 doi:10.20944/preprints202111.0377.v1

Article

Rolling Element Bearing Fault Time Series Prediction Using Optimized MCKD-LSTM Model

Leilei Ma[1], Hong Jiang[1]*, Tongwei Ma[1]*, Xiangfeng Zhang[1]*, Lei Xia[1], Huailing Zeng[2], Yifeng Jiang[2] and Yong Shen[2]*

[1] College of Mechanical Engineering, Xinjiang University, Urumqi 830047, China; 703938263@qq.com(L.M.); 754158555@qq.com(L.X.)
[2] Changsha Institute of Mining Research Co., Ltd., Changsha, 410012，China; 690916572@qq.com(H.Z.); 858872472@qq.com(Y.J.)
* Correspondence: onlyxjjh@xju.edu.cn(H.J.); mtw2772697346@163.com(T.M.); 15276606076@163.com(X.Z.) shenyongvincent@163.com(Y.S.)

Abstract: This paper realizes early bearing fault warning through bearing fault time series prediction, and proposes a bearing fault time series prediction model based on optimized maximum correlation kurtosis deconvolution (MCKD) and long short-term memory (LSTM) recurrent neural network to ensure bearings operation reliability. The model is based on lifecycle vibration signal of the bearing, to begin, the cuckoo search (CS) is utilized to optimize the parameter filter length L and deconvolution period T of MCKD, taking into account the influence and periodicity of the bearing time series, the fault impact component of the optimized MCKD deconvolution time series is improved. Then select the LSTM learning rate α depending on deconvolution time series. Finally, the dataset obtained through various preprocessing approaches are used to train and predict the LSTM model. The average prediction accuracy of the optimized MCKD-LSTM model is 26 percent higher than that of the original time series, proving the efficiency of this method, and the prediction results track the real fault data well, according to the XI'AN JIAOTONG University XJTU-SY bearing dataset.

Keywords: deep learning; time series prediction; long short-term memory; recurrent neural network; maximum correlation kurtosis deconvolution; cuckoo search.

1. Introduction

Rolling element bearing is called "industrial joint", which is widely used in the fields of transmission and hoisting, wind power generation, aerospace and other areas [1-3]. As the core component of the equipment, the frequency of bearing fault is extremely common [4-6], bearing performance has a direct impact on the reliability and safety of heavy machinery, as a result, reliable bearing fault time series prediction is extremely important for industrial production safety [7-9].

When bearing is used as the research object, the bearing outer ring performance monitoring data contains information about bearing reliability [10], but the main problem with relying on bearing performance monitoring data to complete the bearing fault time series prediction is figuring out how to create the bearing fault time series prediction model. Moreover, the bearing fault time series is created by the convolution of vibration signals and different noise signals in the signal transmission process, which will impact the accuracy of the prediction model after training, as a result, the first part of this study looks at how to perform the preprocessing of the original time series. For the research on signal preprocessing, to solve the filtering of nonlinear vibration signals, Dong et al. [11] presented a non iterative denoising method combining spectral wavelet transform and detrended fluctuation analysis. Yan et al. [12] explored the discrete convolution wavelet transform (DCWT) for signal decomposition and reconstruction in signal processing, when the signal changes swiftly, however, the choice of wavelet basis function is severely

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

limited [13]. Many adaptive signal preprocessing approaches have been presented based on wavelet transform denoising, to demodulate the original vibration signal, Vikas et al. [14] employed the variational mode decomposition (VMD) to handle the multi-component modulated non-stationary vibration signal of the transmission. To discover the modal properties of engineering structures, Bagheri et al. [15] proposed a dynamic response decomposition based on variational mode decomposition. VMD was utilized by Mausam et al. [16] to partition the histogram of the input image into many band limited modes, and then the histogram was reconstructed using meaningful modes. Zhang et al. [17] investigated the fractal properties of rolling element bearing vibration signals and developed a bearing defect assessment and diagnosis. Furthermore, the decomposition mode parameter K and penalty coefficient η must be established in order to solve the differing parameter values have a significant impact on the decomposition effect. To calculate the ideal K for water pipe leakage location, Li et al. [18] used the correlation between energy loss coefficient and nearby blade disks as an evaluation index. Zhao et al. [19] employed the envelope nesting approach to direct the potential center frequency, and then used permutation entropy and orthogonality to calculate the K. To alleviate the mode mixing of complicated vibration signals, Zhao et al. [20] presented the based on single-objective salp swarm algorithm to optimize the penalty coefficient η of the VMD. Feng et al. [21] used the whale optimization algorithm (WOA) to optimize VMD parameters in order to achieve adaptive decomposition and noise reduction of vibration signals. The decomposition parameters of VMD, on the other hand, must be set according to the properties of signal. Mode over-decomposition and under decomposition are caused by incorrect parameter selection [22]. McDonald et al. [23] presented a new signal preprocessing method called maximum correlated kurtosis deconvolution (MCKD), which is ideal for processing early bearing fault signals with low signal-to-noise ratio and periodic impact characteristics [24]. To complete composite fault diagnosis, Hong et al. [25] used adaptive MCKD to decouple fault information and noise reduction signal. Zhang et al. [26] suggested a signal noise reduction method based on the teager energy operator and the MCKD. Many scholars have optimised the filter length L and the order of shift M in MCKD on this basis. Lyu et al. [27] optimized the filter length and deconvolution period of MCKD for composite fault diagnosis of gear tooth wear and bearing outer ring fault using a quantum genetic algorithm (QGA). To complete the bearing composite fault diagnosis, Miao et al. [28] used the autocorrelation of the envelope signal to estimate the prior period T. To make MCKD obtain the best noise reduction performance, Yang et al. [29] employed permutation entropy as the measurement index for filter length L selection.

In order to develop a bearing fault time series prediction model, Pan et al. [30] calculates the upper and lower boundaries of unknown elevation on a terrain profile using a double multiplicative neuron (DMN) model and an modified particle swarm optimization (MPSO) technique. For time series prediction, Raubitzek et al. [31] presented a fractal interpolation method. For long term time series prediction, Liu et al. [32] proposed dual-stage two-phase DSTP-based RNN (DSTP-RNN) and DSTP-RNN-Ⅱ. Human fall risk is predicted using the depth neural network model described by Savadkoohi et al. [33]. By input high-level abstract features into an LSTM network, Zhang et al. [34] proposed the CEEMD-PCA-LSTM hybrid prediction model to complete time series prediction. The 1d-CNN model was proposed by Che et al. [35] for regression analysis of time series samples in order to establish a performance deterioration model, finally, bidirectional long short memory (Bi-LSTM) is employed to predict performance decline over time. Dempster-Shafer regression technology was proposed by Niu et al. [36] to perform time series prediction problems.

