RESEARCH PROGRESS IN MECHANICAL AND MANUFACTURING ENGINEERING
e-ISSN: 2773-4765

RPMME

Vol. 6 No. 2 (2025) 185-192
https://publisher.uthm.edu.my/periodicals/index.php/rpmme

A Machine Learning Framework for Predictive Maintenance of Centrifugal Water Pumps using Remaining Useful Life Estimation

M. Syafiq Syazmir Jamal Nasir[1], Ramhuzaini Abd Rahman[1]*

[1] *Department of Mechanical Engineering, Faculty of Mechanical and Manufacturing Engineering Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, Johor Darul Takzim, MALAYSIA*

*Corresponding Author: huzaini@uthm.edu.my
DOI: https://doi.org/10.30880/rpmme.2025.06.02.020

Article Info

Received: 31 July 2025
Accepted: 31 October 2025
Available online: 10 December 2025

Keywords

Predictive maintenance, Remaining Useful Life (RUL), Centrifugal water pump, Machine learning, Long Short-Term Memory (LSTM), Support Vector Machine (SVM), K-Nearest Neighbors (KNN).

Abstract

Centrifugal water pumps are fundamental to industrial plants, municipal water networks, and agricultural irrigation systems, where continuous operation and variable loading expose them to mechanical degradation and unanticipated failure. The financial and operational consequences of pump downtime underscore the importance of intelligent maintenance strategies that can predict failures before they occur. This study presents a comprehensive machine-learning-based predictive maintenance framework for estimating the Remaining Useful Life (RUL) of centrifugal water pumps using multi-sensor operational data. The approach integrates data preprocessing, extraction of ten statistical time-domain condition indicators, and comparative evaluation of three machine learning algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM). The results demonstrate that the LSTM model achieved superior predictive performance, yielding a Mean Absolute Error (MAE) of 0.0356, a Root Mean Square Error (RMSE) of 0.0064, and a Coefficient of Determination (R^2) of 0.9953. The findings reveal distinct degradation patterns across the monitoring period, confirming the utility of statistical features in modelling pump health. Overall, the study demonstrates that integrating statistical indicators and deep learning provides a reliable and scalable approach to predictive maintenance for centrifugal pump systems.

1. Introduction

Centrifugal water pumps are widely used in industrial, agricultural, commercial, and municipal systems due to their robustness, efficiency, and ability to operate continuously. Their reliability is vital because unexpected pump failures can disrupt critical processes, reduce productivity, and incur substantial repair costs. Continuous exposure to varying hydraulic conditions, fluid pressures, and environmental influences leads to the degradation of mechanical components, including bearings, seals, and impellers. Consequently, ensuring the sustained reliability of centrifugal pumps requires maintenance strategies that can effectively detect, predict, and mitigate the progression of failure.

Traditional maintenance strategies — including corrective maintenance, where failures are addressed only after they occur, and preventive maintenance, where servicing is scheduled irrespective of equipment condition — are increasingly unsuitable for modern industrial contexts (Chaibi et al., 2024; Keane et al., 2024). Corrective maintenance often results in unplanned downtime and high repair costs, while preventive maintenance may fail to detect early-stage degradation, often leading to premature replacement of still-functional components. These limitations highlight a critical gap in conventional approaches: neither strategy effectively addresses real-time equipment health, thereby limiting their efficiency and cost-effectiveness.

Predictive maintenance (PdM) seeks to overcome these limitations by leveraging sensor data, statistical indicators, and machine learning (ML) to monitor equipment condition and forecast impending failures. ML techniques, including k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and various deep learning models, have shown high potential in detecting degradation patterns within complex sensor datasets (Afuan & Isnanto, 2025; Kaewsong & Suriya, 2025; Manivarma et al., 2023). Deep learning architectures—particularly Long Short-Term Memory (LSTM) networks—offer advanced capabilities in temporal sequence modelling, making them highly suitable for Remaining Useful Life (RUL) estimation (Dong et al., 2017; Kizito et al., 2021; Wu et al., 2018). Although ML-based prognostics have been successfully applied to rotating machinery such as bearings, motors, and turbines, the literature addressing centrifugal pump RUL prediction remains relatively limited despite the substantial industrial importance of such pumps.

