Journal of Engineering, Project, and Production Management
2026, 16(4), 2026-74

An Intelligent Decision-Support System for Hydroelectric Power Unit Operation and Maintenance Empowered by Digital Twins

Guobiao Zhong[1], Xi'and Zhou[2], and Zhongliang Shao[3]
[1] General Manager and Chief Engineer, R & D Center, Guangdong HongDaXin Electronic Technology Co., Ltd., Guangzhou, 510045, China
[2] Senior Engineer, Hydropower Branch, Guangdong Energy Group Co., Ltd., Heyuan, 517099, China
[3] Associate Professor, School of Automation and Information Technology, Guangdong Polytechnic of Water Resources and Electric Engineering, Guangzhou, 510635, China, E-mail: ShaoZL@126.com (corresponding author)

Project Management
Received January 18, 2026; revised April 21, 2026; accepted May 5, 2026
Available online May 29, 2026

Abstract: With the continuous rise in global energy demand, the Operation and Maintenance (O&M) of large hydropower plant units is becoming increasingly intelligent and digital. However, current O&M management still suffers from low efficiency and limited fault-prediction accuracy under complex operating conditions and equipment failures. To address these challenges, this study proposes an intelligent decision support system that integrates Digital Twin Technology with deep learning techniques, adopting a full life cycle approach for hydropower unit O&M. Numerous experiments have been conducted to demonstrate that the effectiveness of the system is 96.4% concerning fault diagnosis, while the fault reaction time is maintained at twelve minutes. The availability of the unit exceeds 95% under noise conditions of zero and 20 dB. It approaches 100%. All this information indicates that the designed decision support system can monitor the units effectively, manage any faults, and provide optimal control of hydropower units. This work is valuable since it may be used as the basis for implementing intelligent O&M practices in other energy fields.

Keywords: Digital twin technology, hydropower plant units, operation and maintenance, intelligent decision-making, deep learning.

DOI:10.32738/JEPPM-2026-74

1. Introduction

Hydropower, as a green and clean energy source, plays a vital role in ensuring a stable electricity supply during the construction and operation of large-scale hydropower plants (Darani et al., 2023). With continuous technological advancements, various Operation and Maintenance (O&M) management approaches for hydropower units have emerged, based on technologies such as the Internet of Things (IoT), big data analysis, and deep learning methods (Tong et al., 2023; Zhong et al., 2024). However, the intelligence of these technologies is still limited, and there is significant room for improvement in fault prediction, health assessment, and dynamic optimization scheduling. To address the shortcomings of current applications, Digital Twin Technology (DTT) offers a new solution. DTT can comprehensively reflect the operating status of hydropower units through real-time monitoring and virtual model reconstruction, offering data support for fault diagnosis and performance optimization (Don et al., 2025). In deep learning methods, particularly Long Short-Term Memory (LSTM) networks and Depthwise Separable Convolutional (DSC) networks, the excellent ability to process time series data enables precise predictions by capturing long-term dependencies in equipment operation (Putri et al., 2024; Mathai et al., 2024). Therefore, DTT is utilized in this study for real-time monitoring and simulation of hydropower units and applies DSC and LSTM for fault diagnosis and predictive trend forecasting, constructing a Digital Twin and Deep Learning-driven Life Cycle (DTDL) O&M decision support system.

The main contribution of this study is to build the full life-cycle analysis module as an independent decision input. It is treated as equally important as the fault diagnosis and trend prediction modules and is integrated with them in a unified framework. The approach differs from the conventional O&M technique, which considers only one slice of the system's

operational condition. Systematic testing under different operating modes proves that the new approach works better with life-cycle O&M problems.

2. Related Works

DTT creates virtual models of physical devices or systems to monitor and optimize their behavior, performance, and status in real time. DTT has been widely applied in various fields, especially in industries that require highly accurate predictions and optimizations (Starly et al., 2023). For instance, Tang et al. (2023) proposed a collaborative capability optimization model based on DTT to address issues like slow data flow and idle equipment resource waste in smart manufacturing. Jyeniskhan et al. (2023) proposed a framework that integrates machine learning models and DTT. The system achieved an average defect-detection accuracy of 92%, with detection rates of 91% for defective objects and 94% for non-defective objects. Valiollahi et al. (2024) proposed a performance evaluation and modeling method for two complementary real-time positioning systems to improve accuracy in industrial environments. Njoku et al. (2025) developed a strategy to incorporate DTT into battery management systems to solve the problems associated with the processing of vast quantities of information for precise monitoring and control. Özkan et al. (2024) introduced a hybrid of DTT and deep reinforcement learning for optimal microgrid management. Proximal policy optimization and soft actor-critic approaches resulted in an increase in returns by 81.7% and 56.12%, respectively.

Intelligent decision support systems, which combine various data analysis and optimization algorithms, assist O&M personnel in making real-time decisions in complex environments, improving the reliability and cost-effectiveness of equipment operations (Oliveira et al., 2023). For example, Araújo et al. (2023) proposed an intelligent, data-driven decision support method for agricultural systems. Lin et al. (2025) proposed a method combining a circular intuitionistic fuzzy weighted average model with an intelligent decision support system to address complex decision-making problems in container terminals and service quality evaluation. A case study showed that the method provided effective information to terminal managers, policymakers, and stakeholders. Gong et al. (2024) proposed a belief rule-based model based on a multi-layer tree structure to address the uncertainties and lack of standards in complex systems for evaluating aviation engine performance. Corbacho-Abelaira et al. (2024) proposed an intelligent decision support system based on deep learning techniques. Simulation tests proved that the system was able to diagnose high-risk patients using chest X-rays. The method that was designed by Zhu et al. (2024) for developing an intelligent maintenance system aimed at overcoming the difficulties in operating a hydropower station due to the diversity of equipment and rugged terrain. The experimental results indicated that this method digitalized and automated maintenance management and workflows, enabling hydropower station maintenance operations to be conducted safely, efficiently, and reliably.

In summary, DTT improves state observability, and learning-based models improve fault diagnosis and trend prediction, but an integrated, closed-loop O&M framework for large hydropower units remains lacking. This work therefore proposes a DTDL framework that unifies twin synchronization, lightweight deep learning, and life-cycle-aware decision fusion to generate actionable maintenance decisions.

3. Intelligent Decision Support System based on DTT, DSC, and LSTM

3.1. Development of O&M Module Combining DTT and Deep Learning

Intelligent decision-making in hydroelectric unit O&M significantly improves fault prediction accuracy and resource scheduling efficiency, achieving the optimal maintenance timing, cost minimization, and system availability maximization (An et al., 2024). In the data preprocessing stage, to ensure data consistency across different sensors, the raw sensor data is standardized using Eq. (1).

