A Comparative Study of Deep Learning Model Based Equipment Fault Diagnosis and Prognosis

Xianpeng Qiao[1], Hao Yuan Liow[1], Veronica Lestari Jauw[1,*], and Chin Seong Lim[2]

[1]Department of Mechanical, Material & Manufacturing Engineering, University of Nottingham Malaysia, Semenyih, Selangor, 43500, Malaysia

edxxq1@nottingham.edu.my
liowhaoyuan@gmail.com
**Corresponding author: Veronica.Jauw@nottingham.edu.my*

[2]Department of Electrical and Electronic Engineering, University of Nottingham Malaysia, Semenyih, Selangor, 43500, Malaysia

ChinSeong.Lim@nottingham.edu.my

ABSTRACT

Bearing fault diagnosis and prognosis are crucial for the effective management of industrial equipment. Due to the automatic feature extraction of Deep Learning (DL) models, many recent studies have focused on using DL for these tasks. However, most studies address only one of these tasks. This study aims to present DL models and their powerful ML tools capable of both fault diagnosis and prognosis on industrial equipment. To identify the best DL model for both tasks, a comparative study is conducted on various DL models and ML tools, including Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), CNN in parallel with LSTM (CNN-LSTM), Bidirectional LSTM (Bi-LSTM), and transformer models. The ML tools investigated include Recurrent Dropout, Residual Network (ResNet), and Monte Carlo Dropout (MC Dropout). These models are validated using online datasets from Case Western Reserve University (CWRU) and Xi'an Jiao Tong University (XJTU-SY) for the task of fault diagnosis. For fault prognosis, datasets from XJTU-SY and IEEE PHM are used. The results demonstrate the superiority of the ResCNN-LSTM model in both fault diagnosis and prognosis tasks. It achieves prediction accuracy of 99.87% and 96.39% and F1-scores of 0.998 and 0.964 for fault diagnosis on the CWRU and XJTU-SY datasets, respectively. Additionally, it shows a Root Mean Square Error (RMSE) of 8.56 and Mean Absolute Error (MAE) of 12.16 for fault prognosis on the XJTU dataset, and an RMSE of 12.18 using the IEEE PHM bearing dataset. These high performance metrics indicate the model's effectiveness in accurately diagnosing faults and predicting failures.

https://doi.org/10.36001/IJPHM.2025.v16i1.4254

1. INTRODUCTION

Bearings are among the most common and crucial parts in industrial equipment, and the ability to monitor bearing health is essential to ensure safety and minimize maintenance costs of industrial equipment. Fault diagnosis and prognosis of bearings are fundamental aspects of bearing condition monitoring, where fault diagnosis determines the locations and types of faults, and fault prognosis predicts the Remaining Useful Life (RUL) of the bearing (Howard, 1994; Lee & Su, 2024; Qiao et al., 2024; Su & Lee, 2024b). Traditional model-based approaches often rely on interpretations of physical laws, which usually fail to capture the behaviors of complex systems. In contrast, Machine Learning (ML) is a better alternative, which classifies and predicts bearing health states based on historical data (Berghout & Benbouzid, 2022), and Deep Learning (DL) is a field within ML that has gained increasing popularity in recent years due to its automatic feature extraction.

In the field of fault diagnosis, several DL models have been proposed and modified to enhance the extraction of high-quality features, thereby improving the prediction accuracy. Inspired by the improved prediction accuracy of Convolution Neural Network (CNN) in fruit classification and fruit quality detection(Naranjo-Torres et al., 2020), Hoang et al. (Hoang & Kang, 2019) proposed Vibration Image CNN (VI-CNN) to capture the spatial features for the vibration signals. This proposed VI-CNN model attained a remarkable prediction accuracy of 100% on the CWRU dataset. Furthermore, some neural networks have been added to enhance the prediction performance of CNN models. On the one hand, an ensemble of Residual Networks (ResNet) was proposed by Kaiming et al. (He et al., 2016) achieved a remarkable prediction accuracy of 96.43% on the ImageNet test set, securing first place at ILSVRC 2015. Su et al. (Su & Lee, 2024a) proposed a residual-based deep CNN (ResCNN)

to predict degradation level, achieving scores of 574/813 using the PHM North America 2023 Conference Data Challenge. Long et al. (Wen et al., 2020) advocated the use of Transfer CNN (TCNN) with ResNet-50, which exhibited prediction accuracies as high as 98.95% for the CWRU bearing dataset. On the other hand, the integration of Long Shot Term Memory (LSTM) with CNN aimed to extract both spatial-temporal features to enhance predictive performance. For instance, Chen et al. (Chen et al., 2021) proposed a Multi-Scale CNN-LSTM (MSCNN-LSTM) model to predict, and compared its efficiency with the Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), and LSTM. These models were tested using the CWRU bearing dataset to categorize bearing fault type. Results indicated that the MSCNN-LSTM model achieved a prediction accuracy of 98.6%. In addition, the Bidirectional LSTM (Bi-LSTM), an LSTM variant, integrated with CNN was proposed by You et al. (You et al., 2021), to extract the bidirectional temporal feature and spatial features. The proposed CNN-Bi-LSTM model achieved a prediction accuracy of 99.2%.

In the field of fault prognosis, various DL models have been proposed and modified to enhance the extraction of high-quality features, thereby improving the prediction accuracy of RUL. Bingxi Zhao et al. (B. Zhao & Yuan, 2021) introduced Deep CNN (DCNN) to extract spatial features on the XJTU-SY bearing dataset. While CNN adepts at capturing spatial features within time series data but still often struggles to establish temporal dependencies. In order to address this limitation, Maan and Harsha et al. (Rathore & Harsha, 2022) used LSTM, Stacked LSTM, Bi-LSTM, and Bi-LSTM with self-attention mechanism for fault prognosis on self-collected bearing data. As a result, Bi-LSTM models performed better than LSTM models, and Bi-LSTM with self-attention mechanism achieved lowest Mean Absolute Error (MAE) of 0.0176, and Root Mean Square Error (RMSE) of 0.0209. Several methods are proposed to further improve the performance of the LSTM network. On the one hand, Xin Guo et al. (Guo et al., 2023) proposed CNN-Transformer to enhance its ability to improve the long-term temporal features extraction, achieving RMSE of 0.105 and MAE of 0.074. On the other hand, Brain Walker et al. (Walker, n.d.) combined the advantage of MSCNN and LSTM, proposing the Hybrid CNN and LSTM (MSCNN-LSTM) model to capture the spatial-temporal features to enhance the features' quality. This model achieved an RMSE of 0.0645 using the XJTU-SY bearing dataset. In addition, the Monte Carlo dropout (MC dropout) technique is applied by Jinsong Yang (Yang et al., 2022) to randomly dropout the neurons in LSTM for predicting the probability distribution of RUL. This approach resulted in a 36% improvement compared to LSTM in bearing RUL prediction.

