An Inference Spatiotemporal Machinery Prognostics Approach Based on Graph Learning

Ahmed Ayman
School of Engineering and Computing
University of Central Lancashire
Preston, UK
aaabdel-aal@uclan.ac.uk

Omneya Attallah
Department of Electronics and Communications Engineering
Arab Academy for Science, Technology and Maritime Transport
Alexandria, Egypt
o.attallah@aast.edu

Ahmed Onsy
School of Engineering and Computing
University of Central Lancashire
Preston, UK
aonsy@uclan.ac.uk

Hadley Brooks
School of Engineering and Computing
University of Central Lancashire
Preston, UK
hlbrooks1@uclan.ac.uk

Iman Morsi
Department of Electronics and Communications Engineering
Arab Academy for Science, Technology and Maritime Transport
Alexandria, Egypt
drimanmorsi@yahoo.com

***Abstract*— Machinery prognostics facilitates predictive maintenance, minimizing downtime and operational expenses. Nonetheless, challenges persist due to low signal-to-noise ratio and non-stationary signals. Spatiotemporal feature extraction through recurrent and convolutional neural networks has shown promise in addressing these challenges. Nevertheless, the traditional convolutional learning algorithm, which is based on Euclidean distances between the learned features, can increase the model uncertainty. Moreover, traditional feature fusion techniques can weaken the model's performance. This study proposes a novel inferential spatiotemporal approach. Two independent networks based on long short-term memory and a graph convolutional network are designed to extract the influential spatiotemporal features. Then an adaptive neuro-fuzzy inferential network is introduced to calculate the remaining useful life based on the extracted spatiotemporal features. Experimental validation using a benchmark bearing dataset under various operational conditions demonstrates that the proposed approach outperforms existing state-of-the-art methods by 59.34%.**

Keywords— predictive maintenance; machinery prognostics; graph convolutional networks; remaining useful life prediction

I. Introduction

The accurate prediction of the Remaining Useful Life (RUL) of rotating machinery is critical for optimizing operational efficiency and mitigating unexpected failures across various applications, including manufacturing systems, wind turbines, and vehicles [1], [2]. Prognostic approaches to RUL estimation are generally classified into model-based and data-driven methods. Model-based approaches that are based on techniques such as particle filters, Kalman filters, and unscented filters, utilize mathematical models to characterize equipment degradation. Nevertheless, these methods often demand substantial expert knowledge and face challenges when addressing nonlinear and complex degradation dynamics [3]. Conversely, data-driven approaches, which leverage machine learning and deep learning algorithms, have received significant attention due to their capability to autonomously predict equipment lifespan with greater reliability while reducing reliance on domain expertise [4], [5].

Recurrent neural networks (RNN) and convolutional neural networks (CNN) have proven their effectiveness in machinery prognostics applications. However, the progressive degradation of machinery during operation necessitates the evaluation of temporal interdependencies which is more suited to RNN. Nevertheless, their performance is often constrained by issues such as vanishing and exploding gradients during backpropagation [6]. Advanced RNN architectures, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), address these limitations by employing gating mechanisms to preserve important information. For instance, studies by Mao et al. [7] have demonstrated the effectiveness of LSTMs in extracting temporal features for RUL prediction of bearings. While Chen et al. [8] incorporated Empirical Mode Decomposition (EMD) as a preprocessing step to further enhance the ability of LSTMs to learn with characteristics of low signal-to-noise ratio (SNR). Additional advancements include the application of adversarial learning for noise suppression and Bidirectional LSTMs (Bi-LSTMs) for capturing both forward and backward dependencies [9]. In [10], a GRU-based architecture integrated as a machinery health indicator is proposed. Despite the effectiveness of advanced RNN techniques, false sensory readings and poor SNR can mislead the model during training, potentially resulting in the loss of critical features.

