Supplementary Work

1. Related theory and Literature Review

The LSTM network is a special kind of Recurrent Neural Network (RNN) and is capable of solving long-term dependencies problems for which RNN was deficient. An LSTM unit is comprised of a memory cell, a forget gate (f_t), an input gate (i_t), and an output gate (O_t), as shown in Fig. S.1.

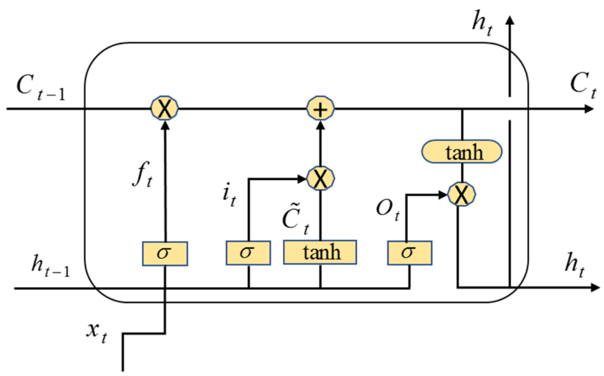


Figure S1. The structure of a single LSTM unit.

$$f_t = \sigma(\mathbf{W}_f \bullet [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) = \sigma(W_{fh} h_{t-1} + W_{fx} x_t + b_f) \tag{S1}$$

Equation (1) denotes the forget gate. It decides that how much information of cell state (C_t) from the previous time step $\mathbf{c}_{t-1}$ is needed to be kept into the current cell state at the current time step. $\mathbf{W}_f$ is the weight matrix of the forget gate, $[\mathbf{h}_{t-1}, \mathbf{x}_t]$. $\mathbf{b}_f$ is the bias of the forget gate, and σ is the sigmoid function. The input gate is defined as follows,

$$i_t = \sigma(\mathbf{W}_i \bullet [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) = \sigma(W_{ih} h_{t-1} + W_{ix} x_t + b_i) \tag{S2}$$

where $\mathbf{W}_i$ is the weight matrix of the input gate, $\mathbf{b}_i$ is the bias of the input gate. Then, the cell state which is used to describe the current input which is given as,

$$\tilde{C}_t = \tanh(\mathbf{W}_c \bullet [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) = \tanh(W_{Ch} h_{t-1} + W_{Cx} x_t + b_C) \quad \text{(S3)}$$

where $\mathbf{W}_c$ is the weight matrix of the current cell state. $\tilde{C}_t$ is the cell state of the current input, and it is calculated from the combination of previous output and the current input. $\tanh$ is a function that can push values between -1 and 1. The cell state C_t at the current time step is the sum of two items. One is the previous cell state C_{t-1} multiplying the forget gate f_t, and the result of the current input cell state $\tilde{C}_t$ multiplying the input gate i_t. This step is known as the cell state updating step, and its mathematical equation is given as,

$$C_t = C_{t-1} * f_t + i_t * \tilde{C}_t \quad \text{(S4)}$$

Therefore, the current (that can be called as short-term) memory $\tilde{C}_t$ and long-term memory C_{t-1} are combined and developed into a new cell state C_t. Due to the control of the forget gate, the LSTM unit can remember the information from a long-time distance, and due to the presence of an input gate, it can keep the current useless information in the memory cell. Finally, the output gate is defined as,

$$O_t = \mathbf{W}_o \bullet [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o = \sigma(W_{Oh} h_{t-1} + W_{Ox} x_t + b_O) \quad \text{(S5)}$$

