MALAYSIAN JOURNAL OF INNOVATION IN ENGINEERING AND APPLIED SOCIAL SCIENCE (MyJIEAS)
e-ISSN: 2811-3780
VOLUME 04, ISSUE 01

Review on Data Integrity in Remaining Useful Life Prediction

Azman Ahmad Bakir[1]*, Mohd Nor Hakim Hassan[1]

[1]Universiti Teknologi MARA,Cawangan Pulau Pinang, Permatang Pauh, Pulau Pinang,13500, Malaysia

Corresponding Author's Email: [a]zmanbakir@uitm.edu.my

Article History: Received 13082024; Revised 25082024; Accepted 07102024;

ABSTRACT – *Maintenance has always been an important function in any manufacturing operation. The main role of maintenance is to ensure that the manufacturing activities run at optimum condition without any interruption. Several maintenance strategies have been introduced such as breakdown maintenance, preventive maintenance, and condition-based maintenance. Condition-based maintenance (CBM) has become one of the popular maintenance strategies in manufacturing industries. CBM is a maintenance program that is triggered by the machine deterioration condition. CBM serves two distinct purposes: diagnostic and prediction. The predictive segment largely focuses on forecasting the remaining useful life (RUL) of components. Presently, the focus of study is primarily on the prediction of Remaining Useful Life (RUL). A point often overlooked, the effectiveness of this maintenance strategy depends on the quality and integrity of observation data. There have been a few review papers on the broad scope of RUL prediction particularly in diagnostic techniques, but they lack focus on the quality and integrity of the maintenance data. This paper provides a review on the significance of data integrity in RUL. This review covers publications from 2005 to 2024. It offers a categorisation of research fields or topics concerning data integrity and methods for data processing. Deep learning neural network models especially LSTM and CNN have been used extensively to treat data integrity issues, particularly in have also gained popularity, especially in newer publications. This review provides evidence on the importance of data integrity in RUL prediction, and it can be useful for new researchers in the area.*

KEYWORDS: Condition Maintenance, Remaining Useful life, Data integrity, Diagnosis, Deep learning neural network, LSTM, CNN

1.0 INTRODUCTION

Maintenance has always been an important element in the manufacturing industry. Machine reliability to produce a constant quality product without interruption is crucial to help the industry maintain its sustainability. Maintenance intervention can prevent unwanted failure. The oldest approach in the maintenance system is known as breakdown maintenance. This method only makes necessary repairs when the machine has a breakdown. As technology progresses, a new maintenance system known as preventive maintenance is introduced. This method uses a periodic time frame to plan the maintenance activity and can improve machine reliability. However, this approach can lead to excessive maintenance and increase the operation cost if not properly planned. A relatively newer approach known as condition-based maintenance (CBM) has been introduced to overcome this limitation. CBM monitors the machine's condition and only performs maintenance if the machine's condition is found to be below healthy condition. In 2006, [1] noted that the CBM approach is more suitable for modern machines. [2] confirmed that the significance of the CBM approach was due to the enhancement in technology. In recent years, there has been an increasing amount of literature on CBM. Currently, CBM also has a predictive maintenance feature, which can determine when the equipment or machine will fail. The predictive maintenance component in CBM focuses on forecasting the Remaining Useful Life (RUL), which denotes the remaining duration of the equipment or machine before it experiences a failure.

The intricacy and the expense of upkeeping the equipment have significantly escalated. Data integrity is a crucial research concern in the estimation of RUL in CBM for predictive maintenance. Data integrity refers to the preservation of the original value or characteristics of data, even when it has undergone different forms of modification, transmission, or disruptions. There have been few review papers on this subject. A recent review work by [3] discussed the data handling in CBM systems.

The significance of maintaining data integrity in the forecast of RUL was emphasized by the individual. The purpose of this work is to provide a comprehensive analysis of research about data integrity in research on RUL prediction. The review presents examples of data, including the classification and frequency of study areas, as well as the type of data and methodologies used for data treatment. This review was conducted using journal publications spanning from 2005 to 2024. This review included a total of 72 periodicals. This work is structured into three primary components. Section 1 provides a concise overview of Maintenance techniques, the past review papers, and the purpose of the current review study. In section 2, the process of reviewing is explained. The method of reviewing the article is described, beginning with the acquisition of information, followed by the classification and analysis of the collected articles. The third section of the article examines problems related to RUL prediction. These concerns are thoroughly analyzed, and the data is shown in a bar chart, illustrating the frequency of each identical problem. Furthermore, this part will examine the root causes of the most prevalent problem. In this section, the author examined the strategies and procedures used to address data integrity challenges.

2.0 METHODOLOGY

2.1 Research Methodology

It is necessary to understand the scope and criteria that were used in the selection of related references. Those criteria are listed and shown in Figure 1. The research on Remaining Useful Life (RUL) prediction involves addressing several key aspects to advance the field effectively. First, it is essential to classify the issues by analyzing and comparing the frequency of challenges in various RUL research areas to identify the most prominent problems, such as data quality, model accuracy, and computational constraints. Once these issues are identified, a thorough investigation into their root causes is necessary, examining factors like noisy sensor data, insufficiently labeled datasets, or mismatched assumptions in models, as derived from extensive literature reviews. Subsequently, the methodologies and techniques employed in previous studies to tackle these challenges must be critically reviewed, focusing on strategies to improve data quality, enhance model performance, and resolve inconsistencies, such as applying noise reduction methods, feature extraction, and advanced machine learning techniques. This holistic approach not only highlights the disparities and commonalities in existing methods but also provides a foundation for researchers to propose innovative solutions to overcome persistent challenges in RUL prediction research.