CNN on the other hand, learns the sensory data based on the grid structure of the influential spatial features in the feature space. Ren et al. [11] employed a CNN architecture to predict the RUL of bearings under specific operational conditions. While to expedite the learning process of CNN models, Majali et al. [6] introduced a signal preprocessing stage prior to training, where the vibration signals were statistically analyzed across time domain and frequency spectrum. A similar preprocessing stage was adopted in [12], though employing LeNet-5 convolutional network to increase the robustness of the CNN model in learning the relevant effective characteristics [13]. Despite the achieved results, relying on a limited set of statistical features can lead to less accurate predictions and the omission of vital information. To address this limitation, the authors of [4] utilized an autoencoder (AE) to identify and focus on the most critical features during training.

The integration of global and local features from lower-level layers known as multi-scale CNN (CNN) has been demonstrated effective in overcoming limitations of signal preprocessing stage and enhance the feature extraction process. In this context, the authors of [14] proposed a MSCNN in which the features of both convolutional and pooling layers are leveraged for the prognostic process. The authors of [15] introduced a dilated MSCNN to incorporate features from various time steps. From the analysis of the previous studies, two key observations can be listed: (1) Convolutional networks primarily excel at extracting the spatial features of acquired machinery signals but may overlook temporal dependencies. (2) In spatial learning, convolutional mechanisms rely on Euclidean distances between spatial instances, which are particularly effective in grid-structured data applications such as images and computer vision.

Consequently, spatiotemporal analysis has proven to alleviate the performance of either temporal or spatial approaches. For instance, Zhao et al. [16] proposed a temporal convolutional approach for spatiotemporal feature learning. The authors demonstrated that spatiotemporal feature extraction can alleviate the accuracy of estimated RUL values by 10-20%. Qiao et al. [17] proposed another spatiotemporal approach based on Conv-LSTM. It aims to learn both temporal and spatial features concurrently by employing convolution operations to replace matrix multiplication within the LSTM unit, focusing on data changes over time steps. However, such an approach is computationally expensive. Moreover, the limitation of traditional convolutional operations which relies on Euclidean distances still persist.

Thus [18] proposed a spatiotemporal prognostic approach based on a multi-scale graph convolutional network (MSGCN). The aim of the GCN is to avoid reliance on Euclidean distances while preserving the reliability of spatial characteristics. The temporal dependencies and fault observations are detected using the sliding windowing technique, while the spatial relations are learned by the GCN. A similar graph structure approach but based on TCN for temporal learning is introduced in [19]. On the other hand, Hua et al. [20] proposed another spatiotemporal approach based on GCN for spatial learning followed by GRU for temporal learning. While in [21], a Bi-LSTM model is employed to extract temporal features, followed by GCN to predict the RUL of the bearing. While these methods demonstrate notable effectiveness, the sequential topology of the network introduces a critical dependency on the performance of the initial network rather than the achieved RUL values. As during backpropagation, the weights and parameters of the initial network remain unaffected by the optimization processes of the subsequent network.

In the context of the above, this paper proposes a spatiotemporal approach based on LSTM and GCN. However, to avoid the influence of the performance of one network on the other, each network is trained independently and in a parallel manner. While a novel adaptive neuro-fuzzy inference system (ANFIS) is proposed to estimate the final RUL based on the inference of the learned temporal and spatial features at each time step. The proposed approach is evaluated using the PRONOSTIA public dataset [22], which comprises 17 full bearings' lifecycle datasets under varying operational conditions.

The structure of this paper is as follows: Section II describe the experimental setup of the employed dataset. Section 3 describes the detailed methodology. Section 4 presents the experimental results and analysis. Finally, Section 5 concludes the study

II. Methods and Materials

A. Data Description

The PRONOSTIA dataset [22] serves as a benchmark in the field of machinery prognostics and is frequently used in the state-of-the-art for evaluation. This dataset comprises 17 run-to-failure experiments involving bearings subjected to three distinct operating conditions, as outlined in Table I. To expedite the degradation process, a radial force exceeding the maximum dynamic load capacity of the bearings was applied, leading to failure within a few hours. Throughout the tests, the rotational speed of the bearings was maintained at a constant value. Data acquisition was performed using two accelerometers, enabling precise measurement of bearing vibrations. A critical failure state was defined as the point where the vibration signal exceeded 20 g, beyond which the bearing was classified as non-operational.

