Journal of Computer and Communications, 2026, 14(7), 110-129
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Towards a Hybrid Digital Twin for Intelligent Monitoring and Predictive Maintenance of Hydraulic Turboalternator Group

Desire Tere Djacba[1], Colince Welba[2], Ruben Torres Martinez[1]

[1]Department of Projects, International Iberoamerican University, Mexico City, Mexico
[2]National Advanced School of Mines and Petroleum Industries, University of Maroua, Maroua, Cameroon
Email: desire.djacba@doctorado.unini.edu.mx, welbacolince@yahoo.fr, ruben.torres@unini.edu.mx

How to cite this paper: Djacba, D.T., Welba, C. and Martinez, R.T. (2026) Towards a Hybrid Digital Twin for Intelligent Monitoring and Predictive Maintenance of Hydraulic Turboalternator Group. *Journal of Computer and Communications*, **14**, 110-129.
https://doi.org/10.4236/jcc.2026.147006

Received: March 23, 2026
Accepted: July 21, 2026
Published: July 24, 2026

Abstract

This study proposes a hybrid digital twin framework for intelligent monitoring and predictive maintenance of Hydraulic Turboalternator Groups (HTG), integrating high-fidelity physical modeling, a Hybrid Stochastic Automaton (HSA), and Long Short-Term Memory (LSTM) neural networks. The framework exploits the strengths of physics-based and data-driven approaches to enhance system reliability under varying operational conditions. The physical model accurately simulates the hydraulic and electromechanical dynamics, while the HSA represents discrete operational states and probabilistic transitions between normal, degraded, and fault conditions. The LSTM network, originally designed for dynamic control, is repurposed to analyze temporal sensor data, enabling early anomaly detection and identification of incipient degradation patterns. By fusing these components, the digital twin provides predictive maintenance insights, allowing interventions to be scheduled proactively, rather than reactively. Validation using real operational datasets demonstrates accurate forecasting of faults, classification of anomaly criticality, and optimization of maintenance intervals. Implementation of this predictive strategy reduces unplanned downtime by 27% and increases overall system availability by 12%, highlighting the operational benefits of integrating hybrid modeling with deep learning. This hybrid digital twin represents a next-generation methodology for asset health management in complex hydraulic energy systems. It bridges the gap between first-principles physics, stochastic modeling, and AI-based prognostics, offering a scalable, robust, and data-intelligent solution for proactive maintenance and operational resilience.

Keywords

Hydraulic Turboalternator Group, Hybrid Digital Twin, Predictive

Maintenance, LSTM, Hybrid Stochastic Automaton

1. Introduction

Hydropower remains one of the most reliable and flexible sources of renewable electricity, playing a crucial role in ensuring the stability and resilience of modern power systems. Hydraulic turbine-generator units, commonly referred to as Hydraulic Turboalternator Groups (HTG), constitute the core components of hydroelectric power plants. Within these systems, the potential energy of water is converted into mechanical energy by the turbine and subsequently into electrical energy by the synchronous generator. Because of their fast dynamic response and operational flexibility, HTG contribute significantly to frequency regulation, voltage control, and grid balancing in large interconnected power systems [1]-[3].

In recent years, the global energy landscape has undergone profound transformations due to the large-scale integration of intermittent renewable energy sources such as wind and solar power. While these technologies support decarbonization goals, their intrinsic variability introduces significant fluctuations in electrical demand and generation. Consequently, hydroelectric facilities are increasingly required to operate under highly dynamic conditions, characterized by frequent start-stop cycles, rapid load variations, and transient hydraulic regimes. These operational stresses can accelerate the degradation of critical mechanical and electrical components, including turbines, bearings, shafts, generators, and excitation systems [4] [5]. Such conditions significantly increase maintenance complexity and raise concerns regarding equipment reliability and long-term availability.

Traditionally, maintenance strategies in hydroelectric power plants have relied on preventive or corrective approaches. Preventive maintenance is typically performed at predetermined intervals based on manufacturer recommendations or historical statistics, whereas corrective maintenance occurs after a failure has already taken place. Although these strategies remain widely used, they present several important limitations in modern industrial environments. In particular, they fail to account for real operating conditions, neglect the large volume of sensor data generated by modern monitoring systems, and provide limited capability for anticipating progressive degradation or early-stage failures [6]-[8]. As a result, these approaches often lead to excessive maintenance costs, unplanned downtime, and suboptimal asset management.

To overcome these limitations, predictive maintenance has emerged as a key paradigm for intelligent industrial systems. Predictive maintenance relies on continuous monitoring, data analytics, and dynamic modeling to estimate the health state of components and anticipate failures before they occur. In this context, digital twin technology has gained considerable attention as a powerful framework for integrating physical modeling, real-time data, and advanced analytics. *A digi-*

tal twin can be defined as a virtual representation of a physical system that evolves synchronously with its real-world counterpart through continuous data exchange and model updating [3] [9]. By combining physical simulations with sensor data and artificial intelligence, digital twins enable advanced monitoring, fault diagnosis, performance optimization, and lifecycle management of complex engineering systems [7] [8] [10]-[14].