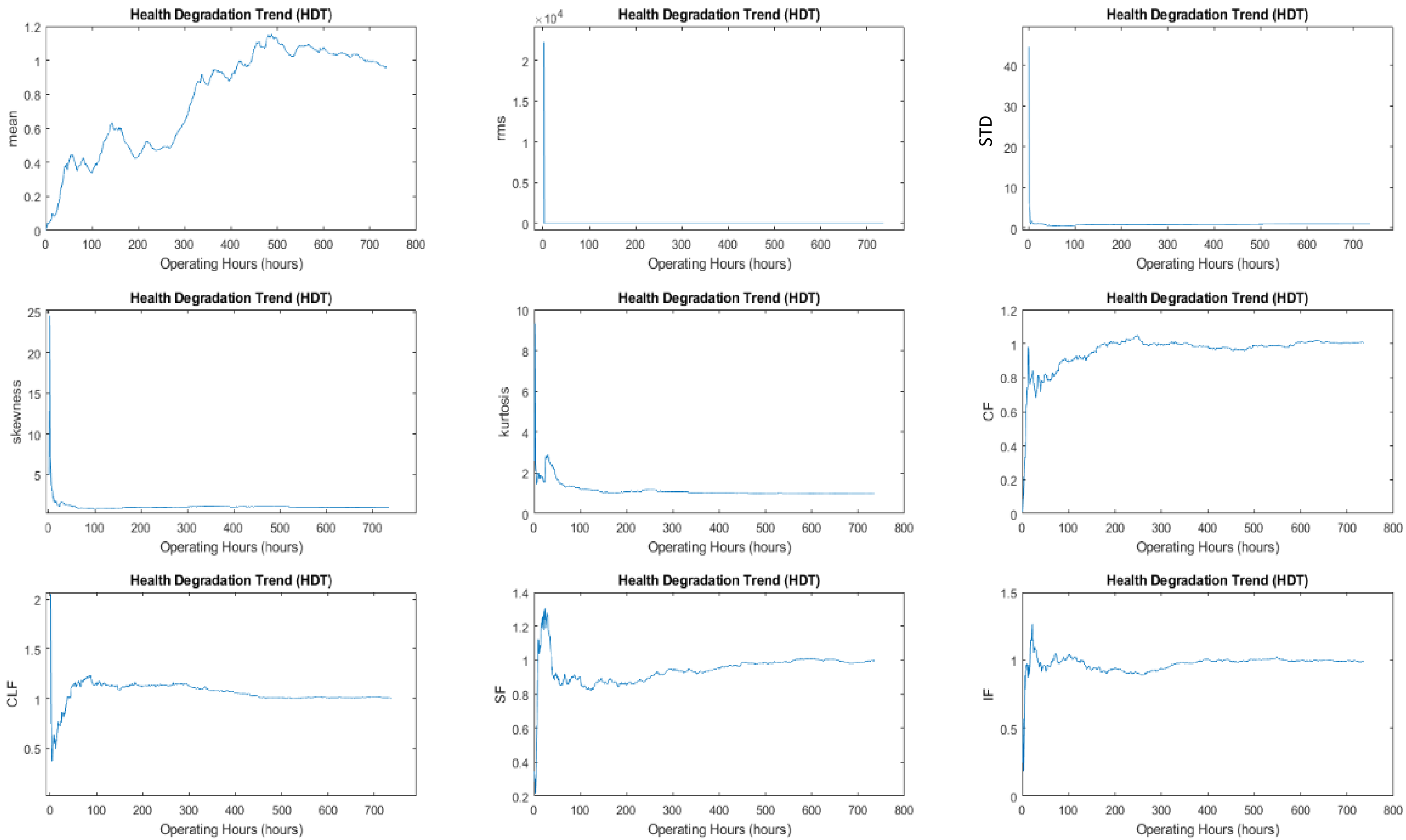
It controls the effect on the current output from the long-term memory. The final output is determined by the output gate and the new cell state i.e.

