information from signals. Compared with other time–frequency analysis methods, such as short-time Fourier transform, Wigner-Ville distribution (WVD), pseudo WVD (PWVD), and smooth pseudo WVD, the CWT offers advantages of low computational complexity and tight support [33]. Considering the basic wavelet function $\psi(t)$, it can be scaled or expanded k times and translated into l steps as follows:

$$\psi_{k,l}(t) = \frac{1}{k}\psi\left(\frac{t-l}{k}\right).$$
(3)

For a signal $s(t)$, the CWT can be expressed as

$$CWT_s(k, l) = \frac{1}{\sqrt{k}} \int_0^{+\infty} s(t)\psi^*\left(\frac{t-l}{k}\right) \mathrm{d}t,$$
(4)

where $\psi^*(t)$ is the conjugate function of the basic $\psi(t)$.

The proposed TFR-subsequence–3DCNN model is illustrated in Figure 3. Combining TFRs in two directions, i.e., the horizontal and vertical directions, into

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 5 of 15

one subsequence with height H, width W, and channel number $C=2$, the TFR-subsequence for the current time can be obtained. The corresponding HI for the training dataset is defined as the degradation percentage at time t, i.e., HI values of 0 and 1 correspond to perfectly healthy and failure, respectively. Hence, a higher health indicator value indicates the worse health and lower RUL of the bearings. Using the run-to-failure time T for the training data, the corresponding HI is expressed as

$$HI(t) = \frac{t}{T}. \tag{5}$$

Details of the overall network architecture are listed in Table 1. The labeled TFR-subsequence was used as the 3D convolution input, which should be a five-dimensional tensor in the form of (N, H, W, L, C), where H and W represent the height and width, respectively; C refers to the channel number of the current TFR-subsequence; L is the overlapping number of the TFR-subsequences (L = 1 in 2D convolution); N refers to the number of samples. The outputs of the model, i.e. the HIs, can effectively reflect the degradation of the bearings. Two fully connected layers exist in the network after the three convolution layers. The sizes of the three 3D convolution kernels were 7×7×2, 3×3×2, and 3×3×2, separately. The stride of each convolution layer was set to 2×2×1. In addition, one dropout layer was added after the last convolution layer to prevent overfitting in the training. The sigmoid function was used as the activation function in the last layer of the network to fit the established HI. The rectified linear unit (ReLU) function was used as the activation function for the other layers to overcome gradient disappearance. For a variate v, the sigmoid and ReLU functions are expressed as follows [34]:

$$\text{Sigmoid}(v) = \frac{1}{1+e^{-v}}, \tag{6}$$

Table 1 Details of proposed model architecture by layer

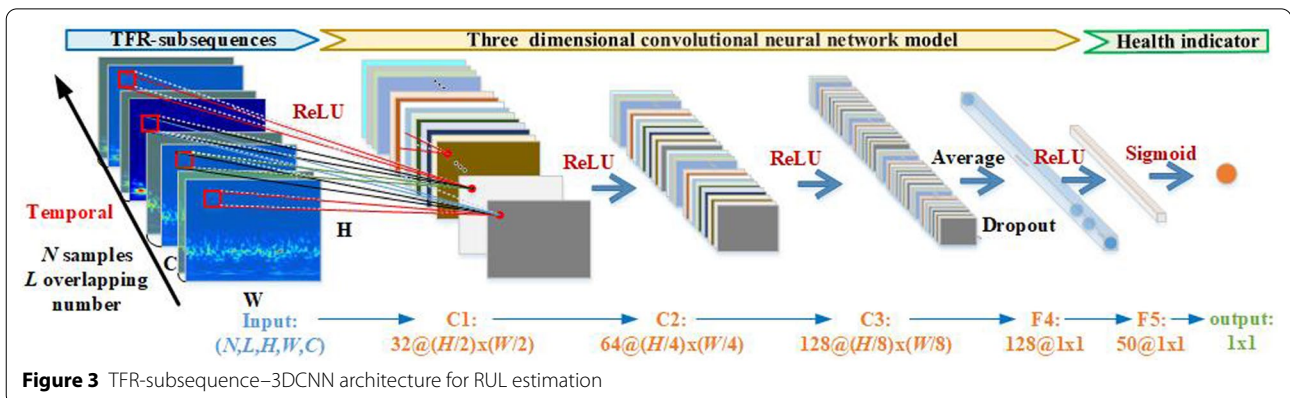| Layer	Details
Input	Size: $N \times L \times W \times H \times C$; channel: $L \times C$
Convolution	3D kernel: $7 \times 7 \times 2$; channel: 32; stride: $2 \times 2 \times 1$
Convolution	3D kernel: $3 \times 3 \times 2$; channel: 64; stride: $2 \times 2 \times 1$
Convolution	3D kernel: $3 \times 3 \times 2$; channel: 128; stride: $2 \times 2 \times 1$
Dropout	Rate: 0.25
Pooling	Average_3Dpooling: $12 \times 12 \times L$
Fully-connected	Channel: 128
Fully-connected	Channel: 50
Output	Channel: 1

$$\text{ReLU}(v) = max(0, v). \tag{7}$$

3 GPR for Prediction

GPR is an important tool for inferring the relationship between inputs and outputs and determining the conditional distribution of the target output. As a nonparametric, statistical, and Bayesian method, GPR was adopted in this study to model the evolution of uncertainty in time-stream data over time to support the failure time prediction of the bearings. For a training set $D = [(t_i, HI_i|\ i=1, 2, ..., n)]$, where t_i and HI_i denote the time number and target outputs, respectively. The prior distribution of the multivariate variance normal function $f(t)$ can be specified through the mean function $\mu(x)$ and covariance matrix $\boldsymbol{K}$.

$$f(t_i) \sim \text{MVN}(\mu(t_i), \boldsymbol{K}), \quad i = 1, 2, \cdots, n \tag{8}$$

Subsequently, the problem is converted to the prediction of the output in D^* using the observed training set.

Figure 3 TFR-subsequence–3DCNN architecture for RUL estimation

Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 6 of 15

The covariance function must satisfy the Mercer condition, i.e., every semi-positive definite symmetric function should be a kernel [35, 36]. Therefore, the optimal solution of the hyperparameters ϕ can be adaptively obtained via the maximum likelihood method. Using the established regression model, the output can be obtained by substituting the input value t_i into the model. The GPR process in the experiment can be summarized as follows:

(1) Establishment of the training dataset: The HIs predicted by the TFR-subsequence–3DCNN model and the corresponding time information were used to construct the training dataset. The GPR model is expected to achieve an accurate and stochastic RUL prediction.

(2) Model training: The GPR-based model was trained using the predicted HIs of the training bearings. The mean value was set to 0. The covariance function was set as a combination of the square exponential and periodic covariance functions.

(3) Prediction: The model was extrapolated in time to obtain the predicted HI values for RUL estimation. The failure time of this bearing was determined based on the physical meaning of the HI, and its RUL was calculated.