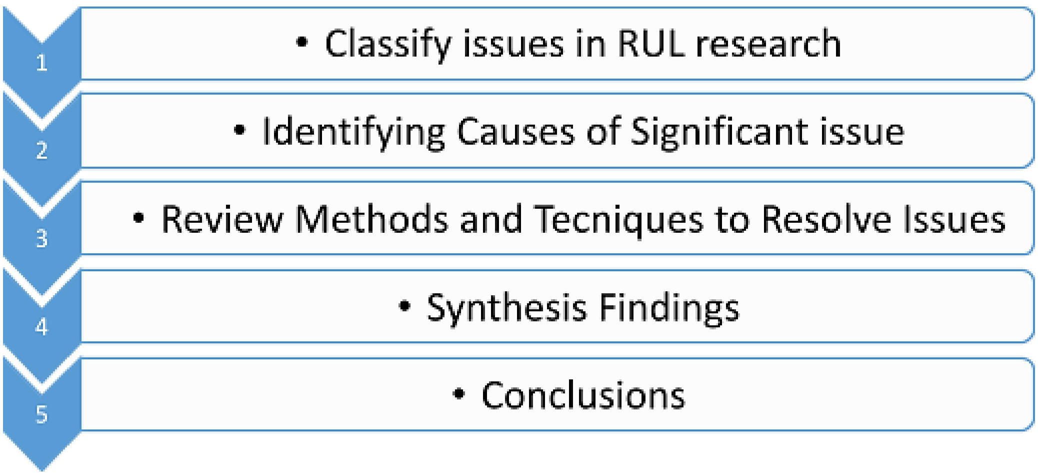

Figure 1. Flowchart outlining the methodology for the review process

3.0 RESULT AND DISCUSSION

3.1 Issues in RUL prediction research

Figure 2 depicts the typical challenges encountered in research projects related to the prediction of RUL. Figure 2 exhibits a comprehensive collection of 17 problems, encompassing a variety of difficulties such as engine malfunctions and health indicators. The challenges differ across multiple study domains, encompassing automobiles, rotating machinery, battery degradation, and others. Each of them is linked to studies focused on predicting the RUL. The data analysis revealed that the most common challenges encountered were data integrity, which was discussed in 52 publications, data mining, which was addressed in 15 papers, and data correlations, which were explored in 12 articles. In addition, Table 1 displays an exhaustive compilation of literature that is linked to problems with RUL prediction research.

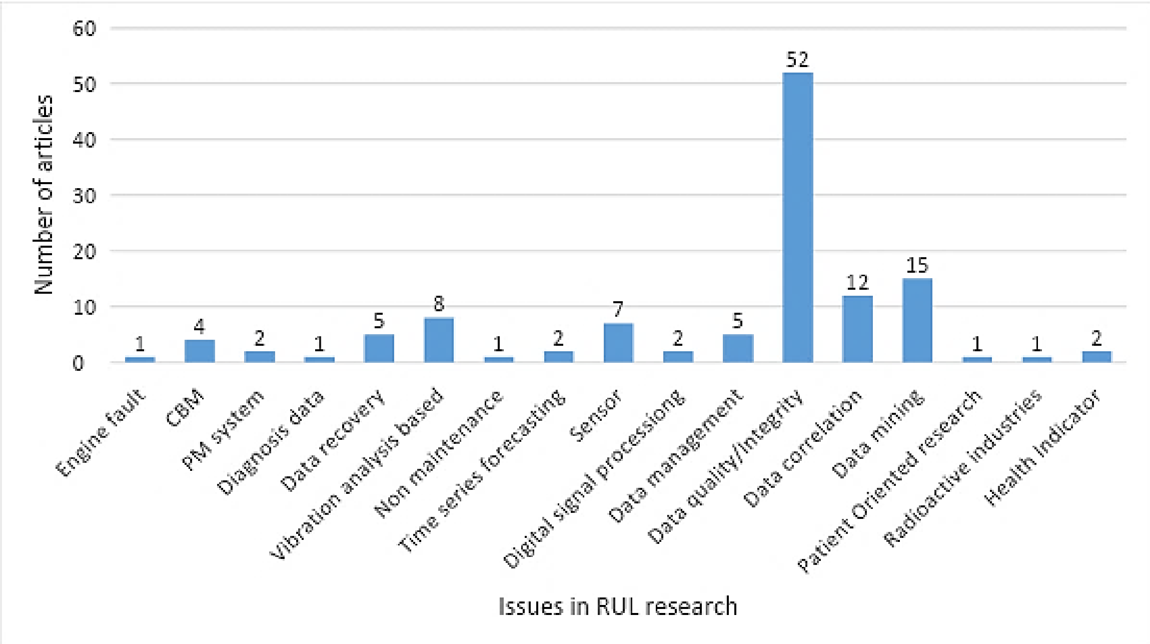

Figure 2. Overall Issues in RUL Research

According to Figure 2, the predominant concern or challenge is the quality and integrity of data. The accuracy of prediction in developing the RUL model heavily relies on the data used. Hence, if the data has quality issues, it will undoubtedly impact the accuracy of the predictions. For instance, researchers [3] investigated the degradation of lithium-ion batteries and encountered difficulties in extracting meaningful characteristics, which ultimately impacted their ability to accurately predict the RUL of the batteries. A further study conducted by [4] on health indicators of rotating machinery, focusing on analyzing vibration data, likewise encountered difficulties in extracting meaningful features from the raw data. The third example of research conducted by [5] found that the raw data contains a significant amount of noise, which has a detrimental impact on the reliability of the RUL prediction model. The subsequent section of 3.2 will provide a detailed explanation of the factors that contribute to data quality.

Data mining is the second most prevalent topic in RUL prediction research. Data mining is the systematic extraction of patterns, correlations, trends, and valuable insights from extensive databases by the application of statistical, mathematical, and computer methods. Some of the difficulties are interconnected and correlate with data quality issues. In a study conducted by [6] on wind turbine prognostics, the researchers encountered challenges related to the limited availability of data and the presence of high data uncertainty, which could impact the acquisition of reliable data. In a separate study conducted by [7], researchers encountered difficulties in data mining due to incomplete data containing missing data.