Despite the proposals of various DL models, such as CNN, LSTM, CNN-LSTM, transformer along with powerful ML tools such as ResNet, dropout, and Monte Carlo Dropout (MC dropout), the prediction performance of DL models and the ML tools for both effective fault diagnosis and prognosis in industrial equipment remains uncertain. However, in complex machinery systems, such as those in the aerospace industry, merely detecting the fault location and mode, or predicting the remaining useful life (RUL) of a failure, is insufficient for conducting efficient maintenance with the lowest maintenance costs. Therefore, this study aims to find optimized DL models in both tasks by comprehensively comparing these DL models and ML tools based on their prediction accuracy using the CWRU bearing dataset for fault diagnosis and the XJTU-SY dataset for fault prognosis. Specifically, this paper is organized as follows: Section 2 outlines the architectures of RUL prediction and introduces the DL models and the ML tools in detail. Section 3 delineates datasets and the results of DL models and ML tools across various tasks. Section 4 synthesizes the findings discussed in this paper.

2. METHODOLOGY

This section introduces the framework for fault diagnosis and prognosis used in this study and explains the working principle of the DL models discussed in this study. Section 2.1 and Section 2.2 introduce the full framework and the structures of DL models, respectively.

2.1. Framework for Fault Diagnosis and Prognosis

The flowchart consists of data pre-processing, model training, and model testing, illustrated in the Figure 1. The data pre-processing phase involves removing outliers, and normalizing data within the range of 0 to 1 to facilitate rapid convergence in subsequent processes. Subsequently, five-fold cross-validation techniques involve splitting the data into K equally sized folds or subsets. The model is trained on K-1 folds and tested on the remaining fold. This process is repeated K times, with each fold used as a test set once, to mitigate overfitting and strive for optimal model performance. In the training process of five-fold cross-validation, the training dataset is further split into training data and validation data. The training data is utilized for model training, the validation data aids in predicting targets and recording losses, and the results are juxtaposed with those from the training data to identify overfitting or underfitting instances. This training process stops until the overfitting happens. Finally, the test data is utilized to evaluate the trained model's performance.

2.2. Deep Learning (DL) Model

2.2.1. Convolution Neural Network (CNN) model

CNN was initially developed for image processing but has found versatile applications in processing complex data. A standard CNN comprises convolutional and pooling layers as features extraction, also referred to as sub-sampling layers,

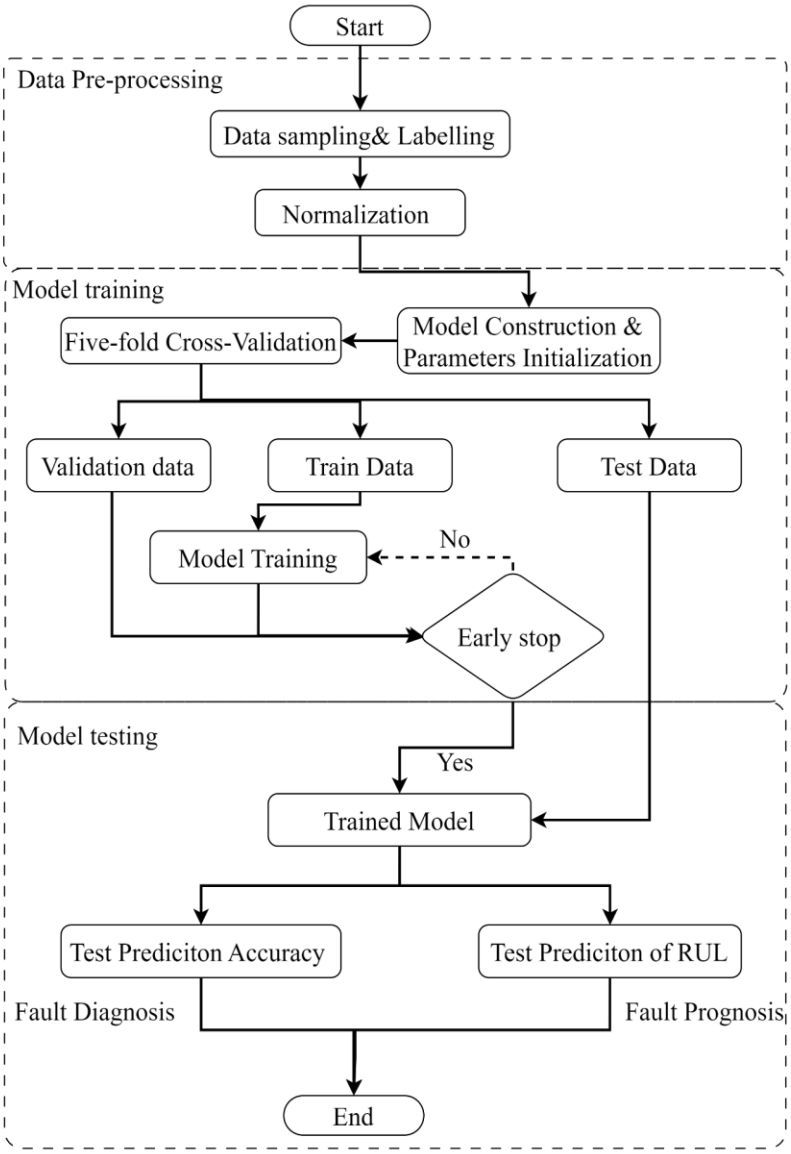

Figure 1. Flowchart of the proposed methodology.

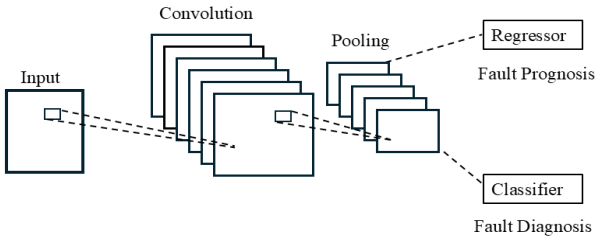

Figure 2. Overview of the CNN model for fault diagnosis and fault prognosis.

where the outputs are subsequently forwarded to a classifier or regressor, as illustrated in Figure 2 (Z. Li et al., 2022). Specifically, Convolutional layers employ sets of kernels as filters to extract features, while pooling layers down-sample the extracted features to decrease computational complexity and mitigate overfitting. The classifier and regressor are the downstream task for fault diagnosis and prognosis, typically implemented as one-layer linear neural networks.