Based on the previous study, this paper takes advantage of the MCKD strong noise reduction effect in periodic signals to denoise the bearing fault time series and acquire the deconvolution time series. After that, deconvolution time series are used to train the long short-term memory recurrent neural networks, and the optimized MCKD-LSTM prediction model is created to predict the bearing fault time series.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

2. Maximum Correlated Kurtosis Deconvolution

Mcdonald et al. proposed maximum correlated kurtosis deconvolution (MCKD) in 2012 [23]. It successfully applies in gear flaking fault diagnosis by taking into account the impact and periodic characteristics of fault information. This algorithm assumes that y represents the impulse signal, h represents the response of y after transit through the transmission path, and x represents the signal convoluted from various signals on the transmission path, which represents the process as shown in formula (1):

$$x = h * y \tag{1}$$

The essence of MCKD is to find a FIR filter to solve the input signal y through the output signal x, i.e.

$$y = f * x = \sum_{k=1}^{L} f_k x_{n-k+1} \tag{2}$$

Where $f = [f_1, f_2, \ldots, f_L]^T$ is the filter factor of length L.

MCKD takes maximum correlation kurtosis as its evaluation criterion and calculates as follows:

$$O(CK_M(T)) = \frac{\sum_{n=1}^{N}(\prod_{m=0}^{M} y(n-mT))^2}{(\sum_{n=1}^{N} y_n^{\,2})^{M+1}} \tag{3}$$

In order to obtain the optimal inverse filter coefficient f, the first derivative of the objective function is zero, such as formula (4):

$$\frac{d}{df_k} CK_M(T) = 0, \quad k = 1, 2, \cdots, L \tag{4}$$

Therefore, the optimum filter coefficient can be obtained.

$$f = \frac{\left\| y^2 \right\|}{2\left\| \beta \right\|^2} (X_0 X_0^T)^{-1} \sum_{m=0}^{M} (X_{mT} \alpha_m) \tag{5}$$

Where:

$$\alpha_0 = [y_1 y_{1-T}^2 \;\; y_2 y_{2-T}^2 \;\; \cdots \;\; y_N y_{N-T}^2]^T \tag{6}$$

$$\alpha_1 = [y_{1-T} y_1^2 \;\; y_{2-T} y_2^2 \;\; \cdots \;\; y_{N-T} y_N^2]^T \tag{7}$$

$$\beta = [y_1 y_{1-T} \;\; y_2 y_{2-T} \;\; \cdots \;\; y_N y_{N-T}]^T \tag{8}$$

$$X_r = \begin{bmatrix} x_{1-r} & x_{2-r} & x_{3-r} & \cdots & x_{N-r} \\ 0 & x_{1-r} & x_{1-r} & \cdots & x_{N-1-r} \\ 0 & 0 & x_{1-r} & \cdots & x_{N-2-r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x_{N-L-r-1} \end{bmatrix}_{L \times N} \tag{9}$$

The specific steps to realize MCKD are as follows:

1) Determine filter length L, The order of shift M and period T of impact signal.
2) Calculate $X_0 X_T^0$ and X_T^m of the original signal $x(n)$.
3) Obtain filtered output signal $y(n)$.
4) Calculate α_m and β according to $y(n)$.
5) Update filter coefficient f.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

If the signal f before and after filtering conforms to $\Delta CK_M(T) < \varepsilon$, stop the iteration and jump back to step (3).

The deconvolution signal y of the actual acquisition signal x can be obtained by substituting the obtained inverse filter coefficients (2).

3. Cuckoo Search

The cuckoo search [37] is a new heuristic search algorithm that integrates the Lévy flights theory with the parasitic behavior of cuckoos. It has the characteristics of few parameters and fast convergence. Three ideal rules are assumed by the cuckoo algorithm.

1) Each cuckoo lays only one egg at a time and places the eggs in a randomly selected nest, which is also known as a host nest.

2) The parasitic nest with the highest quality eggs will be retained for the next generation.

3) The number of possible nests is fixed, and the chance of discovering host eggs in a nest is p.

When the host bird discovers the host egg, it either throws it out or abandons the nest to establish a new one in a new site.

After randomly generating n nest placements, the Lévy flights search strategy illustrated creates a new nest location using formula (10):

$$X_i^{t+1} = X_i^t + \alpha_0 \frac{\phi \times \mu}{|\nu|^{1/\beta}} (X_i^t - X_b^t) \tag{10}$$

Where:

$$\phi = \left(\frac{\Gamma(1+\beta) \times \sin(\pi \times \frac{\beta}{2})}{\Gamma(\frac{1+\beta}{2} \times \beta \times 2^{\frac{\beta-1}{2}})} \right)^{1/\beta} \tag{11}$$

Where, X_{t+}^i and X_t^i signify the ith cuckoo's nest site in the T and T+1 generations, respectively, X_t^b is the optimal nest location currently searched, the value α_0 is used to change the step size, α_0=0.01 in this paper, μ and ν are random values generated using the normal distribution, H set to 0.5 be the default, Γ is the normal gamma function.

The nest with the higher fitness value is kept when the new nest location is found using the Lévy flight search strategy, then, based on the discovery probability p, a portion of the nest positions are eliminated, and a new nest position is constructed using the preferred random walk search strategy shown in formula (12):

$$X_i^{t+1} = X_i^t + r(X_j^t - X_k^t) \tag{12}$$

Where, r is a random number between 0 and 1, and X_t^j and X_t^k are two candidate solutions chosen at random from the current population.

4. Long Short-term Memory Recurrent Neural Network

The gating mechanism is employed in the long short-term memory (LSTM) recurrent neural network [38], which is frequently used to process time series signals, and specific formula of LSTM is as follows:

4.1. Forward Calculation Method of LSTM

For a given time series signal x= (x_1, x_2, …, x_t), and the hidden layer sequence h_{t-1}= (h_1, h_2, …, h_{t-1}), the candidate state value $\tilde{c}_t$, input gate value i_t, forgetting gate value f_t, output gate value o_t, memory cell value c_t, hidden layer sequence h_t, and output sequence y_t=(y_1, y_2, … , y_t) at time t can be determined using the conventional LSTM model (as shown in Figure 1), which is:

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{13}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (14)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (15)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (16)$$

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \quad (17)$$

$$h_t = o_t \odot \tanh(c_t) \quad (18)$$

$$y_t = W_y h_t + b_y \quad (19)$$

Where, W and U are weight matrices for time series (for example, W_c represents the candidate state weight matrix from the input layer to the hidden layer, and U_c represents the candidate state weight matrix from the hidden layer at time t-1 to the hidden layer at time t), the offset vector is b. (for example, b_c represents the offset vector of the candidate state from the input layer to the hidden layer), σ is the sigmoid function, meanwhile, the sigmoid function is the activation function of gating unit, the tanh function is the activation function of candidate state, and the time is represented by the subscript t.

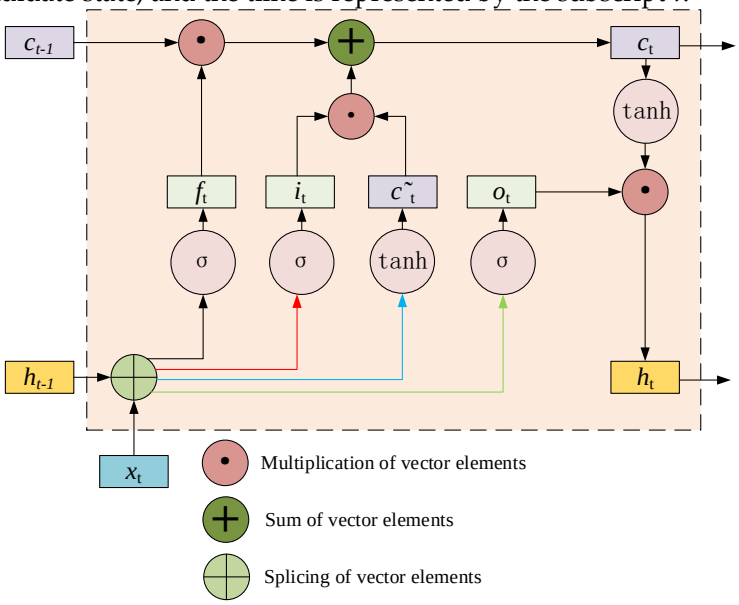


Figure 1. Structure diagram of LSTM.