The final example of the most prevalent difficulty in RUL research is data correlation. In data correlations, the issue arises when one is unable to comprehend the relationship between the provided data. Temporal data correlations are crucial in RUL predictions, as demonstrated by [8] in their study on the prognosis of rolling bearings. In another context, as described in the paper by [9], a distinct scenario is presented where there is diversity in the dataset obtained from several units of equipment. This will also compromise the accuracy of the RUL prediction.

Table 1. Articles related to issues in RUL

Issues	Related articles
Engine fault	[10]
CBM	[11],[12],[13],[14]
PM System	[15],[16]
Diagnosis data	[17]
Data Recovery	[18],[19],[20],[21],[22]
Vibration analysis	[23],[11],[24],[25],[26],[27],[14],[28]
Non-maintenance	[29]
Time series forecasting	[30],[31]
Sensor	[32],[33],[34],[35],[36],[37],[38]
Digital signal processing	[39],[15]
Data management	[40],[41]
Data quality/integrity	[3],[4],[42],[5],[43],[6],[44],[45],[40],[46],[47],[48],[49],[50],[41],[7],[51],[52],[23],[11],[53],[32],[54],[55],[33],[56],[57],[58],[59],[60],[61],[25],[34],[35],[62],[15],[63],[64],[14],[30],[36],[28],[19],[37],[31],[18],[19],[38],[65],[21],[29],[22],[57],[58],[62]
Data Correlation	[8],[9],[39],[11],[32],[55],[66],[33],[16],[57],[58],[62]
Data mining	[3],[6],[41],[7],[8],[52],[32],[54],[55],[67],[66],[56],[10],[68],[31]
Patient-oriented research	[65]
Radioactive industries	[69]
Health indicator	[57],[58]

3.2 Causes of the most significant issue

The bar chart from Figure 3 depicts the several causes concerning data quality and the corresponding number of articles dedicated to each issue. The topic of "lack of significant feature" has been extensively studied in 18 publications, emphasizing its crucial significance. The issue of "missing data" is also a significant one, which is discussed in 16 articles. Both the topics of "data recovery" and "data fusion issues" are addressed in a total of 7 articles each, which highlights their significance in talks on data quality. The topics of "noise" and "data scarcity" are addressed in 6 articles each, indicating a significant degree of attention. On the other hand, "data uncertainty" is discussed in 4 articles, suggesting a modest level of worry. Finally, the topic of "data error" has received the least amount of attention, with only 2 articles dedicated to it. This indicates that it is perceived as a less significant concern in comparison to the other causes.

The cause of the lack of significant features was a serious challenge affecting data integrity. In 2024, a study was undertaken on Turbofan engines, resulting in the development of a model for predicting RUL. During this investigation, [42] encountered challenges in extracting the significant features due to constraints in the available data. The researchers suggested employing Long Short-Term Memory (LSTM) deep learning approaches to enhance the model's capacity to extract valuable information. In a separate instance, a study conducted by [7] encountered comparable concerns with a significant feature. Nevertheless, the occurrence of data information loss during data mining has a detrimental impact on the accuracy of their prediction model. [7], suggested utilizing baseline similarity attention and dual channel techniques to enhance the potential of feature extraction.

Missing data is the second most common cause of data integrity. [32], had a problem with missing data while constructing their RUL prediction model for a piece of equipment during the process of data mining from several sensors. Currently, several devices utilize multiple sensors, and with the progress of the Internet of Things (IoT) market, the frequency of data mining has risen. Nevertheless, a strong data processing system was necessary to prevent any loss of data while doing the data mining procedure. [32] utilized LSTM and Recurrent Neural Network (RNN) to rectify errors caused by incomplete input. Whereas the second article by [70] focused on the issue of missing data in wireless sensor networks. The absence of data will have an impact on the accuracy of the RUL projection. To enhance the scenario, unsupervised fuzzy ART neural networks are utilized to impute the missing data, hence enhancing the accuracy of rule prediction.

Data fusion is the third most prevalent cause of data integrity. As previously stated, contemporary equipment relies on the integration of various sensors and accurate data fusion. The utilization of abundant data collected from many sensors has the potential to enhance the precision of predictions. An investigation conducted by [11] on the failure prognosis of rolling bearings has identified a difficulty with data fusion that is hindering the production of accurate predictions. The study was conducted on real equipment in an industrial setting, where the working environment, noise, and failure modes of non-stationary equipment were observed. The data collected from these varied settings has to be combined. [11] has proposed the use of LSTM, a type of deep learning approach, to extract all the relevant characteristics needed to create an RUL prediction model. This model combines data from various machine states and parameters. To enhance the accuracy of predicting bearing faults and RUL, the subsequent study conducted by [49] suggests the inclusion of additional data elements. The existing data fusion approach was enhanced by incorporating multi-residual feature fusion and an attention mechanism. This improvement led to enhanced data extraction and, consequently, increased accuracy in predicting RUL.