2.2.2. Long Short-Term Memory (LSTM) model

LSTM, an enhanced Recurrent Neural Network (RNN), mitigates the challenges of exploding and vanishing gradient prevalent in traditional RNNs. Mainly employed for time-series data analysis, LSTM excels at capturing sequential dependencies by retaining and processing information from past inputs across extended sequences, followed by downstream tasks of fault diagnosis and prognosis similarly depicted in Figure 3 (Berghout & Benbouzid, 2022; C. Zhao

et al., 2020). LSTM leverages a gate mechanism comprising forget, input, and output gates to manage the flow and retention of features, making it particularly adept at handling long time series data compared to RNN. The architecture of a typical LSTM cell is depicted in Figure 3. Within an LSTM cell, the hidden h_{t-1} and cell states c_{t-1} from the preceding cell are merged with the current input x_t to compute the present cell state h_t and hidden state c_t, which also serves as the current output. Activation functions $tanh$, in conjunction with learned weights and biases, are employed to govern the information flow within the cell.

Furthermore, the Bi-LSTM network is proposed to extract the temporal bi-features from both the past and the future, while the LSTM network extracts temporal features only from the past. A typical Bi-LSTM is shown in Figure 4.

2.2.3. Hybrid Convolution Neural Network and Long Short-Term Memory (CNN-LSTM)

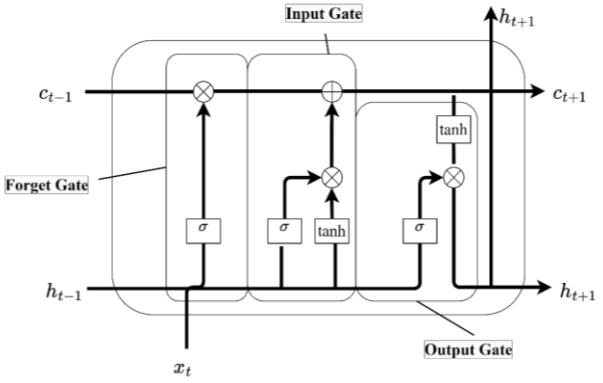

Figure 3. Structure of one LSTM Cell. (Berghout & Benbouzid, 2022; C. Zhao et al., 2020)

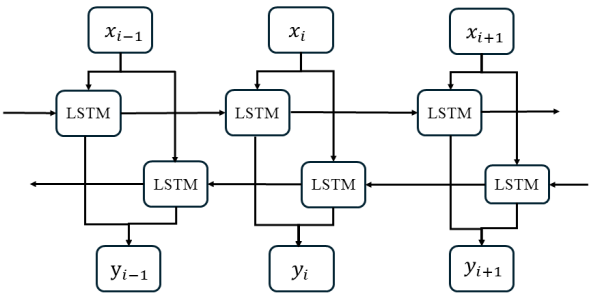

Figure 4. Academic diagram of the Bi-LSTM.

By amalgamating the strengths of the CNN model in capturing spatial features and the LSTM model in capturing temporal features, the CNN-LSTM model effectively processes time series data, concurrently handling and merging the extracted features for fault diagnosis or prognosis.

2.2.4. ML tools

Apart from the predominantly utilized DL models, various ML tools are also leveraged to boost the efficacy of DL models. One ML tool is the Residual Network (ResNet) introduced by Kaiming et al. (Hoang & Kang, 2019) to avoid the challenges of gradient vanishing or exploring that arise with the addition of more neural layers. With an escalation in the number of neural layers, DL models can extract high-level features (C. Zhao et al., 2020). To counter this, a residual connection is established by adding the input to the output within the block, as depicted in Figure 5.

Another ML tool to boost the efficacy of DL models is the regularization technique, dropout. Semeniuta et al. (M. Li et al., 2019) proposed dropout techniques, which extend the dropout concept to Fully Connected Neural Networks (FCNNs). Dropout is a standard regularization technique used in FCNNs to prevent model overfitting, which works by setting a dropout rate to randomly dropout neurons in the hidden layers, as shown in Figure 6(b) to enhance the robustness of FCNNs. Moreover, Monte Carlo Dropout (MC dropout) has been proposed to estimate the uncertainty of DL models (Gal & Ghahramani, 2016; Wang et al., 2023). A CNN operates deterministically by optimizing a single set of fixed model parameters post-training. The output remains consistent when such a network is repeatedly evaluated with the same input. MC dropout addresses this limitation by repeatedly assessing a given input with the trained network, introducing dropout to the layers of the network in each iteration to attain the output distribution.

This study delves into the examination of the CNN, LSTM, and CNN-LSTM models, along with ResNet and MC dropout.

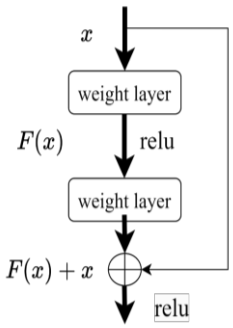

Figure 5. Structure of ResNet (Hoang & Kang, 2019)

3. CASE STUDY

In this article, the RTX 4070 Ti Super graphics card is employed to train the DL model, and PyTorch 2.2.0 is used for developing the DL framework.

3.1. Data Description and Pre-processing

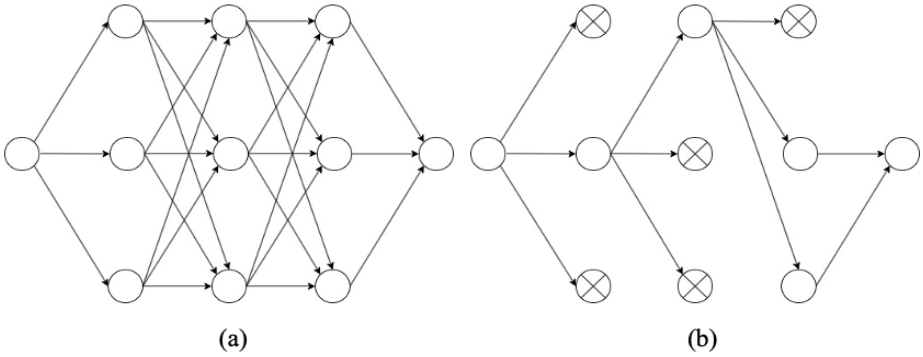

Figure 6. (a) Fully Connected Neural Network; (b) Fully Connected Neural Network with dropout. (M. Li et al., 2019)