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

In Eq. (1), x and x' represent the raw sensor data and the normalized data, respectively. μ is the mean of the raw data, and σ is the standard deviation of the raw data. To enhance the O&M efficiency and decision quality of the hydroelectric unit, the O&M system must handle daily tasks. DTT precisely analyzes the unit's health status and operational trends and provides optimization recommendations. The dynamic system simulation in DTT describes how the system changes over time, as shown in Eq. (2).

$$\frac{dX}{dt} = A \cdot X + B \cdot U \tag{2}$$

In Eq. (2), X represents the state vector, such as unit speed and water level, and U represents the control inputs, such as regulating water flow and power generation load. A is the state transition matrix, which represents the system's inherent dynamics, and B is the control matrix, which represents the impact of control inputs on the state. The output of the DTT model reflects the virtual unit's state at each time step, as shown in Eq. (3).

$$Y = C \cdot X + D \cdot U \tag{3}$$

In Eq. (3), C and D represent the output matrix and control matrix, respectively, and Y is the output, such as power generation and unit efficiency. In the discretization process of the DTT model, the system's discrete dynamics are

simulated by discretizing the state Equation. The discretized state Equation is shown in Eq. (4).

$$X_k = A_d \cdot X_{k-1} + B_d \cdot U_k \tag{4}$$

In Eq. (4), A_d and B_d represent the discrete state transition and control matrices, respectively, X_k is the system state at the discrete time, and U_k is the control input at the discrete time. Eqs. (2) through (4) adopt a discrete linear state-space form to represent a locally linearized subsystem of the hydroelectric unit around a nominal operating point (e.g., quasi-steady head and load). The linearizations of A and B occur within the operating region employed for synchronization; any remaining nonlinear coupling and other unmodeled dynamics are incorporated into a disturbance input that is dealt with through periodic twin updates. Consequently, the state space realization is meant for local simulation and state estimation purposes, not as a global model of the nonlinear system. The application of DTT in the hydroelectric unit O&M is shown in Fig. 1.

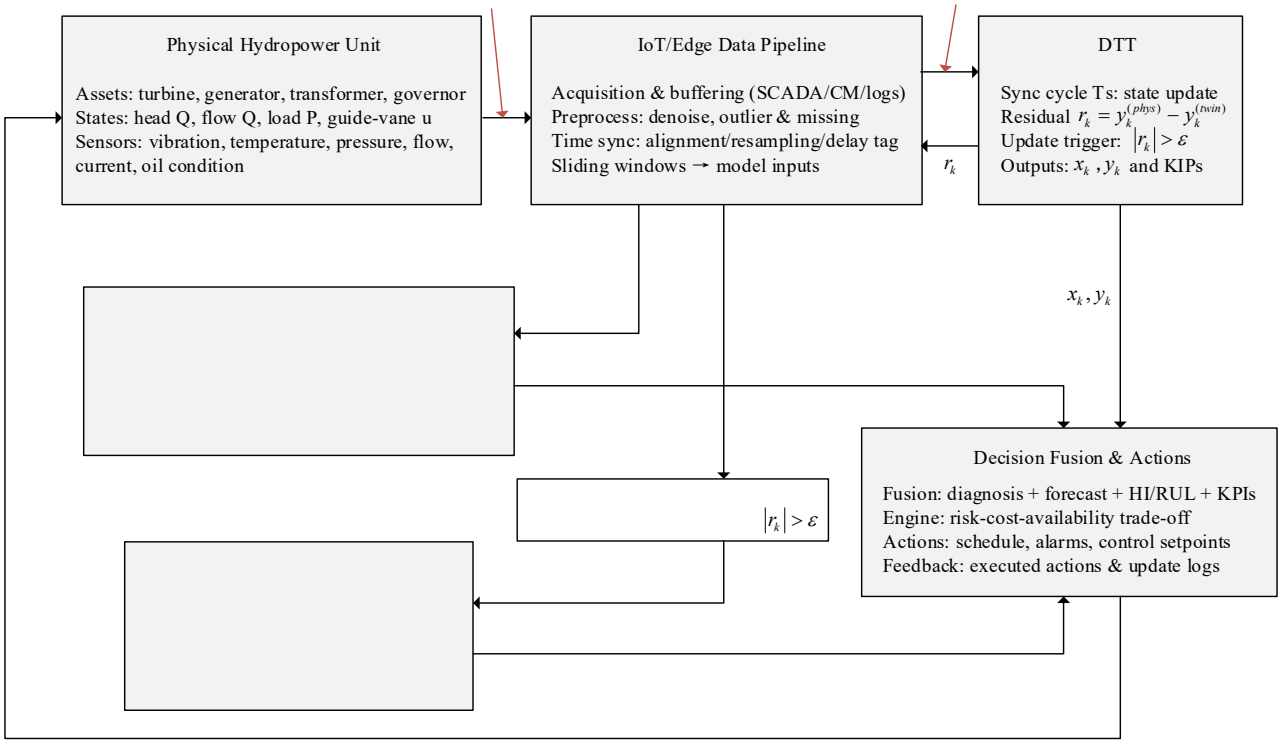


Fig. 1. Schematic diagram of DTT structure

In Fig. 1, the DTDL decision-making framework creates a cyber-physical closed-loop for O&M of hydropower units. The multi-source data stream is collected from an IoT/edge pipeline where buffering, denoising, and time synchronization are conducted prior to window formation. The DTDL system is implemented based on six types of sensors, including pressure/pulsation sensors, vibration sensors, temperature sensors, rotational speed sensors, guide vane opening sensors, and electric sensors. Representative examples of pressure-related sensors are shown in Fig. 2.

Micro MPM 480 PCB 112A

Fig. 2. Pressure sensors

The digital twin synchronizes the physical and virtual states with period Ts and supports event-based updates driven by the consistency residual e, producing twin states and key performance indicators (KPIs). Data quality control is applied before windowing: missing values are interpolated for short gaps (otherwise windows are discarded), outliers are screened using physical limits and 3σ rules, communication delay is handled via timestamp alignment within a tolerance, and packet loss is mitigated by edge buffering (rejecting unrecoverable windows). The synchronized twin outputs are then fed to the deep learning module (DSC-based fault diagnosis and LSTM-based trend prediction), and their outputs are combined with life-cycle information to support downstream decision fusion and maintenance scheduling. DSC can efficiently process large-scale sensor data and accurately detect fault types, while LSTM utilizes historical data to predict future operational states (Akiner et al., 2024). These two neural networks can be used separately to construct fault diagnosis and trend prediction modules, and, after integration, they form the MobileLSTM model. As shown in Eq. (5), DSC provides higher parameter efficiency than standard convolution when processing multi-source sensor channels.

$$N_{\text{param}}^{\text{Conv}} = k^2 C_{\text{in}} C_{\text{out}},\ N_{\text{param}}^{\text{DSC}} = k^2 C_{\text{in}} + C_{\text{in}} C_{\text{out}} \tag{5}$$

In Eq. (5), k denotes the kernel size, C_{in} and C_{out} denote the number of input channels and output channels, and $N_{\text{param}}^{\text{Conv}}$ and $N_{\text{param}}^{\text{DSC}}$ denote the parameter counts of standard convolution and DSC, respectively. Targeted temporal feature extraction is achieved by depth-wise convolution in DSC at the local level in each channel of the sensor, while point-wise convolution captures the channel correlation and achieves multi-source feature fusion. Consequently, DSC is more appropriate for modeling multi-source time series from sensors in hydropower units. Based on this, the fault diagnosis module using DSC is constructed, as shown in Fig. 3.