Within the energy sector, digital twins have recently been applied to various applications, including smart grids [15] [16], wind turbines [17] [18], and hydroelectric power plants [13] [14] [19]. For hydropower systems in particular, digital twin technology allows the reproduction of coupled hydraulic, mechanical, and electrical dynamics, making it possible to analyze transient behaviors and evaluate extreme operating scenarios without exposing physical equipment to risk [3] [19]. Moreover, digital twins facilitate the integration of heterogeneous data sources and provide decision-support tools for operators responsible for maintaining system reliability and efficiency.

Despite these advantages, purely physics-based digital twins often struggle to represent uncertain phenomena and stochastic events occurring in complex industrial systems. Hydropower installations, for instance, are subject to multiple sources of uncertainty, including hydraulic turbulence, cavitation effects, random mechanical vibrations, and grid disturbances. These phenomena cannot always be captured accurately using deterministic models alone. Consequently, recent research efforts have focused on the development of hybrid digital twins (HDT) that combine physical modeling with probabilistic approaches and machine learning techniques [9] [20] [21].

Hybrid digital twins typically integrate three complementary modeling layers. The first layer consists of deterministic physical models, derived from differential equations representing the fundamental laws of fluid mechanics, electromechanics, and energy conversion [5] [11]. The second layer incorporates stochastic or statistical models, capable of representing discrete events, random failures, and operational mode transitions. Hybrid Stochastic Automata (HSA) and piecewise deterministic stochastic processes are particularly suitable for modeling such behaviors in industrial systems [22] [23]. The third layer introduces artificial intelligence techniques, especially deep learning models designed to exploit large volumes of historical operational data.

Among machine learning methods, Long Short-Term Memory (LSTM) neural networks have demonstrated remarkable performance for modeling nonlinear temporal dynamics and predicting complex time-series behaviors. Originally introduced by Hochreiter and Schmidhuber (1997) [24], LSTM networks are particularly well suited for analyzing sequential data, detecting hidden patterns, and forecasting future system states. These capabilities make them highly relevant for predictive maintenance applications, where early identification of degradation signatures is essential. Recent studies have successfully applied LSTM-based models to industrial prognostics, energy systems monitoring, and fault detection tasks [8]

[12] [25].

Building upon these advances, hybrid digital twin architectures integrating physics-based models, stochastic modeling, and deep learning have recently emerged as promising solutions for intelligent monitoring of energy infrastructures. Such architectures enable the continuous synchronization between the physical asset and its virtual counterpart, allowing the system to learn from operational data while maintaining physical interpretability. This synergy provides improved predictive capabilities, enhanced anomaly detection, and more efficient maintenance planning [9] [26] [27].

The present work builds upon the study previously conducted by Djacba *et al.* (2026), entitled "*Towards an Advanced Digital Twin for the Dynamic Simulation of a Hydraulic Turboalternator Group Under Variable Load Constraints*" [28]. In that preliminary research, the conceptual and methodological foundations of a HTD architecture were established by combining physical modeling, HSA, and LSTM neural networks. The proposed framework aimed to simulate the transient behavior of HTG and support adaptive control strategies under variable load conditions.

The objective of the present article is to extend this work by focusing on the practical implementation and experimental validation of the HDT-LSTM architecture applied to a real hydroelectric system. Specifically, the study aims to demonstrate the capability of the proposed digital twin to:

- detect operational anomalies at an early stage,
- anticipate degraded operating states,
- predict critical system dynamics under fluctuating load conditions,
- and improve the overall reliability and maintenance planning of hydroelectric facilities.

By integrating physics-based modeling, stochastic state transitions, and data-driven learning mechanisms, the proposed HDT provides a comprehensive framework for intelligent monitoring and predictive maintenance of HTG. Ultimately, this approach contributes to enhancing the operational efficiency, resilience, and sustainability of modern hydroelectric infrastructures.

Unlike our previous work, which focused primarily on the conceptual design and simulation of a hybrid digital twin, the present study introduces several key contributions:

- Validation using real SCADA data from an operational hydropower plant;
- Explicit coupling between the Hybrid Stochastic Automaton (HSA) and the LSTM model;
- Integration of a predictive maintenance decision layer based on a Health Index (HI);
- Quantitative evaluation of early fault detection performance and prediction horizon.

These additions significantly enhance the practical applicability and experimental validation of the proposed framework.

2. Models

2.1. General Architecture of the Hybrid Digital Twin (HDT)

The Hybrid Digital Twin (HDT) architecture developed in this study aims to reproduce both the physical dynamics and the degradation processes affecting HTG. Unlike conventional digital twins that rely exclusively on deterministic simulations or purely data-driven models, the proposed framework integrates three complementary modeling layers that operate in a coordinated manner.