The vibration data was sampled at a frequency of 25,600 samples per second, ensuring high-resolution observations. Each sample covered a duration of 0.1 seconds, consisting of 2,560 data points. Data recording occurred at regular 10-second intervals, ensuring consistent and accurate monitoring of the degradation process throughout the experiments.

TABLE I. CHARACTERISTICS OF THE PRONOSTIA DATASET [22]

Characteristics	1	2	3
Load (N)	4000	4200	5000
Speed (rpm)	1800	1650	1500
Training	B1_1 B1_2 B1_3	B2_1 B2-_2 B2_3	B3_1 B3_2
Validation/ Testing	B1_4 B1_5 B1_6 B1_7	B2_4 B2_5 B2_6 B2_7	B3_3

III. Proposed Framework

This section outlines the proposed framework which is structured into four interconnected stages: data processing, feature learning, feature mapping, and RUL prediction. In the initial stage, raw vibration signals from the monitored bearings were sampled at a frequency of 25.6 kHz. The raw signals were segmented into equal-length windows, with each segment comprising 2,560 instantaneous accelerometer readings. The standard deviation (Std) of each segment was then computed, resulting in a single representative value for each window. This approach not only addresses the SNR but also facilitates the model training process and enhance the convergence rate. Following segmentation and feature extraction, the data was normalized using the min-max scaling technique.

Next, for the feature learning stage, the LSTM network architecture is utilized to learn the temporal patterns of the acquired vibration signal of the test bearings. During this stage, the LSTM is designed to encounter the influential features at different time steps using the internal memory cell and output gates. The proposed temporal network is configured as a one layer, however, with 32 LSTM cells in a sequential manner. In which the input time sample is trained using 32 LSTM cells, each has its independent memory cell, update gate, and set of weights, before being encountered as an extracted features. This has proven to effectively eliminate the depicted noise in the signal and achieve higher performance. During backpropagation, the model's weights are adjusted after quantifying the error of each training epoch of the 100 epochs, ensuring that the model parameters are refined iteratively. Adam has been selected as the optimizer, with a learning rate of 0.01.

On the other hand, the spatial features are being learned using the proposed GCN. At first, the bearing signal is segmented according to the calculated Std value. Afterwards, each window value is considered as a node feature within the designed graph structure. While the edges of the graph capture the interconnections among the graph nodes. Formally, in this study, the proposed graph can be characterized as an undirected graph and it can mathematically be expressed as:

$$G=(V,E,A) \tag{1}$$

V represents the set of nodes, where $v_i \in V$, and similarly E represents the set of edges, where $e_{ij} = (v_i, v_j) \in E$ which imply the connectivity of nodes v_i and v_j. Afterwards, the adjacency matrix of the graph is introduced to represent the connection structure of the nodes in which:

$$A_{ij} = \begin{cases} 1, & e_{ij} \in E \\ 0, & e_{ij} \notin E \end{cases} \tag{2}$$

Afterwards, each node v_i integrates its own feature x_i with the features of its neighbouring nodes x_j to compute a new representation. The aggregated features are then passed through a nonlinear activation function to generate the final output. The process can be entirely be expressed as follows:

$$H^{(l+1)} = \sigma(\widehat{D}^{\,-1/2}\hat{A}\widehat{D}^{\,-1/2}H^{(l)}W^{(l)}) \tag{3}$$

$\hat{A}$ is equal to $A + I$, where A is the adjacency matrix A with self-loops (identity matrix I). Also, $\widehat{D}$ is the degree matrix with corresponding to $\hat{A}$. For this graph structured network, 16 channels are configured to process the data of each node and aggregate it to the neighboring nodes. Moreover, in this study, the $H^{(0)}$ and $W^{(0)}$ are randomly initiated and they are get optimized during the network training as it is conducted over 100 epochs, allowing the network to converge to an optimal global minimum.