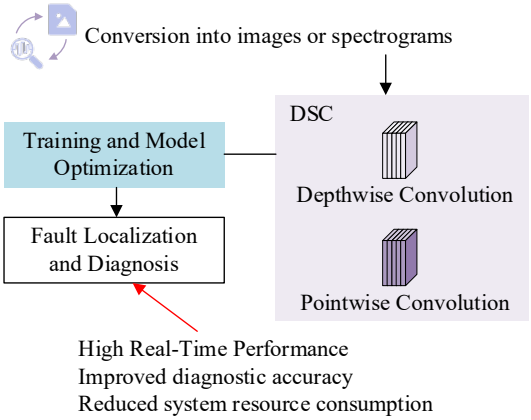


Fig. 3. Schematic diagram of the fault diagnosis module based on DSC

In Fig. 3, the collected signal data is first converted into images or spectrograms to meet the input requirements of the convolutional neural network. Then, the DSC model is used for feature extraction and training optimization, enabling accurate localization and diagnosis of equipment faults. In trend prediction, the LSTM model learns long-term dependencies in time-series data by recursively calculating hidden states. The update process is shown in Eq. (6).

$$h_t = \tanh(W_h \cdot h_{t-1} + W_x \cdot x_t + b_h) \tag{6}$$

In Eq. (6), W_h and W_x are the weight matrices, b_h is the bias term. h_t is the hidden state of LSTM at the current time t, and x_t is the current input data. The LSTM output is shown in Eq. (7).

$$y_t = W_y \cdot h_t + b_y \tag{7}$$

In Eq. (7), h_t is the hidden state at time t, and y_t is the predicted future trend output of the LSTM model. The LSTM gating mechanism controls the flow of information through the forget gate, input gate, and output gate. The forget gate is shown in Eq. (8).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

In Eq. (8), W_f and b_f are the weights and biases, h_{t-1} is the hidden state at time $t-1$, and f_t is the forget gate, which controls the degree of forgetting of the previous hidden state. Three LSTM layers are used to process large-scale hydroelectric station data, with 100 neurons per layer, and a trend prediction module is constructed, as shown in Fig. 4.

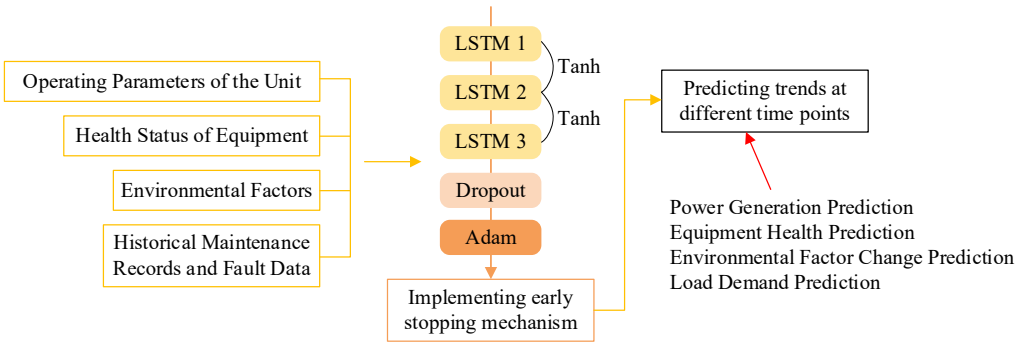


Fig. 4. Schematic diagram of the trend prediction module based on LSTM

As shown in Fig. 4, the trend prediction module includes three LSTM layers. In addition, dropout layers are also adopted to avoid overfitting, and the Adam optimizer is utilized during the training process. To enhance the model's performance during training, early stopping is adopted to predict trends at various points in time. They include generation power prediction, equipment status prediction, environmental trend prediction, and load prediction.

3.2. Intelligent Decision Support System Integrating Full Life Cycle Module

By integrating data from fault diagnosis, trend prediction, and other aspects, the system can make intelligent decisions. To improve the system's decision-making, a full life-cycle module is introduced into the traditional architecture to build a more comprehensive and practical O&M intelligent decision-support system. O&M with full life-cycle management will allow the incorporation of natural elements of state surveillance, health assessment, and recycling upon retirement. By incorporating full life cycle management within the intelligent decision-making process, it becomes possible to optimize decision-making processes, starting from operational to retirement phases, by relying on available data. The application requirements for full life cycle management in the hydroelectric unit O&M are shown in Fig. 5.

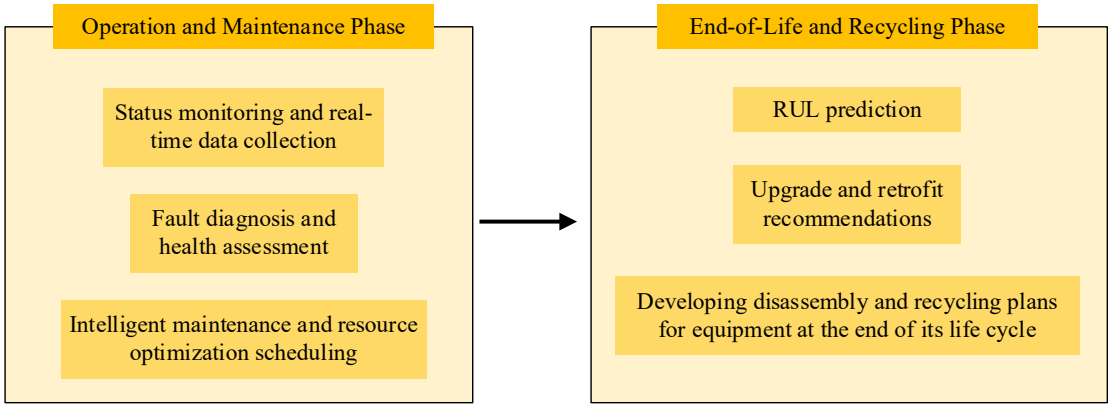


Fig. 5. Application requirements of hydroelectric units throughout their life cycle

As shown in Fig. 5, the application requirements for hydroelectric units in full life cycle management are divided into the O&M phase and the retirement and recycling phase. The O&M phase includes state monitoring, real-time data collection, fault diagnosis, health evaluation, and intelligent operation and resource optimization scheduling. In the retirement and recycling phase, when the equipment's service life is nearing its end, the system predicts its retirement time using remaining-life prediction. Life-cycle data analysis evaluates equipment health status using historical operational data. The data analysis process is shown in Eq. (9).

$$HI_t = \frac{\sum_{i=1}^{t} L_i}{t} \tag{9}$$

In Eq. (9), L_i is the i-th data point in the life cycle data, which includes maintenance records, equipment running time, etc. HI_t is the health index of the equipment at time t. The full life cycle module is set as a part of the intelligent decision-making system to supplement the diagnosis and prediction modules. The module is shown in Fig. 6.