4.2. Reverse Computation Method of LSTM

The three steps of the LSTM training algorithm are as follows:

1) The output value of each neuron is calculated forward $f(y)=f(w^T x)$

2) The cost function is the mean square deviation function J, and the error term δ_j value of each neuron is calculated inversely, as follows:

$$J = \frac{1}{2N}\sum_{i=1}^{N}\left\|y_i - f(y)\right\|^2 \quad (20)$$

$$\delta_t^T = (y_i - f(y)) \quad (21)$$

$$\delta_{o,t}^T = \delta_t^T \tanh(c_t) o_t (1 - o_t) \quad (22)$$

$$\delta_{f,t}^T = \delta_t^T o_t (1 - \tanh(c_t)^2) c_{t-1} f_t (1 - f_t) \quad (23)$$

$$\delta_{f,t}^T = \delta_t^T o_t (1 - \tanh(c_t)^2) \tilde{c}_t i_t (1 - i_t) \quad (24)$$

$$\delta_{f,t}^T = \delta_t^T o_t (1 - \tanh(c_t)^2) i_t f_t (1 - \tilde{c}_t^2) \quad (25)$$

3) Reverse error gradient calculation, i.e.

$$\Delta E = \delta_j f'(y) \quad (26)$$

4) Determine the Δw weight difference, i.e.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

$$\Delta w = \alpha \bullet \Delta E \bullet x \tag{27}$$

Where, α is the learning rate, which in this case is 0.01 in this study.

4.3. Metrics

The mean square error (MSE) is used as the measuring standard in this research to assess the accuracy of the prediction model. MSE may be calculated using the following formula:

$$MSE = \frac{1}{N}\sum_{\hat{t}=1}^{N}(\hat{y}_i - y_i)^2 \tag{28}$$

Where, $\hat{y}_i$ and y_i are the fault time series test and model output values at time $\hat{t}$, respectively, and N is the number of data points, by calculating the MSE of the training and test sets, this research quantitatively analyzes the fitting and prediction accuracy of model.

5. Parameter Optimization Based on Cuckoo Search

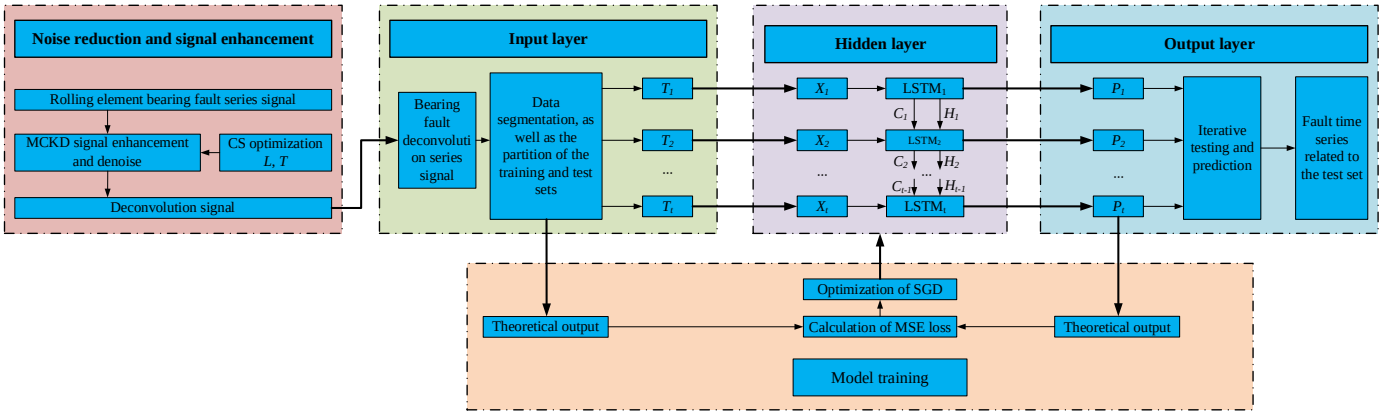
When utilizing the MCKD to minimize the noise in a bearing series signal, the filter length L and deconvolution period T must be set first. The parameter combination [L, T] of the MCKD can be searched by the CS to perform parameter adaptive screening, taking into account the interaction between affecting parameters.

When using the CS to optimize parameters, it is important to choose the right fitness function based on the signal characteristics and take into account the periodic impact signal of the bearing signal, which can differ from the noise signal. In reference [39] proposes a dimensionless crest factor of envelope spectrum (E_c) index that takes into consideration the periodic properties of fault information in vibration signals. The mathematical expression of the index E_c, assuming the signal envelope spectrum amplitude $X(j)$ (j = 1, 2,…, M), and the index E_c is as follows:

$$E_c = \frac{e_{\max}}{e_{rms}} \tag{29}$$

Where, $e_{\max}$ is the highest value in the range [$n \times f_r$, $f_s/2$] of the envelope signal obtained after Hilbert Demodulation, f_r is the bearing signal's frequency conversion, and f_s is the sampling frequency. The effective value is e_{rms}, which is defined as the effective value of signal following Hilbert demodulation. In this research, n = 2 is chosen to avoid the influence of f_r on E_c.

The envelope spectrum peak factor E_c of the envelope signal acquired by Hilbert demodulation is determined using the MCKD operation on the fault signal at any nest X_i location (i.e. the optimum parameters), and the E_c is the fitness value of the bird nest. The envelope spectrum peak factor E_c is significant when periodic impact occurs in the decomposition results, and the decomposition effect is optimal, on the other hand, if the envelope spectrum peak factor E_c is relatively small, the decomposition effect is not well. As a result, the optimization object is set at the greatest E_c. The flow of the fault diagnosis method proposed in this paper is shown in Figure 2.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

Figure 2. Prediction process of rolling element bearing fault time series based on optimized MCKD-LSTM.

6. Experimental Signal Analysis

The experimental data adopts the LDK UER204 rolling element bearings dataset of XJTU-SY bearing dataset [40] of XI'AN JIAOTONG UNIVERSITY, the bearing accelerated life test bed and outer ring crack of bearing are shown in Figure 3, the two unidirectional acceleration sensor models in the vertical and horizontal directions, PCB 352C33, collect the vibration signals through a DT9837 portable dynamic signals collector, the test sampling frequency is 25.6kHz, the sampling interval is 1min, the number of sampling points is 32769, and the sampling time is 1.28s, the horizontal vibration signals in dataset bearing1_1 are selected for analysis.