$$h_t = O_t * \tanh(C_t) \quad \text{(S6)}$$

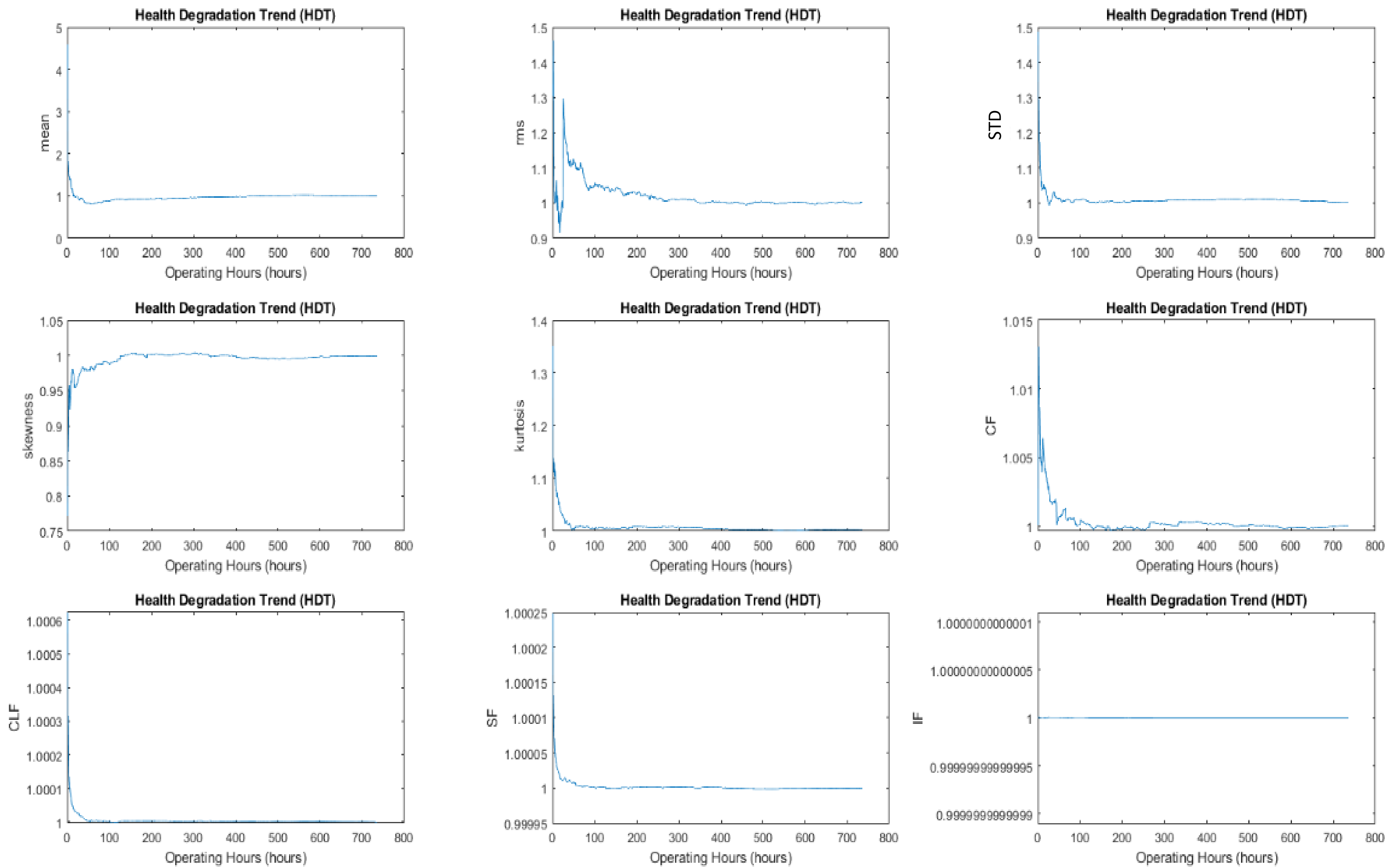
There is a great interest in the development of deep learning neural networks and its application as an alternative solution for solving the unpredictable degradation progression problems of a complex system. In this respect, an overview of the current research work is presented here.