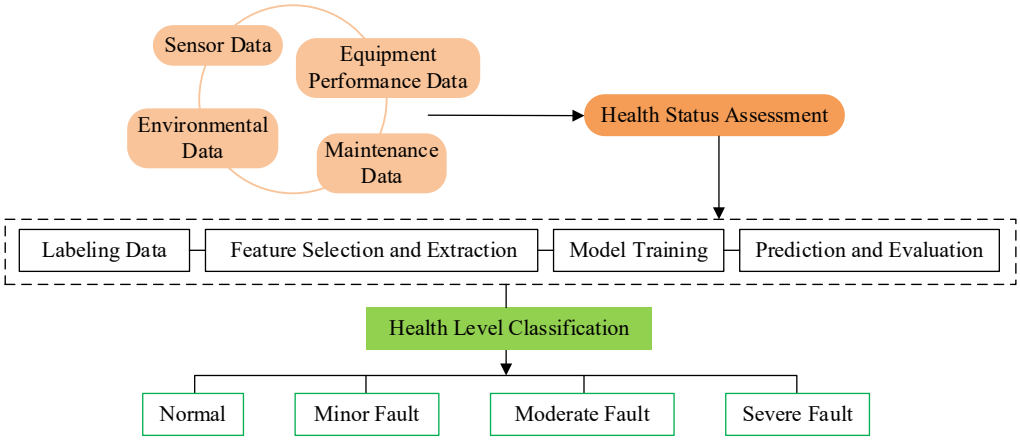


Fig. 6. Application process of the life cycle analysis module

In Fig. 5, the life-cycle analysis module starts from synchronized sensor streams, operating-condition variables, environmental records, alarm logs, and maintenance records. After data cleaning and time alignment, the raw records are reorganized into fixed-length window samples for health-state analysis. In the context of data labeling, a few initial labels are created based on maintenance data, alarm levels, forced outage history, and inspection results. Then, unlabeled windows are clustered based on their similarities of degradation, and only the uncertain boundary points are validated by humans to form a health label set using semi-supervision. As for feature selection and extraction, statistical characteristics related to vibration and temperature, trend features associated with load and efficiency deviations, residuals from twin consistency, alarm levels, and maintenance history variables are utilized. Low-variance and strongly correlated variables are filtered out, and engineered features are merged with latent variables produced by the other two modules. During model training, the labeled windows are divided into training and validation subsets by operating condition, class weights are used to alleviate label imbalance, and early stopping is applied to avoid overfitting. During prediction and evaluation, the trained health-grade classifier outputs the probabilities of four states for each window, and the final class is determined by the maximum posterior probability. The operation status of the hydroelectric unit is divided into four levels: normal, minor fault, moderate fault, and severe fault, achieving a refined determination of the equipment's current health status. The entire intelligent decision-making system bases its decisions on multiple inputs, including fault diagnosis, trend prediction, and life-cycle data. The decision-making process is shown in Eq. (10).

$$D = w_1 \cdot F + w_2 \cdot T + w_3 \cdot L \tag{10}$$

In Eq. (10), w_1, w_2, and w_3 are the weight coefficients, and F, T, and L are the fault diagnosis results, trend prediction results, and life cycle data, respectively. D is the decision, which determines whether the hydroelectric unit needs maintenance or whether to increase the load. In some decisions, a weighted average based on the outputs of different modules may be required. The decision after the weighted average operation is shown in Eq. (11).

$$D_{final} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i} \tag{11}$$

In Eq. (11), w_i is the weight of each module, x_i is the output of each module, and D_{final} is the final output. When predicting unit faults, a fault-rate model estimates the probability of a fault occurring in a unit over a given period. The failure rate calculation method is shown in Eq. (12).

$$\lambda(t) = \frac{1}{\theta} \cdot e^{\frac{t}{\theta}} \tag{12}$$

In Eq. (12), θ represents the equipment's lifetime parameters, and $\lambda(t)$ represents the probability of a fault occurring at time t. The time-to-failure is assumed to follow a Weibull lifetime distribution, where the parameter vector λ (and optionally the shape parameter β) is updated via a regression mapping from the online health indicator HI_t and operating conditions u_t, thereby linking the failure-rate model parameters to the actual degradation state. Accordingly, the RUL at time t is computed from the conditional survival function as $RUL(t) = \mathrm{E}[T - t \mid T > t] = \int_0^{\infty} \frac{S(t+\tau \mid HI_t, u_t)}{S(t \mid HI_t, u_t)} d\tau$, where $S(\cdot \mid HI_t, u_t)$ is determined by the assumed lifetime distribution in Eq. (12) with parameters updated from HI_t and operating conditions u_t. To account for the dynamic changes in the model's state, action, and reward, a Markov decision process is adopted to optimize the hydroelectric unit's maintenance decisions. The study constructs the DTDL

intelligent decision-making system applied to a large-scale hydroelectric station O&M, as shown in Fig. 7.

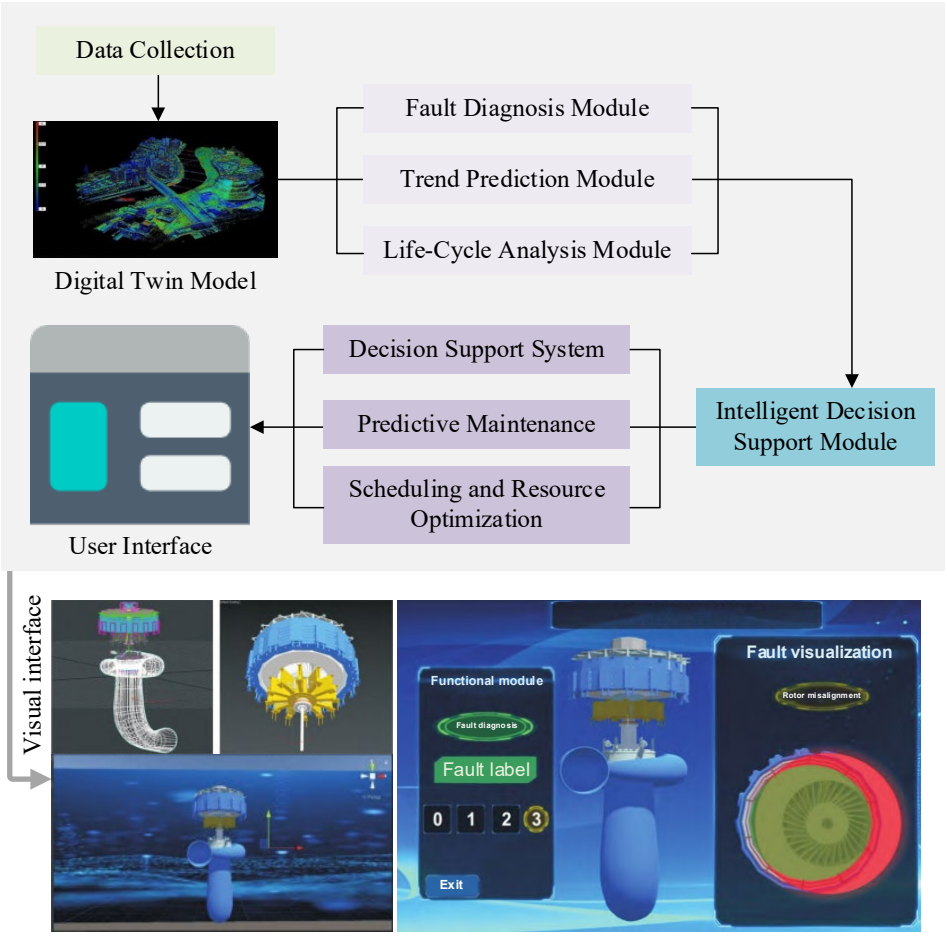


Fig. 7. Workflow of the DTDL intelligent decision-making system

As shown in Fig. 7, the DTDL system integrates fault diagnosis, trend prediction, and full life-cycle data for intelligent decision-making, providing data-driven support for optimal maintenance timing, cost reduction, resource allocation optimization, and improved fault diagnosis accuracy.