This gap highlights the need for a comprehensive framework tailored specifically to centrifugal pump prognostics. While previous research has tested various algorithms (Chen et al., 2021; He et al., 2022; Luo et al., 2025; Panda et al., 2018; Ranawat et al., 2021), few studies incorporate pump degradation mechanisms, vibration-based statistical indicators, and comparative ML evaluation within a unified structure (Luo et al., 2025; Panda et al., 2018; Ranawat et al., 2021). Moreover, real-world deployment pathways—such as graphical interfaces for operational personnel—are rarely included in existing work. Addressing these gaps requires a holistic approach that integrates data preprocessing, feature engineering, model training, and deployment considerations.

The objective of this study is therefore to develop, evaluate, and implement a complete predictive maintenance framework for centrifugal water pumps using multi-sensor data collected over several months of operation. The approach includes systematic data preprocessing, extraction of ten vibration-based time-domain statistical condition indicators, development of KNN, SVM, and LSTM models for RUL regression, comparative performance evaluation, and real-time deployment. By providing both analytical and deployment perspectives, the study demonstrates a replicable pathway for implementing machine-learning-based predictive maintenance solutions in industrial pump systems.

The remainder of this paper is organized as follows. Section 2 discusses the degradation mechanisms of centrifugal pumps and existing ML-based prognostic research, providing the background and related works that motivate the present framework. Section 3 details the proposed predictive maintenance pipeline, including data acquisition, preprocessing techniques, feature extraction, and model development. Section 4 presents experimental results and discusses the behavior of condition indicators, model performance, and RUL prediction accuracy. Finally, Section 5 concludes the study and outlines potential directions for future research.

2. Background and Related Works in Predictive Maintenance of Centrifugal Pumps

Centrifugal pumps serve as critical components in fluid transport systems due to their operational simplicity, structural robustness, and adaptability across diverse applications (Eaton et al., 2022; Ellorde et al., 2021). They operate by converting mechanical energy from a rotating impeller into hydrodynamic pressure, enabling efficient fluid movement. However, prolonged operation exposes these pumps to degradation phenomena such as bearing wear, mechanical seal failure, shaft misalignment, impeller erosion, cavitation, and hydraulic imbalance. These degradation modes alter the pump's vibration signatures, thermal characteristics, and overall mechanical response, making vibration-based monitoring an effective method for capturing early signs of failure.

Predictive maintenance emerged as a response to the limitations of conventional maintenance practices, emphasizing real-time monitoring and data-driven prediction of failures. Early PdM approaches relied on handcrafted features and threshold-based rules, but these methods struggled with complex, nonlinear degradation patterns (Wang & Zhao, 2022; Winkel et al., 2023). The introduction of machine learning provided significant improvements, enabling systems to autonomously learn relationships between sensor data and health equipment. Classical ML models such as KNN and SVM have been used extensively for fault classification and condition estimation, but their performance is constrained by their inability to fully capture temporal dependencies and long-range correlations in time-series degradation data (Iaousse et al., 2023; Sekar, 2025; Zhang et al., 2025).

Deep learning approaches have transformed predictive maintenance research by enabling automated feature extraction and robust temporal modelling. Convolutional neural networks (CNNs) excel at learning spatial patterns from vibration spectrograms (Hendriks & Dumond, 2021), whereas recurrent neural networks (RNNs), particularly LSTM, are effective in modeling the sequential relationships that characterize degradation progression (Xu et al., 2024). Despite the effectiveness of deep learning in related domains, its adoption for centrifugal pump RUL prediction remains limited, with few studies offering a comprehensive examination of pump-specific degradation indicators combined with deep learning-based prognostic modelling.

3. Proposed Predictive Maintenance Framework

The predictive maintenance framework developed in this study, as illustrated in Fig. 1, comprises several interconnected stages designed to process raw sensor data, extract meaningful indicators, train predictive models, and deploy these models for real-time use. The framework begins with the acquisition of multi-sensor actual operational data from the internet (Werth, 2021), which was collected from a centrifugal pump system operating over a five-month period. Sensor readings were recorded at one-minute intervals, including measurements of vibration, temperature, flow rate, and system status from 52 channels. This prolonged monitoring period provided sufficient temporal coverage to capture early degradation, intermediate wear, and final failure states.