3.1.1. CWRU dataset

The CWRU dataset does not contain a complete run-to-failure dataset, it is used exclusively for fault diagnosis. The CWRU dataset is one of the most utilized datasets for this purpose, offering vibration signals of normal bearings and those with faults in the inner raceway, ball, and outer raceway under 3 motor load (Smith & Randall, 2015). Data are labeled according to health states and fault locations: 0 for Healthy, 1 for Inner Race, 2 for Ball, and 3 for Outer Race, facilitating multiclass classification. There are three operating conditions with different motor loads, named HP0, HP1, and HP2. Subsequently, the max-min normalization is applied to range the data from 0 to 1, as described in Eq. (1). Finally, 1000, 400, and 500 data are chosen randomly as training, validation, and testing data.

$$x^* = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

3.1.2. XJTU-SY dataset

The XJTU-SY dataset includes run-to-failure horizontal and vertical vibration signals from 15 bearings under 3 different working conditions, as outlined in Table 1 (Yaguo et al., 2019). This dataset is applicable for both fault prognosis and diagnosis tasks. For fault diagnosis, the vibration data are labeled based on the Fault Occurrence Time (FOT), calculated in (Wang et al., 2023) using the time-varying 3σ criterion on the Root Mean Squared (RMS) values of horizontal vibration signals. Data is labeled as healthy before reaching the FOT and as faulty after the FOT. For fault prognosis, the sliding window technique is employed for data sampling, as shown in Figure 7. This technique, used in computer science and signal processing, involves selecting a fixed-size subset or "window" from a larger dataset and moving this window through the dataset incrementally. The fixed-size subset is used as input, and the last value of this subset serves as the RUL label for DL model training.

3.1.3. IEEE PHM Dataset

The IEEE 2012 PHM bearing dataset is a collection of experimental data from 17 ball bearings provided by the IEEE Reliability Society (Nectoux et al., 2012). The dataset consists of sets of six and 11 bearings used for training and testing purposes, respectively. The data were collected from an accelerated degradation test rig, shown in Figure 8, in which bearings were simultaneously tested until failure under constant radial load and rotational speed values. The bearings were tested for three distinct operating conditions:

Operating condition	Datasets	Working life	Fault Occurrence Time
1	Bearing 1_1	2 h, 3 min	78 min
	Bearing 1_2	2 h, 41 min	37 min
	Bearing 1_3	2 h, 38 min	59 min
	Bearing 1_4	2 h, 2 min	88 min
	Bearing 1_5	52 min	35 min
2	Bearing 2_1	8 h, 11 min	453 min
	Bearing 2_2	2 h, 41 min	47 min
	Bearing 2_3	8 h, 53 min	128 min
	Bearing 2_4	42 min	31 min
	Bearing 2_5	5 h, 39 min	1122 min
3	Bearing 3_1	42 h, 18 min	2341 min
	Bearing 3_2	41 h, 36 min	1232 min
	Bearing 3_3	6 h, 11 min	342 min
	Bearing 3_4	25 h, 15 min	1418 min
	Bearing 3_5	1 h, 54 min	21 min

Table 1. Overview of XJTU-SY bearing dataset.

Figure 7. Sliding window technique.

- Condition 1: A rotational speed of 1,800 r/min and a radial force of 4,000 N,
- Condition 2: A rotational speed of 1,650 r/min and a radial force of 4,200 N,
- Condition 3: A rotational speed of 1,500 r/min and a radial force of 5,000 N.

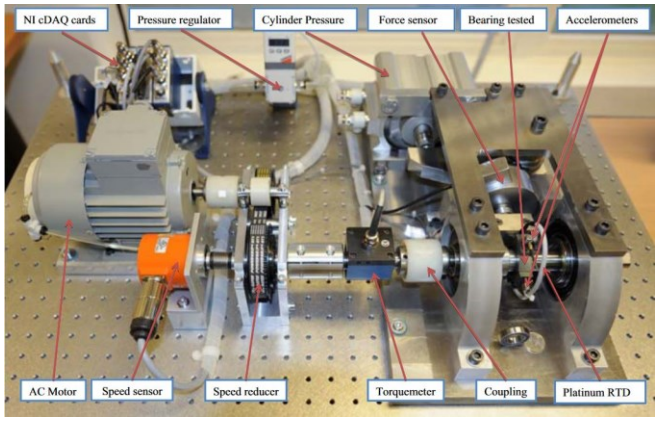

Figure 8. Photograph of the experimental platform used to investigate accelerated bearing degradation (Nectoux et al., 2012).

Datasets	Operating Conditions		
	Condition 1	**Condition 2**	**Condition 3**
Training dataset	Bearing 1_1	Bearing 2_1	Bearing 3_1
	Bearing 1_2	Bearing 2_2	Bearing 3_2
Test dataset	Bearing 1_3	Bearing 2_3	Bearing 3_3
	Bearing 1_4	Bearing 2_4	
	Bearing 1_5	Bearing 2_5	
	Bearing 1_6	Bearing 2_6	
	Bearing 1_7	Bearing 2_7	

Table 2. Overview of the IEEE 2012 PHM bearing dataset.

The experimental platform for the acceleration bearings included three parts: a rotary mechanism, a radial force generator, and data acquisition equipment. An AC motor applied rotational energy to the shaft, while a force actuator generated a force that acted on the bearing. At the measurement point, vibration signals along the vertical and horizontal axes and a temperature signal were collected using vibration and temperature sensors. The collected vibration and temperature data were then divided into 6 training and 11 test datasets, as shown in Table 2. The training dataset was used to train the model, while the test dataset was utilized to verify the trained model's performance. Both the training and test datasets include the horizontal and vertical acceleration and temperature signals collected by the DYRAN 3035B accelerometers and the RTD PT100 PROSENSOR.

3.2. Comparative study

3.2.1. Fault diagnosis

As mentioned in Section 3.1, the CWRU and XJTU-SY bearing datasets are utilized for the task of fault diagnosis. F1 scores and prediction accuracy are the evaluation metrics of the task of fault diagnosis, as described in Eqs. (2~3). Table 3 provides the specifications of hyperparameters for the DL models during training, including the loss function, optimizer, batch size, learning rate, and number of epochs. Below are the results of the comparative study for the CWRU and XJTU-SY bearing datasets.

$$F1 - score = (2 * TP)/(2 * TP + FP + FN) \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Where TP means the true positive, TN denotes the true negative, FP is the false positive, and TN is the true negative.

Hyperparameter	CWRU	XJTU-SY
Loss function	Sparse categorical cross-entropy	
Optimizer	SGD	Adam
Batch Size	256	100
Learning Rate	$1e-1, 1e-2, 1e-3, 1e-4$	
Epoch	10,30	10,20,35

Table 3. Specification of hyperparameters for the task of fault diagnosis.