4. Performance Evaluation of the DTDL System

4.1. Comparison of Fault Prediction Performance of MobileLSTM Model

To evaluate the DTDL system under multi-sensor and multi-regime conditions, comparative experiments were implemented using TensorFlow 2.0/Keras (Python 3.8) and MATLAB/Simulink for twin modeling. Training and inference were run on an RTX 3090 GPU with an i7-10700K CPU and 32 GB RAM. Two public multi-sensor datasets were used as proxies: IEEE PHM 2012 and the Tennessee Eastman Process (TEP). For IEEE PHM 2012, six bearing sequences that operated until failure were selected for the training set, while eleven were selected for the test set, where the training samples were stratified according to operating condition to reserve validation samples for tuning hyperparameters and implement early stopping. For TEP, each simulation run was considered an individual sample (with 52 features), where 500 samples of each mode were used in training and 500 for testing, while about 20 percent of the training data was kept for validation. To make the data organization used in this study more explicit, Table 1 gives an example of structured window samples that follow the same preprocessing, synchronization, and labeling.

To verify the performance of the fault diagnosis and trend prediction module, the MobileLSTM model was compared with fault diagnosis and prediction models based on Support Vector Machine (SVM) and Random Forest combined with LSTM (LSTM-Random). To ensure the reproducibility and fairness of comparisons, the study provides baseline key parameters for SVM and LSTM-Random, and the core hyperparameter settings and selection methods for MobileLSTM, as detailed in Appendix Table A1. The first comparison focused on the accuracy and fault response time during fault diagnosis. The specific results are shown in Fig. 8.

In Fig. 8(a), the MobileLSTM model demonstrates remarkable stability throughout the entire time, achieving a final accuracy of 96.4%, outperforming both LSTM-Random and SVM. As shown in Fig. 8(b), the average response time of MobileLSTM has been decreased to only 12 minutes, resulting in more than 30% improvement compared to the other comparative models. The combination of the benefits of high accuracy and fast response time is due to the interaction of both the DSC and LSTM techniques in the MobileLSTM system. With such a combination, the proposed system can

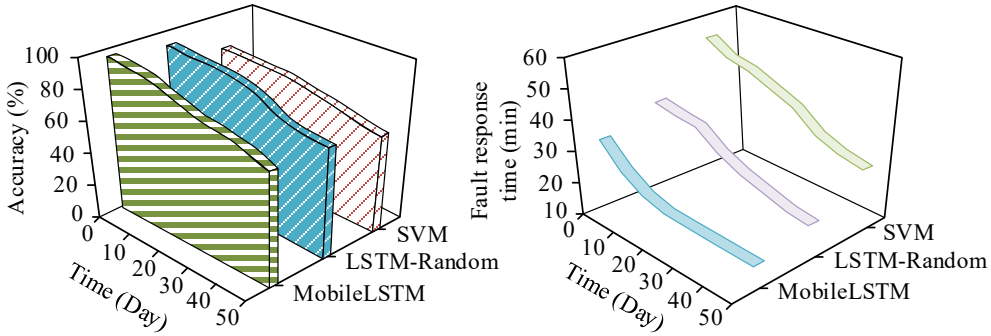
efficiently recognize fault patterns from massive concurrent data without encountering gradient vanishing. To validate the effectiveness of the data quality strategy, the study simulated random packet loss on the test set and evaluated performance as shown in Table 2.

Table 1. Example of structured samples Panel A.

Sample ID	Window length (time steps)	Mean rotor speed (rpm)	Bearing temperature (°C)	Vibration RMS (mm/s)
W-01	256	149.8	63.2	0.21
W-02	256	149.5	71.8	0.36
W-03	256	148.9	79.6	0.58
W-04	256	147.6	89.7	0.92

Panel B.

Sample ID	Load (% rated)	Twin residual	Alarm count in window	Health label	Next-24 h trend target
W-01	81.4	0.018	0	Normal	Stable operation
W-02	83.1	0.042	1	Minor fault	Slight degradation
W-03	78.4	0.087	3	Moderate fault	Degradation continues
W-04	74.8	0.146	5	Severe fault	Maintenance required soon



(a) Accuracy of the models (b) Response time of the models

Fig. 8. Experimental results of accuracy and fault response time

Table 2. Performance under random packet loss

Model	Packet loss rate (%)	Acc (%)	RMSE
CNN	0	93.2	0.49
	5	92.3	0.53
	10	91.0	0.58
	20	88.7	0.67
LSTM	0	94.8	0.37
	5	94.2	0.40
	10	93.1	0.45
	20	91.4	0.53
CNN-LSTM	0	95.4	0.33
	5	94.9	0.35
	10	94.1	0.39
	20	92.8	0.46
MobileLSTM	0	96.4	0.26
	5	96.0	0.28
	10	95.5	0.31
	20	94.4	0.37

In Table 2, higher packet loss rates lead to lower Acc and higher RMSE, but MobileLSTM still shows stable prediction performance at low and medium packet loss rates. This indicates that the proposed data quality control can reduce the impact of input disturbances caused by unstable on-site communication. The study then validated the accuracy of operational trend prediction, with the results for Root Mean Square Error (RMSE) and preventive maintenance rates shown in Fig. 9.

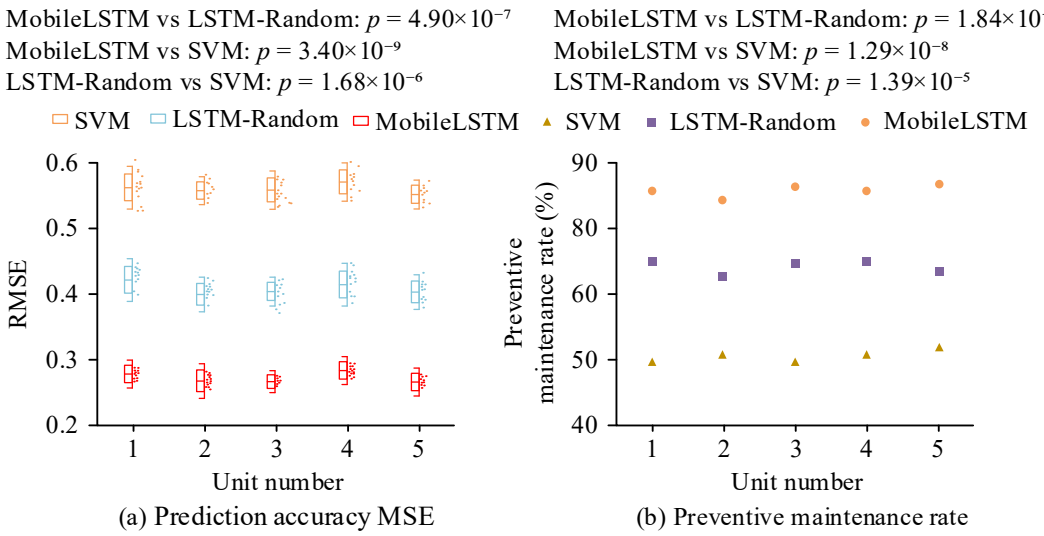


Fig. 9. Operation trend and preventive maintenance rate results

Note: Perform a paired t-test with 5 paired samples (n=5).