Afterwards, the extracted spatial and temporal features of both networks are then fed to the configured ANFIS, defined as the feature mapping stage. By leveraging the IF-THEN rules of the fuzzy logic system in conjunction with the backpropagation mechanism of the neural network, the system effectively calculates the RUL based on multi-dimensional feature representations at each time step. In this scenario, the input matrix of the ANFIS model is of 2 dimensions and the employed membership functions is selected to be Gaussian to facilitate the representation of intricate and non-linear relationships. The inference process utilizes the Sugeno model, enabling the efficient generation of accurate output functions through weighted average computations.

Finally, the overall accuracy of the proposed model is quantified using the root mean square error which is described as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f(x_i)-(y_i))^2}{n}} \tag{4}$$

Where $f(x_i)$ is the output of the model for the i_{th} sample, y_i represents the matching label, and n represents the length of the entire data vector.

IV. Results and discussion

This section discusses the findings of the proposed study. For this study, the first three bearing sets from the first and second operational conditions, as specified in Table I, were used for training, while the remaining sets were reserved for testing. This dataset allocation ensured an 80-20 split between training and testing data, providing a robust evaluation framework for the proposed method. Thus, the procedure was conducted nine times, each focusing on a distinct bearing set: B1_4, B1_5, B1_6, B1_7, B2_4, B2_5, B2_6, B2_7, and B3_3.

Initially, the extracted Std values for each window were input into the GCN to enable spatial learning while preserving temporal mapping. The network weights were initialized randomly and updated iteratively based on the calculated mean squared error (MSE). Figure 1 is a box plot that illustrates the MSE values over 100 epochs for all test bearing sets used in this study. Similarly, the temporal network was trained using the same procedure, with Figure 2 depicting the MSE values recorded after each epoch.

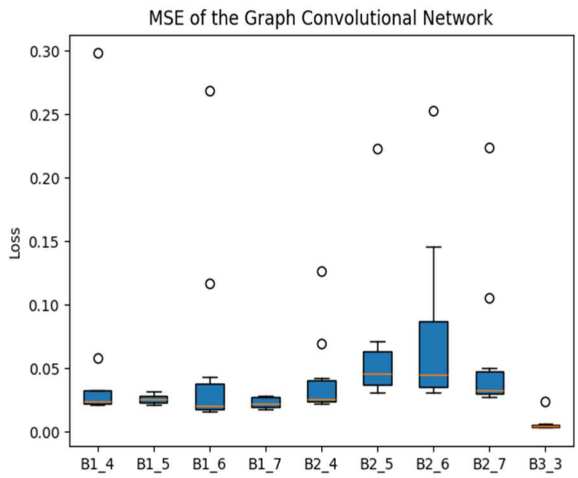

Fig. 1. *MSE values of the proposed spatial network*

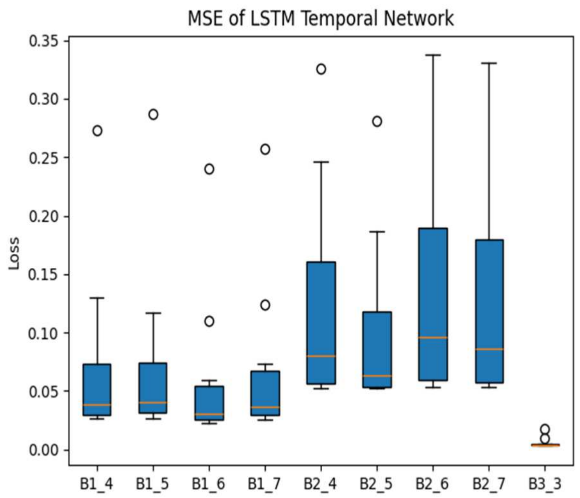

Fig. 2. *MSE values of the proposed temporal network*

It can be noted that in Figure 1, the MSE values stay almost constant across all of the operational sets. This shows that this method works well for capturing the spatial characteristics, which can be very similar for the same experimental setup and bearing manufacturing. The MSE values in Figure 2 change noticeably between operational conditions and slightly within the same operational condition. This shows the temporal dynamics that happened and how well the system learned to capture and learn these dynamics until they reached a global minimum. This emphasizes the critical importance of spatiotemporal analysis, which encounters both characteristics of the extracted features in the estimation of the RUL values.