Zhu [1] presented a new data-driven transferable method based on multiple layer perceptron, which was providing reliable transferable prognostics under various working conditions. Hinchi [2] studied a convolutional LSTM network to predict the rolling bearing element life. Yang [3] used LSTM work to conduct the natural gas pipeline safety classification problem. Zhang [4] used the LSTM, RNN network to predict the remaining useful life of the battery, which was solving the long-term dependencies problem among the degraded capacities of lithium-ion batteries. Wu [5] utilized the vanilla LSTM neural networks to obtain good RUL predictions for the industrial complex engineered systems for avoiding catastrophic failures and minimizing the economic losses. Qu [6] proposed a wind power prediction model based on the LSTM model, which showed a higher prediction accuracy and greater potential of engineering applications. Mao [7] proposed a RUL prediction approach based on the deep feature representation and LSTM method. Fan [8] studied the use of Deep Bidirectional LSTM (DBLSTM) to capture the information in future sequence data. Ozal [9] proposed a bidirectional LSTM network-based wavelet sequence called DBLSTM-WS to classify the electrocardiogram(ECG) signals. Kara Ahmet [10] predicted the RUL of lithium-ion batteries by extracting the spatio-temporal relations between the multivariate time series data and captured nonlinear characteristics. Esfahani [11] predicted the RUL of turbofan jet engines using NASA's commercial modular aero-propulsion system simulation (C-MAPSS) dataset. Unlike the other CNN-LSTM models, where features were typically extracted using CNN, which were then fed to LSTM, his developed model utilized both the algorithms organically for enhancing the prediction ability. Songhao Gao [12] utilized the phase space warping (PSW) and hidden Markov model regression for developing a scale-normalized bearing health indicator. Later on, for predicting the rolling bearing RUL, the developed indicator was utilized as the input for the encoder–decoder LSTM model with an attention mechanism. Jiahang Luo [13] predicted the RUL of rotating machinery by developing a novel convolution-based attention mechanism BiLSTM model. He focused on the cell states of Bi-LSTM. For obtaining the feature Information, the input signal was passed through the CNN. Then, for performing the convolution operation, the obtained features were fed into the Bi-LSTM network with an attention mechanism. Mohamed Marei [14] predicted the RUL of a cutting tool by developing a hybrid LSTM model with an embedded transfer learning mechanism. The novel point of the proposed method was the introduction of the transfer learning mechanism to the volume of datasets that were required for the training of the developed deep learning model. Huaqing Peng [15] developed a novel three-stage fault prediction approach for the analysis of the type of failure and the identification of the degradation period. He made a hybrid LSTM model for extracting the spatiotemporal features of the fault type and degradation period by utilizing the cross-entropy loss function. Then he predicted the type of failure by utilizing the BiLSTM network as the regression model for predicting the feature’s future trend. Satish Kumar [16] developed an approach of fusing the feature selection technique along with deep learning models. He predicted the tool wear by utilizing the NASA milling data sets along with vibration signals. for the feature selection and ranking, multiple steps were taken and different (LSTM) models were used for improving the overall RUL prediction accuracy of the developed model. Jiachen Yao [17] utilized the LSTM network for developing a transfer reinforcement learning (DTRL) network. For tracking the tool states, local features were extracted from consecutive sensor data and the trained network size was adjusted by controlling the time sequence length. For a smooth processing of the temporal information and mining long-term dependencies, the LSTM network was utilized to construct the value function approximation.

Thus, the above-mentioned literature revealed that LSTM modeling has the outstanding potential for capturing the temporal and long-term information [18], that is why a lot of researchers have utilized LSTM models alone or with the combination of some other models for determining the RUL of different engineering equipment. Having a view of the above-given knowledge, the authors also utilized a hybrid deep LSTM based modeling for predicting the online RUL of a slurry pump in the absence of its run to failure data.