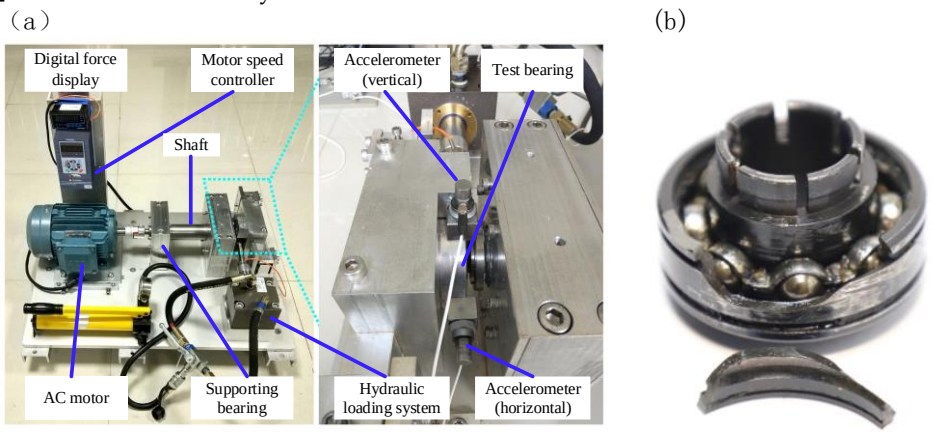
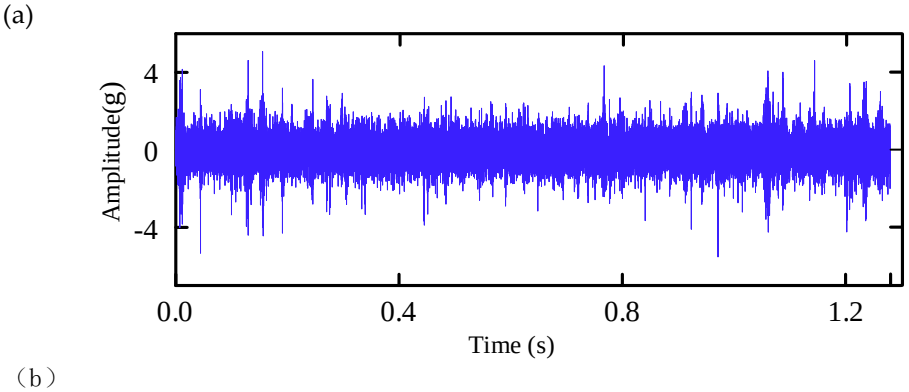


Figure 3. Bearing accelerated life test bed and outer ring crack of bearing. (a) Bearing accelerated life test bed. (b) Outer ring crack of bearing.

6.1. Data Preprocessing

The vibration signal of the 50th series in the horizontal direction of Bearing1_1 is chosen for analysis in order to anticipate the bearing fault time series, the time-frequency domain diagram of a vibration signal is shown in Figures 4 (a) and (b). In the temporal domain, there are many impact components to consider, and there is no rule to follow, the frequency spectrum shows the frequency conversion 34.38Hz and its frequency doubling components, and in the high frequency band, there are several resonance frequency bands, the frequency components are complex, and the bearing outer ring fault has no distinct frequency, as a result, denoising the original time series is required to retrieve the time series including more impact information.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

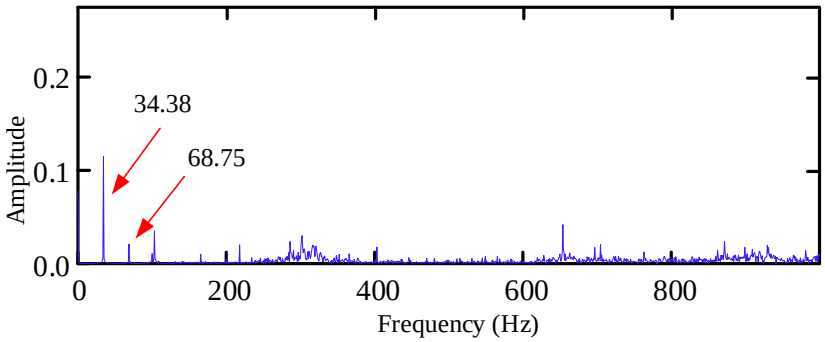
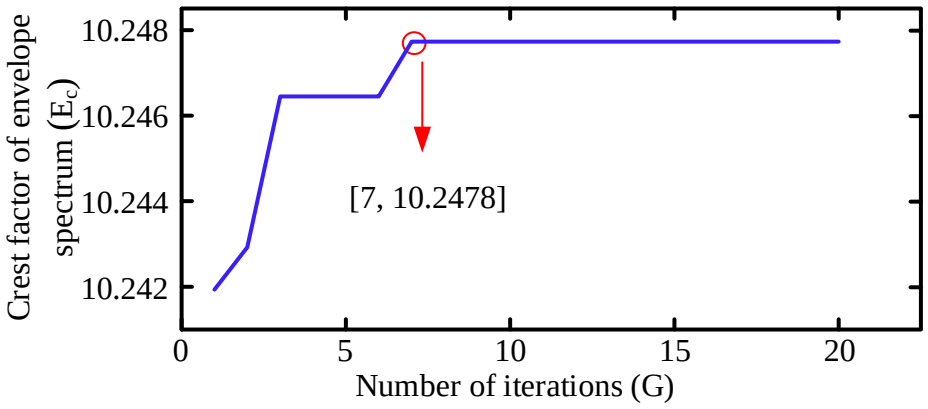


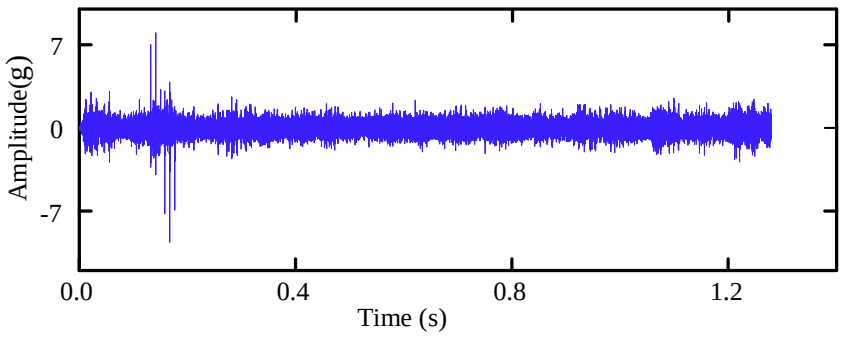
Figure 4. Time frequency domain diagram of original time series. (a) Time domain diagram of original time series. (b) Spectrum of original time series.

The MCKD is used to preprocess the original time series, with the order of shift M set to 1 and the iteration termination times G set to 20, the CS is used to optimize the filter length L and deconvolution period T in MCKD, the parameters of CS are set as follows: the dimension of solution D set to 2, the population size N set to 15, the host bird with a probability P set to 0.1, the upper and lower bounds are searched based on $L>2f_s/f_c$ and $T=f_s/f_c$ [41], where f_s is the sampling frequency and f_c is the characteristic frequency, and setting the optimization range as L=[100, 1500] and T=[50, 1000]. Figure 5(a) depicts the results, the peak factor of the local maximum envelope spectrum converges to 10.2478 at the 7th iteration, and the optimization parameter combination [L, T] corresponding to the peak factor of the local maximum envelope spectrum is [600, 235]. Denoise the original time series signal using MCKD parameters to obtain the deconvolution series signal and envelope spectrum, as illustrated in Figure 5 (b) and (c), it can be seen that the impact component intensity of the deconvolution series signal is increased in the time domain, the noise interference component is greatly reduced, and the frequency conversion component 34.38Hz, 108.6Hz, and its frequency doubling emerge in the envelope spectrum. This frequency is close to the theoretical value of the bearing outer ring crack characteristic frequency of 107.91Hz, resulting in a significant reduction in noise.