Preprocessing was undertaken to ensure data quality and reliability. Initial inspection revealed missing values in several channels, with sensor_15 showing almost complete data loss and was therefore removed entirely. Remaining missing entries were handled using column-wise mean imputation, preserving statistical consistency without discarding large amounts of data. All numerical features were normalized using min–max scaling to ensure uniformity and to enhance model convergence in the training phase.

Vibration data were then transformed into ten statistical condition indicators, which served as health-representative features. These indicators—Root Mean Square (RMS), Kurtosis, Peak-to-Peak amplitude, Shape Factor, Peak Factor, Impulse Factor, Clearance Factor, Mean Absolute Value, Standard Deviation, and Crest Factor—provide insight into signal energy, impulsiveness, waveform structure, and variability. These characteristics correlate directly with mechanical degradation patterns such as bearing wear, cavitation effects, and rotor imbalance. A temporal examination of these indicators confirmed that they reflected meaningful degradation behavior throughout the monitoring period.

Three machine learning models were developed and trained to estimate RUL: KNN, SVM, and LSTM. The dataset was divided into 80% for training and 20% for testing. KNN and SVM models used the extracted statistical indicators as input features, with hyperparameters optimized to improve prediction accuracy. The LSTM model employed sequences of historical data to learn temporal dependencies and long-term degradation trends, leveraging its recurrent architecture to model nonlinear temporal dynamics. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).

Mean Absolute Error (MAE):

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

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|} \tag{2}$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

where y_i, and $\hat{y}_i$ are the actual and predicted values of RUL, respectively. Meanwhile, $\bar{y}_i$ is the mean of actual values of the RUL.

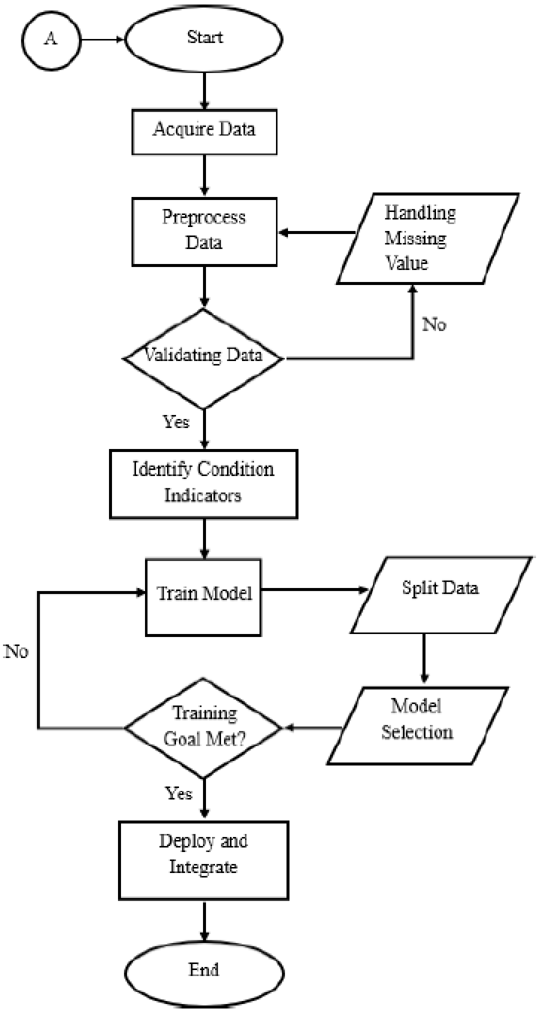

Fig. 1 *Proposed predictive maintenance framework*

4. Results and Discussion

The results of this study provide insight into the degradation patterns of the centrifugal pump and the performance of the various ML models in predicting RUL. The preprocessing stage effectively eliminated all missing values, stabilized feature distributions, and ensured that the dataset was suitable for further analysis. The removal of heavily corrupted sensors and imputation of isolated missing entries preserved the integrity of the training data while maintaining the physical meaning of the recorded signals. Fig. 2 shows the statistical summary of the sensor data.