CWRU bearing dataset

Table 4 lists the abbreviations of the proposed models to facilitate easy comparison. Figure 9 illustrates the prediction accuracy and F1-score using the CWRU dataset for fault diagnosis. Specifically, Figure 9(a) presents prediction accuracy and F1-score under different motor loads—HP1, HP2, and HP3. The orange bars represent the prediction accuracy and yellow bars demonstrate the F1-score. All basic models achieved prediction accuracy above 99% and F1-score above 0.95. The CNN model emerged as the best model, outperforming the LSTM under HP1 and HP3, the CNN-LSTM under HP1, and transformer under HP1, HP2, and HP3. This superiority is attributed to the CNN model's efficiency in extracting spatial features, which is crucial for detecting health and failure in bearings during fault diagnosis. This is because the LSTM and transformer models struggle with extracting temporal features effectively, Additionally, the CNN-LSTM model outperformed LSTM by 0.11% under HP3, showing similar results under HP1 and HP2 compared to the LSTM. This improvement is due to the integration of LSTM with the CNN model, forming the CNN-LSTM model, which compensates for the LSTM's shortcomings in extracting spatial features. Furthermore, the average prediction accuracy and F1-score were calculated, as shown in Figure 9(b). The prediction accuracy and F1-score represent the average results under a fused dataset that includes HP1, HP2, and HP3. Results indicate that the CNN-

LSTM model and its variants demonstrate the highest prediction accuracy among the CNN, LSTM, and transformer models and their respective variants. This is due to the complexity of the fused bearing dataset compared to data from a single motor load. The CNN model alone struggles to extract high-level features effectively, while the CNN-LSTM model addresses this limitation by combining spatial and temporal feature extraction capabilities, which better suits complex datasets. Among the variations of the CNN-LSTM model, the ResCNN-LSTM and ResCNN-Bi-LSTM models achieved the highest prediction accuracy, with 99.87% and 99.85%, respectively. This suggests that the ML tool of ResNet can enhance the performance of the CNN-LSTM and CNN-Bi-LSTM models by solving the vanishing and exploding gradient problem of the CNN-LSTM. Although MC dropout can provide a confidence level when diagnosing faults, it slightly compromises the prediction accuracy and F1-score of the CNN, LSTM, and CNN-LSTM models.

XJTU-SY Dataset

Figure 10 illustrates the prediction accuracy using the XJTU-SY bearing dataset for fault diagnosis. The CNN-LSTM and its variants achieved the highest prediction accuracy, while the transformer model demonstrated the highest F1-score.

This suggests that while CNN-LSTM excels in overall accuracy, the transformer model is better at balancing precision and recall, which can be crucial in applications where false positives and false negatives have different costs. Among all the variants in the CNN-LSTM model, ResCNN-LSTM-MC, ResCNN-Bi-LSTM, and ResCNN-LSTM attained 96.42%, 96.4%, and 96.39% prediction accuracy, respectively. ML tools, such as MC dropout, recurrent dropout, and ResNet, enhance the performance of the CNN, LSTM, and CNN-LSTM models for the XJTU-SY bearing dataset in the task of fault diagnosis. This improvement is due to MC dropout preventing overfitting during the training process, and ResNet mitigating the gradient vanishing and explosion issues in deeper neural layers. In contrast, employing MC dropout in DL models does not improve the performance for fault diagnosis using the CWRU bearing dataset. This discrepancy is likely because the CWRU dataset lacks degradation patterns, resulting in lower performance, whereas the XJTU-SY bearing dataset includes degradation patterns.

3.2.2. Fault Prognosis

For the task of fault prognosis, MAE and RMSE are the main evaluation metrics, as described in Eqs. (4~5).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |RUL_i - \widehat{RUL_i}| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (RUL_i - \widehat{RUL_i})^2} \tag{5}$$

Model	Abbreviation
CNN	C
LSTM	L
CNN-LSTM	C_L
CNN-Bi-LSTM	C_BL
Model with MC Dropout	_MC
Model with Recurrent Dropout	_Rec
Bi-LSTM	BL
ResNet	RC
Transformer	T

Table 4. Abbreviations of proposed models.

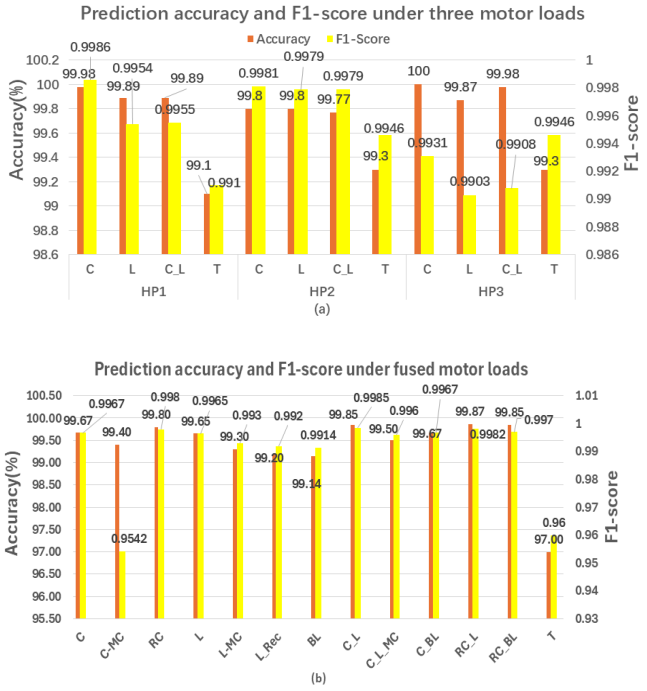

Figure 9. Prediction accuracy and F1-score using CWRU bearing dataset for the task of fault diagnosis:(a) under three motor loads;(b) under fused motor loads.

Where the RUL_i is the real RUL label, and the $\widehat{RUL_i}$ is the predicted RUL label.

XJTU-SY Dataset

Figure 11 illustrates the prediction accuracy using the XJTU-SY bearing dataset for the task of fault prognosis. It shows that the LSTM model and its variants exhibit the highest RMSE and MAE, while the transformer achieves the lowest RMSE of 9.47 and the lowest MAE of 11.07. This performance is attributed to the transformer's ability to effectively handle both spatial and temporal features, making it superior to the CNN and LSTM models in fault prognosis. However, among the variants of the CNN-LSTM, the ResCNN-Bi-LSTM-MC achieved an MAE of 8.52 and RMSE of 12.26, and the ResCNN-LSTM achieved an MAE of 8.56 and RMSE of 12.16. Another model with notable

performance is the ResCNN, which achieved an MAE of 8.61 and an RMSE of 12.32. Similar to fault diagnosis, ResNet is the best ML tool for enhancing the performance of the CNN-LSTM by avoiding gradient vanishing.