(a)



(b)



(c)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

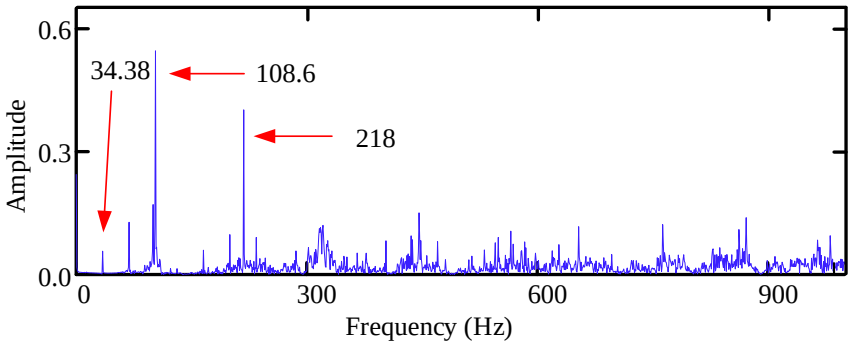


Figure 5. Time frequency diagram of deconvolution time series. (a) Variation curves of different E_c indexes with iteration times. (b) Time domain diagram of deconvolution series. (c) Spectrum of deconvolution time series.

Then, the deconvolution time series are taken as one- dimensional time series to train the bearing fault time series prediction model, all at once, one-dimensional vibration signals are selected based on the 50th original time series every ten series, six groups of time series are taken as the training set, and seven groups of time series are taken as the test set from 102 series every two sequences, the optimized MCKD is then utilized to denoise, after establishing the whole data set, it is input the LSTM model, which is used to train the model and predict bearing fault time series.

To demonstrate the benefits of optimized MCKD-LSTM in bearing fault time series prediction, this study denoises the original fault series using EMD and optimized VMD, then completes the LSTM model training and bearing fault time series prediction comparison with it. Because the impact signal will be included in the bearing lifecycle signal, the fault impact information of the bearing will be contained in some IMF components after the signal is processed by EMD, resulting in the kurtosis diagram of ten IMF components IMF1-IMF10 according to the kurtosis criterion shown in Figure 6. The five components with the highest kurtosis, IMF9, IMF6, IMF1, IMF2 and IMF10 are chosen for signal rearrangement as one-dimensional time series. Both the training and test sets are EMD processed at the same time.

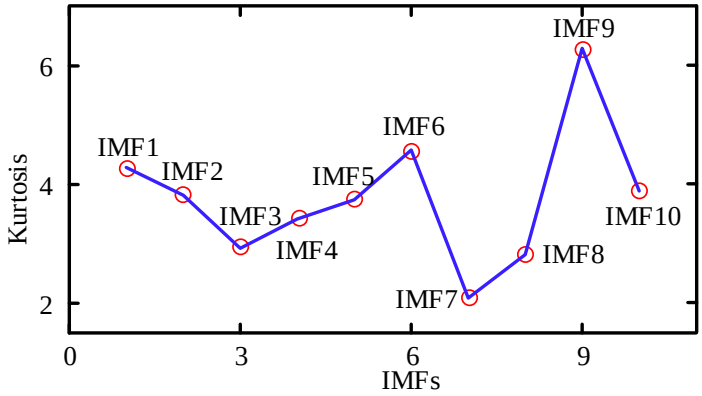


Figure 6. Kurtosis diagram of IMF components of time series signals (EMD)

Similarly, VMD denoises the data set, but when it analyzes the signals, it must take into account the decomposition mode parameter K and the penalty term coefficient α, in general, the central frequency observation method [42] and EMD-VMD can be used to select the decomposition mode parameter K, and it is easy to ignore the relationship between K value and penalty term coefficient α. The CS can search the influence parameter combination [K, α] of VMD to perform adaptive parameter selection, taking into account the interaction between the influence parameters, finally, as shown in Figure 7(a), (b), and (c), the kurtosis diagrams corresponding to the three approaches of VMD-C, VMD-EMD, and VMD-CS are obtained, in Figure 7(a), the five components IMF8, IMF6, IMF5, IMF7, and IMF3 with the highest kurtosis are selected in order for signal reformation as one-dimensional time series, in Figure 7(b), the five components IMF10, IMF9, IMF7, IMF6, and IMF4 with the highest kurtosis are selected in order for signal reformation as one-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

dimensional time series, in Figure 7(c), the three components IMF4, IMF3, and IMF2 with the highest kurtosis are selected one by one for signal reformation as one-dimensional time series. Finally, all data set have undergone noise reduction processing, and the data set listed in Table I have been determined.

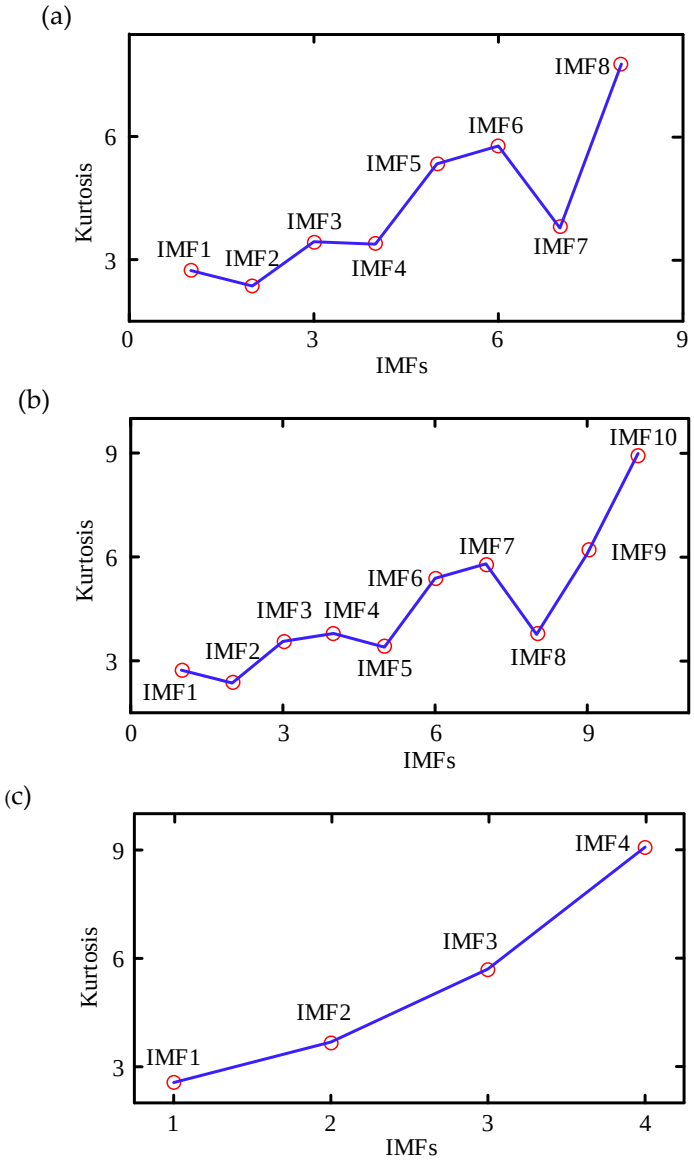


Figure 7. Kurtosis diagram of IMF components of original time series. (a) Kurtosis diagram of IMF components of time series (VMD-C). (b) Kurtosis diagram of IMF components of time series (VMD-EMD). (c) Kurtosis diagram of IMF components of time series (VMD-CS)

