Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
©2023 ESREL2023 Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P164-cd

Prediction of Remaining Useful Life of Bearings using a Parallel Neural Network

Sajawal Gul Niazi
School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China. E-mail: sajawalgul@outlook.com

Ali Nawaz
School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China. E-mail: alinawaz.sanjrani@muetkhp.edu.pk

Tudi Huang
School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China. E-mail: huangtudi@std.uestc.edu.cn

Song Bai
School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China. E-mail: baisong@std.uestc.edu.cn

Hong-Zhong Huang*
School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China. E-mail: hzhuang@uestc.edu.cn

** Corresponding author. Tel: +86-28-61831252; Fax: +86-28-61830227.*

This study advocates the utilization of a parallel neural network (PNN) architecture for the estimation of remaining useful life (RUL) of bearings. The use of conventional machine learning and deep learning techniques has been inadequate in terms of accuracy and computation time, because of huge input data sizes and the time-dependent nature of the output. To address this limitation, the PNN architecture incorporates multiple parallel processing paths with multiple input neurons that take in data from condition detectors of mechanical machines and output neurons that predict RUL. The PNN structure provides better accuracy and computation time by efficiently handling vast amounts of data and integrating both spatial and temporal information simultaneously. Additionally, time-transformer and recurrent neural network (RNN) are used to handle complex time series data. Improvement methodologies like positional encoding with self-attention mechanism and ConvLSTM neural network are utilized to leverage multidimensional time-frequency data to process spatial and temporal dependencies present in the extracted features, further increasing model's efficiency. A case study is conducted on XJ-SY rolling element-bearing dataset to validate the proposed methodology, where PNN performed exceptionally in terms of accuracy and efficiency. It is concluded that PNNs exhibit potential for predicting RUL of bearings and can be applied to other machinery types.

Keywords: Remaining useful life (RUL), Parallel neural networks (PNNs), Bearings, Time-Transformer, ConvLSTM, Positional encoding, Self-attention.

1. Introduction

Condition-based maintenance (CBM) is a widely accepted maintenance strategy that focuses on status monitoring of physical assets to implement maintenance actions only when the asset's performance is unacceptable. This approach empowers industries to conserve resources and reduce maintenance costs. Ali et al., (2022).

The effectiveness of CBM depends on implementing accurate data monitoring, fault detection, diagnosis, and prognosis processes, along with precise estimation of the Remaining Useful Life (RUL) for maintenance decision-making. Machinery degradation and failure can result in substantial economic losses and safety hazards. Therefore, Accurate prediction of the RUL of machinery components is crucial. Rolling element bearings are the most significant element in rotating machinery, and also are a common cause of operation failures. Heng et al., (2009). Hence, precise RUL prediction for the bearings can obviously improve the operation safety and overall reliability of the rotating machinery.

RUL prediction approaches are categorically divided into model-based and data-driven. Model-based methods use prior expert knowledge and model failure mechanisms to mathematically model equipment degradation. However, to mathematically model the machinery efficiently is a difficult task, when the components are complex, resulting in simplifications to the key attributes, which compromises the model accuracy. Alaswad et al.,(2011). Alternatively, data-driven approaches rely on historical run-to-failure data to estimate the RUL using different machine learning techniques. Lei et al., (2018).

To this end, various machine learning (ML) techniques have been proposed for RUL prediction, including deep belief networks (DBN) by Peng et al., (2018), convolutional neural networks (CNN) by Ren et al., (2018), long short-term memory neural networks (LSTM) by Jiang et al., (2019), attention mechanism-based models (Jiang et al., (2019); Chen et al., (2020)), and boot-strap fusion technique by Huang et al., (2021). These techniques have shown promising results in RUL prediction for various machinery components, such as bearings, lithium-ion batteries, aircraft engines, and wind turbines .Li et al., (2022). Despite the success of these techniques, traditional ML models have certain limitations because big data input is complex, and also time dependent. This paper is aimed to solve these two problems by employing a parallel neural network structure processing a lot more data simultaneously. And also using the models such as time-transformers and ConvLSTM neural networks which map the non-linear time-dependencies from the input data effectively.