In Fig. 9(a), MobileLSTM has a very low average value of RMSE at 0.26, which is much less than that of LSTM-Random and SVM at 0.43 and 0.56, respectively. In addition, as illustrated in Fig. 9(b), MobileLSTM performs well in preventive maintenance of hydroelectric units, always achieving more than 85.4%, and occasionally exceeding 92.3%. This reduction in RMSE is attributed to the processing capability of the LSTM network on the nonlinear and non-stationary nature of hydroelectric unit data. The forget gate mechanism effectively filters environmental noise interference while preserving critical historical state information, enabling precise fitting of unit degradation trends. The study further supplements the comparison and ablation analysis of conventional CNN, LSTM, CNN-LSTM, and MobileLSTM, presenting quantitative results from both complexity and performance perspectives as shown in Table 3.

Table 3. Comparison of conventional structures and ablation outcomes

Model	Params (M)	FLOPs / window (MFLOPs)	Fault diagnosis Acc (%)	Trend prediction RMSE
CNN	0.107	54.854	93.2	0.49
LSTM	0.252	128.614	94.8	0.37
CNN-LSTM	0.360	183.468	95.4	0.33
MobileLSTM	0.266	135.391	96.4	0.26

Table 3 shows that MobileLSTM demonstrates that DSC's lightweight feature extraction, combined with LSTM's temporal memory modeling, achieves a superior trade-off between efficiency and accuracy.

4.2. Simulation Application Analysis of the DTDL System

After validating the MobileLSTM model's performance, the study further simulated and validated the DTDL system. The overall performance of the large-scale hydroelectric units equipped with the DTDL system was evaluated, with unit and system availability shown in Fig. 10.

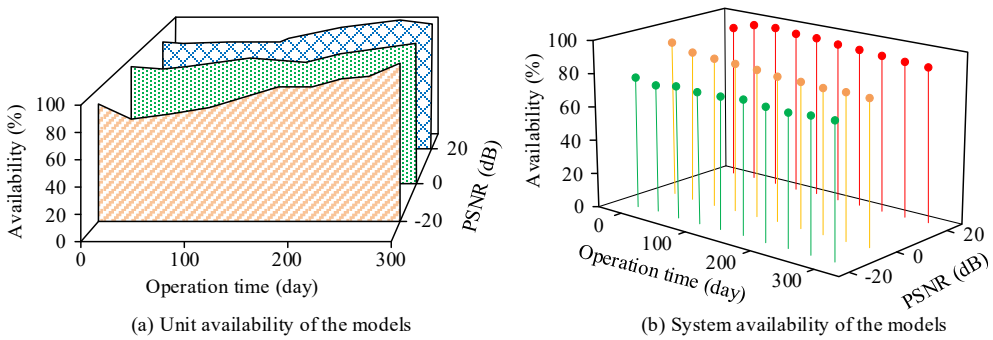
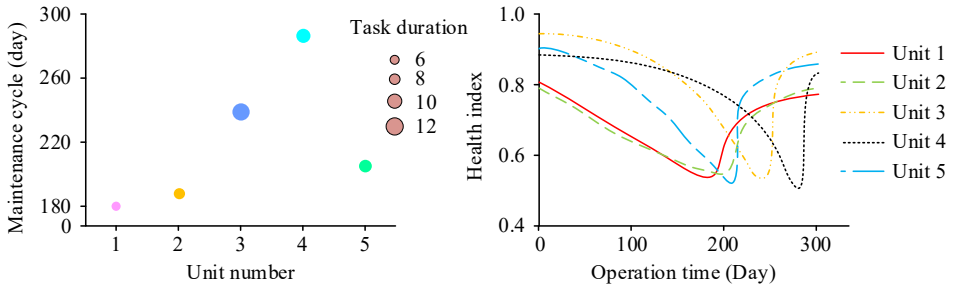


Fig. 10. Unit availability and system availability of the DTDL system

Note: Signal-to-Noise Ratio (SNR) denotes the signal-to-noise ratio of the injected monitoring signal, and Peak Signal-to-Noise Ratio (PSNR) denotes the peak signal-to-noise ratio between the reconstructed sequence and the clean reference. Higher SNR or PSNR values indicate less signal distortion.

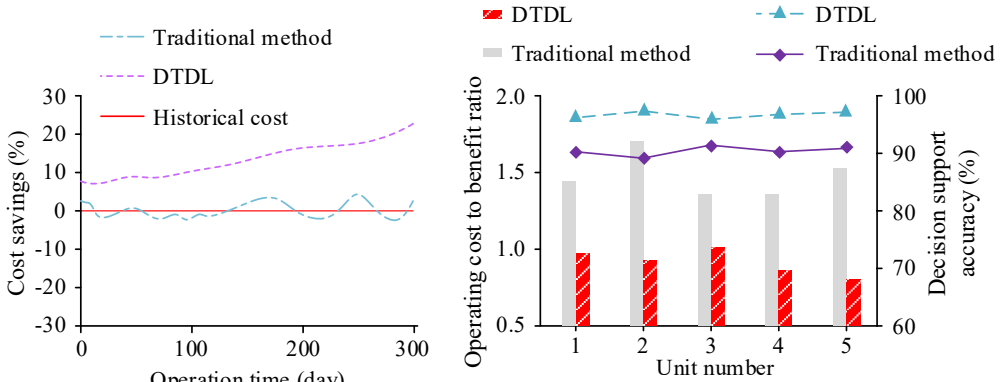
In Fig. 10(a), under different SNR conditions, the DTDL system maintained high availability, with overall availability above 82% even in a -20dB environment and steadily increasing to over 95% in the 0dB and 20dB conditions. As shown in Fig. 10(b), the DTDL system also performed well in terms of overall system availability. As the runtime progressed, in the extreme interference scenario of -20dB, the system-maintained availability above 80%, demonstrating the DTDL system's ability to integrate data in complex multi-source environments and its accurate grasp of the global system operation status. The study then further validated the maintenance cycles and the units' health index, which ranged from 0 to 1. The specific results are shown in Fig. 11.



(a) Maintenance cycle of DTDL system (b) Health index of DTDL system

Fig. 11. Maintenance cycle and health index of DTDL system units

Note: Units 1-5 refer to five generation units belonging to the same hydroelectric power plant, and they can be regarded as independent objects of maintenance during scheduling and health monitoring. The maintenance period of a unit refers to the duration of time between two consecutive maintenance decisions produced by the decision-making support system. Each unit's health condition can be monitored independently through time. Health index refers to a normalized value representing the composite operating status within [0, 1]. It can be calculated based on vibration, temperature rise, efficiency loss, alarm rate, and maintenance history after being normalized.



(a) Cost savings rate (b) The benefits of the system and the maintenance effectiveness

Fig. 12. O&M cost comparison results

Note: To make the economic indicators more explicit, Tables A2 provides examples of the cost-item structures in appendix.