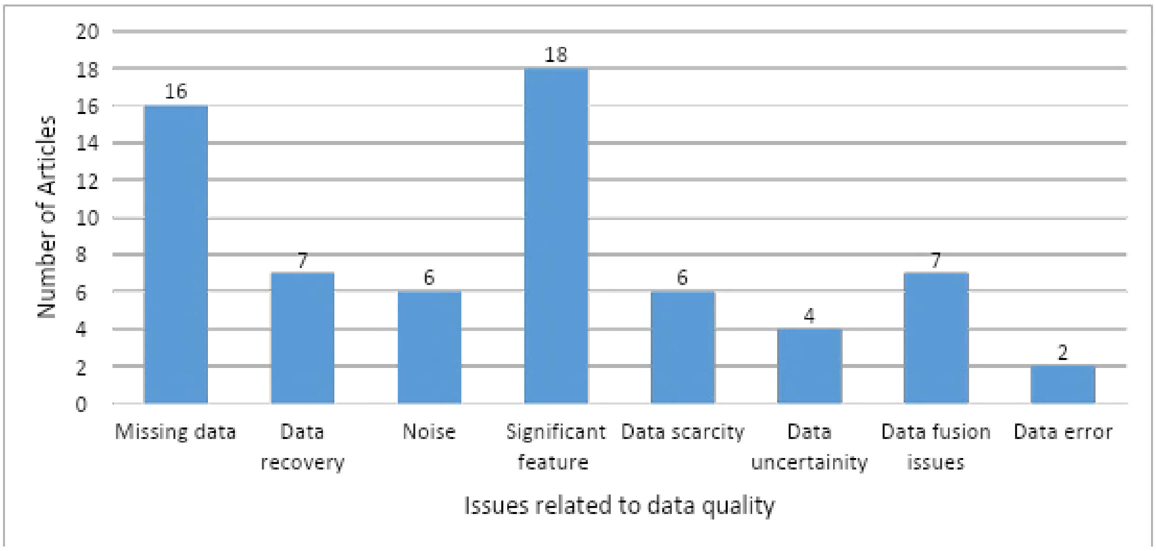

Figure 3. Issues related to data integrity in RUL prediction research

3.3 Techniques used by other researchers in data integrity issue

The bar chart Figure 4 presents a summary of the many methodologies utilized by researchers to tackle data quality concerns, together with the corresponding quantity of articles describing each methodology. The method The LSTM model is the most frequently utilized and is mentioned in a total of seven publications. Convolution Neural Network (CNN) and Autoregressive Integrated Moving Average (ARIMA) are subsequently mentioned, with six and five stories respectively. Each of the techniques, Baseline Similarity, Prognostics and Health Management (PHM), Markov, and Kernel density, is discussed in three separate articles. Two articles examine several techniques such as the Transfer function, Expectation Maximization (EM) Algorithm, Cubic smoothing, Gaussian Mixture Model (GMM), Health State, Statistical, Bayesian, Self-supervised, Attention-based, Autoencoder, and Multibranch. Each article focuses on a certain set of techniques. One article mentions several alternative methods, including LSTM, CIS Cormack, Joint Posterior, Supervisory Control and Data Acquisition (SCADA), and Fuzzy logic, each of which is addressed only once. This graphic demonstrates the variety of methods that researchers employ to address data quality concerns, with a significant inclination towards sophisticated machine learning techniques such as LSTM and CNN.

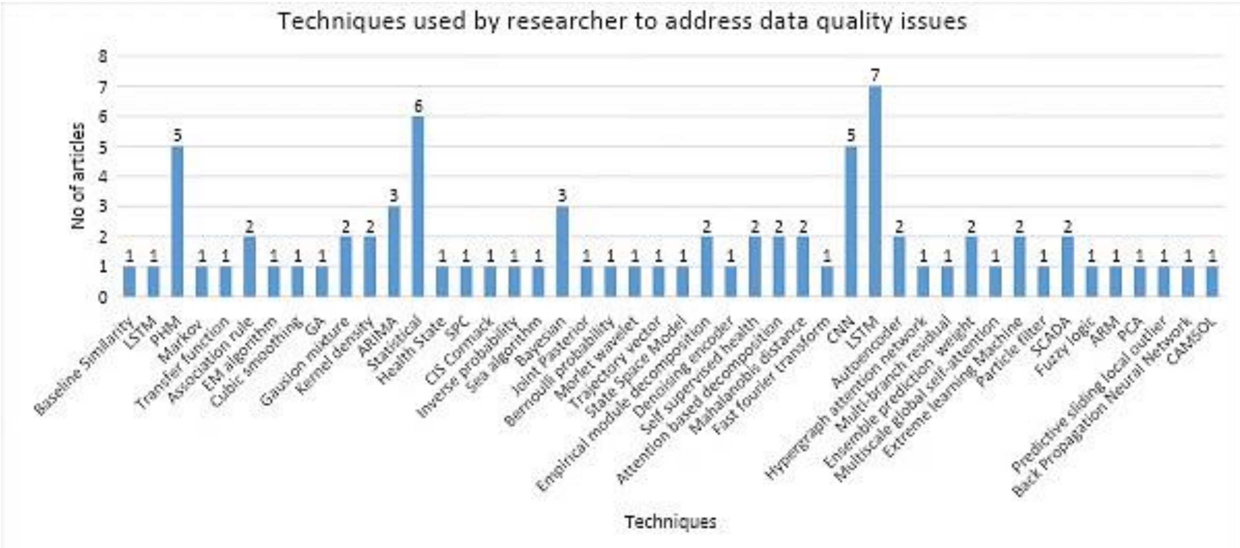

Figure 4. Techniques used by researchers to address data integrity issues

From Figure 4, LSTM is the most popular method applied in RUL prediction specifically related to data integrity issues. LSTM is a specialized variant of recurrent neural network RNN that is specifically developed to effectively model sequences and capture their long-range dependencies. LSTMs, in contrast to conventional RNNs, employ a sequence of gates (input, output, and forget) to regulate the information flow. This enables them to preserve and modify cell states over time, effectively resolving the issue of disappearing gradients. The architecture of LSTM allows for the retention of crucial information over extended sequences, rendering them highly advantageous for tasks like language modeling, speech recognition, and time series prediction. In a study performed by [3] Predicting the RUL of a battery is a tough task due to the complicated and nonlinear nature of battery degradation prognosis. To tackle this issue, a novel hybrid data-driven model named PCA-CNN-BiLSTM was introduced. This model combines three different techniques: Principal Component Analysis (PCA), CNN, and bi-directional long short-term memory (Bi-LSTM). PCA decreases the correlation between features and isolates significant lifespan features. CNN finds local patterns and minimizes processing requirements. Bidirectional Long Short-Term Memory (Bi-LSTM) forecasts RUL by considering both past and future data.