IEEE PHM Dataset

Figure 12 illustrates the prediction accuracy using the IEEE-PHM bearing dataset for the task of fault prognosis. It shows that the CNN model and its variants exhibit the highest RMSE and MAE, while the transformer achieves the lowest RMSE of 11.94 and the lowest MAE of 12.13. This performance is attributed to the transformer's ability to effectively handle both spatial and temporal features, making it superior to the CNN and LSTM models in fault prognosis. However, among the variants of the CNN-LSTM, the ResCNN-Bi-LSTM achieved an MAE of 12.89 and an RMSE of 9.923. Similar to fault diagnosis, ResNet is the best tool for enhancing the performance of the CNN-LSTM by avoiding gradient vanishing.

4. CONCLUSION

The fault diagnosis and prognosis of industrial equipment are vital in the field of Prognosis Health Management (PHM) for effective and reliable operations, reducing downtime costs. However, there is a lack of comprehensive comparative studies addressing both fault diagnosis and prognosis. This study compares CNN, LSTM, CNN-LSTM, and transformer models for these tasks. Additionally, ML tools such as MC dropout, Recurrent dropout, and ResNet are integrated with

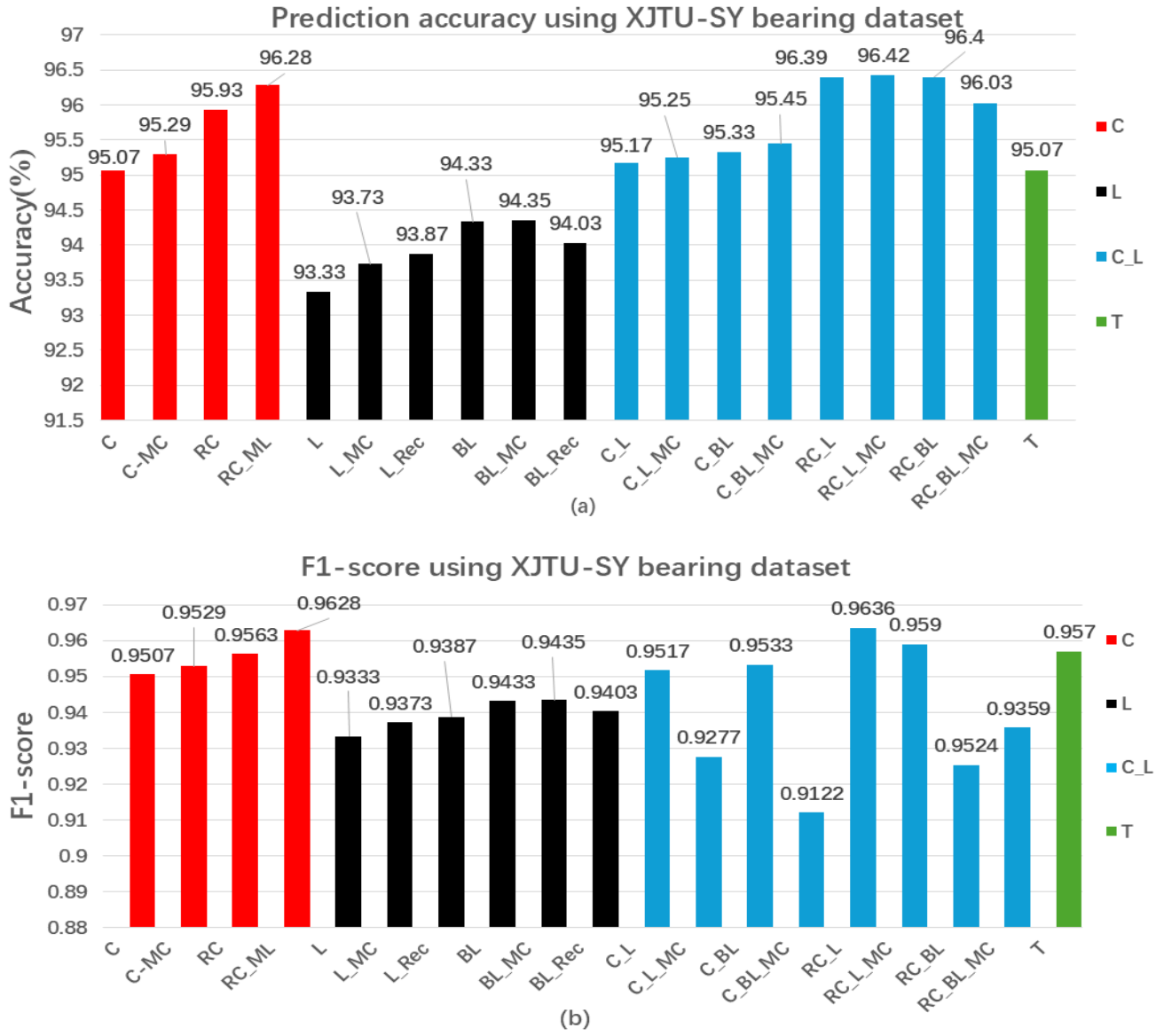

Figure 10. Prediction accuracy(a), and F1-Score (b) using XJTU-SY dataset for the task of fault diagnosis.

DL models to investigate their effects on fault diagnosis and prognosis.

Results indicate that the transformer model achieves its lowest Mean Absolute Error (MAE) of 11.07 and 12.13 when using the XJTU-SY and IEEE PHM bearing datasets, respectively. However, it performs poorly in fault diagnosis tasks. In contrast, the ResCNN-LSTM model excels in both fault diagnosis and prognosis. It achieves high prediction accuracy of 99.87% and 96.39% for diagnosis using the CWRU and XJTU datasets, respectively. Additionally, the ResCNN-LSTM model shows an RMSE of 8.53 for prognosis using the XJTU dataset, highlighting its robust performance across these tasks. Furthermore, the ResNet technique enhances the performance of DL models in both fault diagnosis and prognosis, while MC dropout improves DL model performance in fault prognosis and fault diagnosis using the XJTU-SY dataset. However, due to the limited length of the article, other DL models such as, autoencoder, Generative Adversarial Networks (GAN), and diffusion models were not explored. In the future, real-time applications in the aerospace industry will leverage the ResCNN-LSTM model to further demonstrate its superiority in both fault diagnosis and prognosis. Specifically, detecting failure modes such as compressor stalls or turbine blade damage, alongside predicting engine failure times, can significantly enhance safety and reduce maintenance costs by utilizing advanced DL tools to analyze real-time data from aircraft engines.