In this paper, a two-stage prognostic approach for RUL prediction of rolling element bearings using a PNN structure with Time transformers and ConvLSTM neural network is proposed. This approach aims to overcome the limitations of traditional ML models and improve the accuracy and efficiency of RUL prediction. A case study is also conducted on rolling element bearings to verify the effectiveness of the proposed methodology.

2. Theoretical Background

In this section, first deep learning models of RNNs and Time-Transformers are reviewed briefly. Then, specific improved methodologies for both approaches naming: ConvLSTM neural network, positional encoding integrated with self-attention are introduced. These networks can effectively handle highly complex and non-linear time-series data, making them suitable for capturing patterns and dependencies to predict the RUL of bearing with higher accuracy.

2.1.*Recurrent neural network*

Recurrent Neural Networks (RNNs) are designed to incorporate sequential dependencies in the data, making them an effective approach to handle such data where input data in later stages of the time are highly dependent on previous events present in the data. RNNs process sequential data by propagating information step by step using a loop that enables information to persist over time. Although, RNNs struggle with long-term dependencies due to vanishing gradients, to avoid this hurdle LSTM (Long Short-Term Memory) neural network were introduced which overcome this with a gating mechanism that selectively controls the flow of information and gradients through the network.

2.1.1.*Convolutional long short-term memory*

ConvLSTM is a special case of LSTM neural network that integrates convolutional layers into the LSTM structure. The ConvLSTM network is particularly useful for processing multidimensional sequential data. It captures spatiotemporal patterns present in feature images utilizing the convolutional operations. The convolutional layers in the network help capture spatial dependencies within the input data, while the LSTM layers help capture the temporal dependencies. Ma and Mao (2020).

The ConvLSTM structure can be represented mathematically as follows:

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \quad (4)$$

$$h_t = o_t \circ \tanh(c_t) \quad (5)$$

where x_t is the input at time step t, h_{t-1} is the hidden state at the previous time step, c_{t-1} is the cell state at the previous time step, i_t, f_t, o_t are the input, forget, and output gates respectively, σ and $\circ$ represent sigmoid function and element-wise multiplication, and W and b are the learnable parameters of the network.

The Fig. 1 illustrates the architecture of a ConvLSTM network, which consists of convolutional layers and LSTM units. The fed input data is processed using convolutional layers to extract the spatial features. The resulting feature maps are then fed into the LSTM component of the network to map the temporal dependencies over time.

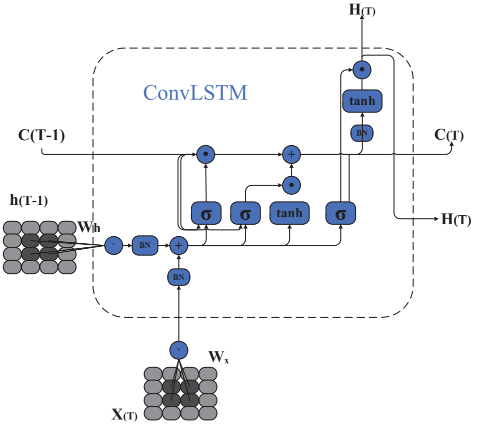

Fig. 1. ConvLSTM neural network structure

2.2. *Time transformers*

Time Transformers architecture uses self-attention mechanisms to model temporal relationships in sequential data.

The Time Transformer architecture can be represented mathematically as follows:

$$q_t = W_q x_t \quad (6)$$

$$k_t = W_k x_t \quad (7)$$

$$v_t = W_v x_t \quad (8)$$

$$a_t = \left(\frac{q_t k_t^T}{\sqrt{d_k}}\right) v_t \quad (9)$$

$$o_t = W_o a_t \quad (10)$$

where x_t is the input at time step t, q_t, k_t, v_t are the query, key, and value vectors respectively, a_t is the attention vector, W_q, W_k, W_v, W_o are the learnable weight matrices, and d_k is the dimensionality of the key vectors.