The first layer corresponds to the *physical modeling layer*, which describes the hydraulic, mechanical, and electromagnetic processes governing the operation of the HTG. These models are derived from fundamental physical laws and provide a deterministic representation of system dynamics under varying operating conditions.

The second layer is the *behavioral modeling layer*, implemented through an HSA. This layer represents discrete operational states and probabilistic transitions between them. It captures events such as component degradation, abnormal vibrations, and emergency shutdown conditions that cannot be fully described by deterministic equations.

The third layer corresponds to the *cognitive layer*, which incorporates artificial intelligence techniques, particularly LSTM neural networks. This layer processes historical and real-time sensor data to detect hidden temporal patterns associated with early degradation processes.

By combining these three layers, the HDT is capable of simultaneously modeling continuous physical dynamics, discrete operational states, and data-driven predictive insights. This integrated architecture enables a comprehensive representation of the HTG lifecycle and provides a robust platform for intelligent monitoring and predictive maintenance.

The architecture is based on three main layers (Table 1):

Table 1. Constituent layers of the HDT.

Layer	Description	Objective
Cognitive	Long Short-Term Memory (LSTM) model	Anticipate performance drifts and predict failures
Behavioral	Hybrid Stochastic Automaton (HSA)	Represent discrete events and state transitions
Physical	Multi-physics model (hydraulic, mechanical, electromagnetic)	Reproduce the actual dynamics of the HTG

This integrated architecture enables the dynamic fusion of real and simulated data to compute a Health Index (HI) that is continuously updated in real time.

2.2. Multi-Physics Modeling of the HTG

Hydraulic Subsystem

The turbine operation, specifically that of a Kaplan turbine, is modeled using a physics-based approach rather than a purely data-driven representation. The model relies on the fundamental principles of fluid mechanics, namely the continuity equation and the *Bernoulli equation*, extended to account for hydraulic head losses and cavitation effects:

$$J_t \frac{\mathrm{d}\omega}{\mathrm{d}t} = T_h\left(Q, H, \omega\right) - T_m, \tag{1}$$

where:

- J_t is the turbine moment of inertia,
- T_h is the hydraulic torque, which depends on the flow rate Q, the net head H, and the rotational speed ω.

The hydraulic parameters are calibrated based on manufacturer curves and efficiency tests.

The transmission shaft is modeled as a single-shaft torsional system with viscous damping:

$$J_m \frac{\mathrm{d}\omega_m}{\mathrm{d}t} = T_h - T_e - D_m\left(\omega_m - \omega_t\right), \tag{2}$$

where

- J_m is the moment of inertia,
- T_h is the hydraulic torque,
- T_e is the electromagnetic torque,
- D_m represents the damping coefficient.

This equation models the balance between mechanical input and electrical output, including damping effects that ensure system stability.

Vibration and mechanical load measurements are utilized to detect stiffness drifts and dynamic imbalances.

Electromagnetic Subsystem

The synchronous generator is described using a reduced-order model based on the *Park equations*:

$$\begin{cases} V_d = R_s i_d - \omega\psi_q + \dfrac{\mathrm{d}\psi_d}{\mathrm{d}t} \\ V_q = R_s i_q - \omega\psi_d + \dfrac{\mathrm{d}\psi_q}{\mathrm{d}t} \end{cases}, \tag{3}$$

where:

- V_d, V_q are the stator voltages,
- i_d, i_q are the stator currents,
- ψ_d, ψ_q are the flux linkages,
- R_s is the stator resistance,
- ω is the electrical angular speed.

These equations allow reproducing rapid voltage and current variations during load fluctuations.

Validation of the Physical Model

The equations are implemented in Python (Anaconda) for time integration. Simulated results (speed, torque, voltage, current) are compared with historical measurements from a reference hydroelectric plant to validate the model's fidelity.

2.3. Hybrid Stochastic Automaton (HSA)

Knowledge Base Creation

The HSA is developed and calibrated using historical data collected from the SCADA system of the Mbakaou Small Hydropower Plant (SHP), operated by IED Invest Cameroon, located in the Djérem department of the Adamaoua region. The plant was initially commissioned with an installed capacity of 1480 kW (Phase 1), with a planned expansion to 2800 kW in Phase 2. IED (Innovation Énergie Développement), a French engineering company founded in 1988 in Francheville, specializes in renewable energy and electrical engineering projects across Africa and Asia.

Representative load data are summarized in Table 2.

Table 2. Measurement and control categories.