Table 1. The central frequencies of IMF components corresponding to different *K* values

Method	EMD	VMD	VMD-EMD	VMD-CS	MCKD
Training Set	6×1000	6×1000	6×1000	6×1000	6×1000
Test Set	7×100	7×100	7×100	7×100	7×100

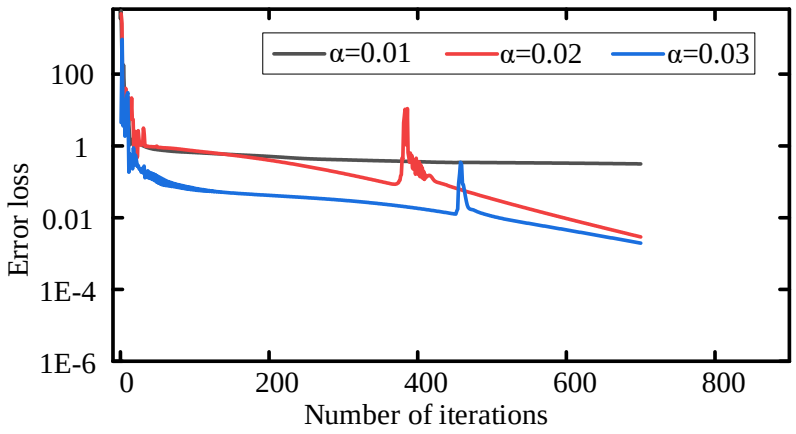
6.2. Parameter Selection

Because the learning rate has such a large impact on the LSTM neural network model, the experimental results are analyzed when the learning rates are 0.01, 0.02, and 0.03 respectively, to determine the learning rate of the LSTM model based on the bearing time

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

series, and to obtain the error loss and model accuracy of the LSTM model shown in Figure 8, the LSTM model shows an over fitting phenomenon at α set to 0.02, resulting in severe swings in prediction accuracy, but the prediction accuracy of the LSTM model is steady when the learning rate between 0.01 and 0.03, as seen in the picture. Table II shows a comparison of mean square error experimental data acquired at various learning rates. The mean square error of the LSTM model on test time series 1, 2, 3, 4, 5, 6, and 7 at α set to 0.01 is a respectively, and the prediction accuracy of each time series is greater than that α set to 0.03, hence this paper selects the learning rate α set to 0.01 as the training rate for the LSTM model.

(a)



(b)

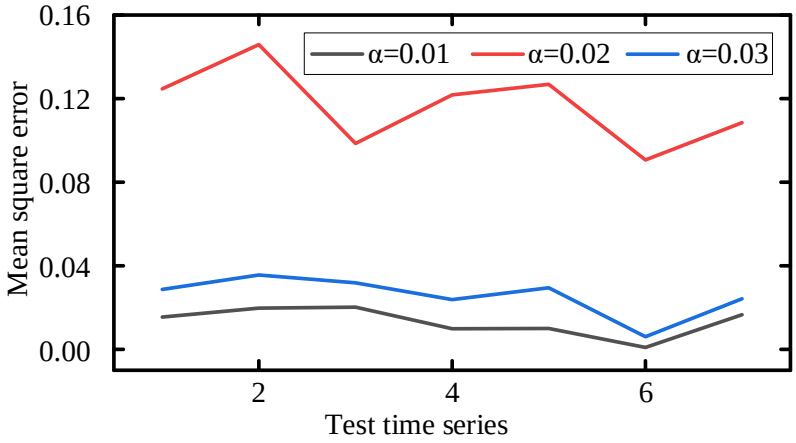


Figure 8. Variation of error loss and mean square error comparison of LSTM models at different learning rates. (a) Error loss of LSTM models at different learning rates. (b) Mean square error of LSTM models at different learning rates.

Table 2. The central frequencies of IMF components corresponding to different *K* values

Model Learning Rate	Test Time Series						
	1	2	3	4	5	6	7
	Mean Square Error						
0.01	0.01544	0.01972	0.02019	0.00986	0.01002	0.00089	0.01660
0.02	0.12468	0.14582	0.09857	0.12179	0.12682	0.09063	0.10852
0.03	0.02869	0.03561	0.03181	0.02381	0.02946	0.00608	0.02420

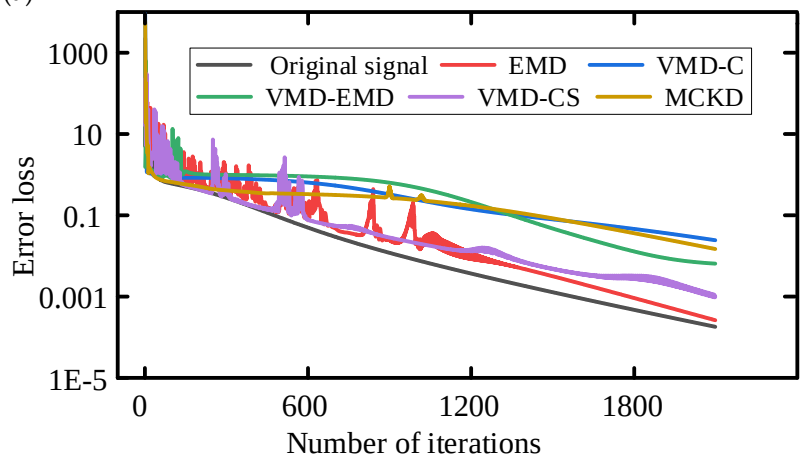
6.3. Prediction Model

The dataset produced from various data preprocessing are input into the LSTM model for prediction, and the error loss changes under various models are calculated as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

shown in Figure 9(a), the error loss of the prediction result of the original time series input LSTM model is the minimum, as can be seen in the image, the loss obtained by the model is often greater than that acquired by the original time series after different preprocessing procedures are applied. In test time series 1, 3, 4, 5, and 6, the prediction accuracy of MSE obtained by optimizing MCKD-LSTM model is the highest (MSE is the lowest), as shown in the accuracy comparison results of prediction results under different models in Figure 9(b).