2.2.1. *Positional encoding and self-attention*

Positional encoding is used to incorporate information about the sequential ordering of the input data into the network architecture. This is achieved by adding fixed sinusoidal functions of different frequencies and phases to the input embeddings of the network. The positional encoding function can be represented mathematically as follows:

$$\mathrm{PE}(pos, 2i) = \sin(pos / 10000^{2i/d_{model}}) \quad (11)$$

$$\mathrm{PE}(pos, 2i+1) = \cos(pos / 10000^{2i/d_{model}}) \quad (12)$$

where pos is the position of the input element in the sequence, i is the dimension of the embedding vector, and d_{model} is the dimensionality of the input embeddings. Huang et al., (2022).

Self-attention is additional component of the Time Transformer architecture, which enables the network to selectively attend to different parts of the input sequence at different times. The self-attention mechanism can be represented mathematically as follows:

$$\mathrm{Attention}(Q, K, V) = \left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (13)$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

In the Time Transformer architecture, the self-attention mechanism is applied multiple times, using different sets of query, key, and value matrices in each layer. The output of each layer is then passed through a feedforward neural network, and residual connections and layer normalization are applied to improve the stability of the training process.

The structure of the time transformer is illustrated in Fig. 2.

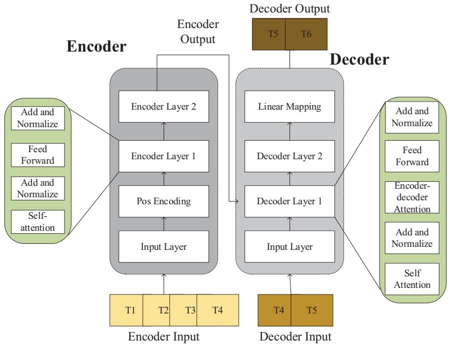

Fig. 2. Time Transformer with positional Encoding structure

3. Proposed Prognostics Approach

The proposed technique in this paper encompasses the practice of a parallel neural network architecture that combines ConvLSTM and Time Transformer networks. The input to the ConvLSTM network is the time-frequency representations (TFRs) of sensor data which are extracted from the vibration signal by performing the Morlet continuous wavelet transforms over a sequence of the data, while the Time Transformer network takes in one-dimensional extracted features from the sensor vibration data with positional encoding and self-attention.

3.1.*Data acquisition and pre-processing*

In the case of bearing vibration data, the first step is to acquire the data from sensors installed on the bearing. The input consists of vibration data collected from bearings undergoing accelerated failure testing. Once the data has been acquired, the next step is to pre-process it to remove any noise and prepare it for use it for RUL prediction. Second step of proposed technique is to measure the first prediction time (FPT) of the dataset. FPT of the vibration signal is the point at which the fault in the signal occurs and the bearing moves from a healthy stage to the unhealthy degradation stage.

In Fig. 3 the raw vibration signal of a bearing is displayed and the red dotted line represents the FPT in mins, which divides the healthy and degradation stage of the bearings. This is done by calculating the root mean square (RMS) of the signal. When the RMS of the signal deviates from healthy stage mean value by a magnitude of 3 times the standard deviation of the signal $|\mu_i - \mu| > 3\sigma$ at that point fault has occurred and all the data after that is taken for the analysis, which represent the degradation stage of the bearings.

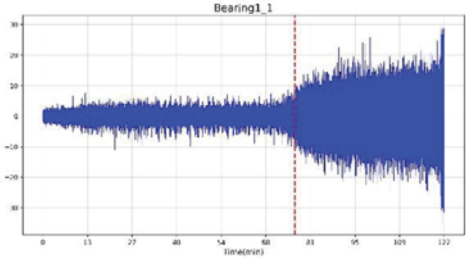

Fig. 3. FPT determination in the vibration time-series data

3.2.*Health indicator extractions*

After FPT calculation of the vibration signal the dataset in the degradation stage is converted into number of input samples using a sliding window and labelled with RUL from 1 to 0. In this approach the number of samples are set to 2048, it is worth noting that the after the FPT determination dataset for all the bearing are not same in length. But a clever data augmentation strategy is employed, which uses the sliding window of size 1024 data instances to divide the rest of the dataset into same number of samples by variational stride for respective dataset. Which enables putting more focus on data where the degradation process is fast, but skipping the input data where redundancy occurs and the degradation of the bearing is slow. After acquiring the number of samples features extraction is performed. The details of the 29 extracted features are provided in Table 1.