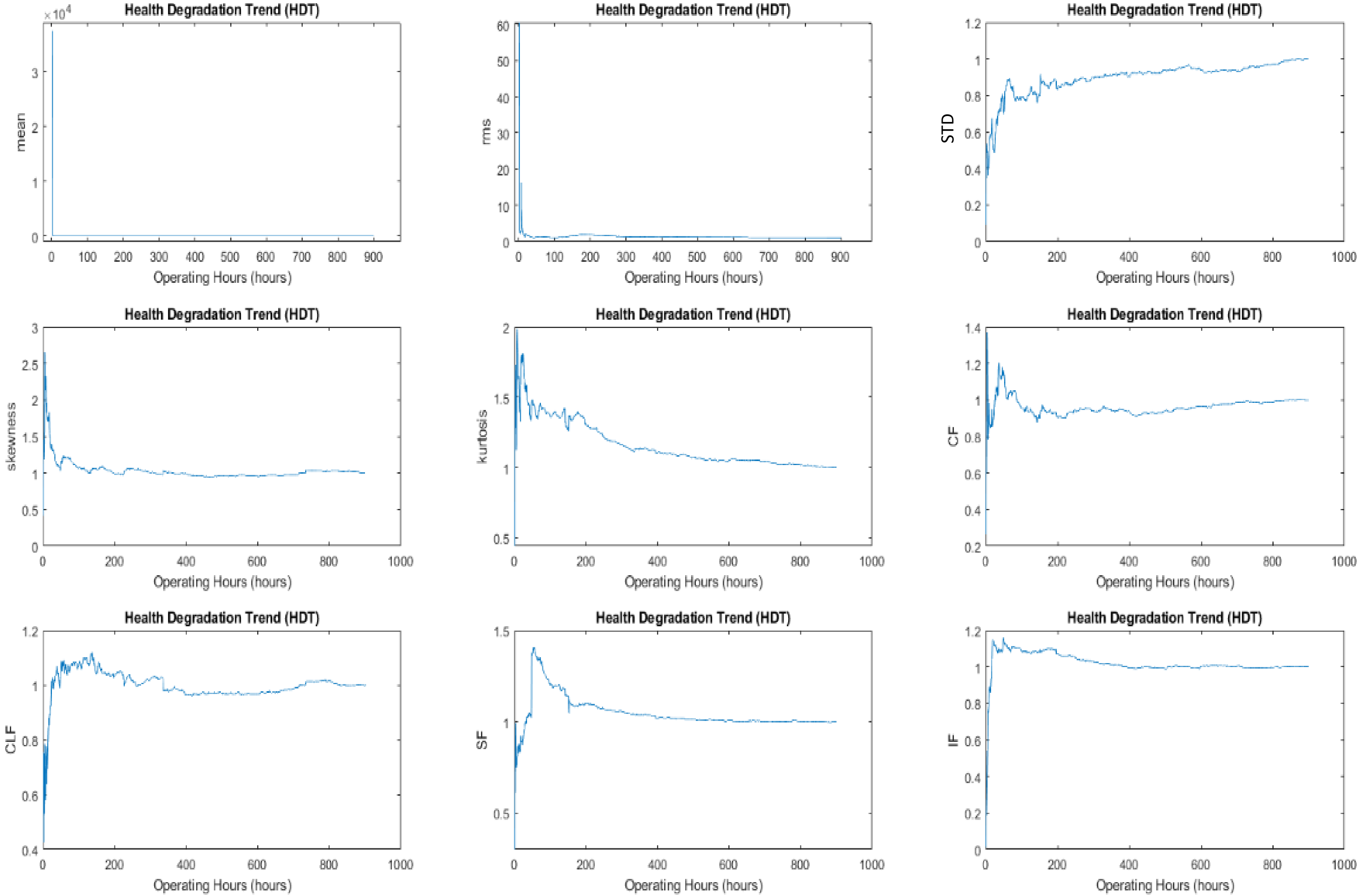

Health Degradation Trend (HDT)
mean
rms
×10^4
STD
skewness
kurtosis
CF
CLF
SF
IF
Operating Hours (hours)

(a)