4 Case Study

4.1 Experimental Setting
The experimental data were obtained from the PRONOSTIA platform [37], and two accelerometers were mounted to measure the vibration signals. Figure 4 shows the PRONOSTIA platform. The sampling rate of the vibration signals was 25.6 kHz, and 0.1 s of data were recorded every 10 s. A total of 17 datasets under different conditions are presented in Table 2, and each data group contained signals in both the horizontal and vertical directions.

4.2 Vibration Data and TFR-Subsequence
Figure 5 shows the horizontal signal and its corresponding TFRs for two samples (the first and last samples) of bearing 1-1. The accuracy of RUL prediction can be improved by correctly describing the nonlinear degradation features of the bearing vibration signals with different faults. In CWT-based analysis, it is crucial to use appropriate parameters. Many studies have shown that reasonable parameter selection can effectively improve the signal parsing ability of the CWT model, effectively counteract the interference of background noise, and improve the model robustness. The complex Morlet wavelet was proven to be effective in analyzing bearing degradation. According to the criterion of maximum energy-to-Shannon entropy ratio, the complex Morlet wavelet [38] function was selected in the CWT with a bandwidth parameter of 3 and a center frequency of 3. The length of the CWT wavelet function scale was set to 256. Figures 5a and b show the bearing operating under stable conditions, and the center frequency was approximately 4000 Hz. Figures 5c and d show the bearing at the end of the life-recession phase with severe failure. The impact components are shown in Figure 5c. The high energy concentrations of these impact components in the time–frequency domain are shown in Figure 5d. Therefore, the CWT-based TFRs can effectively highlight fault-related information in the bearing signals.

Furthermore, to improve the computational efficiency, the bilinear interpolation algorithm [39] was used to perform digital zooming on the TFRs. The dimensions of the original CWT TFRs, which were 256×2560, incurred a heavy computational burden. The inverse transform of the pixel in the bilinear interpolation was calculated as follows:

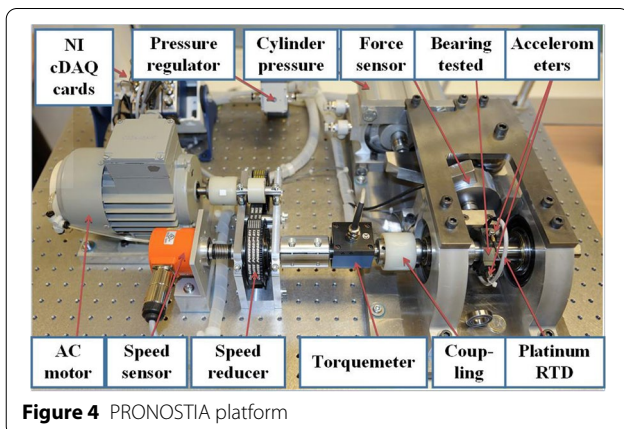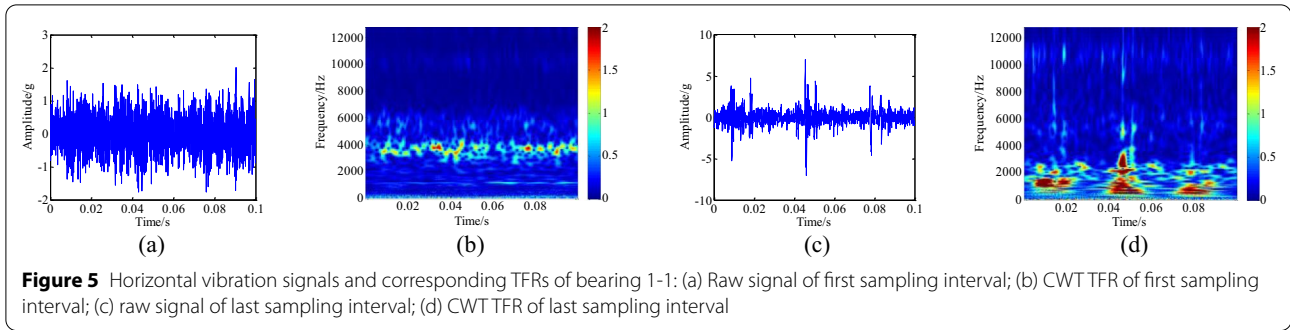**Figure 4** PRONOSTIA platform

Table 2 Bearing dataset with operational conditions

Condition	1	2	3
Speed (r/min)	1800	1650	1500
Load (N)	4000	4200	5000
Training dataset	Bearing 1-1 Bearing 1-2	Bearing 2-1 Bearing 2-2	Bearing 3-1 Bearing 3-2
Testing dataset	Bearing 1-3 Bearing 1-4 Bearing 1-5 Bearing 1-6 Bearing 1-7	Bearing 2-3 Bearing 2-4 Bearing 2-5 Bearing 2-6 Bearing 2-7	Bearing 3-3

Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 7 of 15

Figure 5 Horizontal vibration signals and corresponding TFRs of bearing 1-1: (a) Raw signal of first sampling interval; (b) CWT TFR of first sampling interval; (c) raw signal of last sampling interval; (d) CWT TFR of last sampling interval

$$I^*(x,y) =$$
$$(1 - x_f)(1 - y_f)A + x_f(1 - y_f)B + y_f(1 - x_f)C + x_f y_f D, \quad (9)$$

where $A = I(x_i, y_i)$, $B = I(x_i + 1, y_i)$, $C = I(x_i, y_i + 1)$, and $D = I(x_i + 1, y_i + 1)$. In addition $x = x_i + x_f$ and $y = y_i + y_f$, with integer components x_i and y_i as well as corresponding floating-point coordinates x_f, y_f ($0 < x_f, y_f < 1$). Using the raw TFR corresponding to Figure 5b as an example, which was reduced with dimension $[R_{row}, R_{column}]$, the unified value was set as $R_{row} = R_{column} = R$ for convenience of calculation. Figure 6 shows the TFR visualization images for the four cases, where the reduced dimensionality R was set to 500, 250, 100, and 50. As shown, aliasing occurred in the TFR image when the reduced dimension was extremely small; this shows that bilinear interpolation results in the loss of high-frequency components in TFRs. To maintain sufficient information in the reduced TFRs, R was set to 100. Subsequently, the reduced TFRs were arranged in the chronological order for the network input. Figure 7 shows the process in which the signal is converted into TFR-subsequences. The dislocated overlapping method can ensure the validity of temporal information extraction for RUL estimation.

