

AI-Driven Predictive Maintenance Technology for Intelligent Manufacturing: A Review

Yanze Liu

*Lester International College, Dalian University of Technology, Wuhan, China
lyz0817@mail.dlut.edu.cn*

Abstract. Based on the development trends and application cases of artificial intelligence-powered predictive maintenance technology in intelligent manufacturing, a general analysis will be carried out in this paper. The old way of fault repair and time-based preventive maintenance has the following problems: there will be significant business disruption due to unexpected failure; maintenance will be carried out unnecessarily; and resources will not be used efficiently. With the help of the new IoT platform and AI algorithm, condition monitoring for the maintenance of industrial equipment has begun to be used in practice. The whole closed-loop technical architecture of "sensing - evaluation - determination - actuation" and the operating logic of the five basic stages of data acquisition, improvement, analysis, strategic planning and feedback are introduced here, along with the research and application results of some representative machine learning and deep learning architectures for fault diagnosis and remaining useful life estimation. At the same time, it is pointed out that there are still issues with the small-sample generalisation ability and model transparency of this method, and this provides a necessary direction for further improvement and practical application of industrial smart operation and maintenance platforms.

Keywords: Predictive maintenance, intelligent manufacturing, industrial ai, machine learning, remaining useful life prediction

1. Introduction

Given the current changes in global industry, which are progressing towards intelligence and digitalisation, "Smart Manufacturing" has emerged as an essential new form of enterprise to improve efficiency and productivity in the new era of the Industrial Internet of Things. Predictive maintenance technology has been added to the whole technical system of "smart manufacturing" gradually, and thus, instead of the old reactive repair and preventive maintenance mode, intelligent monitoring is now employed. According to Susto GA and others, the three general categories of maintenance are reactive maintenance, preventive maintenance and predictive maintenance [1]. Reactive maintenance is carried out only after a single component fails; as a result, the entire production line must be stopped, production output declines, there is an increase in breakdowns of related parts in the surrounding area, and finally, the management of the factory is disturbed. Based on the past empirical data and relevant industrial experience, regular maintenance of the equipment

will be carried out. Nevertheless, due to the unpredictable and irregular nature of the production cycle, this way may lead to unnecessary maintenance operations and thus reduce the financial efficiency of the equipment.

In recent years, with the development and spread of technologies such as the Internet of Things and artificial intelligence, it has become possible to collect and store a large amount of data from the operation of machinery in real time to create a rich data foundation for condition monitoring. Algorithms such as machine learning, deep learning and transfer learning are also good at complex pattern recognition, construction of non-linear correlation models, time series prediction, etc., and can be applied to learn the deterioration trends of equipment based on historical and current data to accurately identify early-stage faint fault indicators and reasonably predict the remaining useful life of the equipment. Therefore, deep integration of artificial intelligence with cyber-physical systems for predictive maintenance is now one of the first applications of artificial intelligence [2].

Based on the above system, the current paper will examine all new achievements and applications of artificial intelligence technology in predictive maintenance for smart factories. First, introduce the basic definition and structure of prognostic maintenance. Then it will be explored how all sorts of artificial intelligence technologies can be applied to investigate the reasons for diagnostic failure and lifespan prediction problems and improve the repair strategy. It also addresses the main problems that have arisen so far and future directions for research. Based on the analysis of the current literature, this paper aims to provide some reference points for scholars and practitioners in this area to promote the intelligent modernisation and creative development of intelligent manufacturing maintenance architecture together.

2. Overview and technical framework of predictive maintenance in intelligent manufacturing

2.1. Overview

Predictive Maintenance (PdM) in Smart Manufacturing is a type of production and maintenance mode that belongs to the Industry 4.0 system. At the IoT sensing tier, many types of data from the machine, such as oscillation, heat, sound, electrical current, etc., are collected. Signal filtering, statistics and machine/deep learning technology can be used to monitor the changes in some parameters of the equipment in real time and predict the remaining useful life (RUL) and approximate failure time intervals. Therefore, the best maintenance plan can be developed before the function of the equipment drops to a serious level [3]. PdM is no longer a problem to be handled at a fixed time or in a regular manner; instead, it will be triggered by changes in the operating environment of the equipment and will therefore be in the form of condition-based maintenance. Reduce the cost of unexpected idle time and redundant service expenses to extend the life of equipment. Functionally closed systems have been put forward as expansions of condition-based maintenance (CBM), and the five stages of this are information acquisition, data processing, analysis, strategy development and feedback application. Machine learning is also used in the above architecture. In accordance with the learning logic, many machine learning frameworks have been used, such as: (1) Supervised learning is employed to build regression or classification models based on labeled data for RUL prediction and surface defect recognition; (2) Unsupervised learning is used for clustering, feature extraction and outlier detection of raw data to find hidden fault patterns and track changes in the production process; (3) Semi-supervised learning and deep learning can be applied to build deep neural networks that can learn high-order non-linear features and solve problems of sparse data and complex working states; (4) Reinforcement learning is used to obtain feedback through interaction with the environment and optimise dynamic planning and flexible

control strategies. First of all, one of the advantages of machine learning in smart manufacturing is that it can change the old way of static production based on experience into a closed-loop smart structure for data-oriented, self-perceptive and self-improving operation to achieve goals such as predictive maintenance, adjustment of process parameters, real-time quality inspection, intelligent coordination, etc. [4].