Table 1. Detailed features used for neural network training

Feature Types	Feature Names
2D-Time Frequency Domain	1.Morlet Continuous Wavelet transform
1D-Time Domain Statistical Features	2.Root mean square 3.Kurtosis 4.Skewness 5.Shape factor

1D-Time Domain Impulsive features	6.Mean absolute value 7.Minimum value 8.Impulse Factor 9.Crest Factor 10.Clearance Factor 11.Standard deviation
1D-Time Domain Trigonometric features	12.Standard deviation of inverse hyperbolic cosine 13.Standard deviation of inverse hyperbolic sine
1D-Frequency domain features	14-29.Energies of sixteen band between 1hz and 2000hz

3.3.*Parrallel neural network*

After the health indicator extraction of the features the inputs are fed in to the deep neural network. The ConvLSTM network is able to capture both spatial and temporal dependencies in the input data, making it well-suited for modelling 2D TFRs. Meanwhile, the Time Transformer network are designed to model sequential data and can capture long-term dependencies between time-series features, frequency features and trigonometric feature which are all one-dimensional input data. By combining the strengths of these two networks, the proposed architecture is able to effectively model the complex relationships between sensor data and predict the remaining useful life of industrial assets.

3.4.*Remaining useful life prediction*

In real-world circumstances, noise is present in the input data which results in uncertainty and can lead to inaccurate predictions. To address this issue, a Kalman filter is used to smooth the RUL predictions and improve their accuracy.

The Kalman filter is a mathematical algorithm that uses a series of equations to estimate the state of a system based on noisy input data. The Kalman filter equations are given by:

$$\hat{x}k \mid k = \hat{x}k \mid k-1 + K_k(z_k - H\hat{x}_{k|k-1}) \qquad (14)$$

$$P_{k|k} = P_{k|k-1} - K_k H P_{k|k-1} \qquad (15)$$

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1} \qquad (16)$$

Where $\hat{x}k \mid k$ is the state estimate at time step k , z_k is the noisy measurement at time step k , H is the measurement matrix, $Pk \mid k$ is the error covariance matrix at time step k , R is the measurement noise covariance matrix, and K_k is the Kalman gain at time step . By incorporating the Kalman filter into the proposed parallel neural network architecture, the accuracy of the RUL predictions is improved further. The overall structure of the proposed technique is illustrated in Fig. 4.

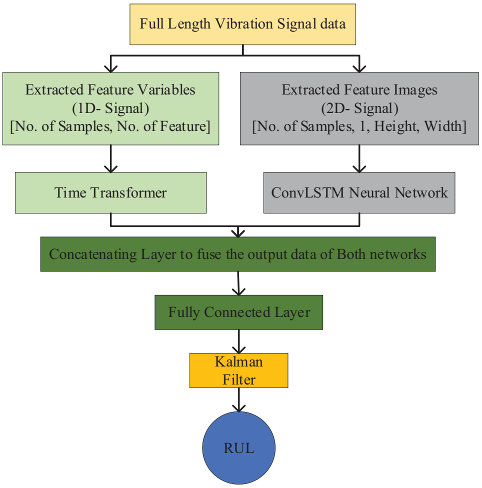
Full Length Vibration Signal data
Extracted Feature Variables
(1D- Signal)
[No. of Samples, No. of Feature]
Extracted Feature Images
(2D- Signal)
[No. of Samples, 1, Height, Width]
Time Transformer
ConvLSTM Neural Network
Concatenating Layer to fuse the output data of Both networks
Fully Connected Layer
Kalman Filter
RUL

Fig. 4. Framework of the proposed technique

4. Case Study

To extensively validate the efficiency and generalization capability of the proposed technique widely used dataset, i.e., XJTU-SY rolling bearings dataset is adopted as case study in this paper (Wang et al., 2018).