4.3 RUL Estimation
The key factors pertaining to the proposed model were investigated comprehensively, including the parameter optimization method selected and the effect of the overlapping number. The superiority of the GPR-based RUL prediction was analyzed.
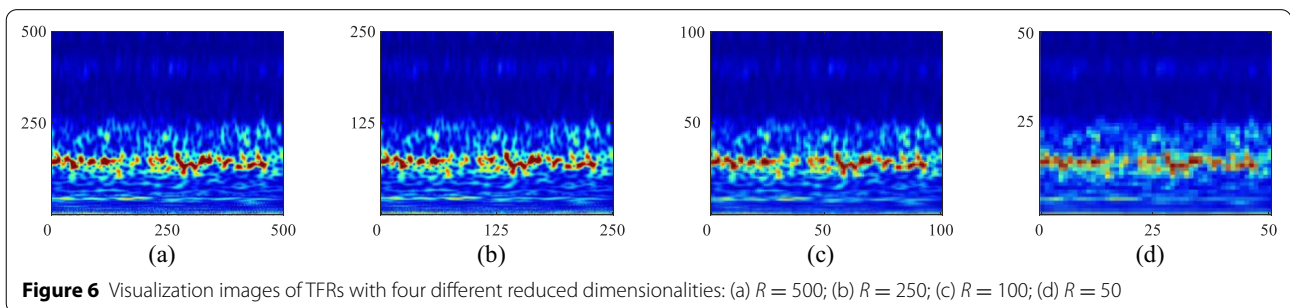
4.3.1 Parameter Optimization Method
The proposed TFR-subsequence–3DCNN model was built using the Keras framework and implemented in an NVIDIA Tesla K80 GPU with a mini-batch of 128 and an epoch of 100. To avoid overfitting in the training process, 20% of the training data were randomly shuffled as the validation dataset to determine the network parameters. Finally, the mean square error (MSE) was applied as the loss function in the 3DCNN model. By setting $y \in \mathbb{R}^{1 \times n}$ as the true label and $\hat{y} \in \mathbb{R}^{1 \times n}$ as the predicted label, the MSE loss is expressed as

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \quad (10)$$

For a training set $Tr = [(x_i, y_i)|i = 1, 2, \cdots, n]$, where x_i is the TFR subsequence and y_i is the predicted value, the structural risk minimization rule is used as the objective function, as follows:

$$\arg\min_{\theta} O(Tr; \theta) = \sum_{i=1}^{n} L(y_i, f(x_i); \theta), \quad (11)$$
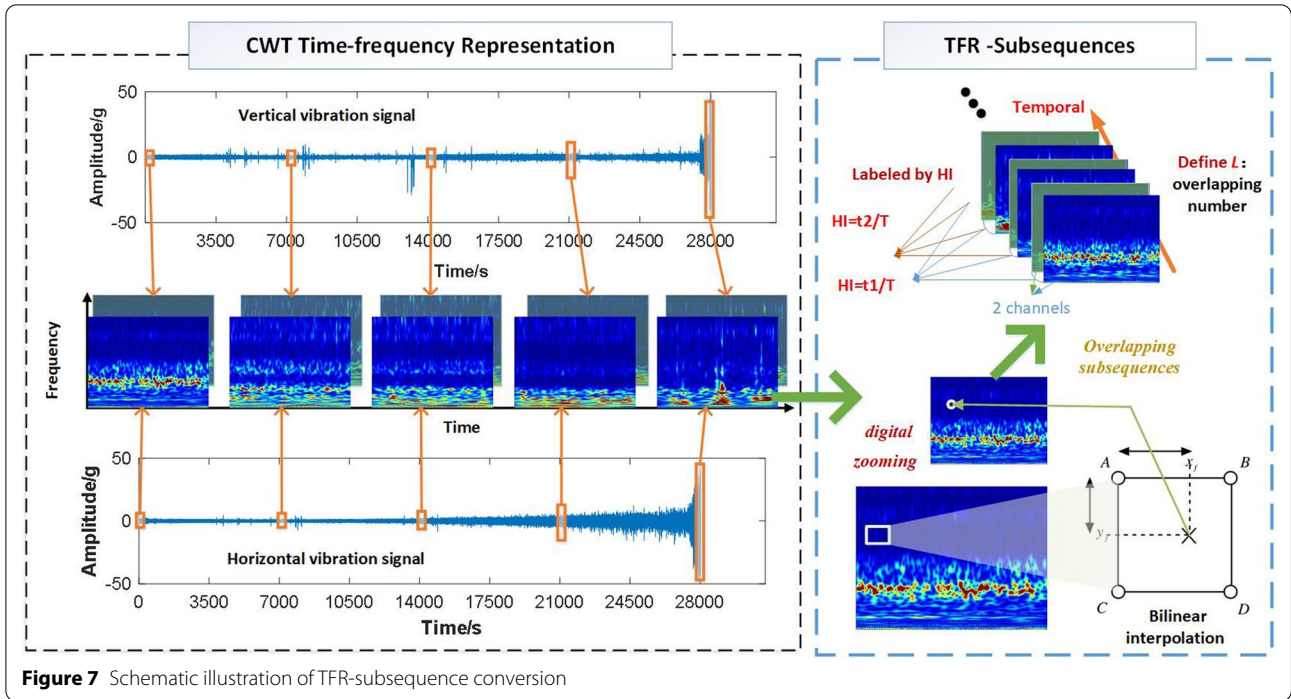

Figure 6 Visualization images of TFRs with four different reduced dimensionalities: (a) $R = 500$; (b) $R = 250$; (c) $R = 100$; (d) $R = 50$

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 8 of 15

Figure 7 Schematic illustration of TFR-subsequence conversion

where θ is the whole model parameter, $f(x_i)$ the model output, and $L(y_i, f(x_i); \theta)$ the MSE loss function. To adaptively minimize loss during model training, the root mean square prop (RMSProp) [40] method was applied to the model. RMSProp uses differential squared weighted averages for the gradients of weights W and bias b to overcome significant swings in the update of the loss function, as well as to accelerate the convergence. During the jth iteration, the RMSProp is solved as follows:

$$S_{dw} = \beta S_{dw} + (1-\beta)\mathrm{d}w^2, \tag{12}$$
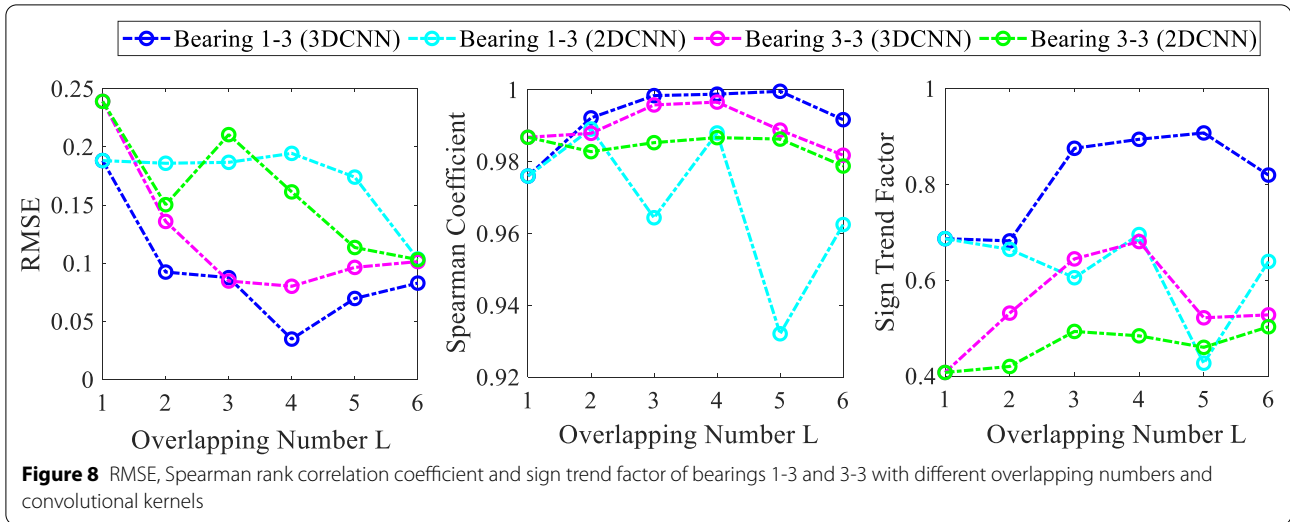
$$S_{dw} = \beta S_{dw} + (1-\beta)\mathrm{d}w^2, \tag{13}$$