Condition indicators derived from vibration data displayed identifiable degradation trends. RMS and Mean Absolute Value gradually decreased as the pump approached failure, reflecting diminishing energy transfer and progressive mechanical deterioration. Conversely, indicators such as Kurtosis, Peak Factor, Impulse Factor, and Clearance Factor exhibited pronounced spikes indicative of impact events or abrupt structural disturbances. Standard Deviation (see Fig. 2) showed a progressive increase in variability, consistent with unstable system behavior during the final degradation phase. These observations align with established mechanical failure patterns in centrifugal pumps and confirm the suitability of the selected indicators for health modelling.

The comparative evaluation of machine learning models revealed distinct performance differences, as shown in Table 2. The KNN model exhibited moderate accuracy but struggled with noise sensitivity due to its distance-based nature. The SVM model achieved improved robustness and stability but lacked the ability to capture long-term temporal dependencies inherent in the degradation sequence. The LSTM model delivered the best performance, achieving MAE = 0.0356, RMSE = 0.0064, and R^2 = 0.9953. Its ability to retain historical information and model nonlinear temporal patterns contributed to its superior accuracy in predicting RUL.

The LSTM-based RUL prediction curve as shown in Fig. 3 exhibited a smooth decline during early-stage degradation and a steeper decline near the failure point, closely matching the actual degradation timeline. The absence of oscillations or irregular prediction behavior indicates strong model generalization and stability. This confirms the suitability of LSTM for real-world deployment in pump prognostics applications.

Table 1 *Statistical summary of the sensor data*

	NumMissing	Min	Median	Max	Mean
amp	0	2018-04-01 00:00:00	2018-06-16 11:59:30	2018-08-31 23:59:00	2018-06-16 11:59:30
_02	19	33.1597	51.6493	56.0330	50.8674
_03	19	31.6406	44.2274	48.2205	43.7525
_04	19	2.7980	632.6389	800	590.6739
_05	19	0	75.5768	99.9999	73.3964
_10	19	0	44.2913	76.1069	41.4703
_11	19	0	45.3631	60	41.9183
_12	19	0	32.5158	45	29.1370
_13	19	0	2.9298	31.1876	7.0789
_14	21	32.4096	420.1062	500	376.8600
_19	16	0	665.6724	878.9179	590.8298
_20	16	0	399.3670	448.9079	360.8052
_21	16	95.5277	879.6976	1.1075e+03	796.2259
_23	16	0	981.9250	1.2276e+03	922.6093
_24	16	0	625.8735	1000	556.2354
_26	20	43.1548	861.8696	1.2144e+03	786.4118
_27	16	0	494.4684	2000	501.5066
_28	16	4.3193	967.2799	1.8411e+03	851.6903
_31	16	23.9583	917.7083	1800	863.3231
_33	16	6.4606	512.2717	1.5786e+03	486.4060
_34	16	54.8824	226.3560	425.5498	234.9718
_35	16	0	473.3493	694.4791	427.1298
_36	16	2.2610	709.6680	984.0607	593.0339
_37	16	0	64.2955	174.9012	60.7874
_38	27	24.4792	49.4792	417.7083	49.6559
_39	27	19.2708	35.4167	547.9166	36.6104
_40	27	23.4375	66.4062	512.7604	68.8445
_41	27	20.8333	34.8958	420.3125	35.3651