Category	Variables	Notes / Usage
Electrical line measurements	Line_Active_Power ✅ Line_Reactive_Power Phase1_Current ✅ Phase2_Current Phase3_Current ✅ Line_Frequency ✅	Frequency can be used as feedback input
Turbine power measurements	GR1_Active_Power ✅ GR2_Active_Power ✅	Direct measurement of turbine output
Regulation setpoints (PID)	GR1_Distributor_Setpoint ✅ GR1_Pitch_Setpoint ✅ GR2_Distributor_Setpoint ✅ GR2_Pitch_Setpoint ✅ GR1_TSSSP, GR2_TSSSP GR1_TSSPD	Control variables to adjust turbines
Internal mechanical measurements	GR1_Speed ✅ GR2_Speed ✅ GR1_TE, GR2_TE	Internal measurements; some may be filtered if too numerous
Environmental measurements	ODP_Upstream_Level GR1_BMC_Water_Level GR1_VT200, GR2_VT200	Levels and environmental sensors
Hydraulic controls	ODP_M900/1_Valve_1 ODP_M900/2_Valve_2	Hydraulic valve commands

HSA States and Transitions (PyCATSHOO)

The Hybrid Stochastic Automaton (HSA) is implemented using the simulation platform PyCATSHOO (PythoniC Object Oriented Hybrid Stochastic AuTomata), a specialized tool designed for modeling complex industrial systems that exhibit both continuous dynamics and discrete stochastic events. PyCATSHOO provides an efficient framework for representing hybrid systems in which deterministic physical processes interact with probabilistic state transitions, making it particularly suitable for reliability analysis and dynamic risk assessment in energy infrastructures.

Within the proposed hybrid digital twin architecture, the HSA is used to model the different operating conditions of the HTG. The system behavior is represented through a set of discrete states corresponding to normal operation, degraded operating modes, and critical failure conditions. Transitions between these states are governed by probabilistic rates that depend on monitored indicators such as vibration levels, temperature variations, and electrical parameters.

By integrating both continuous system dynamics and discrete operational events, the HSA provides a realistic representation of the degradation process affecting the turboalternator group. This hybrid modeling approach allows the digital twin to capture not only the physical evolution of the system but also the stochastic nature of faults and operational disturbances that may occur during plant operation.

The different operational states defined in the HSA model are summarized in Table 3.

Table 3. Description of operational states.

State	Description
S_0	Normal operation
S_1	Degraded mode (high vibrations, misalignment)
S_2	Imminent failure
S_3	Emergency shutdown

Transitions $S_i \to S_j$ are governed by probabilistic rates λ_{ij} that depend on measured indicators (RMS vibration, bearing temperature, stator current). The LSTM is trained to recognize temporal signatures that precede a transition to an abnormal state, enabling preventive system responses.

2.4. Cognitive Layer: Deep Learning and Predictive Maintenance

Data and Preprocessing

The HSA is developed and calibrated using historical SCADA data collected from an operational hydroelectric turbo-generator system. The primary dataset consists of measurements recorded at a sampling rate of *1 minute* over a continuous one-year period (2023), providing a representative multivariate time series of the plant's hydraulic, mechanical, and electrical behavior. No resampling was applied, ensuring that the original temporal resolution of the monitoring system is preserved.

To enable temporal learning and sequential modeling, the data were structured into overlapping sliding windows of 30 time steps (*i.e.*, 30 minutes). After preprocessing, this resulted in approximately 25,000 sequences used for model development. The dataset was split chronologically into training (80%) and testing (20%) subsets to preserve temporal causality and avoid data leakage.

The dataset includes 56 *variables* covering turbine and generator vibrations, bearing, winding, and stator temperatures, active and reactive power, rotational speeds, actuator positions (guide vanes and runner blades), phase currents, upstream water level, and other plant-relevant measurements. This multi-domain information is used to define probabilistic state transitions within the HSA, such as the likelihood of moving from normal operation to degraded or failure states.

It should be noted that earlier references to high-frequency sampling (200 Hz) and short time windows (5 seconds) corresponded to a preliminary simulation dataset and have been removed to avoid confusion, ensuring full consistency with the real SCADA-based analysis presented in this study.

Table 4 below summarizes all variables used in the model, distinguishing inputs (features) from the various HTG subsystems and the output (target) to predict.

Table 4. Model inputs (Features) and output (Target).

Category	Variables	Role in Model
Turbine Power	GR1_Active_Power, GR2_Active_Power	Input features
Turbine Speed	GR1_Speed, GR2_Speed	Input features
PID Setpoints	GR1_Distributor_Setpoint, GR1_Pitch_Setpoint, GR2_Distributor_Setpoint, GR2_Pitch_Setpoint	Input features
Grid	Line_Active_Power, Phase1_Current, Phase2_Current, Phase3_Current, Line_Frequency	Input features
Output (Target)	Failure	Target variable to predict

LSTM Model Architecture

The cognitive layer constitutes the data-driven component of the Hybrid Digital Twin (HDT), designed to extract meaningful temporal patterns from SCADA data and provide predictive insights for maintenance and adaptive control. By leveraging Long Short-Term Memory (LSTM) networks—well suited for sequential data due to their memory cells and gating mechanisms—the model captures long-term dependencies associated with slow degradation processes while filtering out noise.

The dataset used in this layer is identical to that employed for the Hybrid System Automaton (HSA) calibration, ensuring consistency within the HDT framework. After appropriate preprocessing, including data cleaning, normalization, and temporal structuring, the model is trained to estimate the probability of impending failures based on multiple sensor signals such as turbine speed, electrical currents, active power, and control setpoints.