$$w = w - \alpha\frac{\mathrm{d}w}{\sqrt{s\mathrm{d}w}+\varepsilon}, \tag{14}$$

$$b = b - \alpha\frac{\mathrm{d}b}{\sqrt{s\mathrm{d}b}+\varepsilon}. \tag{15}$$

The gradient momenta, S_{dw} and S_{db}, were accumulated by the loss function during the previous (j-1)th round of iteration. β is the index of gradient accumulation. When dw or db is a significant value, it is divided by the square root of the gradient accumulated during update to obtain a smaller update amplitude.

4.3.2 Effect of Overlapping Number L

The overlapped TFR-subsequences contain more useful degradation information and may be more advantageous than TFRs in improving the prediction performance when used as inputs of the 3DCNN model. In theory, the higher the value of L, the more time-related information is contained in the constructed TFR-subsequences, and the more conducive is the model in modeling the time dependence of the sample. However, if L is extremely high, then the effect of a single sample will be weakened, and the distinguishability of the samples at different times will be weakened, which is not conducive to the capturing of fault-related transient information in the signal. Owing to the effect of the overlapping number L on the prediction, experiments were performed on bearings 1-3 and 3-3 with different kernels (3DCNN and 2DCNN) and overlapping numbers (L was set as 1, 2, 3, 4, 5, and 6). As the interval time for each TFR-generating signal was 0.1 s, the contrast signal time was set to 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6 s in the 2DCNN for comparison, respectively. It is noteworthy that when $L = 1$, the 3DCNN was the same as the 2DCNN. However, when $L \geq 2$, they were different. For the 2DCNN, inputting TFRs corresponding to multiple times can increase the amount of information that can be obtained; however, the change information between adjacent TFRs at consecutive times cannot be extracted. The 3DCNN model can simultaneously extract the variation in consecutive TFRs. As criteria, the Spearman rank correlation coefficient, sign trend factor,

Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 9 of 15

Figure 8 RMSE, Spearman rank correlation coefficient and sign trend factor of bearings 1-3 and 3-3 with different overlapping numbers and convolutional kernels

and root mean squared error (RMSE) were compared to investigate the prediction performance, as shown in Figure 8.

For the predicted HIs $\hat{y} \in \mathbb{R}^{1 \times n}$ of the CNN model and true HIs $y \in \mathbb{R}^{1 \times n}$, the RMSEs of y and $\hat{y}$ are

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2}, \tag{16}$$

where y_i represents the measurement vector of a feature on the ith system, and $\bar{y}$ refers to the mean of y. The Spearman rank correlation coefficient [41] and sign trend factor [42] of the predicted HIs are calculated as follows:

$$\rho(\hat{y}, t) = \frac{1}{n} \sum_{i=1}^{n} \left| \mathrm{corr}\big(\mathrm{rank}(\hat{y}_i), \mathrm{rank}(t_i) \big) \right|, \tag{17}$$

$$tf(\hat{y}, t) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathrm{sgn}\big(\hat{y}_{i+1} - \hat{y}_i \big), \tag{18}$$

where $t \in \mathbb{R}^{1 \times n}$ in Eq. (17) represents the measured feature on the ith system, $\mathrm{rank}(\hat{y}_i)$ is equal to the average of their positions in the ascending order of value $\hat{y}_i$, and $\mathrm{corr}(\cdot)$ refers to the correlation coefficient. In Eq. (18), $\mathrm{sgn}(\cdot)$ refers to the Sgn function.

As shown in Figure 8, the calculation error of the 3DCNN-based approach was less than that of the 2DCNN-based approach. The RMSE decreased first and then increased as L increased. Within a certain range, increasing L enables more information to be extracted. However, when L is extremely large, the difference in data at different times will be reduced. The TFRs during the stable operation of the bearing did not differ

significantly, rendering it difficult to extract subtle fault-related changes in the signal. Spearman's rank correlation coefficient can be used as an indicator of the dependency between two variables, e.g., the HI and runtime. It uses a monotone equation to evaluate the correlation between two statistical variables. For bearings 1-3, the coefficient value obtained by the 3DCNN was significantly higher than that by the 2DCNN, and the maximum value was achieved when L was 3, 4, or 5. For bearing 3-3, the coefficient value obtained by the 3DCNN was higher than that by the 2DCNN, and the maximum value was achieved when L was 3 or 4. The sign trend factor represents the monotonicity of the data, whose trend was the same as that of the Spearman coefficient.

Furthermore, Figure 9 shows the HIs of bearing 1-3 calculated using the 3DCNN model with different overlapping numbers. As shown, when $t \geq 10000$ s, the linearity of the HIs in the six states was extremely high; however, when $t < 10000$ s, the HIs fluctuated significantly. When $L \leq 4$, the fluctuation weakened. However, when $L = 6$, the prediction effect worsened. It was discovered that an increase in the overlapping number within a certain range resulted in a more accurate HI prediction. However, a high overlapping number resulted in more memory storage and calculation burden for the model. Therefore, the overlapping number L was set to 4.