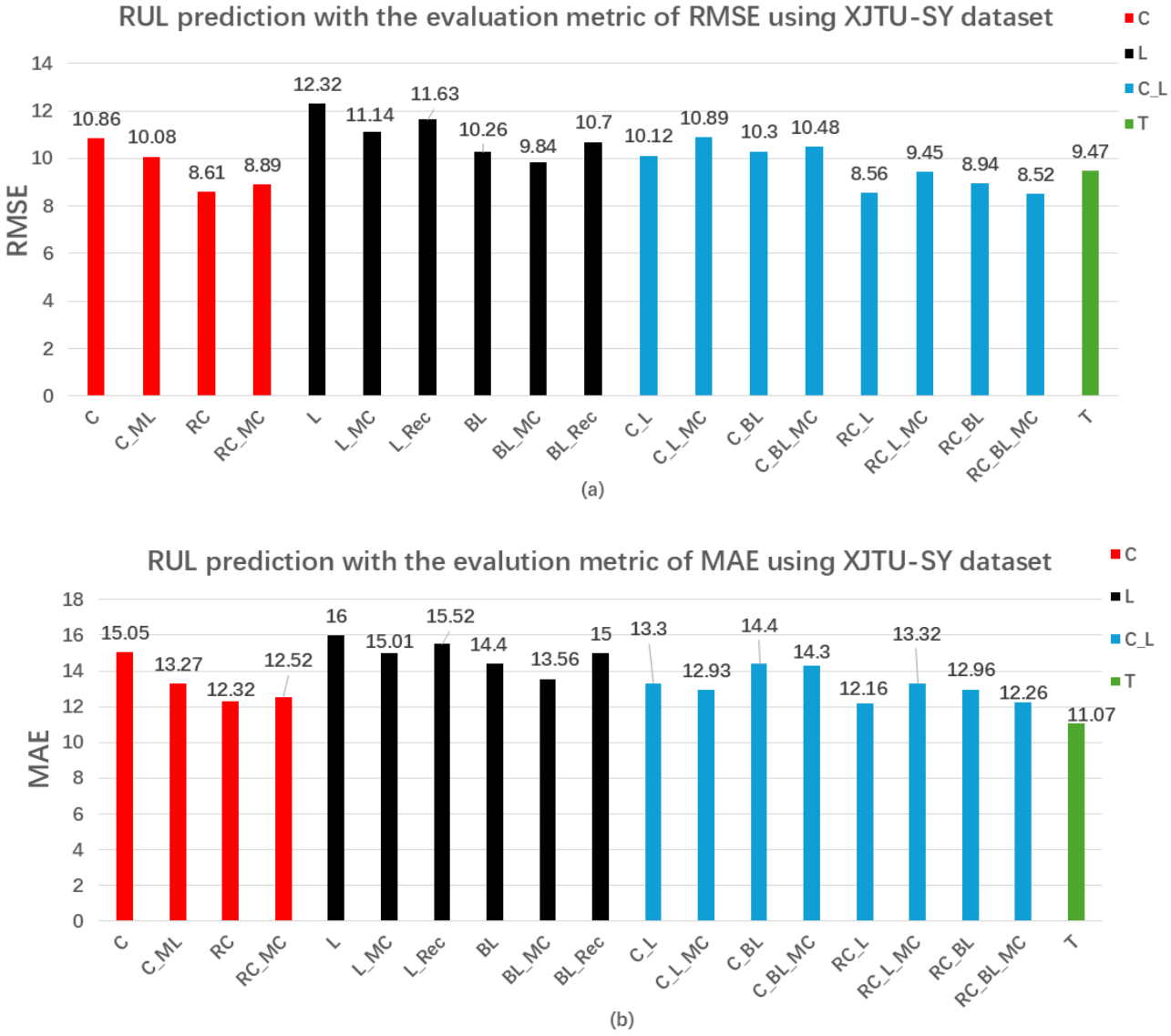

Figure 11. RUL prediction using the XJTU-SY bearing dataset for the task of fault prognosis with the evaluation metric of: (a). RMSE; (b) MAE.