Within the HDT architecture, the predicted failure probability is coupled with the HSA, enabling anticipation of transitions between discrete operating states. This integration allows proactive system responses, such as maintenance alerts or adaptive adjustment of control parameters, thereby enhancing system reliability and supporting intelligent predictive maintenance strategies.

The overall workflow of the cognitive layer, from data preprocessing to integration within the HDT framework, is summarized in **Table 5**.

Table 5. Main steps of the proposed program.

Step	Description
1. Data Import and Preparation	Loads the SCADA dataset used for both HSA calibration and the cognitive layer of the HDT. Raw data include heterogeneous formats (mixed numeric and string values), which are converted to float, standardized, and cleaned using NaN handling and interpolation.
2. Fault Label Creation	A fault (Panne = 1) is defined when frequency < 49.8 Hz and active power < 50% of its mean value; this represents a proxy SCADA-based indicator of abnormal or degraded operating states rather than strictly confirmed failures.
3. Class Rebalancing	Due to severe imbalance (≈0.2% faults), oversampling is applied only to the training set after a strict chronological split (80/20), achieving a 1:3 ratio (fault:normal) and preventing any time-series leakage.
4. Data Scaling (Normalization)	All variables are normalized to the range [0, 1] using MinMax scaling to ensure stable and efficient model training.
5. Temporal Sequence Construction	The time series is segmented into overlapping sliding windows (length = 30). Each sequence includes multiple sensor signals and is used to predict the probability of failure.
6. Data Splitting and Weighting	The dataset is split into training (80%) and testing (20%) sets. Class weights are computed to further mitigate imbalance.
7. LSTM Modeling	The model is defined as a sequential neural network with the following architecture: LSTM (128, return_sequences=True), Dropout (0.1), LSTM (64), Dense (32, ReLU), and Dense (1, sigmoid). LSTM layers capture temporal dependencies, Dropout reduces overfitting, and the sigmoid activation outputs a failure probability between 0 and 1.
8. Model Training	Training uses the Adam optimizer (learning rate = 10^{-3}) and binary cross-entropy loss function. This configuration ensures efficient convergence and accurate binary classification of system states.
9. Evaluation	The model predicts failure probabilities on test data. Performance is evaluated using precision-recall analysis, and an optimal detection threshold is selected. Temporal filtering is applied to reduce false alarms.
10. Visualization	Visualization includes real vs predicted faults, probability distributions, filtered predictions, and precision-recall curves.
11. Integration within HDT Framework	The predicted failure probability is integrated with the HSA, enabling anticipation of transitions between system states and supporting predictive maintenance and adaptive control.

The detailed architecture of the LSTM network used for fault prediction is summarized in **Table 6**, highlighting each layer, its type, and associated parameters. This configuration captures temporal dependencies from the SCADA data, reduces overfitting, and outputs the predicted probability of system failure.

Table 6. LSTM network layer structure.

Layer	Type	Parameters / Notes
1	LSTM	128 neurons, return_sequences = True (captures temporal dependencies)
	Dropout	Dropout
2	LSTM	64 neurons (captures temporal dependencies)
3	Dense (ReLU)	32 neurons (hierarchical feature extraction)
4	Output (Sigmoid)	1 neuron, outputs failure probability between 0 and 1 $P_f(t+\Delta t)$

HDT-LSTM Coupling

The LSTM model is integrated into the hybrid digital twin loop as follows:

- Signals from the physical layer (torque, temperature, current) feed the LSTM in real time.
- The failure probability $P_f(t)$ is transmitted to the HSA to anticipate transitions between system states (e.g., $S_1 \to S_2$).

To ensure reproducibility, the transition rate from state S_i to S_j is explicitly modulated by the LSTM output:

$$\lambda_{ij}(t) = \lambda_{ij}^0 \cdot \left(1 + \alpha \cdot P_f(t)\right) \tag{4}$$

where:

λ_{ij}^0 is the nominal transition rate,

$P_f(t)$ is the failure probability predicted by the LSTM,

α is a sensitivity coefficient.

This formulation explicitly defines how the cognitive layer dynamically influences the behavioral layer. The HDT then issues a conditional maintenance recommendation based on the updated system state.

2.5. Global Health Index (HI)

A **Health Index (HI)** is computed as a weighted combination of normalized operational parameters (temperature, vibration, speed, etc.):

$$\text{HI} = \sum_i \omega_i \cdot \frac{x_i}{x_{i,\max}}, \tag{5}$$

where:

- x_i is the value of parameter i,
- $x_{i,\max}$ is the reference maximum value of this parameter,
- ω_i is the weight associated with each parameter, reflecting its relative importance.

The weights ω_i are determined using a combination of expert knowledge (e.g., the critical importance of vibration and temperature) and statistical analysis of SCADA data, including variance-based evaluation of parameter influence.