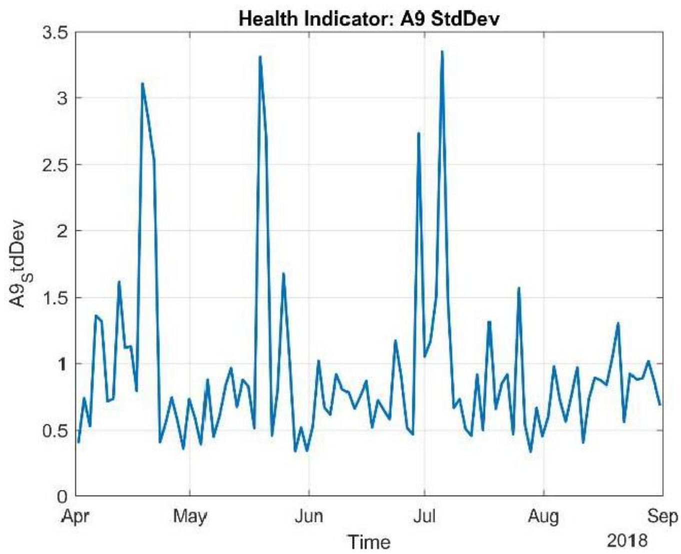

Fig. 2 *Standard deviation of the vibration data*

Table 2 *Model performance for RUL prediction*

Metric	KNN	LSTM	SVM
Mean Absolute Error (MAE)	0.0411	0.0356	0.0916
Root Mean Squared Error (RMSE)	0.0078	0.0064	0.0147
Coefficient of Determination (R^2)	0.9942	0.9953	0.9892

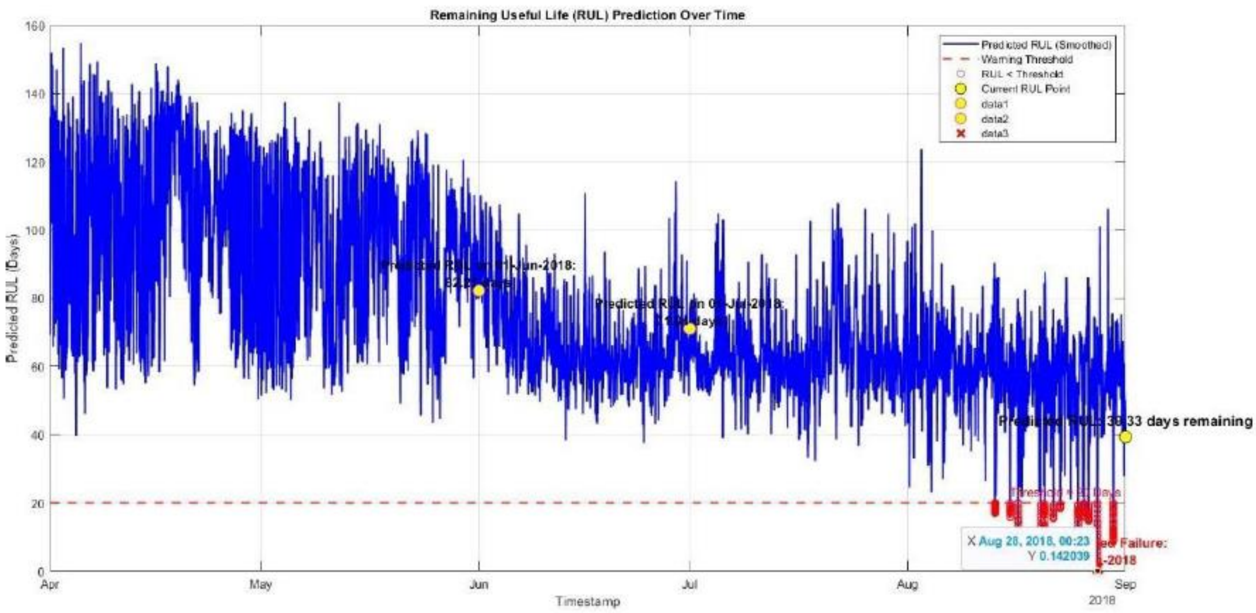

Fig. 3 *LSTM-based RUL prediction over time*

5. Conclusion and Future Work

This study presented a comprehensive machine-learning-based framework for the predictive maintenance of centrifugal water pumps. Through multi-sensor data acquisition, rigorous preprocessing, extraction of ten statistical condition indicators, and the development of three regression models, the research established a systematic approach to predicting Remaining Useful Life. The statistical indicators successfully captured degradation behavior, and the LSTM model provided the most accurate and reliable predictions, outperforming classical methods such as KNN and SVM.