4.1.*Dataset description*

This dataset is comprised of 15 rolling bearings elements vibration data which were gathered while performing the accelerated degradation tests. The platform used to generate the dataset is depicted in Fig. 5. The detailed information about the experimental platform can be obtained from ref (Wang et al., 2018). To record the whole degradation process from normal condition to bearing failure, each of the accelerated degradation test was conducted until the maximum amplitude of the horizontal or vertical vibration signals surpassed a threshold of 10 times the maximum amplitude of vibration signals in the healthy stage of the tests.

The detailed information of the run-to-failure vibration data of 15 rolling element bearings are given in Table 2. Note that the vibration signals from Bearing1_4 and Bearing3_2 are not used due to the sudden failure characteristic and great variation in vibration signal resulting from multiple connective failures respectively. Therefore, for XJTU-SY rolling bearings datasets, overall horizontal vibration signal from 13 bearings for 3 different operating conditions are used to verify the proposed method. The columns in Table 2 represent the operating condition, bearing number, total lifetime and first prediction time, and also fault types of datasets OR, IR, C and B denote the outer race, inner race, cage and ball faults respectively.

Table 2. Detailed information of case study dataset

OC	Bearing Number	LT (mins)	FPT (mins)	Fault Type
35Hz & 12kN	Bearing1_1	123	75	OR
	Bearing1_2	160	45	OR
	Bearing1_3	157	56	OR
	Bearing1_4	121	112	C
	Bearing1_5	51	31	IR, OR
37.5Hz & 11kN	Bearing2_1	490	450	IR
	Bearing2_2	160	47	OR
	Bearing2_3	532	290	C
	Bearing2_4	41	24	OR
	Bearing2_5	338	145	OR
40 Hz & 10kN	Bearing3_1	2537	2358	OR
	Bearing3_2	2495	2069	IR, B, C, OR
	Bearing3_3	370	338	IR
	Bearing3_4	1514	1428	IR
	Bearing3_5	113	8	OR

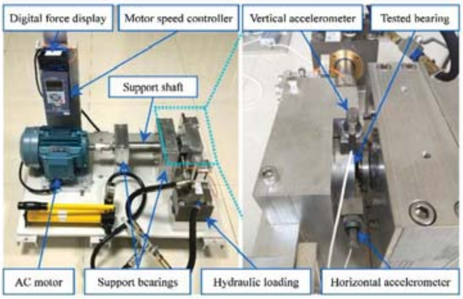

Fig. 5. Platform used to generate the XJ-SY Bearing Dataset

4.2.*Results*

For training the network, bearings of all conditions are used separately, because at different operating conditions the degradation stage behave differently. The final prediction for each bearing is generated based on leave-one out strategy for training and testing bearings. The RUL prediction analysis is performed using PyTorch programming language and the case study is conducted on Intel CPU i7-7700K @ 4.2 GHz, 32 GB RAM and Nvidia T1000, 16GB memory platform.

The training data is first divided into 90% training and 10% validation, batch size of 32, sequence length of 20, loss function mean square error (MSE) and Adam optimizer are used as the training parameters for the network, all the parameters used for the training of the network are first optimized based on the grid search methodology.

Two commonly used evaluation metrics are adopted naming: mean absolute error (MAE), root mean square error (RMSE) in this study. Deng et al., (2021). Which are represented mathematically as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} | \hat{y}_i - y_i | \tag{17}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (\hat{y}_i - y_i)^2} \tag{18}$$

Where $\hat{y}_i, y_i$ represent the predicted and actual RUL prediction respectively.

The final results are compared with other deep learning models including multi-layer perceptron (MLP) using only the 1D features, LSTM model, and another parallel neural network technique comprised of MLP and multi-scale convolutional neural network (MSCNN) which takes in both one dimensional and multi-dimensional inputs. Table 3 shows that the proposed method performs better in almost all test cases than the other deep learning methodologies of similar nature. Moreover, Fig 6 illustrates the graphical representation of the RUL prediction using the proposed technique and proves its generalization capabilities on different bearings under dissimilar working conditions.