4.3.3 Effect of TFR Generation Method Selection
An effective TFR generation method should be able to fully reflect the degradation information of the bearing; in fact, this is a prerequisite for the established model to fully utilize the temporal and spatial features in the bearing vibration signal. The CWT method has been proven to be effective for the time–frequency analysis of bearing
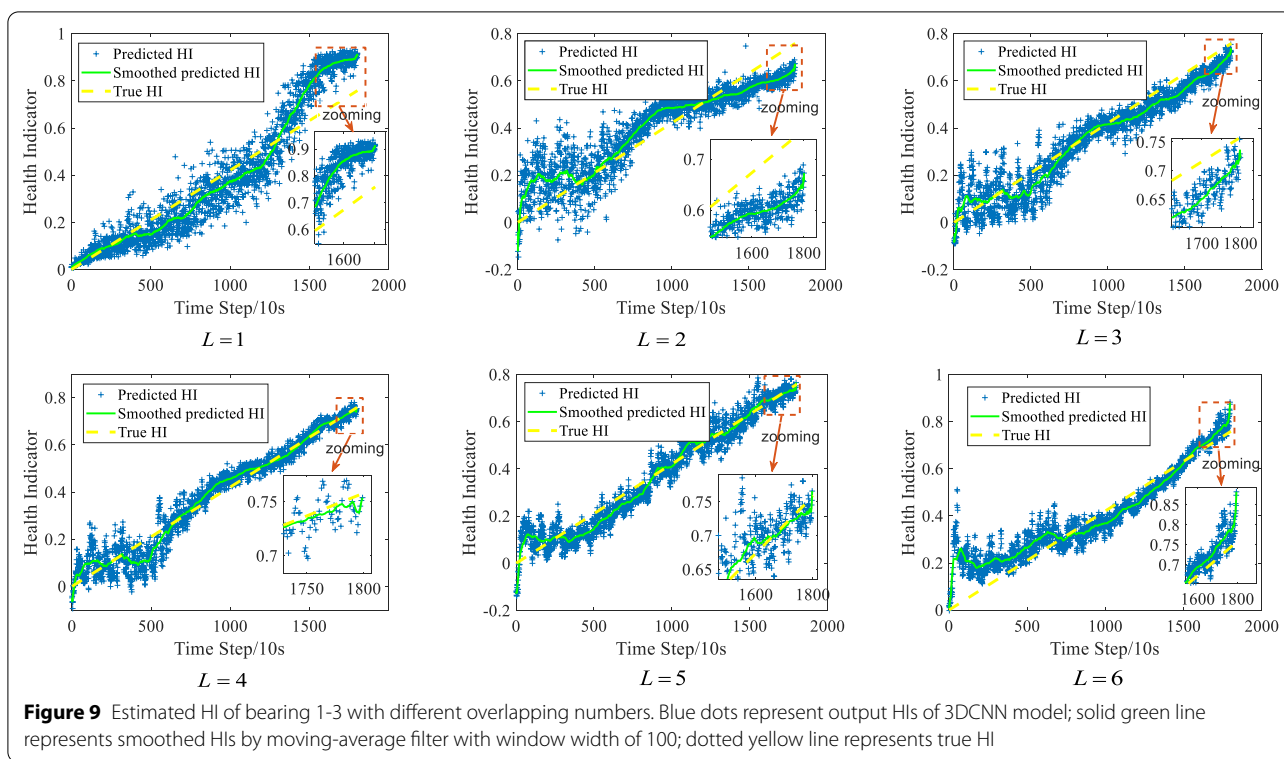
Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 10 of 15

Figure 9 Estimated HI of bearing 1-3 with different overlapping numbers. Blue dots represent output HIs of 3DCNN model; solid green line represents smoothed HIs by moving-average filter with window width of 100; dotted yellow line represents true HI

Table 3 RUL prediction performance of bearing 1-3 based on different TFR generation methods

TFR method	RMSE	Spearman coefficient	Sign trend factor
CWT	0.0523	0.8732	0.9980
WVD	0.1913	0.6596	0.9552
PWVD	0.1355	0.9422	1.0000

Table 4 Time consumption of different TFR generation methods

TFR method	Initial	Middle	Terminus	Mean
CWT	0.071	0.076	0.091	0.079
WVD	0.257	0.269	0.286	0.271
PWVD	0.158	0.160	0.170	0.163

vibration signals, and its parameter determination and use methods are relatively mature; therefore, we selected it as the TFR generation method in this study. An experiment was performed to compare and analyze the effects of typical time–frequency analysis methods, i.e., WVD and PWVD, on the model performance when used in the established model. Table 3 presents the RUL prediction performance of bearings 1-3 based on different TFR generation methods Among them, the Hamming window function with a window length of 25 was selected for the WVD and PWVD methods. Overall, the TFRs generated by the CWT method yielded the smallest prediction error when used for HI construction, which is conducive to obtaining higher-precision RUL estimation performance. The prediction performance obtained by the PWVD-based TFRs indicated a clear advantage in terms of the monotony of the constructed HIs. The model performance corresponding to the TFRs generated by the

WVD method was not comparable to that of the PWVD method, as it demonstrated a certain inhibition effect on the cross-interference terms. The high-quality TFR method is beneficial for accurately reflecting the degradation state of bearings and improving RUL prediction accuracy.

Table 4 shows the time consumption of different TFR generation methods based on the initial, middle, and terminus intervals of the time–frequency analysis for bearing 1-1. As shown, the CWT method is advantageous in terms of computational efficiency compared with the other methods. For RUL prediction tasks with a large sample size, high computational efficiency is extremely important. Considering the results show in Tables 3 and 4, we used the CWT method for the time–frequency analysis of bearing vibration signals.
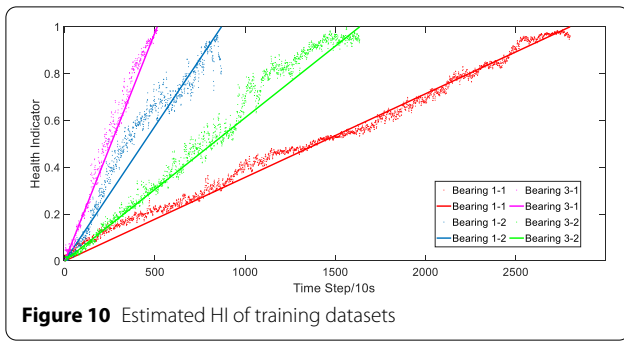
Wang *et al. Chin. J. Mech. Eng.* (2021) 34:62

Page 11 of 15

Figure 10 Estimated HI of training datasets

4.3.4 Superiority of GPR-Based RUL Prediction

Figure 10 shows the estimated HIs of the training bearings. The solid lines and dots correspond to the actual and estimated HIs, respectively. The *Y* and *X*-axes correspond to the HI and time, respectively. The HIs estimated using the proposed model can be used to track the actual degradation trends of the four training bearings. The test data used in the experiment were obtained from the terminate-before-failure bearings. The trained model was used to predict the HIs corresponding to the testing datasets.

The RUL can be obtained directly based on the definition of the HI. In Eq. (5), *t* refers to the current time, and *T* the run-to-failure time of the bearings. For the testing datasets, the unknown *T* can be calculated using the known HI. This is the same as using the RUL as the model output directly, which has been reported previously. However, the predicted HI at every moment should be nearly accurate. When the model was built, it was regarded as a regression problem rather than a classification problem, and the overall objective was the minimum MSE of the entire MSE. The RUL estimation result

will be extremely volatile if calculated directly using the discrete HI output value. As shown in Figure 9, the smoothed HI was easily affected by the individual points. Therefore, the GPR algorithm was used in this study for the RUL prediction of bearings. The algorithm considers not only a single HI, but also the mixed distribution form of all existing HIs and calculates the distribution function corresponding to an individual bearing. For the independent variable time *t*, we assume that $HI = f(t)$ as an unknown distribution form. For a specified value, the HI was modeled as a normal distribution. The mean vector of this joint normal distribution was assumed to be zero. The covariance matrix was defined as the similarity measure of each element corresponding to two HIs and was determined based on the squared exponential kernel. In addition, using the Bayesian method, not only the distribution of each point can be obtained via the GPR method, but also the unknown function is inferred.