A threshold of $HI < 0.3$ is used to indicate a degraded system state. This threshold is selected based on historical degradation patterns and empirical validation using SCADA data, and corresponds to operating conditions preceding observed abnormal events. It is therefore considered a conservative decision boundary for early maintenance intervention.

2.6. Software Tools and Experimental Environment

The computational tools used for the modeling, simulation, and implementation of the proposed framework are summarized in Table 7, along with their respective functions in the development process.

Table 7. Computational tools and their functions.

Tool	Role
Python (TensorFlow/Keras)	Multi-physics simulation, LSTM model training and integration
PyCATSHOO	HSA modeling and hybrid event management

This setup constitutes a coherent HDT environment, combining physical modeling, deep learning, and decision logic for predictive maintenance of HTG.

3. Results and Discussion

3.1. Fault Detection Analysis

To evaluate the fault detection capabilities of the proposed hybrid digital twin framework, Figure 1 illustrates the temporal alignment between actual failure events recorded in the hydroelectric plant and the corresponding predictions generated by the model. This comparison provides insight into the model's ability to capture the onset and progression of degradation phenomena in real operating conditions.

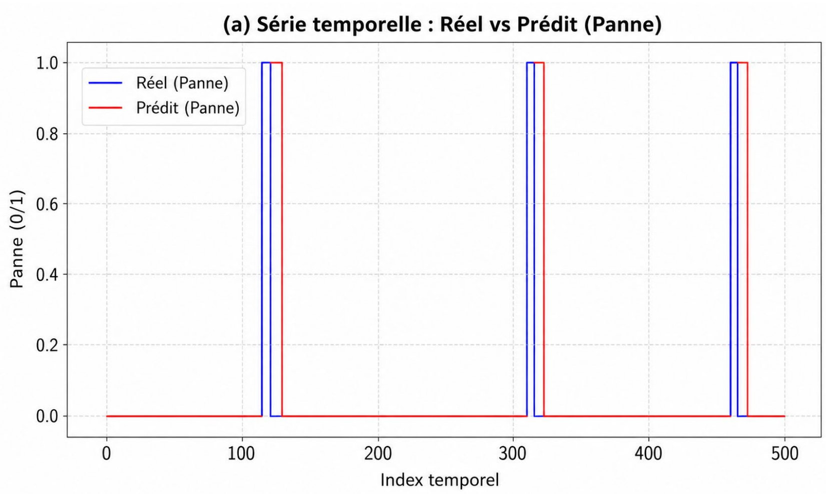

Figure 1. Model performance in fault detection.

3.1.1. Description

A strong visual agreement is observed between the predicted and actual failure events. Peaks in the measured failure signals correspond closely with the model's predictions, indicating that the system is capable of accurately identifying critical fault occurrences. Minor deviations, where predicted values slightly precede or lag the observed events, reflect the model's anticipatory behavior, which is typical in predictive maintenance frameworks aiming to provide early warnings.

3.1.2. Interpretation

The obtained results demonstrate the strong predictive capability of the hybrid digital twin. The close alignment between predicted and observed failure events confirms that the model successfully captures the underlying dynamics leading to system degradation. Importantly, the model exhibits a conservative behavior, assigning low probabilities to most operating conditions while effectively highlighting critical situations. This balance ensures that unnecessary maintenance interventions are minimized without compromising safety or system reliability.

3.2. Predicted Fault Probabilities

Figure 2 presents the distribution of fault probabilities predicted by the LSTM-based cognitive layer of the hybrid digital twin. This visualization allows for the assessment of how the model evaluates the likelihood of failures across the dataset and helps identify whether the predictions are concentrated near specific thresholds or more widely dispersed.

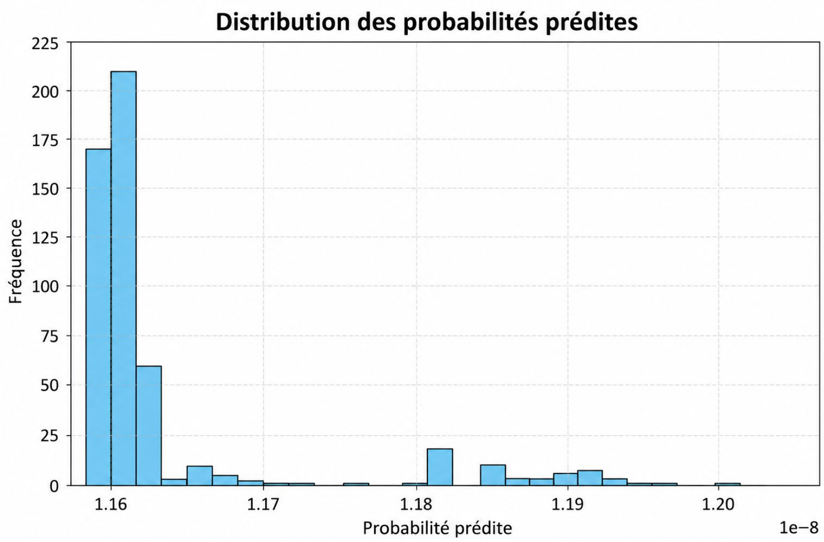

Figure 2. Distribution of probability scores associated with faults.