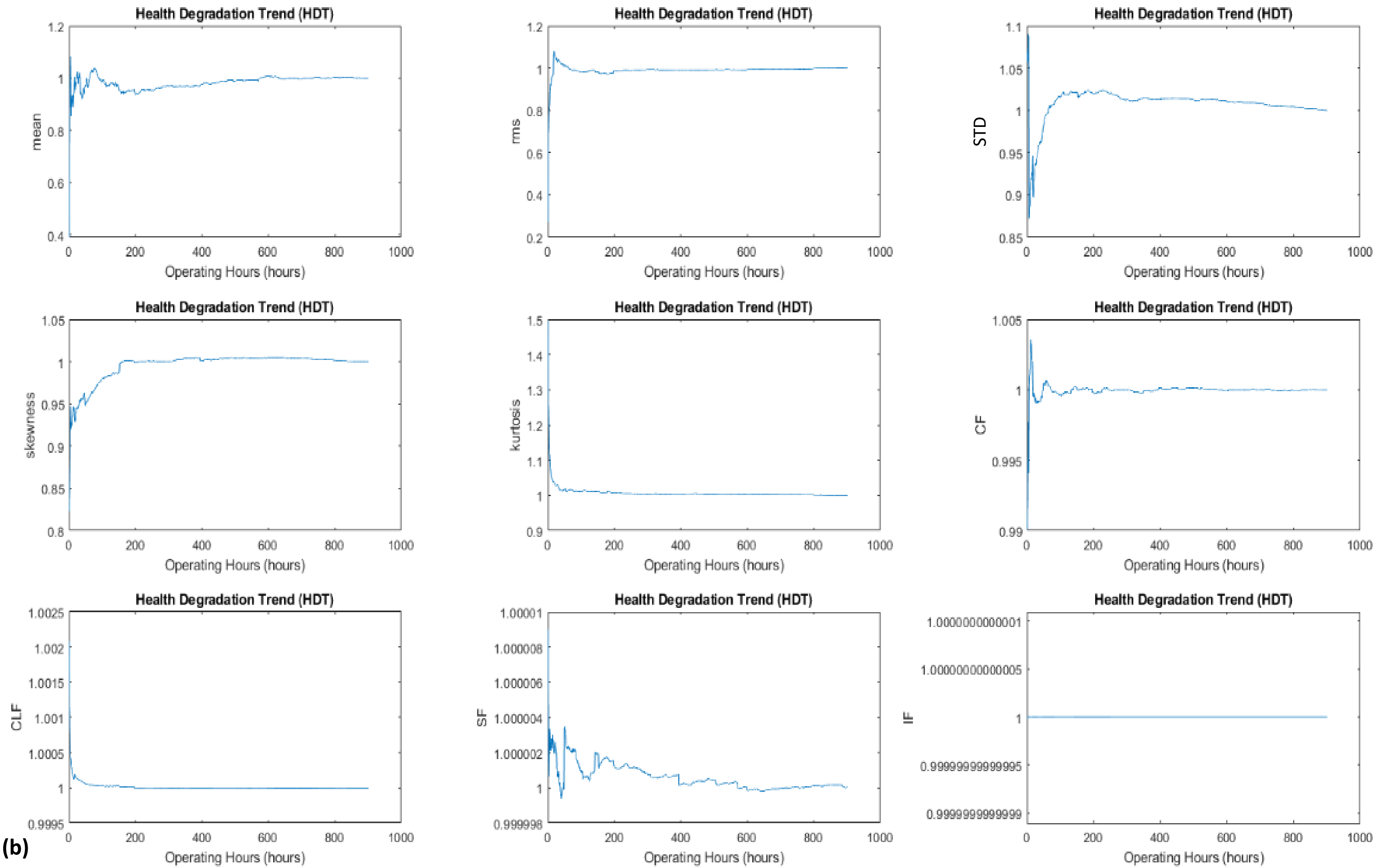


(b)

Figure S2. Developed HDTs for 18 traditional statistical features (a) nine in the time domain, and (b) nine in the frequency domain, for channel 2.


Health Degradation Trend (HDT)
mean
×10^4
rms
STD
skewness
kurtosis
CF
CLF
SF
IF
Operating Hours (hours)

(a)



(b)

Figure S3. Developed HDTs for 18 traditional statistical features (a) nine in the time domain, and (b) nine in the frequency domain, for channel 4.

Table S1. 18 Statistical features extracted in time and frequency domains (X_n represents the valid vibration data in temporal and spectral domains)

(1) Mean:	(2) Standard deviation (STD):	(3) Root mean square:
$\mu = \frac{1}{N}\sum_{n=1}^{N} X_n$	$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(X_n-\mu)^2}$	$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n)^2}$
(4) Skewness:	(5) Kurtosis:	(6) Crest Factor:
$SK = \frac{\sum_{n=1}^{N}(X_n-\mu)^3}{(N-1)\sigma^3}$	$KU = \frac{\sum_{n=1}^{N}(X_n-\mu)^4}{(N-1)\sigma^4}$	$CF = \frac{\max(\|X_n\|)}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n)^2}}$
(7) Clearance Factor:	(8) Shape Factor:	(9) Impulse Factor:
$CLF = \frac{\max(\|X_n\|)}{\left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\|X_n\|}\right)^2}$	$SF = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n)^2}}{\frac{1}{N}\sum_{n=1}^{N}\|X_n\|}$	$IF = \frac{\max(\|X_n\|)}{\frac{1}{N}\sum_{n=1}^{N}\|X_n\|}$