Using bearings 1-3 and 3-3 as examples, Figure 11 shows the estimation result and the corresponding confidence interval. The Gaussian maximum likelihood estimation function was used as the optimal solution method, the superposition of a linear mean value and a constant mean value was used as the mean value function, and the Materniso function was used as the covariance function. The convergence of the model was achieved in 50 iterations. The prediction of the GPR model showed a 95% confidence interval. The mean of the confidence interval bound, namely the red line in the figure, represents the predicted HI value at an unknown time. Using $HI = 1$ as the failure threshold, the intersection points of the middle line of the GPR prediction, and the failure threshold as the failure point, the bearing RUL can be further calculated.
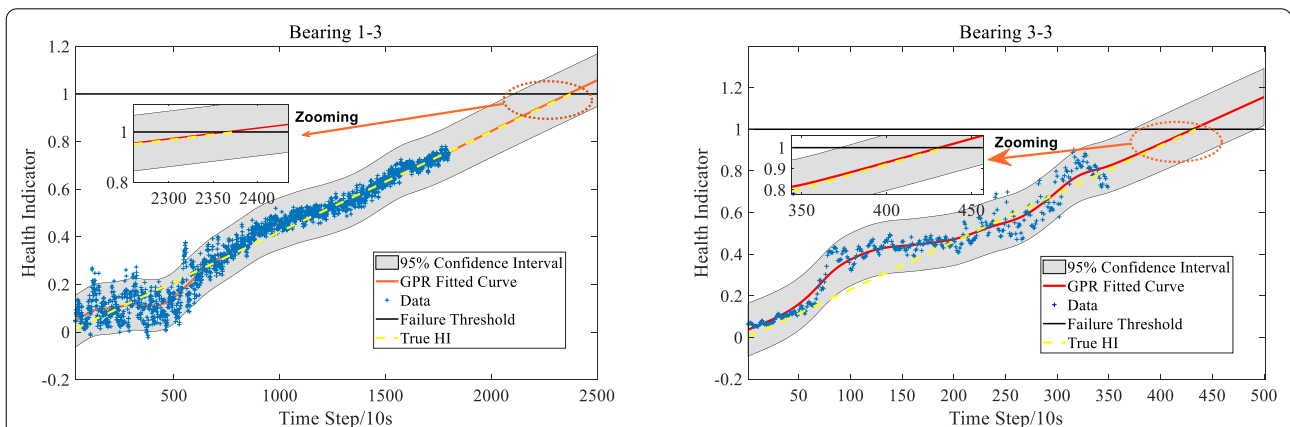

Figure 11 RUL estimation results of bearings 1-3 and 3-3 (Solid black line refers to failure threshold; blue dots represent output data of 3DCNN model; gray interval shows 95% confidence interval predicted by GPR, and solid red line shows mean of this confidence interval. Intersection of black and red lines represents failed node. Dotted yellow line represents true HI)

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 12 of 15

The GPR method is a Bayesian approach that can be used to model streaming data. Because of the advantages of the GPR method in uncertainty prediction, it was selected as a regression method for RUL estimation in this study. The difference between GPR and other regression algorithms is that in the latter, only the output value corresponding to the input data is expected to be obtained in general regression algorithms, such as linear, logistic, multiple, and Lasso regressions. By contrast, the GPR model obtains the distribution of the function based on the input data. The estimated distribution is advantageous when it is used to predict the RUL estimation uncertainty, which is difficult to realize using general regression methods.

As shown in Figure 11, the estimation ability of GPR for model errors is well demonstrated. The GPR method is beneficial for the RUL estimation of bearings in terms of two aspects: 1) It offers better anti-interference ability for the mutation points in HIs. Because GPR obtains a 95% confidence interval by calculating the probability distribution of existing data, abnormal mutation points are naturally excluded from the confidence interval, thereby reducing the effect of mutation points on the prediction result; 2) compared with point estimation, the result obtained is more stable. Using the mean of the confidence interval as the predicted value reduces the effect of data fluctuation on the results. Using bearing 3-3 as an example, a significant fluctuation remained in the HIs near the end of the test data, thereby increasing the prediction error significantly. The predicted value represented by the red line was similar to the true value represented by the yellow line. In the future, HIs can be predicted based on the distribution form of the existing data, and the result be consistent with the true value. The experimental results indicate that effectiveness of using the GPR model to predict the RUL of bearings based on the HIs constructed using the 3DCNN model.

4.4 Comparison with Other Related Studies

The proposed method was compared with four recent state-of-the-art methods: 1) the LSTM model proposed in Ref. [21], 2) the MSCNN model proposed in Ref. [22], 3) the DBN–DP model proposed in Ref. [18] based on the DBN and diffusion process; 4) the RNN–AM model proposed in Ref. [20] based on the RNN and attention mechanism; 5) the hybrid prognostics method based on systematically combining relevance support vector regressions, exponential degradation models, and the Fréchet distance proposed in Ref. [29]; and 6) an enhanced encoder–decoder method proposed in Ref. [17].

By defining $R\hat{U}L_i$ and $ActRUL_i$ as the ith estimated and actual RULs, respectively, the prediction error is expressed as

$$Er_i = 100 \times \frac{ActRUL_i - R\hat{U}L_i}{ActRUL_i}. \tag{19}$$

The RUL estimation score of the ith experiment is

$$A_i = \begin{cases} \exp\left(-\ln\left(0.5\right) \cdot \left(Er_i/5\right)\right), & Er_i \leq 0, \\ \exp\left(\ln\left(0.5\right) \cdot \left(Er_i/20\right)\right), & Er_i > 0. \end{cases} \tag{20}$$

The score of the entire model is expressed as

$$\text{Score} = \frac{1}{6}\sum_{i=1}^{6} A_i. \tag{21}$$

Table 5 shows the prediction performance of the five methods. The current time, actual RUL, predicted RUL, estimation error, standard deviation (SD) of the errors, and mean score are presented in the table. As shown, the prediction SD of the prediction model based on LSTM and the RNN was high, indicating the low robustness of the model. Both models use raw vibration signals as the input data of the network for analysis. As mentioned

