Response to Reviewer 2 Comments

Point 1: In Section 1, the paper provided the comparison of computational resource requirements from few published articles as presented in Table 1. I am curious to know how the Authors calculated the computational complexity?

Response 1: There are two main criteria to measure the complexity of the algorithm:

1. Number of parameters
   The parameter length of each model is 32 bit floating point number, which occupies the storage space of 4Byte. The calculation methods of the number of parameters in the convolutional layer and the fully connected layer are shown in Eqs. (9) and (10), respectively. The number of parameters is the sum of the number of parameters of each network layer, and the number of parameters of the model represents the size of storage space occupied by all parameters.

2. The computation of floating numbers
   The computation of floating numbers is expressed by the number of floating-point operations that each network layer propagates forward to process one data. The realization of convolution or full connection is actually one multiplication and one addition. The calculation methods are shown in Eqs. (11) and (12). The computation of floating numbers of the model is the sum of all network layers.

Point 2: In Section 2, the WPT method basically decomposes a raw vibration signal into two different levels namely approximate and detail. Please indicate these component in Figure 1.

Response 2: The reviewer’s comments are correct, the information description about Approximate and Detail of signals have been supplemented in Figure 1.

Point 3: Please enhance the resolution of Figure 4. The input signal can be revised to better resolution and increase the font size of Figure 4.

Response 3: According to reviewer’s comment, Figure 4 has been revised to better resolution and increase the font size.
Point 4: It would be better of the WPT-CNN structure described in Line 213 to 253 is also presented in flowchart. This will make reader easily understand the flow process.

Response 4: Lines 213 to 253 describe the technical details of the WPT-CNN network structure, the purpose is to focus on how to select parameters in the process of algorithm implementation. For example, 2. The lengths of the conv_1 to conv_5 convolutional layers of the network are different, and their values are set to 11, 3, 3, 3 and 3, respectively. These technical details are difficult to express in the form of flow charts.

Point 5: Name the features presented in Figure 10. Different color is refer to particular feature.

Response 5: The reviewer's comments are correct, different color is refer to particular feature. Figures 10 and 11 have been revised, and the fault categories corresponding to each color are supplemented.
**Point 6:** The Conclusion need to be added with more sentences. Probably the future work such as validate the proposed method using the real fault bearing data can also be added.

**Response 6:** According to reviewer's comment, the conclusion has been rewritten, and the future work has also been added.

**Conclusions:**
This study investigates fault prediction in rotating machinery. Based on the ability of the wavelet packet transform to extract more frequency-domain information and the low computational load of 1D convolutions, a convolutional network-based lightweight deep learning fault prediction algorithm is proposed. This network can be regarded as two layers. The first layer performs the wavelet packet transform (WPT), with an aim to extract finer information from a frequency-domain perspective, and the second layer is the designed relatively lightweight CNN (as shown in Figure 4). To validate the effectiveness of the proposed WPT-CNN network structure, the bearing dataset collected by the Case Western Reserve University (CWRU)'s Bearing Data Center were used to implement experiments. The comparison results demonstrate that the proposed algorithm is superior to the available algorithms in noise resistance (5-58% higher) and transfer-learning ability (0.5-44% higher). In addition, the proposed algorithm also has the lowest computational complexity and memory space requirement (72.5% and 88.5% lower than those of the available algorithms, respectively). Therefore, the proposed algorithm can effectively improve the identification accuracy and is relatively lightweight.

Future research will consist of developing a new model for the deep convolution neural network optimization problem by the improvement of regularization term and optimization strategy with the results of this paper to predict the fault characteristics of the rotating machinery. Also, to meet the actual working conditions of industrial production, future research will also focus on improving the generalization ability and the lightweight problem of the model.