2.2. Overall architecture

At the bottom is the basic concept and technical structure of smart manufacturing predictive maintenance; that is to say, it uses industrial big data and applies artificial intelligence technology to mechanical knowledge to build a "perception-analysis-decision-execution" architecture system. The bottom tier is the data collection module; it will use sensors, Industrial IoT and other equipment to obtain current data from the machinery and collect information on its condition, temperature, vibration, etc. After cleaning, normalisation and feature selection have been carried out to obtain a small set. The middle tier is the intelligent analysis core, and usually statistics and machine learning frameworks are used here. Based on the correlation of equipment failure data, a degradation framework has been established statistically, and then machine learning algorithms such as Neural Networks and Random Forest are used to predict the trend of equipment failure according to this information. Combine the two to give more importance to prediction accuracy and interpretability, and then obtain the RUL and failure probability. At the top is the decision and application module; according to the status of the repaired equipment and production plans, flexible maintenance blueprints will be generated to organise inventory and schedule platforms for spare parts during the repair period, and the algorithm will be constantly optimised based on feedback. It has several levels in its architecture, and can be divided into edge-side instantaneous surveillance and cloud-side comprehensive optimisation; it can meet the different requirements of various levels, such as individual machines and factory line networks, and finally achieve the combined goals of reduced costs, increased productivity, improved quality and security, etc.

3. Basic structure of predictive maintenance in intelligent manufacturing

According to the cycle of the technical framework for intelligent manufacturing anticipatory maintenance, the five consecutive essential stages to achieve this goal can be organised as follows: data acquisition, data refinement, data analysis, decision-making and response transmission. All the stages are interconnected and together constitute the whole "Perception-Analysis-Decision-Execution" repair loop. To be precise, the data collection level is Industrial Internet of Things; it can gather a large number of different pieces of data from all over and provide normalised primary data for the upper-tier programme. The Data Transformation Layer corrects the abnormal situation and then filters and alters the attributes to obtain the wear health index of the equipment. The first part of the data analysis layer is the physical model and empirical algorithm for fault diagnosis and life prediction. Determine a level for optimisation of the maintenance plan based on the condition of assets. At the time of carrying out task orders and returning data, the response implementation tier will adaptively optimise the model and carry out circular optimisation.

3.1. Data acquisition

The main way to collect data is through Industrial IoT and distributed sensor systems that can gather all kinds of information about the working conditions of machinery and provide instant access.

Collected materials are the mechanical sensing indicators for oscillation, heat, strain and power, as well as the working records of the PLC, apparatus operation logs and past repair documents [2, 3]. At the edge device, all the protocol decoding and standardisation of packaging will be done to build an original database covering the entire life cycle of the machinery and provide a continuous source of high-precision surveillance data for later analysis.

Now, a distributed sensor network based on the Internet of Things has been established to collect data from the production equipment. A typical type is a piezoelectric accelerometer to obtain vibration acceleration data, a thermocouple that converts heat into a voltage signal, and a Hall effect current sensor which is a kind of electrical current sensor. Hardware Functions are the "sensory organs" of machines, and they can turn mechanical parameters into measurable electrical signals. Collect the data, use the protocol conversion function at the edge gateway to convert it, synchronize all the information streams according to time to keep the order of creation, construct a diverse dataset with multiple types covering the entire life cycle of the equipment, and prepare standardised feed material for the next evaluation [5].

3.2. Data processing

To solve the problems of acoustic interference, missing data and different scales in the first-time monitoring data, a regularisation pipeline for information will be employed here. Eliminate the outliers in a statistical way to reduce noise and then normalise the values; reduce the number of features by feature elimination. Attribute identification and aggregation are then carried out on the time-domain metric, such as the mean and kurtosis, etc., and also in the tools for frequency-domain and time-frequency domain (wavelet analysis, Empirical Mode Decomposition, etc.) to convert low Signal-to-Noise Ratio (SNR) primitive signals into diagnostic indicators of machinery wear patterns for predictive modelling [4].