(a)



(b)

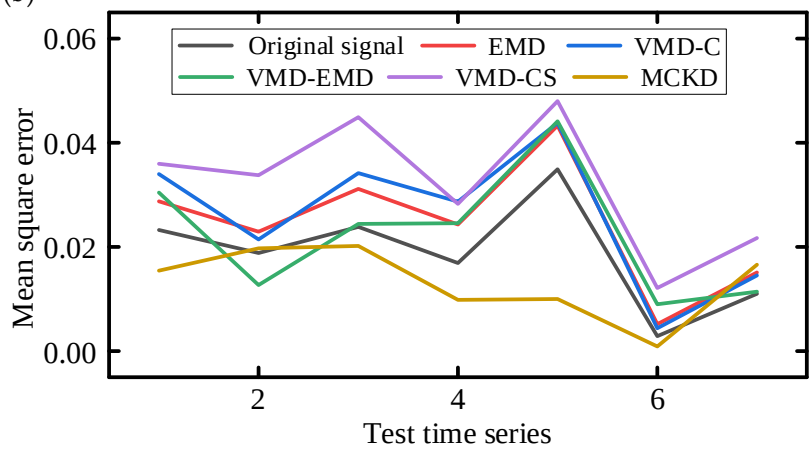


Figure 9. Error loss and mean square error comparisons for different models. (a) Error loss for different models. (b) Mean square error for different models.

In order to compare the validity of the proposed model, the prediction results of the original time series and the optimized MCKD-LSTM model are compared, as can be seen from the comparison between Figure 10 (a) and (b), the original time series prediction results have some deviation in the whole time series, but the optimized MCKD-LSTM model tracks the real fault data well, it can also be seen from the comparison of experimental results of mean square error under different prediction models in Table III: on test series 1, 3, 4, 5 and 6, the mean square error of the original time series is 0.02327, 0.02384, 0.01691, 0.0349 and 0.00287 respectively. The prediction results of the optimized MCKD-LSTM model are 0.01544, 0.02019, 0.00986, 0.01002 and 8.95153e-4 respectively, and the average prediction accuracy is improved by 26%.

(a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 November 2021 doi:10.20944/preprints202111.0377.v1

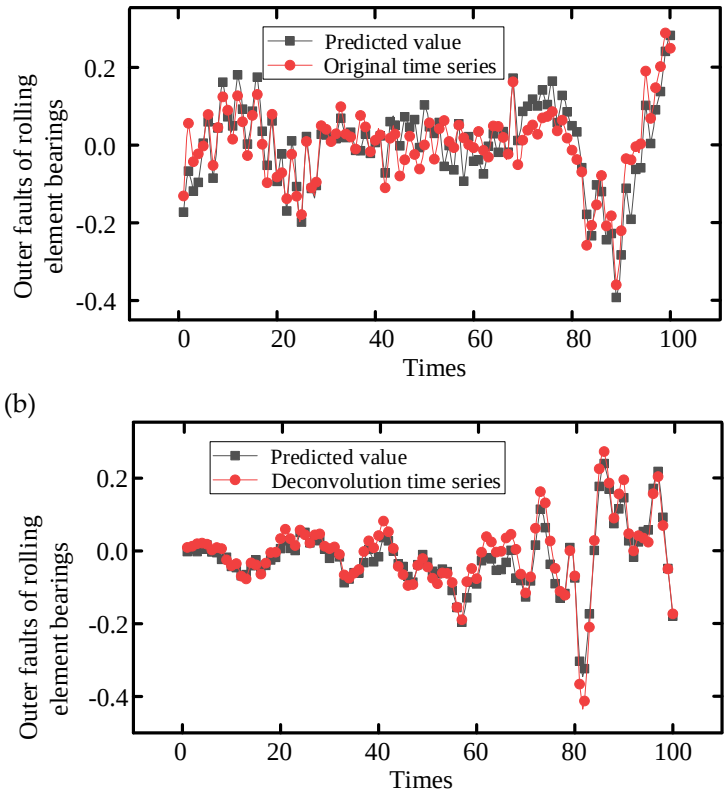


Figure 9. Time series prediction results. (a) Prediction results of original time series. (b) Prediction results of deconvolution time series

Table 3. The central frequencies of IMF components corresponding to different K values

7. Conclusion

Optimized MCKD-LSTM model for bearing series prediction is proposed in this research. The model combines optimizing MCKD preprocessing of the original series and time series prediction using deconvolution signals, the effectiveness of this method is verified by XJTU-SY bearing dataset of XI'AN JIAOTONG UNIVERSITY, and the conclusions are as follows:

1) When comparing the results of EMD, VMD-C, VMD-EMD, VMD-CS and MCKD on original time series, it can be seen that the impact component of the deconvolution time series obtained by optimizing MCKD is enhanced, and the fault characteristic frequency of the bearing outer ring is extracted.

Model	Test Time Series						
	1	2	3	4	5	6	7
	Mean Square Error						
Original signal	0.02327	0.01883	0.02384	0.01691	0.0349	0.00287	0.01101
EMD	0.02875	0.02292	0.03114	0.0243	0.04327	0.0052	0.01509
VMD-C	0.02756	0.02089	0.02828	0.02296	0.03895	0.00481	0.01344
VMD-EMD	0.03043	0.01268	0.02442	0.02456	0.04411	0.00899	0.01143
VMD-CS	0.03596	0.03376	0.0449	0.02826	0.04799	0.01213	0.0217
MCKD	0.01544	0.01972	0.02019	0.00986	0.01002	8.95153E-4	0.0166

2) The accuracy and loss change of the model are affected by the learning rate of the neural network. Overfitting difficulties will occur if the change rate is too high or too low, affecting efficiency and prediction ability of the model, as a result of the experimental investigation, the learning rate of the LSTM prediction model of bearing time series is determined to be α=0.01.

3) After deciding that the learning rate α set to 0.01, the optimized MCKD-LSTM model has the highest prediction accuracy, which is 26% higher than the original time series prediction, and the prediction results track the real fault data well.

Author Contributions: Conceptualization, L.M., T.M. and Y.S.; methodology, T.M., H.Z., L.X. and Y.J.; formal analysis, L.M., T.M., H.J., X.Z. and Y.S.; writing—original draft preparation, L.M., T.M. and Y.S.; writing—review and editing, L.M., T.M., Y.S., L.X. and Y.J.; supervision, H.J., X.Z. and H.Z.; project administration, T.M. and Y.S.; funding acquisition, H.J., X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Natural Science Foundation of China under Grant 51865054 and Grant 51765061, in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2018D01C043, in part by the Natural Science Foundation of Xinjiang University under Grant BS180216..