Table 5 RUL prediction performance

Testing bearing	Actual RUL (s)	RUL (s) / (*Er* (%))						
		LSTM (2019)	MSCNN (2019)	DBN-DP (2019)	RNN-AM (2020)	Hybrid (2020)	Encoder–decoder (2021)	Proposed method
1–3	5730	400/(93.02)	4731/(17.43)	5600/(2.27)	5294/(7.61)	5440/(5.06)	3650/(36.30)	5630/(1.75)
1–4	339	−310/(191.18)	259/(10.67)	260/(23.53)	874/(−157.81)	260/(23.30)	2500/(−637.46)	350/(−2.94)
1–5	1610	−6490/(503.11)	3996/(−148.19)	2950/(−83.23)	2778/(−72.55)	1540/(4.35)	1570/(2.48)	1660/(−3.11)
1–6	1460	−6460/(542.47)	1744/(−19.45)	1163/(20.34)	1446/(0.96)	1450/(0.68)	1640/(−12.33)	1530/(−4.79)
1–7	7570	7310/(3.43)	6090/(19.55)	5834/(22.93)	1061/(85.98)	10790/(−42.53)	3210/(57.60)	7510/(0.79)
3–3	820	990/(−20.73)	1060/(−29.27)	544/(33.66)	803/(2.07)	850/(−3.66)	770/(6.10)	790/(3.66)
SD of *Er*/%		247.36	52.97	43.58	83.24	21.83	268.80	3.31
Score		0.16	0.30	0.44	0.45	0.62	0.39	0.77

above, it is difficult to extract complex fault-related information using raw signals. The CNN-based method uses TFRs as the network input and is more stable than LSTM and RNN models in prediction, as reflected by its lower SD of the error. However, the prediction accuracy of the CNN-based model is not ideal. Furthermore, the DBN–DP model considers the instability of the prediction results. After the DBN network, the DP method was added to model the HIs to improve the prediction stability. Therefore, the DBN–DP model yielded a lower SD of the prediction error compared with the three abovementioned models. The TFR-subsequence–3DCNN–GPR method achieved a significantly higher score of 0.77 than the other methods. In both operating conditions 1 and 3, the RUL estimation error was within 5%, which was significantly lower than those of the other methods. The results fully demonstrate the feasibility and superiority of the proposed prognostics approach. The highest prediction accuracy was due to the simultaneous consideration of the spatial characteristic information of the TFR and the temporal correlation of the temporal correlation of adjacent TFRs. The lowest SD of the prediction error was attributed to the GPR method since it can effectively model the HIs. The experimental results demonstrated the effectiveness of the proposed method.

5 Conclusions

To address the three challenges in existing data-driven bearing RUL estimation methods, namely, the management of TFRs, temporal dependency of degradation states, and prediction volatility and uncertainty, a novel TFR-subsequence–3DCNN–GPR prognostics approach was proposed herein. The proposed model uses novel TFR-subsequences as inputs for end-to-end estimation, i.e., a new attempt to provide more spatiotemporal information for a more accurate RUL estimation of bearings. Subsequently, the 3DCNN was leveraged to model the HIs of different adjacent TFR subsequences as the first attempt for the RUL estimation of bearings. The proposed model demonstrated superiority in the state tracking of bearings by extracting spatial TFR information and temporal interrelated information of data synchronously. Furthermore, the GPR model was applied to estimate the failure time and uncertainty bounds. By deriving the HI value of the unknown point, it can be directly used to estimate the RUL of bearings. Using the mean value of the distribution as the predicted HI value, the proposed model was more resistant to abnormal and fluctuating points than other prediction methods. This study provides insights into the comprehensive use of bearing vibration signals, particularly degradation-related spatiotemporal information in the data, to achieve a highly accurate RUL prediction with uncertainty quantification. A comparison with other state-of-the-art methods demonstrated the superiority of the proposed approach.

Authors' contributions
TW and FC were in charge of the whole trial; XW wrote the manuscript; AM, QH, WZ, and AL assisted with sampling and laboratory analyses. All authors read and approved the final manuscript.

Authors' Information
Xu Wang, born in 1992, is currently a Ph.D. candidate of mechanical engineering at *Department of Mechanical Engineering, Tsinghua University, China*. He is also co-cultured with aerospace science and technology at *High-Tech Research Institute of Xi'an, China*. He received his M.S. degree in ordnance science and technology from *High-Tech Research Institute of Xi'an, China*, in 2017. His research interests include condition monitoring and fault diagnosis of machinery.

Tianyang Wang, born in 1985, is currently an assistant professor of mechanical engineering at *Department of Mechanical Engineering, Tsinghua University, China*. He received the B.S. and Ph.D. degrees in mechanical engineering from *Beijing Jiaotong University, China*, in 2008 and 2015, respectively. His recent research interests include signal processing and fault diagnosis of rotating machinery.

Anbo Ming, born in 1986, is currently an assistant professor at *High-Tech Research Institute of Xi'an, China*. He received the B.S. degree and Ph.D. degree in aeronautical science and technology from *High-Tech Research Institute of Xi'an, China*, in 2007 and 2014, respectively. He was an assistant professor at *High-Tech Research Institute of Xi'an, China*. His current research interests include condition monitoring and fault diagnosis of mechanical equipment.

Qinkai Han, born in 1982, is currently an assistant professor of mechanical engineering at *Department of Mechanical Engineering, Tsinghua University, China*. He received the B.S. and Ph.D. degrees in aviation propulsion engineering from *Beihang University, China*, in 2004 and 2010, respectively. His research interests include rotor dynamic analysis, machinery condition monitoring and fault diagnosis, and the development of self-powered system for intelligent fault diagnosis.

Fulei Chu, born in 1959, is currently a professor of mechanical engineering at *Department of Mechanical Engineering, Tsinghua University, China*. He received the Ph.D. degree in mechanical engineering from *Southampton University* in 1993. He is the Vice President of the *Chinese Society for Vibration Engineering (CSVE)*, the Executive Director of the Fault Diagnosis Chapter in CSVE, and a member of the Rotor Dynamics Chapter in the *International Federation for the Promotion of Mechanism and Machine Science (IFToMM)*. His research interests include rotating machinery dynamics, machine fault diagnosis, nonlinear vibration, and vibration control.

Wei Zhang, born in 1963, is currently a professor at *High-Tech Research Institute of Xi'an, China*. He received the B.S. degree in power machines from the *High-Tech Research Institute of Xi'an, China*, in 1985, the M.S. degree in mechanics from *Xi'an Jiaotong University, China*, in 1995, and the Ph.D. degree in ordnance science and technology from the *High-Tech Research Institute of Xi'an, China* in 2003. His research interests include nondestructive testing, condition monitoring, and fault diagnosis of machinery.

Aihua Li, born in 1966, is currently a professor at *High-Tech Research Institute of Xi'an, China*. His research interests include condition aluminum air battery development, computer vision, and fault diagnosis of machinery.

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 14 of 15

Author Details
[1] State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China. [2] High-Tech Research Institute of Xi'an, Xi'an 710025, China.