To address the noise disturbance (unrelated random perturbation), data gap and scale variation in the acquired data, signal processing and statistics will be employed in this step [6]. To reduce high-frequency interference, a low-pass filter can be employed to keep only the low-frequency parts of the signal, or a wavelets transformation is used to split the signal into many components based on different frequencies. Interpolation or the Expectation-Maximisation (EM) method is used to complete the missing data. Standardise (or normalise) the mean and standard deviation of the Z-score, or use Min-Max scaling to rescaled it to a certain range. Based on the above, degradation characteristics can be obtained in the form of time-domain metrics (mean, kurtosis, etc.), frequency-domain metrics (FFT spectral peaks) and time-frequency analysis (wavelet packet decomposition). PCA is to reduce the number of variables and to form a feature vector with information about the health of the equipment [7].

3.3. Data analysis

As the foundation of the technology, this layer includes degradation models for equipment, such as mechanistic degradation models and data-oriented technology tools, such as Artificial Neural Networks, Support Vector Machines, Random Forest or LSTM [2], as well as Paris' law, Wiener process, Weibull distribution [4], etc. One way to address the above problems is as follows: (1) Early detection of faults and abnormal situations; (2) Prognostic prediction of the remaining useful life (RUL); (3) Precise assessment of the current health status of the equipment and calculation of a confidence interval for the prediction to balance prediction accuracy with the actual requirements of engineering [2, 4].

According to wear-out theory, a few anomalies have occurred recently, and thus an estimation of remaining useful life can be made. The physical method is Paris' law for the development of fatigue cracks under stress cycles and the Wiener process to model the path of functional deterioration by means of stochastic calculus; only then can the causes of failure be accurately predicted. According to the direction of data analysis, Random Forest, Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks have been used to find changes in the archived data over time. In order to improve the accuracy and conciseness of the results, an integrated modelling strategy can be employed; that is to say, the output of the formulas for physics-based laws can serve as the input for a neural network, or Physics-Informed Neural Networks (PINNs) are used to add mechanical constraints to the objective function and guarantee the correctness of the forecast while also meeting the demand for technical interpretability [2].

3.4. Decision-making

Judgment is based on the asset's deterioration state and the probability density function of RUL from the forecast in the analytical tier. Based on the production schedule, inventory levels of parts, availability of repair staff, financial conditions, etc., a repair optimisation system will be set up to reduce overall repair costs and improve equipment reliability. It will take the initiative to arrange regular maintenance plans, decide on service grades and asset distribution schemes, etc., to help directors make financially prudent preservation choices [4].

Obtain the condition data of the machinery and its corresponding RUL probability curves; then, based on this, consider the production schedule, spare part stock, repair resource availability, operating downtime cost, etc., and formulate an optimisation framework at the executive level. Generally speaking, dynamic programming can be used to choose an option with the minimum total cost; opportunistic maintenance schemes will add the time until the next failure of regular maintenance, etc., and many-objective optimisation methods can be employed to consider both cost and reliability/hazard simultaneously [8]. Many heuristic methods have been applied to the problem; for example, Genetic Algorithms imitate the process of natural selection to find a good solution, and Particle Swarm Optimisation uses the cooperation of a large number of particles to find an optimal answer, so a suitable repair period, level of maintenance (routine maintenance, part replacement or comprehensive overhaul) and resource allocation plan will be obtained to enhance the overall effect of maintenance.

3.5. Feedback execution

The realisation feedback tier is based on the maintenance schedule given by the analytical guidance layer, and then it will be implemented as field operation tasks. At the same time, it also records and supervises the repair process (service period, replaced parts, maintenance results). Send the operation result and post-maintenance verification data back to the database, update and adjust the coefficient of the algorithm in real time to form a closed-loop improvement process of "Sensing - Interpretation - Planning - Acting - Re-sensing", and thus continuously improve the accuracy of forecasts and strategies on the platform over time [3].

At this time, the preservation choice will be turned into an operation task and all important data during the execution process, such as the ID of the replaced part, maintenance start and end times, correction efficiency, etc., will be recorded. Operational information is transmitted to the main repository via the manufacturing network for the purpose of model improvement. To deal with concept drift (changes in the pattern of data over time) caused by alterations in the operating

conditions of the equipment, incremental learning and transfer learning can be used to extend the learning process based on new data and apply some knowledge gained from a previously trained model to the current problem. All of the above will be adjusted continuously in accordance with the actual operating situation and to meet changes in the demand for services or other reasons for maintenance, etc.

4. Machine learning and deep learning models for prediction

Research on predictive frameworks based on machine learning and deep learning is the main driving force for the development of the technology infrastructure of prognostic maintenance in smart manufacturing. Data-driven thinking does not need to find the exact mathematical formula of a physical model; it can also determine the condition of equipment and predict its remaining useful life (RUL) accurately by extracting implicit patterns from a large amount of old and new data without knowing the structure of a complex system. The three main directions of research in this field are feature abstraction learning, temporal model architecture and generalisation to all kinds of working conditions.