The statistical approach is the second most used technique for resolving data integrity issues. To summarize, a statistical method offers a systematic framework for making well-informed judgments using data. Data collection, analysis, and interpretation techniques are combined in this process, which is supported by probability theory, to obtain valuable insights and predictions. An example of an article that applies a statistical technique is the study by [71], which examines the prognosis in a vibration dataset. Vibration analysis is essential for diagnosing machine malfunctions. However, in intricate machines, vibrations from many components and external noise might disrupt the desired signal. To tackle this issue, a study conducted by [71] explores a super-exponential algorithm (SEA) that relies on a single-input single-output (SISO) convolution mixing model to deconvolve the vibration signal. The paper investigates the efficacy of skewness and kurtosis schemes in recovering signals with diverse statistical distributions by utilizing both simulated and real industrial machine signals.

Another popular technique used by researchers is the RUL prediction model is CNN. CNNs are a specific category of deep learning models that excel in the analysis of visual input. Their purpose is to autonomously and flexibly acquire knowledge about spatial hierarchies of characteristics from input images, using a sequence of layers that execute convolutions, followed by pooling and fully connected layers. Convolutional layers utilize a collection of adaptable filters to analyze the input image and detect specific local features, such as edges, textures, and forms. Pooling layers decrease the dimensionality, enhancing computational efficiency and increasing the model's resistance to minor input translations. Fully linked layers consolidate the gathered features to generate conclusive forecasts. CNNs have demonstrated exceptional performance in several computer vision applications, such as picture categorization, object identification, and image segmentation. An example of a researcher applying CNN is presented in a study by [72]. The article introduces a novel attention-based Deep Convolutional Neural Network (DCNN) architecture to enhance RUL forecasts by efficiently utilizing multivariate temporal information. The validation conducted on a benchmark dataset demonstrates that this strategy surpasses the performance of current methods, with a Root Mean Square Error (RMSE) of 11.81 and a score of 223. This clearly illustrates its efficacy in accurately forecasting the RUL of turbofan engines.

4.0 CONCLUSION

This report provides a comprehensive analysis of data integrity in research relevant to RUL prediction. The review has demonstrated that maintaining data integrity is crucial in predicting RUL. Data integrity is essential as it directly impacts the process of decision-making. The focus of RUL prediction research typically centers around diagnostic data acquired from sensors. This review has demonstrated several ways, with the most widely used being the deep learning approach, specifically LSTM and CNN. There is a commonality in the methodologies employed, however the models given in the framework differ. The discoveries in this analysis are likely to be valuable for novice researchers in the field. Further investigation is required to enhance the accuracy of RUL prediction, with a specific emphasis on ensuring the integrity of the data.

ACKNOWLEDGMENT

The author would like to express deep appreciation to the organizers of ICAES2024 for their remarkable endeavors in establishing a stimulating and fruitful atmosphere. Their unwavering commitment and thorough preparation ensured that this conference was a highly rewarding experience for all attendees. The author also would like to express gratitude to Universiti Teknologi MARA Malaysia (UiTM) Cawangan Pulau Pinang for their consistent support and encouragement during this research endeavor. The university's resources and guidance have been indispensable in completing this assignment.

REFERENCES

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.

[2] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Comput. Ind. Eng.*, vol. 63, no. 1, pp. 135–149, Aug. 2012, doi: 10.1016/j.cie.2012.02.002.

[3] J. Feng, F. Cai, H. Li, K. Huang, and H. Yin, "A data-driven prediction model for the remaining useful life prediction of lithium-ion batteries," *Process Saf. Environ. Prot.*, vol. 180, pp. 601–615, Dec. 2023, doi: 10.1016/j.psep.2023.10.042.

[4] Y. Duan, X. Cao, J. Zhao, M. Li, and X. Yang, "A Spatiotemporal Fusion Autoencoder-Based Health Indicator Automatic Construction Method for Rotating Machinery Considering Vibration Signal Expression," *IEEE Sens. J.*, vol. 23, no. 20, pp. 24822–24838, Oct. 2023, doi: 10.1109/JSEN.2023.3309013.

[5] Z. Wang, Y. Liu, F. Wang, H. Wang, and M. Su, "Capacity and remaining useful life prediction for lithium-ion batteries based on sequence decomposition and a deep-learning network," *J. Energy Storage*, vol. 72, p. 108085, Nov. 2023, doi: 10.1016/j.est.2023.108085.

[6] J. Cuesta, U. Leturiondo, Y. Vidal, and F. Pozo, "Challenges on prognostics and health management for wind turbine components," *J. Phys. Conf. Ser.*, vol. 2745, no. 1, p. 012003, Apr. 2024, doi: 10.1088/1742-6596/2745/1/012003.

[7] Y. Hu, X. Li, H. Wang, Y. Liu, X. Liu, and Z. Cao, "A Baseline Similarity Attention-Based Dual-Channel Feature Fusion Network for Machine Remaining Useful Life Prediction," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024, doi: 10.1109/TIM.2024.3398070.

[8] G. Zheng, Y. Li, Z. Zhou, and R. Yan, "A Remaining Useful Life Prediction Method of Rolling Bearings Based on Deep Reinforcement Learning," *IEEE Internet Things J.*, vol. 11, no. 13, pp. 22938–22949, Jul. 2024, doi: 10.1109/JIOT.2024.3363610.

[9] N. Li, Y. Lei, N. Li, and J. Lin, "Machine remaining useful life prediction considering unit-to-unit variability," in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, Jun. 2017, pp. 103–108. doi: 10.1109/ICPHM.2017.7998313.