References
[1] M F Li, T Y Wang, F L Chu, et al. Component matching Chirplet transform via frequency-dependent chirp rate for wind turbine planetary gearbox fault diagnostics under variable speed condition. *Mechanical Systems and Signal Processing*, 2021, 161: 107997.
[2] T Han, C Liu, L J Wu, et al. An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems. *Mechanical Systems and Signal Processing*, 2019, 117: 170-187.
[3] F B Zhang, J F Huang, F L Chu, et al. Mechanism and method for the full-scale quantitative diagnosis of ball bearings with an inner race fault. *Journal of Sound and Vibration*, 2020, 488: 115641.
[4] J Lee, F J Wu, W Y Zhao, et al. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 2014, 42: 314-334.
[5] Y F Li, M J Zuo, K Feng, et al. Detection of bearing faults using a novel adaptive morphological update lifting wavelet. *Chinese Journal of Mechanical Engineering*, 2017, 30: 1305-1313.
[6] Y G Lei, N P Li, L Guo, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 2018, 104: 799-834.
[7] K S Singleton, E G Strangas, S Aviyente. The use of bearing currents and vibrations in lifetime estimation of bearings. *IEEE Transactions on Industrial Informatics*, 2017, 13: 1301-1309.
[8] S Kerst, B Shyrokau, E Holweg. A model-based approach for the estimation of bearing forces and moments using outer ring deformation. *IEEE Transactions on Industrial Electronics*, 2020, 67: 461-470.
[9] M Cerrada, R V Sánchez, C Li, et al. A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 2018, 99: 169-196.
[10] T S Liu, K P Zhu, L C Zeng. Diagnosis and prognosis of degradation process via hidden semi-Markov model. *IEEE/ASME Transactions on Mechatronics*, 2018, 23: 1456-1466.
[11] L L Cui, X Wang, Y G Xu, et al. A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing. *Measurement*, 2019, 135: 678-684.
[12] X S Si, W B Wang, C H Hu, et al. A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation. *Mechanical Systems and Signal Processing*, 2013, 35: 219-237.
[13] Y G Lei, N P Li, S Gontarz, et al. A model-based method for remaining useful life prediction of machinery. *IEEE Transaction on Reliability*, 2016, 65: 1314-1326.
[14] D Wang, K L Tsui, Q Miao. Prognostics and health management: A review of vibration based bearing and gear health indicators. *IEEE Access*, 2017, 6: 665-676.
[15] D Z Zhao, L Y Li, W D Cheng, et al. Generalized demodulation transform for bearing fault diagnosis under nonstationary conditions and gear noise interferences. *Chinese Journal of Mechanical Engineering*, 2019, 32: 7.
[16] R N Liu, B Y Yang, E Zio, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 2018, 108: 33-47.
[17] L Liu, X Song, K Chen, et al. An enhanced encoder–decoder framework for bearing remaining useful life prediction. *Measurement*, 2021, 170: 108753.
[18] C H Hu, H Pei, X S Si, et al. A prognostic model based on DBN and diffusion process for degrading bearing. *IEEE Transactions on Industrial Electronics*, 2020, 67 (10): 8767-8777.
[19] H Wang, M J Peng, Z Miao, et al. Remaining useful life prediction techniques for electric valves based on convolution auto encoder and long short term memory. *ISA Transactions*, 2021, 108: 3333-3342.
[20] Y H Chen, G L Peng, Z Y Zhu, et al. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Applied Soft Computing Journal*, 2020, 86: 105919.
[21] B Zhang, S H Zhang, W H Li. Bearing performance degradation assessment using long short-term memory recurrent network. *Computers in Industry*, 2019, 106: 14-29.
[22] J Zhu, N Chen, W Peng. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, 2019, 66: 3208-3216.
[23] M F Li, T Y Wang, F L Chu, et al. Scaling-basis Chirplet transform. *IEEE Transactions on Industrial Electronics*, 2021, doi: https://doi.org/10.1109/TIE.2020.3013537.
[24] H Huang, N Baddour, M Liang. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction. *Journal of Sound and Vibration*, 2018, 414: 43-60.
[25] R Klein, E Masad, E Rudyk, et al. Bearing diagnostics using image processing methods. *Mechanical Systems and Signal Processing*, 2014, 45: 105-113.
[26] J J Shi, M Liang, D S Necsulescu, et al. Generalized stepwise demodulation transform and synchrosqueezing for time – frequency analysis and bearing fault diagnosis. *Journal of Sound and Vibration*, 2016, 368: 202-222.
[27] B Wang, Y G Lei, T Yan, et al. Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery. *Neurocomputing*, 2020, 379: 117-129.
[28] Z S Chen, X T Tu, Y Hu, et al. Real-time bearing remaining useful life estimation based on the frozen convolutional and activated memory neural network. *IEEE Access*, 7, 2019: 96583-96593.
[29] B Wang, Y G Lei, N P Li, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 2020, 69(1): 401-412.
[30] D Wang, P W Tse. Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method. *Mechanical Systems and Signal Processing*, 2015, 56: 213-229.
[31] K P Zhu, T S Liu. Online tool wear monitoring via hidden semi-Markov model with dependent durations. *IEEE Transactions on Industrial Informatics*, 2018, 14: 69-78.
[32] S W Ji, W Xu, M Yang, et al. 3D Convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35: 221-231.
[33] Ł Jedliński, J Jonak. Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. *Applied Soft Computing Journal*, 2015, 30: 636-641.
[34] S Qian, H Liu, C Liu, et al. Adaptive activation functions in convolutional neural networks. *Neurocomputing*, 2018, 272: 204-212.
[35] F Kang, S X Han, S Salgado, et al. System probabilistic stability analysis of soil slopes using Gaussian process regression with Latin hypercube sampling. *Computers and Geotechnics*, 2015, 63: 13-25.
[36] L L Kang, R S Chen, N X Xiong, et al. Selecting hyper-parameters of Gaussian process regression based on non-inertial particle swarm optimization in Internet of Things. *IEEE Access*, 7, 2019: 59504-59513.

Wang *et al. Chin. J. Mech. Eng.*   (2021) 34:62

Page 15 of 15

[37] P Nectoux, R Gouriveau, K Medjaher, et al. PRONOSTIA : An experimental platform for bearings accelerated degradation tests. *PHM'12*, United States: Denver, Colorado, 2012, 1–8.

[38] T P Le. Use of the Morlet mother wavelet in the frequency-scale domain decomposition technique for the modal identification of ambient vibration responses. *Mechanical Systems and Signal Processing*, 2017, 95, 488-505.

[39] M Mastyło. Bilinear interpolation theorems and applications. *Journal of Functional Analysis*, 2013, 265: 185-207.

[40] J J Wang, J X Yan, C Li, et al. Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction. *Computers in Industry*, 2019, 111: 1-14.

[41] K J Ashok, S Abirami. Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method. *Information Sciences*, 2018, 460-461: 23-41.

[42] J Coble, J W Hines. Identifying optimal prognostic parameters from data: A genetic algorithms approach. *Annual Conference of the Prognostics and Health Management Society, San Diego*, CA, 2009, 1–11.