First of all, in terms of the stages of feature abstraction and degradation simulation, it is necessary to develop Health Indicators that can accurately reflect the wear-out process of the asset. Both SVM and GPR are also very sensitive to the quality of feature engineering in shallow machine learning models based on the traditional method. For the most part, scholars have taken various ways in signal processing to extract features from the time and frequency domains of non-linear data, such as oscillations and electric currents. However, handcrafted features have difficulty covering all kinds of fault patterns and are not easily generalisable. Therefore, deep learning can learn the features of all kinds of data automatically by means of a few layers of non-linear mapping. For instance, a one-dimensional Convolutional Neural Network (1D-CNN) can be used to process the raw signal; by means of convolution, it will automatically extract local frequency interval power and impact traits and thus solve the problem of manual feature engineering and improve the descriptive ability of degradation indicators [2].

In terms of the sequential nature of RUL estimation, more research has been conducted to obtain long-range correlations in the operating conditions of the machinery. Recurrent Neural Networks (RNNs) and their gated versions, Long Short-Term Memory (LSTM) framework and Gated Recurrent Units (GRUs), have been used to solve the problem of vanishing gradients in the time-series data of instrument deterioration [9]. In recent years, mixed deep learning frameworks have been used to add space and time information; for example, a CNN and an LSTM can be combined to extract the spatial dependency of multi-sensor data with a CNN, and then the temporal change of this information can be determined by an LSTM [10]. The Transformer framework is also used to solve the problem of long-sequence data, and self-attention can be employed to give varying weights to different parts of time; therefore, changes in vital signs can be responded to more promptly.

Finally, there are problems in model explainability and cross-domain applicability in the current research. The industrial environment has a small amount of fault data (data imbalance) and many different operating conditions (covariate shift). To address this problem, research has started to use Physics-Informed Neural Networks (PINNs) and add the mechanical degradation laws of machinery as regularisation terms in the loss function to make data-driven models consistent with physical laws and easier to interpret. At the same time, to address the few-shot problem, Transfer Learning and Domain Adaptation methods have been introduced. The aim of this transfer is to use the knowledge obtained in experimental setups or similar machine categories (source domains) and introduce it to the actual industrial environment (target domain) by reducing the distribution gap between the two;

thus, it can provide reliable predictions in data-scarce circumstances and ensure the stable operation of intelligent maintenance frameworks for all kinds of manufacturing facilities [11].

5. Conclusion

The combination of artificial intelligence and predictive maintenance is now in the stages of conception and application. Based on the above data and physical examination, both the number of unplanned shutdowns and repair costs of the fragmented production line and high-end equipment have dropped considerably. However, there are still problems with the enhancement of algorithm transparency by means of physics-based methods and the improvement of the generalisation ability for small-data and changing-environment problems. In the future, with the help of edge nodes and virtual replicas, smart monitoring will be autonomous and able to make continuous decisions to advance the electronic and environmental friendly development of industries around the world.

References

- [1] Susto, G. A., Schirru, A., Pampuri, S., et al. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3), 812-820.
- [2] Li, J. Q., & Hu, L. B. (2020). A review of equipment predictive maintenance methods based on machine learning. *Computer Engineering and Applications*, 56(21), 11-19.
- [3] Sun, H. B., Chen, C., & Xu, J. K. (2025). Research on big data-driven predictive maintenance technology for smart manufacturing. *Shanghai Light Industry*(1), 129-131.
- [4] Yuan, Y., Zhang, Y., & Ding, H. (2020). Key technologies of industrial artificial intelligence and their application status in predictive maintenance. *Acta Automatica Sinica*, 46(10), 2013-2030.
- [5] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics using condition-monitoring techniques. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- [6] Caesarendra, W., & Tjahjowidodo, T. (2017). A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation. *Machines*, 5(4), 25.
- [7] Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, 96, 1-15.
- [8] De Jonge, B., & Scarf, P. A. (2020). A review on maintenance optimization. *European Journal of Operational Research*, 285(3), 805-824.
- [9] Wang, J. J., Yang, S. P., Liu, Y. Q., et al. (2021). Bearing remaining useful life prediction method based on spatial convolutional long short-term memory neural network. *Journal of Mechanical Engineering*, 57(21), 88-95.
- [10] Zhao, R., Yan, R., Chen, Z., et al. (2019). Deep learning and its applications to machine health monitoring: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 30(4), 967-987.
- [11] Ragab, M., Chen, Z., Wu, M., et al. (2023). A survey on deep transfer learning for machinery remaining useful life prediction under different working conditions. *Reliability Engineering & System Safety*, 230, 108947.