[10] K. Choi *et al.*, "Novel classifier fusion approaches for fault diagnosis in automotive systems," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 3, pp. 602–611, 2009, doi: 10.1109/TIM.2008.2004340.

[11] R. K. Jha and P. D. Swami, "Failure prognosis of rolling bearings using maximum variance wavelet subband selection and support vector regression," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 44, no. 1, p. 49, Jan. 2022, doi: 10.1007/s40430-021-03345-2.

[12] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.

[13] P. Charongrattanasakul and A. Pongpullponsak, "Minimizing the cost of integrated systems approach to process control and maintenance model by EWMA control chart using genetic algorithm," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5178–5186, May 2011, doi: 10.1016/j.eswa.2010.10.044.

[14] W. Guo, P. W. Tse, and A. Djordjevich, "Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition," *Measurement*, vol. 45, no. 5, pp. 1308–1322, Jun. 2012, doi: 10.1016/j.measurement.2012.01.001.

[15] D. Pandey, M. S. Kulkarni, and P. Vrat, "A methodology for joint optimization for maintenance planning, process quality and production scheduling," *Comput. Ind. Eng.*, vol. 61, no. 4, pp. 1098–1106, Nov. 2011, doi: 10.1016/j.cie.2011.06.023.

[16] M. Colledani and T. Tolio, "Integrated quality, production logistics and maintenance analysis of multi-stage asynchronous manufacturing systems with degrading machines," *CIRP Ann.*, vol. 61, no. 1, pp. 455–458, 2012, doi: 10.1016/j.cirp.2012.03.072.

[17] Y. Liu, L. Guo, Q. Wang, G. An, M. Guo, and H. Lian, "Application to induction motor faults diagnosis of the amplitude recovery method combined with FFT," *Mech. Syst. Signal Process.*, vol. 24, no. 8, pp. 2961–2971, Nov. 2010, doi: 10.1016/j.ymssp.2010.03.008.

[18] M. N. Noor, a. S. Yahaya, N. a. Ramli, and A. M. M. Al Bakri, "Filling Missing Data Using Interpolation Methods: Study on the Effect of Fitting Distribution," *Key Eng. Mater.*, vol. 594–595, pp. 889–895, Dec. 2013, doi: 10.4028/www.scientific.net/KEM.594-595.889.

[19] M. N. Noor, a. S. Yahaya, N. a. Ramli, and A. M. M. Al Bakri, "Mean Imputation Techniques for Filling the Missing Observations in Air Pollution Dataset," *Key Eng. Mater.*, vol. 594–595, pp. 902–908, Dec. 2013, doi: 10.4028/www.scientific.net/KEM.594-595.902.

[20] M. Wu, L. Tan, and N. Xiong, "Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications," *Inf. Sci. (Ny).*, vol. 329, pp. 800–818, Feb. 2016, doi: 10.1016/j.ins.2015.10.004.

[21] Z. Li and Q. He, "Prediction of Railcar Remaining Useful Life by Multiple Data Source Fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2226–2235, Aug. 2015, doi: 10.1109/TITS.2015.2400424.

[22] J. Farhan and T. F. Fwa, "Effects of Missing Runway Performance Data on Maintenance Management Decision Making," *J. Transp. Eng.*, vol. 141, no. 10, Oct. 2015, doi: 10.1061/(ASCE)TE.1943-5436.0000794.

[23] Z. Chen *et al.*, "Prediction Reliability Assessment Based on Mahalanobis Distance and GRU in the Application of Bearing RUL Analysis," 2023, pp. 554–567. doi: 10.1007/978-3-031-26193-0_50.

[24] P. W. Tse, S. Gontarz, and X. J. Wang, "Enhanced eigenvector algorithm for recovering multiple sources of vibration signals in machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 7, pp. 2794–2813, Oct. 2007, doi: 10.1016/j.ymssp.2007.02.010.

[25] Y. Li, P. W. Tse, and X. Wang, "Recovery of vibration signal based on a super-exponential algorithm," *J. Sound Vib.*, vol. 311, no. 1–2, pp. 537–553, Mar. 2008, doi: 10.1016/j.jsv.2007.09.036.

[26] M. C. Carnero, R. González-Palma, D. Almorza, P. Mayorga, and C. López-Escobar, "Statistical quality control through overall vibration analysis," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1138–1160, May 2010, doi: 10.1016/j.ymssp.2009.09.007.

[27] W. He, Z.-N. Jiang, and K. Feng, "Bearing fault detection based on optimal wavelet filter and sparse code shrinkage," *Measurement*, vol. 42, no. 7, pp. 1092–1102, Aug. 2009, doi: 10.1016/j.measurement.2009.04.001.

[28] D. Wang, W. Guo, and X. Wang, "A joint sparse wavelet coefficient extraction and adaptive noise reduction method in recovery of weak bearing fault features from a multi-component signal mixture," *Appl. Soft Comput.*, vol. 13, no. 10, pp. 4097–4104, Oct. 2013, doi: 10.1016/j.asoc.2013.05.015.

[29] N. M. Noor, M. M. Al Bakri Abdullah, A. S. Yahaya, and N. A. Ramli, "Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set," *Mater. Sci. Forum*, vol. 803, pp. 278–281, Aug. 2014, doi: 10.4028/www.scientific.net/MSF.803.278.

[30] S. Chiewchanwattana, C. Lursinsap, and C.-H. H. Chu, "Time-series data prediction based on reconstruction of missing samples and selective ensembling of FIR neural networks," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.*, 2002, pp. 2152–2156 vol.5. doi: 10.1109/ICONIP.2002.1201873.

[31] J. Sun, H. Liao, and B. R. Upadhyaya, "A Robust Functional-Data-Analysis Method for Data Recovery in Multichannel Sensor Systems," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1420–1431, Aug. 2014, doi: 10.1109/TCYB.2013.2285876.