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Y. Yizhou, Y. Wenguang, and J. Dongxiang, "Simulation and experimental analysis of rolling element bearing fault in rotor-bearing-casing system," The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Efa(2018), doi:10.1016/j.engfailanal.2018.04.053.
2. A. Dhanola, and H. C Garg, "Tribological challenges and advancements in wind turbine bearings: a review-sciencedirect," Eng. Failure Anal., 118.
3. L. Shaowei, J. Hongkai, W. Zhenghong, "Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis," Mech. Syst. Signal Process., 163(Jan.), 108139.
4. M. Wentao, F. Wushi, L. Yamin, "A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis," Mech. Syst. Signal Process., vol. 150, no. 12, 107233.
5. S. Hao, X. Ling, H. Aijun, "A novel hybrid method based on KELM with SAPSO for fault diagnosis of rolling bearing under variable operating conditions," Measurement, vol. 177, no. 6, 109276.
6. Sachin P. Patel, S.H. Upadhyay, "Euclidean distance based feature ranking and subset selection for bearing fault diagnosis," Expert Syst. Appl., 154, 113400.
7. Z. Lihui, L. Qichen, F. Jinzhi, "Research on prediction method of hub-bearing service life under random road load," J. Mech. Eng., vol. 57, no. 10, pp. 77-86, 2021
8. W. Heng, N. Guangxian, C. Jinhai, "Research on Rolling Bearing State Health Monitoring and Life Prediction Based on PCA and Internet of Things with Multi-sensor," Measurement (2020), doi: https://doi.org/10.1016/j.measurement.2020.107657.
9. Z. Zhibin, Q. Baijie, W. Shibin, "A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis," J. Sound Vib., vol. 446, pp. 429-452.
10. W. Xin，W. Ji，L. Chao, "Exploring LSTM based recurrent neural network for failure time series prediction," J. Beijing Univ. Aeron. Astron., vol. 44, no, 4, pp. 772-784, 2018.
11. D. Xin, L. Guolong, J. Yachao, "Non-iterative denoising algorithm for mechanical vibration signal using spectral graph wavelet transform and detrended fluctuation analysis," Mech. Syst. Signal Process., 149.
12. Y. Zhonghong, C. Peipei, M. Jingxuan, "Discrete convolution wavelet transform of signal and its application on BEV accident data analysis," Mech. Syst. Signal Process., 159.
13. W. Xi，T. Muqin，S. Jiancheng, "Feature extraction of vibration signals based on empirical wavelet transform," J. Vib. Shock, vol. 40, no. 16, pp. 261-266, 2021.
14. V. Sharmaa, A. Parey, "Extraction of weak fault transients using variational mode decomposition for fault diagnosis of gearbox under varying speed," Eng. Failure Anal., 107, 104204.
15. A. Bagheri, O. E. Ozbulut, D. K. Harris, "Structural system identification based on variational mode decomposition," J. Sound Vib., vol. 417, pp. 182-197.
16. M. Chouksey, and R. K. Jha, "A multiverse optimization based colour image segmentation using variational mode decomposition," Expert Syst. Appl., 171, 114587.
17. Z. Yunqiang, R. Guoquan, W. Dinghai, "Rolling bearing fault diagnosis utilizing variational mode decomposition based fractal dimension estimation method," Measurement, vol. 181, no. 1-2, 109614.
18. F. Juan Li, C. Yu, L. Changgang, "Application of an improved variational mode decomposition algorithm in leakage location detection of water supply pipeline," Measurement, 173(8), 108587.

19. Z. Yujie, L. Chaoshun, F. Wenlong, "A modified variational mode decomposition method based on envelope nesting and multi-criteria evaluation," J. Sound Vib. (2019), doi: https://doi.org/10.1016/j.jsv.2019.115099.
20. Z. Xinlong, W. Pengfei, and Y. Xiuxing, "A quadratic penalty item optimal variational mode decomposition method based on single-objective salp swarm algorithm," Mech. Syst. signal Process., 138(Apr.), 106567.1-106567.12.
21. F. Guorui, W. Huiru, Q. Tingye, "A transient electromagnetic signal denoising method based on an improved variational mode decomposition algorithm, Measurement, vol. 184, no. 2, pp.109815.
22. W. Jin hua，H Jia wei，C. Jie, "Multi-fault diagnosis of rolling bearing based on adaptive VMD and IELM," J. Jilin Univ. (Eng. Tech. Edition), https://doi.org/10.13229/j.cnki.jdxbgxb20200856.
23. G. L. Mcdonald, Q. Zhao, M. J. Zuo, "Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection," Mech. Syst. signal Process., vol. 33, pp. 237-255,2012.
24. S. Wei, L. Xinmin, J. Xiaoqiang, "Fault diagnosis method for helicopter swash-plate rolling bearings based on the MCKD and envelope cepstrum," J. Vib. Shock, vol. 38, no. 2, pp. 159-163, 2019.
25. H. Lianhuan, L. Xiaobo Liu, and Z. Hongyan, "Compound faults diagnosis based on customized balanced multiwavelets and adaptive maximum correlated kurtosis deconvolution," Measurement, vol. 146, pp. 87-100.
26. Z. Qing，J. Weizhe，L. Heng, "Combined MCKD-Teager energy operator with LSTM for rolling bearing fault diagnosis," J. Harbin Inst. tech., vol. 53, no. 7, pp. 68-76+83, 2021.
27. Y. Xuan L, H. Zhanqi, Z. Haili, "Application of improved MCKD method based on QGA in planetary gear compound fault diagnosis," Measurement (2019), doi: https://doi.org/10.1016/ j.measurement.2019.02.071.
28. M. Yonghao, Z. Ming, L. Jing, "Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings," Mech. Syst. Signal Process, 92(AUG.), 173-195.
29. Y. Bin，Z. Jiawei，F. Gairong, "Application of OPMCKD and ELMD in bearing compound fault diagnosis," J. Vib. Shock, vol. 38, no. 11, pp. 59-67, 2019
30. P. Wenping, F. Liuyang, Z. Limao, "Time-series interval prediction under uncertainty using modified double multiplicative neuron network," Expert Syst. Appl., vol. 184, no. 1, 115478.
31. S. Raubitzek, T. Neubauer, "A fractal interpolation approach to improve neural network predictions for difficult time series data," Expert Syst. Appl., vol. 169, no. 1, 114474.
32. L. Yeqi, G. Chuanyang, Y. Ling, "DSTPRNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction," Expert Syst. Appl., (2019), doi: https://doi.org/10.1016/j.eswa.2019.113082.
33. M. Savadkoohi, T. Oladunni, and L. A. Thompson, (2021). "Deep neural networks for human's fall-risk prediction using force-plate time series signal," Expert Syst. Appl., vol. 182, no. 1, 115220.
34. Z. Yong'an, Y. Binbin, A. Memon, "A novel deep learning framework: prediction and analysis of financial time series using CEEMD and LSTM," Expert Syst. Appl. (2020), doi: https://doi.org/10.1016/j.eswa.2020.113609.
35. C. Changchang, W. Huawei, N. Xiaomei, "Residual life prediction of aeroengine based on 1D-CNN and Bi-LSTM," J. mech. Eng., https://kns.cnki.net/kcms/detail/11.2187.TH.20210617.1133.026.html.
36. N. Gang, Y. Bo-Suk, "Dempster-Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis," Mech. Syst. Signal Process., vol. 23, no. 3, pp. 740-751, 2009.
37. X. S. Yang, and S. Deb, "Cuckoo Search via Lévy Flights," 2009 World Congress Nature. Biol. Inspired Comput., pp. 210-214, 2009.
38. S. Hochreiter, J. Schmidhuber. "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.
39. Z. Long, X. Guoliang, H. Wenyi, "New procedure and index for the parameter optimization of complex wavelet based resonance demodulation," J. mech. Eng., vol. 51, no. 3, pp. 129-138, 2015.
40. W. Biao, L. Yaguo, L. Naipeng, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," IEEE Transactions on Reliability, pp. 1-12, 2018.
41. M. Tongwei, X. Zhang, H. Jiang, "Early fault diagnosis of shaft crack based on double optimization maximum correlated kurtosis deconvolution and variational mode decomposition," IEEE Access, vol. 9, pp. 14971-14982, 2021.
42. S. Yong, Z Xiangfeng, J Hong, "Comparative study on dynamic characteristics of two-stage gear system with gear and shaft cracks considering the shaft flexibility," IEEE Access, vol. 8, pp. 133681-133699, 2020.