Future research may explore the integration of frequency-domain and time–frequency indicators derived from Fourier and wavelet transforms to enhance feature richness. Hybrid deep learning architectures, including CNN-LSTM combinations and transformer-based models, may further improve prognostic accuracy. Real-time IoT-based deployment of edge devices could facilitate continuous monitoring with minimal latency. Additional studies examining multiple pump types and operating conditions would strengthen generalizability. Coupling RUL prediction with maintenance optimization algorithms could provide enterprise-level decision support.

Acknowledgement

This research is made possible through the assistance of the Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *M. Syafiq Syazmir Jamal Nasir and Ramhuzaini Abd Rahman;* ***data collection, analysis and interpretation of results:*** *M. Syafiq Syazmir Jamal Nasir;* ***draft manuscript preparation:*** *M. Syafiq Syazmir Jamal Nasir and Ramhuzaini Abd Rahman. All authors reviewed the results and approved the final version of the manuscript.*

References

Afuan, L., & Isnanto, R. R. (2025). A comparative study of machine learning algorithms for fall detection in technology-based healthcare system: analyzing SVM, KNN, decision tree, random forest, LSTM, and CNN. *E3S Web of Conferences*, *605*. https://doi.org/10.1051/e3sconf/202560503051

Chaibi, A., Stitou, A., Lagrat, I., Serrou, D., El Mhamedi, A., Tchoffa, D., Oussama, B., & Labiba, B. (2024). IIOT and Deep Learning for Decision-Making: Heading Towards Intelligent Maintenance. *10th Edition of the International Conference on Optimization and Applications, ICOA 2024 - Proceedings*. https://doi.org/10.1109/ICOA62581.2024.10754342

Chen, Y., Yuan, J., Luo, Y., & Zhang, W. (2021). Fault Prediction of Centrifugal Pump Based on Improved KNN. *Shock and Vibration*, *2021*. https://doi.org/10.1155/2021/7306131

Dong, D., Li, X.-Y., & Sun, F.-Q. (2017). Life prediction of jet engines based on LSTM-recurrent neural networks. *2017 Prognostics and System Health Management Conference, PHM-Harbin 2017 - Proceedings*. https://doi.org/10.1109/PHM.2017.8079264

Eaton, A., Ahmed, W. H., & Hassan, M. (2022). Evaluating the Performance Degradation of Centrifugal Pumps Using the Principal Component Analysis. *Journal of Pressure Vessel Technology, Transactions of the ASME*, *144*(2). https://doi.org/10.1115/1.4052522

Ellorde, K. J. A. T., Ignacio, K. J. I. E., Sanchez, E. V. A., Sesima, A. A. A., Umali, R. D., Manuel, M. C. E., Cruz, J. C., Verdadero, M. S., & Tud, R. C. (2021). Numerical Investigation on the Effects of Varying Centrifugal Impeller Blade Number on Pump Head and Efficiency Using ANSYS 2020 R2. *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, 219–224. https://doi.org/10.1109/I2CACIS52118.2021.9495917

He, D., Li, R., Sun, S., & Guo, P. (2022). Prediction of gas-liquid two-phase pressure increment of a centrifugal pump based on machine learning. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *38*(4), 33–41. https://doi.org/10.11975/j.issn.1002-6819.2022.04.004

Hendriks, J., & Dumond, P. (2021). Vibration-based prognosis of rolling element bearings with a 3D convolutional neural network. *"Advances in Acoustics, Noise and Vibration - 2021" Proceedings of the 27th International Congress on Sound and Vibration, ICSV 2021*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117462039&partnerID=40&md5=29f675f86929fc002917eddd318f82fb

Iaousse, M., Jouilil, Y., Bouincha, M., & Mentagui, D. (2023). A Comparative Simulation Study of Classical and Machine Learning Techniques for Forecasting Time Series Data. *International Journal of Online and Biomedical Engineering*, *19*(8), 56–65. https://doi.org/10.3991/ijoe.v19i08.39853

Kaewsong, A., & Suriya, A. (2025). Comparative Analysis of Machine Learning Techniques for Effective Anomaly Detection in Humidity and Temperature Data in WSNs. *Proceedings - IEECON 2025: 2025 13th International Electrical Engineering Congress: Carbon Neutrality: Challenges and Solutions Based on Sustainable Power of Nature*. https://doi.org/10.1109/iEECON64081.2025.10987809