Cost saving = $\frac{(C_{\text{trad}} - C_{\text{DTDL}})}{C_{\text{trad}}} \times 100\%$, where C_{trad} and C_{DTDL} denote the total O&M cost under the traditional strategy and the proposed DTDL strategy, respectively. The operating cost to benefit ratio = $\frac{B_{\text{op}}}{C_{\text{op}}}$, where B_{op} includes energy-output benefit and avoided outage loss, and C_{op} denotes the total operating cost of the corresponding unit. A larger ratio indicates better economic return per unit operating expenditure.

As depicted in Fig. 12(a), the DTDL system was able to save up to 24.8% of costs during its operation. As indicated in Fig. 12(b), across units of various types, the DTDL system has performed exceptionally well when evaluated against operational cost-benefit ratios and decision support accuracy. The average operational benefit ratio for the DTDL system has consistently been above 1.5, whereas decision support accuracy was consistently above 96.1% across all five units. The cost advantage that comes from the DTDL system lies in the comprehensive management module along with MDP optimization. The DTDL system analyzes unit lifespan using its comprehensive lifecycle information and makes intelligent planning regarding the optimal periods for maintenance based on intelligent decision-making algorithms. Through such an analysis, the DTDL system can save costs by eliminating any unnecessary shutdowns for inspections and increasing the

efficiency of machines during their good period via resource management.

5. Discussion

The DTDL was tested based on its ability for fault diagnosis, trend forecasting, and decision-making. In an engineering sense, this means that the model predictions have better stability and are less likely to generate both false positive alarms and missed alarms, thus providing reliable information for the timing of maintenance as well as better distribution of the maintenance efforts. For hydroelectric generator unit O&M, this implies that lower errors lead to lower variance in degradation trends and mean that alarms will be triggered with higher consistency, thus avoiding maintenance activities based on incorrect information and associated outages. At the same time, better responsiveness will facilitate online operation of the algorithm and quick feedback to the O&M side.

The superior performance of the DTDL system cannot be attributed solely to a simple combination of DTT and DL. The key role of DTT in this study is to provide a consistent data-flow organization and a basis for aligning the physical system and the virtual model for multi-source monitoring data, thereby providing a practical engineering entry point for state estimation and model updating. Deep learning, such as MobileLSTM, provides efficient feature extraction for multi-channel time series and long-term dependency modeling, which explains the improved diagnosis and prediction performance under the same data conditions. However, the key part that brings real O&M benefits is the full life-cycle module. This module adds a time dimension and builds a continuous decision chain by weighted decision fusion.

Based on this, the study's limitations and future directions should also be stated. First, even though the architectural design of the model is concerned with lightness, its performance with respect to latency, computational requirements, and robustness when implemented at the edge side and station side still requires more rigorous evaluations and improvements. Second, while the employed datasets in this research are representative enough, there are still some domain disparities in them, hence, generalization of the model in terms of station-wise, unit-wise, and operational variability remains an issue and can be addressed via techniques such as transfer learning or federated learning. Third, for engineering purposes, interpretability and explainability are essential, future research can incorporate additional elements of interpretability and uncertainty visualization in a simple manner.

6. Conclusion

This study develops DTDL, an intelligent decision support system for hydroelectric unit O&M. It integrates real-time monitoring, fault prediction, performance optimization, and dynamic scheduling. The results show that DTDL can improve fault awareness and maintenance decision quality, reduce unplanned interruptions and O&M costs, and increase unit availability through better resource allocation. In this system, the full life-cycle analysis module is treated as an independent decision input, together with fault diagnosis and trend prediction. As a result, long-term health degradation information can be directly included in the final maintenance strategy.

In terms of management practice, this approach affects the management process in several ways. First, maintenance is conducted through condition-based scheduling rather than fixed-interval inspections. Second, managers will not rely on one single signal nor diagnose. It is necessary to evaluate fault classification, predicted trends, and the system's health status simultaneously before making decisions about whether to continue with operation, increase observation, conduct preventive maintenance, or implement emergency actions. Third, maintenance opportunities, spare parts, and manpower can be allocated according to their respective remaining lifetimes and expected cost savings and availability.

The significance lies in demonstrating that the importance of the intelligent operation and maintenance solution does not solely lie in increasing model precision but also in enhancing the process of translating data into action. It is not the fact that the digital twin reflects all information available but rather that it minimizes the distance between reality and the consequences of the situation. This also means that future hydropower O&M should be evaluated not only by prediction metrics, but also by whether the system can support auditable, life-cycle-aware, and economically consistent decisions in real operational workflows. In practice, data reliability remains important. Sensor faults, missing values, and noise can affect health assessment results, so they should be handled through robust data cleaning, anomaly processing, and fail-safe mechanisms. Future work will include better data quality management, more sophisticated fault-detection procedures, and cross-station generalization using multiple units per station and updating twin parameters for each season. Real-world implementation will benefit from integration of the DTDL system with SCADA/CMMS, which may help with simplifying work order processing and safety constraint enforcement.

Author Contributions

Guobiao Zhong, Xi'and Zhou, and Zhongliang Shao contributed to conceptualization and methodology. Guobiao Zhong and Xi'an Zhou contributed to software, analysis, and draft preparation. Zhongliang Shao contributed to manuscript editing.

Funding

The research is supported by Guangdong Provincial Department of Education Research Project: Research and Design of High-Performance Multi-Protocol Edge Controller Based on EtherCAT (2024KTSCX338).

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors used AI tools solely for language editing and readability improvement. The authors reviewed and verified all content and take full responsibility for the accuracy and integrity of the manuscript.

Reference

Akiner, M. E., Kartal, V., Guzeler, A. C., and Karakoyun, E. (2024). Exploring the applicability of the experiment-based ANN and LSTM models for streamflow estimation. *Earth Science Informatics*, 17(4), 3111-3135.

An, R., Lin, P., Li, Z., Zhang, L., Cheng, F., Xia, Y., and Liu, H. (2024). Intelligent ventilation-on-demand control system for the construction of underground tunnel complex. *Journal of Intelligent Construction*, 2(2), 1-16.

Araújo, S. O., Peres, R. S., Filipe, L., Manta-Costa, A., Lidon, F., Ramalho, J. C., and Barata, J. (2023). Intelligent data-driven decision support for agricultural systems-ID3SAS. *IEEE Access*, 11(2), 115798-115815.

Corbacho-Abelaira, D., Casal-Guisande, M., Corbacho-Abelaira, F., Arnaiz-Fernández, M., Trinidad-López, C., Sánchez-Gracián, C. D., and Fernández-Villar, A. (2024). Proposal and definition of an intelligent decision-support system based on deep learning techniques for the management of possible COVID-19 cases in patients attending emergency departments. *IEEE Access*, 12(4), 95035-95046.

Darani, S. H., Rabbanifar, P., Aliabadi, M. H., and Radmanesh, H. (2023). Frequency response model of power system in presence of thermal, wind and hydroelectric units. *Journal of Electrical Engineering and Technology*, 18(4), 2599-2607.

Don, M. G., Liyanarachchi, S., and Wanasinghe, T. R. (2025). A digital twin development framework for an electrical submersible pump (ESP). *Archives of Advanced Engineering Science*, 3(1), 35-43.

Gong, A., He, W., Ma, N., and Cao, Y. (2024). Research on the application of multi-layer tree-structured belief rule base system in aviation engine performance evaluation. *IEEE Access*, 12(3), 139469-139485.