Description

The distribution indicates that the majority of predicted probabilities are extremely low, around 1.16×10^{-8}, suggesting that actual failures are rare events in the system. A small number of higher probability predictions correspond to instances where the model detects increased risk of failure.

Interpretation

This conservative behavior is advantageous in industrial contexts, as it reduces the risk of false alarms while maintaining sensitivity to true faults. The few elevated probability predictions correspond to early warning signals of potential failures, demonstrating that the model can successfully discriminate between normal operating conditions and abnormal states. This ability is critical for predictive maintenance strategies, ensuring interventions are timely yet not excessive.

3.3. Precision-Recall Analysis

To further evaluate the classification performance of the proposed model, **Figure 3** depicts the precision-recall curve, which illustrates the trade-off between precision (the proportion of correctly identified fault events) and recall (the ability to capture all actual faults). This analysis provides a quantitative assessment of model reliability in detecting faults under varying confidence thresholds.

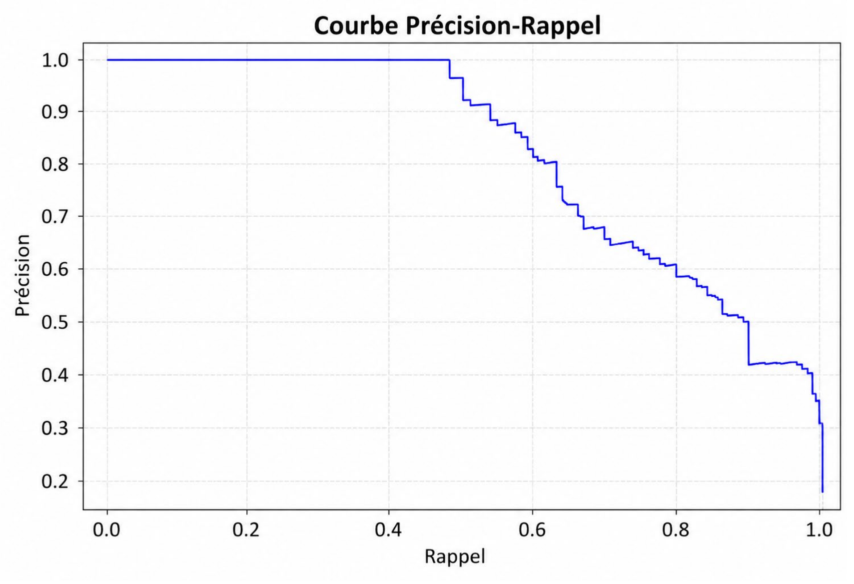

Figure 3. Model performance according to the precision-recall curve.

Description

The curve shows very high precision (≈1.0) when recall is low, indicating that the model initially detects only the most certain fault events. As recall increases, precision gradually declines, reflecting the classical trade-off between sensitivity and false alarm rate inherent in predictive maintenance systems.

Interpretation

Overall, the precision-recall analysis confirms that the hybrid digital twin achieves a well-balanced detection performance. High precision at low recall ensures that critical faults are flagged with minimal false alarms, while the gradual reduction in precision at higher recall levels is an expected outcome when attempting to capture all fault events. These results highlight the model's robustness, confirming that the integration of physics-based simulations, stochastic automata, and LSTM-based cognitive layers effectively enhances both the reliability and predictive capability of the monitoring framework.

3.4. LSTM Model Performance for Degradation Prediction

To quantitatively evaluate the predictive capabilities of the cognitive layer of the hybrid digital twin, the performance of the LSTM network was assessed using the SCADA dataset described in Section 2.4. The dataset consists of multivariate time-series measurements recorded at a 1-minute sampling rate over a continuous one-year period (2023), ensuring consistency between model development and real operating conditions.

Following preprocessing and temporal structuring, the data were segmented into overlapping sliding windows of 30 time steps (*i.e.*, 30 minutes), resulting in approximately 25,000 sequences. These sequences capture the temporal evolution of the system under various operating conditions, including nominal regimes, progressive degradation, and pre-fault states.

The dataset was divided using a strictly chronological split, with 80% of the sequences used for training and the remaining 20% reserved for testing. This strategy preserves temporal causality and ensures a realistic evaluation of the model's predictive performance in a forward-looking scenario.

The LSTM model was trained to learn temporal dependencies associated with degradation dynamics and to estimate the probability of failure based on SCADA measurements. Training was performed using the Adam optimizer with a learning rate of 10^{-3}, a batch size of 128, and over 80 epochs. These hyperparameters were selected to ensure stable convergence and effective learning of temporal patterns present in the data.

The predictive performance of the model is summarized in Table 8, which reports key evaluation metrics commonly used in predictive maintenance applications.