References

1. Zhu, J.; Chen, N.; Shen, C. A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. *Mech. Syst. Signal Process.* **2020**, *139*, 106602, doi:10.1016/j.ymssp.2019.106602.
2. Hinchi, A.Z.; Tkiouat, M. Rolling element bearing remaining useful life estimation based on a convolutional long-short-Term memory network. *Procedia Comput. Sci.* **2018**, *127*, 123–132, doi:10.1016/j.procs.2018.01.106.
3. An, Y.; Wang, X.; Chu, R.; Yue, B.; Wu, L.; Cui, J.; Qu, Z. Event classification for natural gas pipeline safety monitoring based on long short-term memory network and Adam algorithm. *Struct. Heal. Monit.* 2020, *19*, 1151–1159,

doi:10.1177/1475921719879071.

4. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705, doi:10.1109/TVT.2018.2805189.

5. Wu, Y.; Yuan, M.; Dong, S.; Lin, L.; Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, doi:10.1016/j.neucom.2017.05.063.

6. Kisvari, A.; Lin, Z.; Liu, X. Wind power forecasting – A data-driven method along with gated recurrent neural network. *Renew. Energy* **2021**, *163*, 1895–1909, doi:10.1016/j.renene.2020.10.119.

7. Mao, W.; He, J.; Tang, J.; Li, Y. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Adv. Mech. Eng.* **2018**, doi:10.1177/1687814018817184.

8. Fan, Y.; Qian, Y.; Xie, F.; Soong, F.K. TTS synthesis with bidirectional LSTM based Recurrent Neural Networks. *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH* **2014**, 1964–1968, doi:10.21437/interspeech.2014-443.

9. Yildirim, Ö. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202, doi:10.1016/j.compbiomed.2018.03.016.

10. Kara, A. A data-driven approach based on deep neural networks for lithium-ion battery prognostics. *Neural Comput. Appl.* **2021**, *33*, 13525–13538, doi:10.1007/s00521-021-05976-x.

11. Esfahani, Z.; Salahshoor, K.; Farsi, B.; Eicker, U. A New Hybrid Model for RUL Prediction through Machine Learning. *J. Fail. Anal. Prev.* **2021**, *21*, 1596–1604, doi:10.1007/s11668-021-01205-8.

12. Gao, S.; Xiong, X.; Zhou, Y.; Zhang, J. Bearing remaining useful life prediction based on a scaled health indicator and a lstm model with attention mechanism. *Machines* **2021**, *9*, doi:10.3390/machines9100238.

13. Luo, J.; Zhang, X. Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. *Appl. Intell.* **2022**, *52*, 1076–1091, doi:10.1007/s10489-021-02503-2.

14. Marei, M.; Li, W. Cutting tool prognostics enabled by hybrid CNN-LSTM with transfer learning. *Int. J. Adv. Manuf. Technol.* **2022**, *118*, 817–836, doi:10.1007/s00170-021-07784-y.

15. Peng, H.; Li, H.; Zhang, Y.; Wang, S.; Gu, K.; Ren, M. Multi-Sensor Vibration Signal Based Three-Stage Fault Prediction for Rotating Mechanical Equipment. *Entropy* **2022**, *24*, 164, doi:10.3390/e24020164.

16. Kumar, S.; Kolekar, T.; Kotecha, K.; Patil, S.; Bongale, A. Performance evaluation for tool wear prediction based on Bi-directional, Encoder–Decoder and Hybrid Long Short-Term Memory models. *Int. J. Qual. Reliab. Manag.* **2022**, doi:10.1108/IJQRM-08-2021-0291.

17. Yao, J.; Lu, B.; Zhang, J. Tool remaining useful life prediction using deep transfer reinforcement learning based on long short-term memory networks. *Int. J. Adv. Manuf. Technol.* **2022**, *118*, 1077–1086, doi:10.1007/s00170-021-07950-2.

18. Xia, M.; Zheng, X.; Imran, M.; Shoaib, M. Data-driven prognosis method using hybrid deep recurrent neural network. *Appl. Soft Comput. J.* **2020**, *93*, 106351, doi:10.1016/j.asoc.2020.106351.