Finally, the ANFIS network was trained to estimate the RUL by leveraging the inferential relationships derived from the extracted spatial and temporal features. Table II summarizes the RMSE results obtained by the ANFIS network for the first operational condition, while Table III presents the corresponding RMSE results for the second and third operational conditions.

TABLE II. EVALUATION RESULTS OF THE PROPOSED METHOD FOR THE FIRST OPERATIONAL CONDITION SETTINGS

Test set	**B1_4**	**B1_5**	**B1_6**	**B1_7**
RMSE	0.0394	0.0571	0.025	0.015

TABLE III. EVALUATION RESULTS OF THE PROPOSED METHOD FOR THE SECOND AND THIRD OPERATIONAL CONDITION SETTINGS

Test set	**B2_4**	**B2_5**	**B2_6**	**B2_7**	**B3_3**
RMSE	0.022	0.046	0.01	0.020	0.022

To further evaluate the effectiveness of the proposed method, recent state-of-the-art techniques were reviewed to establish a comparative benchmark for the obtained results. Zhu et al. [14] employed an MSCNN-based approach to predict the RUL of bearings in the PRONOSTIA dataset. This study transformed the one-dimensional signal array into a two-dimensional matrix to leverage the grid structure of spatial features during MSCNN training. Similarly, Huang et al. [23] integrated spatial and temporal features using a multi-layer perceptron (MLP) to estimate bearing lifespan values. Furthermore, Rathore et al. [24] proposed a prognostic framework utilizing LSTM combined with an attention mechanism to identify the most critical features and address the low SNR of the monitored signals.

TABLE IV. EVALUATION RESULTS OF THE PROPOSED AND STATE-OF-THE-ART METHODS

Set/ RMSE	[14]	[23]	[24]	**Proposed method**
B1_4	0.515	0.39	0.0969	0.0394
B1_5	0.366	0.32	0.2499	0.0571
B1_6	0.480	0.34	0.2414	0.025
B1_7	0.170	0.35	0.2636	0.015
B2_4	-	-	0.0799	0.022
B2_5	-	-	0.1678	0.046
B2_6	-	-	0.0761	0.01
B2_7	-	-	0.0224	0.020
B3_3	-	-	0.0209	0.022

Table IV presents the RMSE values of all reviewed studies alongside those of the proposed method. The results demonstrate that the proposed method significantly outperforms the existing approaches, underscoring the effectiveness of the multi-scenario feature space designed with graph neural networks and temporal networks. This advantage is attributed to the inference capabilities across multi-scenario data, which surpass traditional fusion strategies such as MLP.

V. CONCLUSION

Recent advancements have demonstrated the efficacy of deep learning techniques in machinery prognostics. However, conventional approaches, such as CNNs, may constrain the accuracy of RUL predictions due to their reliance on Euclidean distances between features in the latent space. Additionally, the low SNR and non-stationary nature of machinery signals can hinder the ability of deep learning models to effectively capture critical features. To address these challenges, this paper introduced a multi-scenery spatiotemporal method that employed a graph-based structure for spatial feature representation, circumventing the limitations of grid-based learning, and utilized LSTM networks for temporal feature extraction. Subsequently, an ANFIS was applied to infer the lifespan of the target machinery at each time step within the multi-scenery space. Experimental results on a real benchmark bearing dataset demonstrated that the proposed approach effectively mitigated the impact of low SNR, resolved the challenges associated with spatial feature learning in traditional CNNs, and achieved superior accuracy in RUL estimation.

REFERENCES

[1] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," J Manuf Syst, vol. 43, pp. 248–256, 2017.

[2] M. A. Djeziri, S. Benmoussa, and E. Zio, "Review on health indices extraction and trend modeling for remaining useful life estimation," Artificial Intelligence Techniques for a Scalable Energy Transition: Advanced Methods, Digital Technologies, Decision Support Tools, and Applications, pp. 183–223, 2020.