Jyeniskhan, N., Keutayeva, A., Kazbek, G., Ali, M. H., and Shehab, E. (2023). Integrating machine learning model and digital twin system for additive manufacturing. *IEEE Access*, 11(2), 71113-71126.

Lin, S. Q., Zhao, J., Yang, W., Wu, X. C., and Zhuang, X. H. (2025). The assessment of terminals for containers and service quality evaluation using a circular intuitionistic fuzzy intelligent algorithm. *IEEE Access*, 13(2), 40824-40843.

Mathai, M., Liu, Y., and Ling, N. (2024). A hybrid transformer-LSTM model with 3D separable convolution for video prediction. *IEEE Access,* 12(7), 39589-39602.

Njoku, J. N., Nkoro, E. C., Medina, R. M., Nwakanma, C. I., Lee, J. M., and Kim, D. S. (2025). Leveraging digital twin technology for battery management: A case study review. *IEEE Access*, 13(1), 21382-21412.

Oliveira, F. R. D. S., and Neto, F. B. D. L. (2023). Method to produce more reasonable candidate solutions with explanations in intelligent decision support systems. *IEEE Access*, 11(3), 20861–20876.

Özkan, E., Kök, İ., and Suat, Ã. (2024). DeepTwin: A deep reinforcement learning supported digital twin model for micro-grids. *IEEE Access*, 12(3), 196432-196441.

Putri, T. H., Caraka, R. E., Toharudin, T., Kim, Y., Chen, R. C., Gio, P. U., and Pardamean, B. (2024). Fine-tuning of predictive models CNN-LSTM and CONV-LSTM for nowcasting PM2.5 level. *IEEE Access*, 12(2), 28988-29003.

Starly, B., Koprov, P., Bharadwaj, A., Batchelder, T., and Breitenbach, B. (2023). "Unreal" factories: Next generation of digital twins of machines and factories in the industrial metaverse. *Manufacturing Letters*, 37(2), 50-52.

Tang, Q., Wu, B., Chen, W., and Yue, J. (2023). A digital twin-assisted collaborative capability optimization model for smart manufacturing system based on Elman-IVIF-TOPSIS. *IEEE Access,* 11(2), 40540–40564.

Tong, K., Zhang, G., Huang, H., Qin, A., and Mao, H. (2023). A novel combined model for vibration trend prediction of a hydropower generator unit. *Insight - Non-Destructive Testing and Condition Monitoring*, 65(1), 43-51.

Valiollahi, S., Rodriguez, I., Zhang, W., Sharma, H., and Mogensen, P. (2024). Experimental evaluation and modeling of the accuracy of real-time locating systems for industrial use. *IEEE Access*, 12(3), 75366-75383.

Zhong, Z., Fan, N., and Wu, L. (2024). Multistage stochastic optimization for mid-term integrated generation and maintenance scheduling of cascaded hydroelectric system with renewable energy uncertainty. *European Journal of Operational Research*, 318(1), 179-199.

Zhu, H., Tang, X., and Hu, X. (2024). A method for constructing an intelligent maintenance assistance system for hydroelectric stations. *Procedia Computer Science*, 243(26), 948–954.

Guobiao Zhong was born in August 1973 and earned his bachelor's degree in Electrical Automation Technology from Guangdong University of Technology. He is a senior engineer and currently serves as General Manager and Chief Engineer of Guangdong Hong Da Xin Electronic Technology Co., Ltd. He led the 5G+ Smart Energy Management Big Data Platform Project (2022) and the Smart Construction Site Project of a first-class DC pumped-storage power station in Guangdong (2023). He has published articles in several internationally reputed peer-reviewed journals and conference proceedings. His research interests include digital twins for hydropower stations, pattern recognition, and information security.

Xi'an Zhou was born in March 1973 and earned his master's in Engineering from Huazhong University of Science and Technology. He is a senior engineer and currently works at Guangdong Energy Hydropower Branch. He led the generator-unit retrofit of a Class-A power plant (2020) and the construction project of a first-class DC pumped-storage power station (2022). He has published articles in several internationally reputed peer-reviewed journals and conference proceedings. His research interests include hydropower engineering and pumped-storage research.

Dr. Zhongliang Shao was born in November 1979. He currently works as an Associate Professor at Guangdong Polytechnic of Water Resources and Electric Engineering. He led the research and design of a high-performance multi-protocol edge controller based on EtherCAT (2024) and the study of a general-purpose industrial controller based on the EtherCAT bus (2018). Dr. Shao has published articles in more than 20 internationally reputable peer-reviewed journals and conference proceedings. His research interests include machine learning, image processing, pattern recognition, and information security.

Appendix

Table A1. Key configurations of models

Model	Architecture/inference-related settings	Training hyperparameters/ Optimization settings	Hyperparameter selection protocol
SVM	Kernel: RBF; Input: the same windowed samples as MobileLSTM (flattened or after the same unified feature extraction); $C \in \{1,10,100\}$, $\gamma \in \{\text{scale}, 0.1, 0.01\}$	No gradient-based training; Feature scaling: mean/variance computed from the training set only	Grid search on the validation set (best accuracy/macro-F1)
LSTM-Random	LSTM encoder: 2–3 layers, hidden size $\in \{64,100\}$; Downstream Random Forest: $n_{trees} \in \{64,100\}$, $\max_{depth} \in \{10, 20, None\}$	Adam, $l_r = 1\times10^{-3}$; batch size=64; max epochs=200; early stopping patience=10; dropout=0.2	Stage-wise tuning: first select the LSTM depth/hidden size on the validation set, then fix the encoder and train RF, choosing the best RF configuration
MobileLSTM	Depthwise separable convolution (DSC) feature extractor+LSTM temporal modeling; LSTM: 3 layers, 100 units per layer, with dropout; DSC: depthwise $k=3\times3$, pointwise channels $\in \{32,64\}$; Fusion: concatenation/attention (per implementation)	Adam, $l_r = 1\times10^{-3}$ (optionally with decay); batch size=64; max epochs=200; early stopping patience=10; weight decay=1×10^{-5}; dropout=0.2–0.5	Validation-driven staged search: (1) fix window length/stride; (2) tune network width and dropout; (3) fine-tune learning rate and regularization based on the validation set

Table A2. Economic analysis of hydroelectric unit O&M **Panel A.** Cost composition by operational stage

Operational phase	Method	Planned maintenance cost	Corrective repair cost	Downtime loss	Labor and material cost	Total cost	Cost saving (%)
Early phase	Traditional	118	96	142	54	410	-
	DTDL	122	73	129	51	375	8.5
Middle phase	Traditional	126	121	167	59	473	-
	DTDL	129	86	136	53	404	14.6
Late phase	Traditional	132	158	201	74	565	-
	DTDL	138	97	150	40	425	24.8

Panel B. Cost-benefit indicators by unit

Unit	Operating cost	Energy-output benefit	Avoided outage loss	Total operating benefit	Operating cost-benefit ratio	Decision support accuracy (%)
Unit 1	188	245	56	301	1.60	96.3
Unit 2	176	231	56	287	1.63	96.7
Unit 3	194	246	62	308	1.59	96.1
Unit 4	201	267	64	331	1.62	96.5
Unit 5	183	238	57	295	1.61	96.8