[32] M. D. Anis, S. Taghipour, and C.-G. Lee, "Optimal RUL Estimation: A State-of-Art Digital Twin Application," in *2020 Annual Reliability and Maintainability Symposium (RAMS)*, Jan. 2020, pp. 1–7. doi: 10.1109/RAMS48030.2020.9153669.

[33] S. Jose, R. H. Ngouna, K. T. P. Nguyen, and K. Medjaher, "Solving Time Alignment Issue of Multimodal Data for Accurate Prognostics with CNN-Transformer-LSTM Network," in *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, May 2022, pp. 280–285. doi: 10.1109/CoDIT55151.2022.9804090.

[34] Y. Li and L. E. Parker, "A spatial-temporal imputation technique for classification with missing data in a wireless sensor network," pp. 22–26, 2008.

[35] C. Alippi, G. Boracchi, and M. Roveri, "On-line reconstruction of missing data in sensor/actuator networks by exploiting temporal and spatial redundancy," *Proc. Int. Jt. Conf. Neural Networks*, pp. 10–15, 2012, doi: 10.1109/IJCNN.2012.6252689.

[36] R. King, "R EVIEW A review of Bayesian state-space modelling of capture – recapture – recovery data," no. January, pp. 190–204, 2012.

[37] C. Alippi, S. Ntalampiras, and M. Roveri, "Model ensemble for an effective on-line reconstruction of missing data in sensor networks," *Proc. Int. Jt. Conf. Neural Networks*, 2013, doi: 10.1109/IJCNN.2013.6706761.

[38] N. Nower, Y. Tan, and A. O. Lim, "Efficient Temporal and Spatial Data Recovery Scheme for Stochastic and Incomplete Feedback Data of Cyber-physical Systems," *2014 IEEE 8th Int. Symp. Serv. Oriented Syst. Eng.*, pp. 192–197, Apr. 2014, doi: 10.1109/SOSE.2014.29.

[39] S. Siahpour, X. Li, and J. Lee, "A Novel Transfer Learning Approach in Remaining Useful Life Prediction for Incomplete Dataset," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3162283.

[40] J. Wu *et al.*, "Temporal multi-resolution hypergraph attention network for remaining useful life prediction of rolling bearings," *Reliab. Eng. Syst. Saf.*, vol. 247, p. 110143, Jul. 2024, doi: 10.1016/j.ress.2024.110143.

[41] Y. Wang, L. Shen, Y. Zhang, Y. Li, R. Zhang, and Y. Yang, "Self-supervised Health Representation Decomposition based on contrast learning," *Reliab. Eng. Syst. Saf.*, vol. 239, p. 109455, Nov. 2023, doi: 10.1016/j.ress.2023.109455.

[42] H. Guo *et al.*, "Remaining Useful Life Prediction via Frequency Emphasizing Mix-Up and Masked Reconstruction," *IEEE Trans. Artif. Intell.*, pp. 1–10, 2024, doi: 10.1109/TAI.2024.3396422.

[43] X. Song, J. Sun, and C. Li, "An entire life-cycle rolling bearing remaining useful life prediction method using new degradation feature evaluation indicators," *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.*, Jan. 2024, doi: 10.1177/09544089241228943.

[44] L. Zheng, W. Jia, and R. Yang, "Improved adaptive war strategy optimization algorithm assisted-adaptive multi-head graph attention mechanism network for remaining useful life of complex equipment," *AIP Adv.*, vol. 14, no. 5, May 2024, doi: 10.1063/5.0206984.

[45] S. Zhang, Y. Li, D. Zhao, Z. Cui, and Q. Zhang, "Remaining Useful Life Prediction of a Lithium-Ion Battery Based on AE and Modified Transformer," in *2023 China Automation Congress (CAC)*, Nov. 2023, pp. 6216–6221. doi: 10.1109/CAC59555.2023.10450600.

[46] T. Luo, L. Guo, Y. Yu, H. Gao, B. Zhang, and T. Deng, "Enabling the Running Safety of Bearings: A Federated Learning Framework for Residual Useful Life Prediction," in *2023 IEEE 2nd Industrial Electronics Society Annual On-Line Conference (ONCON)*, Dec. 2023, pp. 1–6. doi: 10.1109/ONCON60463.2023.10431274.

[47] R. Bai, K. Noman, K. Feng, Z. Peng, and Y. Li, "A two-phase-based deep neural network for simultaneous health monitoring and prediction of rolling bearings," *Reliab. Eng. Syst. Saf.*, vol. 238, p. 109428, Oct. 2023, doi: 10.1016/j.ress.2023.109428.

[48] C. D. Nguyen and S. J. Bae, "Equivalent circuit simulated deep network architecture and transfer learning for remaining useful life prediction of lithium-ion batteries," *J. Energy Storage*, vol. 71, p. 108042, Nov. 2023, doi: 10.1016/j.est.2023.108042.

[49] Y. Yao, T. Liang, J. Tan, and Y. Jing, "A bearing fault detection and remaining useful life prediction method based on a multi-branch residual feature fusion mechanism and optimized weight allocation," *Meas. Sci. Technol.*, vol. 35, no. 2, p. 025906, Feb. 2024, doi: 10.1088/1361-6501/ad0b67.

[50] B. Huang *et al.*, "DAE-Transformer-based Remaining Useful Life Prediction for Lithium-Ion Batteries in Energy Storages," in *2023 3rd International Conference on New Energy and Power Engineering (ICNEPE)*, Nov. 2023, pp. 125–130. doi: 10.1109/ICNEPE60694.2023.10429188.

[51] Z. Wang, W. Shangguan, B. Cai, and C. Peng, "A weight-allocation-based ensemble remaining useful life prediction approach for a single device," *Measurement*, vol. 224, p. 113885, Jan. 2024, doi: 10.1016/j.measurement.2023.113885.