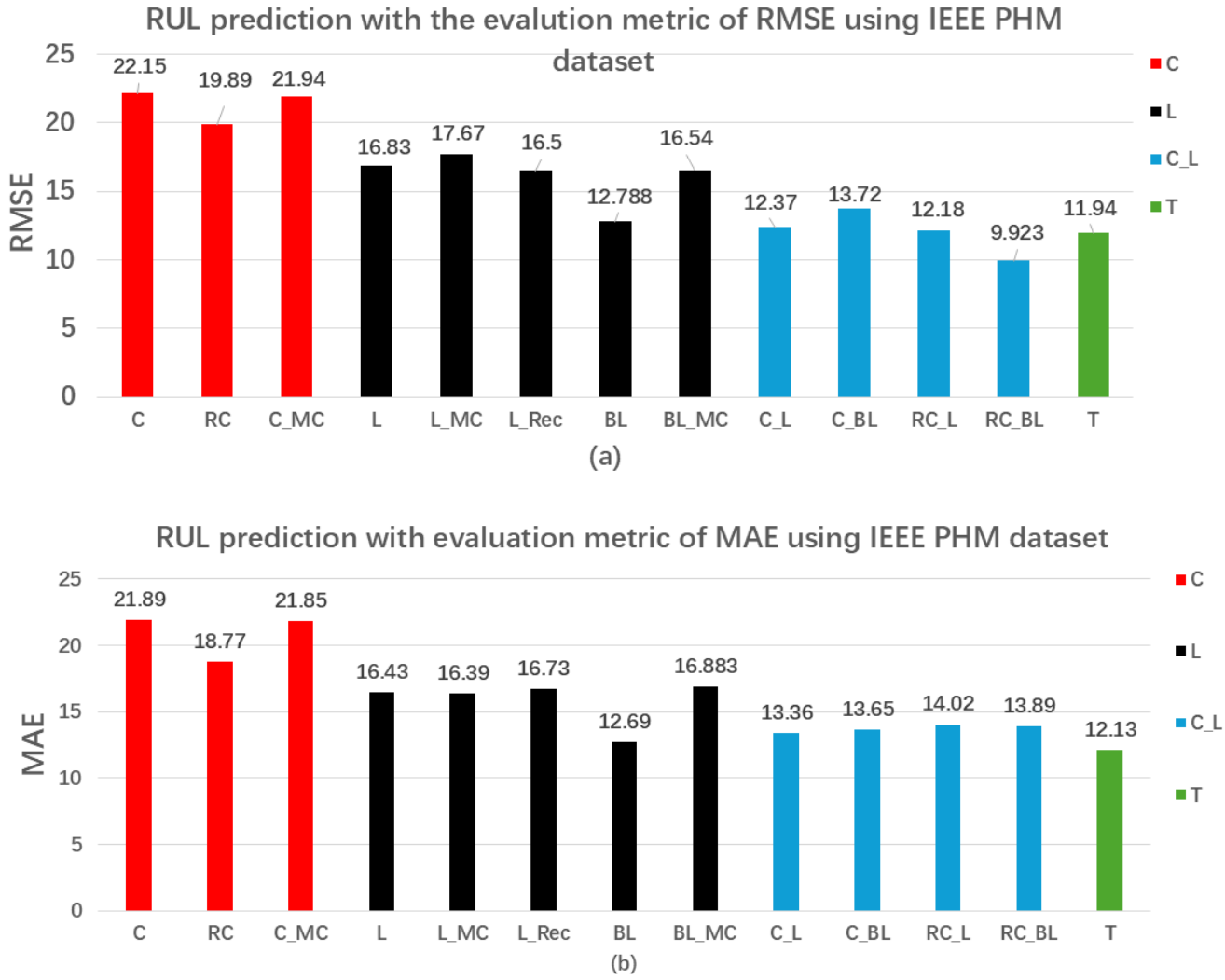

Figure 12. RUL prediction using the IEEE PHM bearing dataset for the task of fault prognosis with the evaluation metric of: (a). RMSE; (b) MAE.

ACKNOWLEDGEMENT

None

REFERENCES

Berghout, T., & Benbouzid, M. (2022). A Systematic Guide for Predicting Remaining Useful Life with Machine Learning. *Electronics*, *11*(7), 1125. https://doi.org/10.3390/electronics11071125

Chen, X., Zhang, B., & Gao, D. (2021). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, *32*(4), 971–987. https://doi.org/10.1007/s10845-020-01600-2

Gal, Y., & Ghahramani, Z. (2016). *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning* (No. arXiv:1506.02142). arXiv. http://arxiv.org/abs/1506.02142

Guo, X., Tu, J., Zhan, S., Zhang, W., Ma, L., & Jia, D. (2023). A novel method for online prediction of the remaining useful life of rolling bearings based on wavelet power spectrogram and Transformer structure. *Engineering Research Express*, *5*(4), 045074. https://doi.org/10.1088/2631-8695/ad08fc

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hoang, D.-T., & Kang, H.-J. (2019). Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research*, *53*, 42–50. https://doi.org/10.1016/j.cogsys.2018.03.002

Howard, I. (1994). *A Review of rolling element bearing vibration 'detection, diagnosis and prognosis'*.

Lee, J., & Su, H. (2024). A Unified Industrial Large Knowledge Model Framework in Industry 4.0 and Smart Manufacturing. *International Journal of AI for Materials and Design*, *1*(2), 41. https://doi.org/10.36922/ijamd.3681

Li, M., Wei, Q., Wang, H., & Zhang, X. (2019). Research on fault diagnosis of time-domain vibration signal based on convolutional neural networks. *Systems Science & Control Engineering*, *7*(3), 73–81. https://doi.org/10.1080/21642583.2019.1661311

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A. (2020). A Review of Convolutional Neural Network Applied to Fruit Image Processing. *Applied Sciences*, *10*(10), 3443. https://doi.org/10.3390/app10103443

Nectoux, P. (2012). *PRONOSTIA: An experimental platform for bearings accelerated degradation tests.* 1–8. https://doi.org/10.1109/PHM.2012.6227845

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N., & Varnier, C. (2012). *PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests*.

Qiao, X., Jauw, V. L., Chin Seong, L., & Banda, T. (2024). Advances and limitations in machine learning approaches applied to remaining useful life predictions: A critical review. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-024-14000-0

Rathore, M. S., & Harsha, S. P. (2022). Prognostics Analysis of Rolling Bearing Based on Bi-Directional LSTM and Attention Mechanism. *Journal of Failure Analysis and Prevention*, *22*(2), 704–723. https://doi.org/10.1007/s11668-022-01357-1

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*, *64–65*, 100–131. https://doi.org/10.1016/j.ymssp.2015.04.021

Su, H., & Lee, J. (2024a). Advanced Diagnostic Model for Gearbox Degradation Prediction Under Various Operating Conditions and Degradation Levels. *Annual Conference of the PHM Society*, *16*(1). https://doi.org/10.36001/phmconf.2024.v16i1.3869

Su, H., & Lee, J. (2024b). Machine Learning Approaches for Diagnostics and Prognostics of Industrial Systems Using Open Source Data from PHM Data Challenges: A Review. *International Journal of Prognostics and Health Management*, *15*(2). https://doi.org/10.36001/ijphm.2024.v15i2.3993

Walker, B. (n.d.). *RUL Prediction for Bearings Using MSCNN and LSTM Networks*.

Wang, L., Cao, H., Ye, Z., & Xu, H. (2023). Bayesian large-kernel attention network for bearing remaining useful life prediction and uncertainty quantification. *Reliability Engineering & System Safety*, *238*, 109421. https://doi.org/10.1016/j.ress.2023.109421

Wen, L., Li, X., & Gao, L. (2020). A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Computing and Applications*, *32*(10), 6111–6124. https://doi.org/10.1007/s00521-019-04097-w

Yaguo L., Tianyu H., Biao W., Naipeng L., Tao Y., & Jun Y. (2019). XJTU-SY Rolling Element Bearing Accelerated Life Test Datasets: A Tutorial. *Journal of Mechanical Engineering*, *55*(16), 1. https://doi.org/10.3901/JME.2019.16.001

Yang, J., Peng, Y., Xie, J., & Wang, P. (2022). Remaining Useful Life Prediction Method for Bearings Based on LSTM with Uncertainty Quantification. *Sensors*, *22*(12), 4549. https://doi.org/10.3390/s22124549

You, D., Chen, L., Liu, F., Zhang, Y., Shang, W., Hu, Y., & Liu, W. (2021). Intelligent Fault Diagnosis of Bearing Based on Convolutional Neural Network and Bidirectional Long Short-Term Memory. *Shock and Vibration*, *2021*(1), 7346352. https://doi.org/10.1155/2021/7346352

Zhao, B., & Yuan, Q. (2021). A novel deep learning scheme for multi-condition remaining useful life prediction of rolling element bearings. *Journal of Manufacturing Systems*, *61*, 450–460. https://doi.org/10.1016/j.jmsy.2021.10.004

Zhao, C., Huang, X., Li, Y., & Yousaf Iqbal, M. (2020). A Double-Channel Hybrid Deep Neural Network Based on CNN and BiLSTM for Remaining Useful Life Prediction. *Sensors*, *20*(24), 7109. https://doi.org/10.3390/s20247109