Table 3. Performance comparisons of different models for RUL estimation

Testing bearing	MLP		LSTM		MLP-MSCNN		Proposed Method	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Bearing1_1	0.274	0.240	0.242	0.213	0.206	0.176	**0.180**	**0.159**
Bearing1_2	0.313	0.270	0.262	0.229	0.240	0.207	**0.190**	**0.161**
Bearing1_3	0.261	0.221	0.184	0.155	0.178	0.151	**0.124**	**0.109**
Bearing1_5	0.318	0.265	0.215	0.181	0.184	0.155	**0.115**	**0.094**
Bearing2_1	0.203	0.172	0.148	0.126	0.117	0.099	**0.113**	**0.094**
Bearing2_2	0.266	0.214	0.232	0.194	0.122	0.102	**0.101**	**0.098**
Bearing2_3	0.230	0.204	0.199	0.164	0.158	0.126	**0.142**	**0.120**
Bearing2_4	0.251	0.213	0.231	0.195	0.177	0.141	**0.114**	**0.087**
Bearing2_5	0.234	0.202	0.108	0.090	**0.0918**	**0.075**	0.112	0.095
Bearing3_1	0.305	0.262	0.247	0.214	0.244	0.204	**0.120**	**0.094**
Bearing3_3	0.318	0.276	0.191	0.156	0.158	0.129	**0.139**	**0.108**
Bearing3_4	0.252	0.220	0.165	0.139	**0.132**	0.107	0.137	**0.103**
Bearing3_5	0.376	0.310	0.267	0.225	0.266	0.219	**0.216**	**0.183**