[52] Z. Zhang, W. Song, Q. Li, and H. Gao, "Multiscale global and local self-attention-based network for remaining useful life prediction," *Meas. Sci. Technol.*, vol. 34, no. 12, p. 125154, Dec. 2023, doi: 10.1088/1361-6501/acf401.

[53] T. Berghout, L.-H. Mouss, O. Kadri, L. Saïdi, and M. Benbouzid, "Aircraft Engines Remaining Useful Life Prediction with an Improved Online Sequential Extreme Learning Machine," *Appl. Sci.*, vol. 10, no. 3, p. 1062, Feb. 2020, doi: 10.3390/app10031062.

[54] C.-G. Huang, C. Men, M. Yazdi, Y. Han, and W. Peng, "Transfer fault prognostic for rolling bearings across different working conditions: a domain adversarial perspective," *Int. J. Adv. Manuf. Technol.*, Jun. 2022, doi: 10.1007/s00170-022-09452-1.

[55] A. Muneer, S. M. Taib, S. M. Fati, and H. Alhussian, "Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine," *Symmetry (Basel).*, vol. 13, no. 10, p. 1861, Oct. 2021, doi: 10.3390/sym13101861.

[56] H. Wang, D. Wang, H. Liu, and G. Tang, "A predictive sliding local outlier correction method with adaptive state change rate determining for bearing remaining useful life estimation," *Reliab. Eng. Syst. Saf.*, vol. 225, p. 108601, Sep. 2022, doi: 10.1016/j.ress.2022.108601.

[57] H. Yang, Z. Sun, G. Jiang, F. Zhao, and X. Mei, "Remaining useful life prediction for machinery by establishing scaled-corrected health indicators," *Measurement*, vol. 163, p. 108035, Oct. 2020, doi: 10.1016/j.measurement.2020.108035.

[58] Z. Ye, Q. Zhang, S. Shao, T. Niu, and Y. Zhao, "Rolling Bearing Health Indicator Extraction and RUL Prediction Based on Multi-Scale Convolutional Autoencoder," *Appl. Sci.*, vol. 12, no. 11, p. 5747, Jun. 2022, doi: 10.3390/app12115747.

[59] S. Ramezani, A. Moini, M. Riahi, and A. C. Marquez, "Predicting the remaining useful life in the presence of the regime-switching behaviour of health index using distance-based evidence theory," *Insight - Non-Destructive Test. Cond. Monit.*, vol. 63, no. 1, pp. 37–46, Jan. 2021, doi: 10.1784/insi.2021.63.1.37.

[60] H. Feng, "An Online Prediction System of Lithium-ion Battery Safety," in *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Aug. 2021, pp. 399–404. doi: 10.1109/AEECA52519.2021.9574221.

[61] A. H. C. Tsang, W. K. Yeung, A. K. S. Jardine, and B. P. K. Leung, "Data management for CBM optimization," *J. Qual. Maint. Eng.*, vol. 12, no. 1, pp. 37–51, Jan. 2006, doi: 10.1108/13552510610654529.

[62] H. Wang and S. Wang, "Discovering patterns of missing data in survey databases: An application of rough sets," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6256–6260, Apr. 2009, doi: 10.1016/j.eswa.2008.07.010.

[63] D. Li, J. Zhang, and H. Wu, "Spatial data quality and beyond," *Int. J. Geogr. Inf. Sci.*, vol. 26, no. 12, pp. 2277–2290, Dec. 2012, doi: 10.1080/13658816.2012.719625.

[64] T.-P. Hong and C.-W. Wu, "Mining rules from an incomplete dataset with a high missing rate," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3931–3936, Apr. 2011, doi: 10.1016/j.eswa.2010.09.054.

[65] N. Nower, Y. Tan, and Y. Lim, "Incomplete Feedback Data Recovery Scheme with Kalman Filter for Real-time Cyber-Physical Systems," 2015.

[66] R. HUANG, J. LI, Z. WANG, J. XIA, Z. CHEN, and W. LI, "Intelligent diagnostic and prognostic method based on multitask learning for industrial equipment," *Sci. Sin. Technol.*, vol. 52, no. 1, pp. 123–137, Jan. 2022, doi: 10.1360/SST-2021-0314.

[67] S. Wang *et al.*, "A Remaining Useful Life Prediction Model Based on Hybrid Long-Short Sequences for Engines," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 1757–1762. doi: 10.1109/ITSC.2018.8569668.

[68] H. FANG, K. A. ESPY, M. L. RIZZO, C. STOPP, S. A. WIEBE, and W. W. STROUP, "PATTERN RECOGNITION OF LONGITUDINAL TRIAL DATA WITH NONIGNORABLE MISSINGNESS: AN EMPIRICAL CASE STUDY," *Int. J. Inf. Technol. Decis. Mak.*, vol. 08, no. 03, pp. 491–513, Sep. 2009, doi: 10.1142/S0219622009003508.

[69] H.-P. Nguyen, W. Fauriat, E. Zio, and J. Liu, "A Data-Driven Approach for Predicting the Remaining Useful Life of Steam Generators," in *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, Nov. 2018, pp. 255–260. doi: 10.1109/ICSRS.2018.8688716.

[70] YuanYuan Li and L. E. Parker, “A spatial-temporal imputation technique for classification with missing data in a wireless sensor network,” in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2008, pp. 3272–3279. doi: 10.1109/IROS.2008.4650774.

[71] Y. Li, P. W. Tse, and X. Wang, “Recovery of vibration signal based on a super-exponential algorithm,” *J. Sound Vib.*, vol. 311, no. 1–2, pp. 537–553, Mar. 2008, doi: 10.1016/j.jsv.2007.09.036.

[72] A. Muneer, S. M. Taib, S. M. Fati, and H. Alhussian, “Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine,” *Symmetry (Basel).*, vol. 13, no. 10, p. 1861, Oct. 2021, doi: 10.3390/sym13101861.