Keane, B., Lim, Z. M., Elkilany, A. R., Kong, P. H., Teoh, J. K., Yeoh, Z. Y., & Chu, Y. B. (2024). Integration of Big Data and Machine Learning Solutions as Preventive Maintenance Strategies to Improve Manufacturing Productivity. *Lecture Notes in Electrical Engineering*, *1253 LNEE*, 11–18. https://doi.org/10.1007/978-981-97-6937-7_2

Kizito, R., Scruggs, P., Li, X., Devinney, M., Jansen, J., & Kress, R. (2021). Long Short-Term Memory Networks for Facility Infrastructure Failure and Remaining Useful Life Prediction. *IEEE Access*, *9*, 67585–67594. https://doi.org/10.1109/ACCESS.2021.3077192

Luo, H., Zhou, P., Cui, J., Wang, Y., Zheng, H., & Wang, Y. (2025). Energy performance prediction of centrifugal pumps based on adaptive support vector regression. *Engineering Applications of Artificial Intelligence*, *145*. https://doi.org/10.1016/j.engappai.2025.110247

Manivarma, D., Akilandeswari, A., & Radhika, S. (2023). Performance Analysis of Vehicle Detection using Support Vector Machine Comparing with K-Nearest Neighbor Algorithm. *Proceedings of the 2nd IEEE International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2023*. https://doi.org/10.1109/ACCAI58221.2023.10199912

Panda, A. K., Rapur, J. S., & Tiwari, R. (2018). Prediction of flow blockages and impending cavitation in centrifugal pumps using Support Vector Machine (SVM) algorithms based on vibration measurements. *Measurement: Journal of the International Measurement Confederation*, *130*, 44–56. https://doi.org/10.1016/j.measurement.2018.07.092

Ranawat, N. S., Kankar, P. K., & Miglani, A. (2021). Fault diagnosis in centrifugal pump using support vector machine and artificial neural network. *Journal of Engineering Research (Kuwait)*, *9*, 99–111. https://doi.org/10.36909/jer.EMSME.13881

Sekar, M. (2025). Machine learning based fault detection and classification for predictive maintenance of gas turbine engines: a comprehensive benchmarking analysis on various models. *Aircraft Engineering and Aerospace Technology*. https://doi.org/10.1108/AEAT-04-2025-0160

Wang, Y., & Zhao, Y. (2022). Multi-Scale Remaining Useful Life Prediction Using Long Short-Term Memory. *Sustainability (Switzerland)*, *14*(23). https://doi.org/10.3390/su142315667

Winkel, F., Wallscheid, O., Scholz, P., & Bocker, J. (2023). Pseudolabeling Machine Learning Algorithm for Predictive Maintenance of Relays. *IEEE Open Journal of the Industrial Electronics Society*, *4*, 463–475. https://doi.org/10.1109/OJIES.2023.3323870

Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, *275*, 167–179. https://doi.org/10.1016/j.neucom.2017.05.063

Xu, R., Wang, Y., Peng, M., Cui, H., Wang, H., & Kuang, C. (2024). PREDICTIONS OF COMPONENT REMAINING USEFUL LIFE USING BAYESIAN LSTM FOR REACTOR COOLANT PUMP. *Proceedings of 2024 31st International Conference on Nuclear Engineering, ICONE 2024*, *1*. https://doi.org/10.1115/ICONE31-135439

Zhang, S., Wang, X., Jiao, F., Zeng, Z., Yuan, G., Zhang, R., & Zhao, C. (2025). Fault prediction and health status evaluation method of GIS equipment based on time series neural network. *Proceedings of SPIE - The International Society for Optical Engineering*, *13689*. https://doi.org/10.1117/12.3071705

Werth, J. (2021, August 10). *LSTM for predictive maintenance on pump sensor data*. Towards Data Science. https://towardsdatascience.com/lstm-for-predictive-maintenance-on-pump-sensor-data-b43486eb3210/