[3] H. Liu, W. Song, Y. Niu, and E. Zio, "A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes," Mech Syst Signal Process, vol. 153, p. 107471, 2021.

[4] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," J Manuf Syst, vol. 48, pp. 71–77, 2018.

[5] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," IEEE Trans Syst Man Cybern Syst, vol. 48, no. 1, pp. 11–20, 2017.

[6] A. Majali, A. Mulay, V. Iyengar, A. Nayak, and P. Singru, "Fault identification and remaining useful life prediction of bearings using

Poincare maps, fast Fourier transform and convolutional neural networks," Mathematical Models in Engineering, vol. 8, no. 1, pp. 1–14, 2022.

[7] Z. Zhao, J. Wu, T. Li, C. Sun, R. Yan, and X. Chen, "Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review," Chinese Journal of Mechanical Engineering, vol. 34, no. 1, p. 56, 2021.

[8] Y. Chen, G. Peng, Z. Zhu, and S. Li, "A novel deep learning method based on attention mechanism for bearing remaining useful life prediction," Appl Soft Comput, vol. 86, p. 105919, 2020.

[9] B.-L. Lu, Z.-H. Liu, H.-L. Wei, L. Chen, H. Zhang, and X.-H. Li, "A deep adversarial learning prognostics model for remaining useful life prediction of rolling bearing," IEEE Transactions on Artificial Intelligence, vol. 2, no. 4, pp. 329–340, 2021.

[10] W. Mao, J. He, J. Tang, and Y. Li, "Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network," Advances in Mechanical Engineering, vol. 10, no. 12, p. 1687814018817184, 2018.

[11] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," IEEE access, vol. 6, pp. 13041–13049, 2018.

[12] Q. Wang, B. Zhao, H. Ma, J. Chang, and G. Mao, "A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion," Journal of Mechanical Science and Technology, vol. 33, pp. 2561–2571, 2019.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[14] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," IEEE Transactions on Industrial Electronics, vol. 66, no. 4, pp. 3208–3216, 2018.

[15] R. Wang, R. Shi, X. Hu, and C. Shen, "Remaining useful life prediction of rolling bearings based on multiscale convolutional neural network with integrated dilated convolution blocks," Shock and Vibration, vol. 2021, no. 1, p. 6616861, 2021.

[16] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," Sensors, vol. 17, no. 2, p. 273, 2017.

[17] H. Qiao, T. Wang, P. Wang, S. Qiao, and L. Zhang, "A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series," Sensors, vol. 18, no. 9, p. 2932, 2018.

[18] X. Yang, X. Li, Y. Zheng, Y. Zhang, and D. S.-H. Wong, "Bearing remaining useful life prediction using spatial-temporal multiscale graph convolutional neural network," Meas Sci Technol, vol. 34, no. 8, p. 085009, 2023.

[19] S. Lv and S. Liu, "A new method for remaining useful life prediction by implementing joint learning of sensor dynamic graph and spatio-temporal features," Meas Sci Technol, vol. 34, no. 9, p. 095123, 2023.

[20] J. Hua, Y. Zhang, D. Zhang, J. He, J. Wang, and X. Fang, "Dynamic Feature-Aware Graph Convolutional Network with Multi-sensor for Remaining Useful Life Prediction of Turbofan Engines," IEEE Sens J, 2024.

[21] G. Gao, B. Zhao, X. Chang, and X. Ming, "Remaining useful life prediction of bearings based on GCN and Bi-LSTM," 2022.

[22] P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests.," in IEEE International Conference on Prognostics and Health Management, PHM'12., IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.

[23] C.-G. Huang, H.-Z. Huang, Y.-F. Li, and W. Peng, "A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing," J Manuf Syst, vol. 61, pp. 757–772, 2021.

[24] M. S. Rathore and S. P. Harsha, "Prognostics analysis of rolling bearing based on bi-directional LSTM and attention mechanism," Journal of Failure Analysis and Prevention, vol. 22, no. 2, pp. 704–723, 2022.