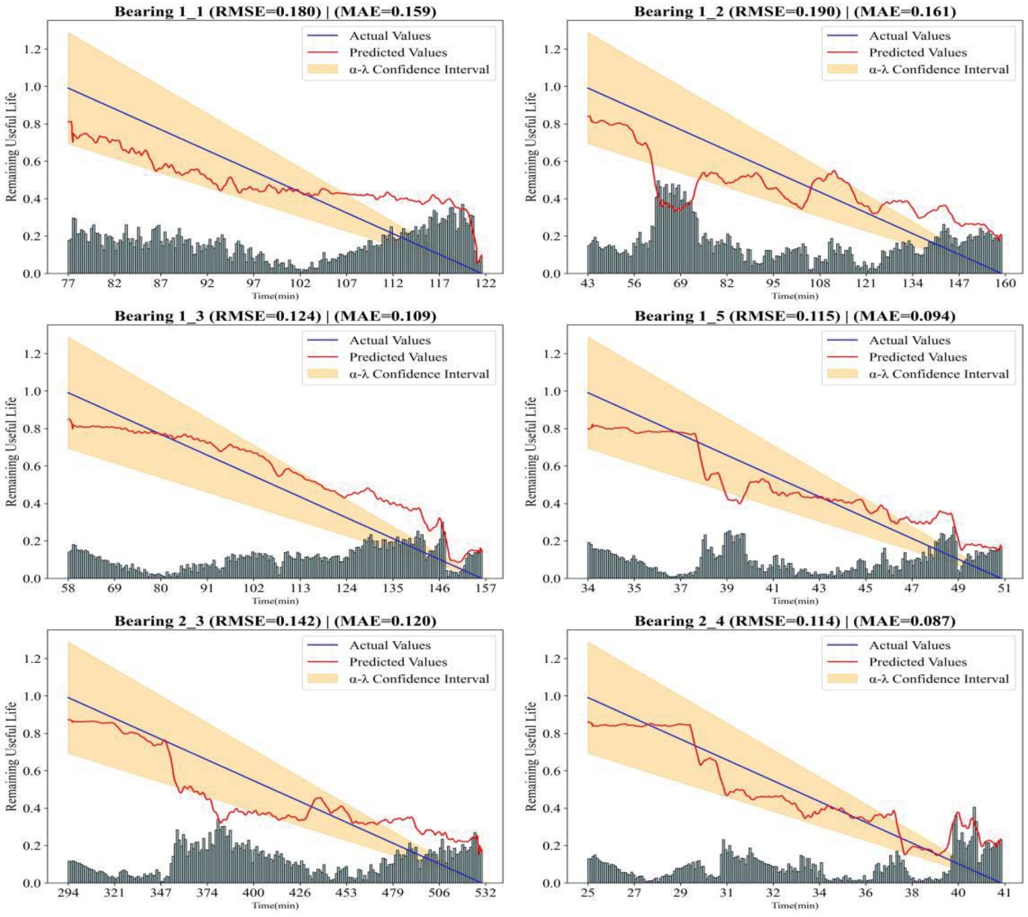

Fig. 6. Prediction results for (a) Bearing1_1, (b)Bearing1_2, (c)Bearing1_3, (d) Bearing1_4, (e) Bearing2_3, (f) Bearing2_4

In Fig. 6, the α-λ performance metric is delineated by upper and lower confidence intervals, with the majority of predictions residing within these bounds. Deviations from these intervals primarily occur in the latter part of the forecast due to the increased irregularity of bearing vibrations which is also reported by Saxena et al., (2008).

5. Conclusions and Future Developments

In this study, a PNN was proposed as an accurate and efficient methodology to obtain the RUL prediction. The conclusions from the application of proposed technique are summarized as follows:

(i) Both 1D time series-based features and 2D image-based feature images are taken simultaneously as inputs, which are fed to the parallel structure for boosting the computational efficiency and accuracy of the network

(ii) The data augmentation technique used in this methodology, using the sliding window uses variational stride to generate a homogeneous dataset for better performance.

(iii) The case study is performed for the rolling element bearings, and verifies the capabilities of the proposed approach.

Further research will focus on the use of transfer learning techniques to achieve improvements in computational time and accuracy. And also, to model the degradation stage in a non-linear manner to better represent the real-world scenario which is not linear in nature.

Acknowledgement

This research is supported by the Natural Science Foundation of Sichuan under the grant number 2023NSFSC0369.

References

Ali, A., & Abdelhadi, A. (2022). Condition-based monitoring and maintenance: state of the art review. Applied Sciences, 12(2), 688.

Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical systems and signal processing, 23(3), 724-739.

Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliability engineering & system safety, 157, 54-63.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical systems and signal processing, 104, 799-834.

Peng, K., Jiao, R., Dong, J., & Pi, Y. (2019). A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. Neurocomputing, 361, 19-28.

Ren, L., Sun, Y., Wang, H., & Zhang, L. (2018). Prediction of bearing remaining useful life with deep convolution neural network. IEEE access, 6, 13041-13049.

Jiang, J. R., Lee, J. E., & Zeng, Y. M. (2019). Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life. Sensors, 20(1), 166.

Chen, Y., Peng, G., Zhu, Z., & Li, S. (2020). A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. Applied Soft Computing, 86, 105919.

Huang, C. G., Huang, H. Z., Li, Y. F., & Peng, W. (2021). A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. Journal of Manufacturing Systems, 61, 757-772.

Li, N., Xu, P., Lei, Y., Cai, X., & Kong, D. (2022). A self-data-driven method for remaining useful life prediction of wind turbines considering continuously varying speeds. Mechanical Systems and Signal Processing, 165, 108315.

Ma, M., & Mao, Z. (2020). Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Transactions on Industrial Informatics, 17(3), 1658-1667.

Huang, L., Mao, F., Zhang, K., & Li, Z. (2022). Spatial-temporal convolutional transformer network for multivariate time series forecasting. Sensors, 22(3), 841.

Ding, H., Yang, L., Cheng, Z., & Yang, Z. (2021). A remaining useful life prediction method for bearing based on deep neural networks. Measurement, 172, 108878.

Saxena, A., Celaya, J., Balaban, E., Goebel, K. , Saha, B., Saha, S. and Schwabacher, M. (2008). Metrics for evaluating performance of prognostic techniques. PHM 2008. International Conference on, p.1.