Table 8. Model performance metrics for fault detection and prediction.

Metric	Value
Overall Accuracy (test)	96.7%
Fault Detection Rate (Recall)	94.3%
False Positive Rate (FPR)	3.2%
Average Prediction Horizon Before Fault	9.6 hours

Description

The results demonstrate strong predictive performance of the LSTM-based cognitive layer. The model achieves an overall accuracy of 96.7%, indicating that most operational states are correctly classified. The recall of 94.3% confirms the model's high capability to detect actual degradation events, while maintaining a low false positive rate of 3.2%, which is critical for practical deployment in industrial environments.

A key result is the average prediction horizon of approximately 9.6 hours before fault occurrence. This indicates that the model can identify early degradation patterns well in advance, allowing sufficient time for maintenance planning and intervention.

Interpretation

These results highlight the effectiveness of the proposed hybrid digital twin framework in capturing degradation dynamics from real SCADA data. The use of minute-level measurements combined with 30-minute temporal windows allows the LSTM model to learn both short-term fluctuations and longer-term degradation trends.

The strict alignment between the data preprocessing described in Section 2.4 and the evaluation protocol ensures methodological consistency and reproducibility. Furthermore, the chronological data split reinforces the reliability of the reported performance by avoiding any form of temporal leakage.

The strong predictive capability of the model, combined with its early warning horizon, demonstrates its suitability for real-world predictive maintenance applications in hydroelectric systems.

3.5. Key Insights

The results obtained from the experimental evaluation highlight several important insights regarding the effectiveness of the proposed hybrid digital twin framework.

First, the fault detection mechanism demonstrates a balanced behavior that combines high sensitivity with a conservative decision strategy. The model successfully detects most fault events while maintaining a low false positive rate. This characteristic is particularly important in industrial monitoring environments, where excessive false alarms could reduce operator confidence and lead to unnecessary maintenance interventions.

Second, the LSTM-based degradation prediction model provides reliable early warning signals for potential failures. The obtained predictive horizon of approximately ten hours offers sufficient time for maintenance teams to plan corrective actions without disrupting the normal operation of the hydroelectric plant. Such anticipatory capabilities are essential for implementing effective predictive maintenance strategies in complex energy infrastructures.

Third, the integration of physics-based modeling, stochastic automata, and deep learning techniques proves to be highly effective for representing the multiscale dynamics of hydraulic turboalternator groups. The hybrid digital twin is capable of capturing both continuous physical processes and discrete operational events, while simultaneously learning hidden temporal patterns from real sensor data. As a result, the proposed architecture demonstrates robust performance across multiple operating regimes and remains resilient to measurement noise and load variability.

Overall, these findings confirm that hybrid digital twin architectures represent a promising solution for improving the reliability, availability, and maintenance efficiency of modern hydroelectric power systems.

3.6. Discussion

The results obtained in this study demonstrate the effectiveness of the proposed hybrid digital twin framework for fault detection and predictive maintenance in hydroelectric systems. The integration of physics-based modeling, hybrid stochastic automata, and data-driven learning enables an accurate representation of the complex and multi-scale dynamics of hydraulic turboalternator units.

A key strength of the proposed approach lies in its hybrid nature. By combining physical knowledge with data-driven techniques, the framework overcomes the limitations of purely data-driven methods, particularly in situations where fault data are scarce or highly imbalanced. This characteristic is especially relevant in hydropower systems, where failure events are infrequent but critical.

The strong agreement between predicted and observed fault events confirms the capability of the model to capture degradation dynamics effectively. The slight temporal shifts between predictions and actual events highlight the anticipatory nature of the framework, providing early warning signals that are essential for predictive maintenance strategies.

Moreover, the probabilistic distribution of the predictions reflects a well-balanced detection mechanism. The predominance of low fault probabilities under normal conditions reduces the occurrence of false alarms, while higher probabilities successfully identify abnormal operating states. This trade-off is crucial in industrial applications, where both reliability and operational continuity must be ensured.

The results also emphasize the importance of the hybrid data strategy. By combining simulated data derived from the physical model with real SCADA measurements, the framework improves its generalization capability and robustness to noise, missing data, and operational variability. This hybridization enhances the practical applicability of the model in real-world environments.

4. Conclusions

This study proposed a hybrid digital twin framework for fault detection and predictive maintenance in hydroelectric power systems, combining physics-based modeling, hybrid stochastic automata, and LSTM-based learning. The results demonstrate high detection accuracy, low false alarm rates, and the ability to anticipate faults several hours in advance, thereby enabling more efficient and proactive maintenance strategies.

Despite these promising results, some limitations should be acknowledged. The performance of the proposed framework depends on the quality and representativeness of the SCADA data, as well as on the accuracy of the underlying physical model. In addition, the validation has been conducted under a specific set of operating conditions, which may not fully capture all possible extreme or transient scenarios encountered in real power systems.

Future work will focus on extending the proposed approach toward real-time implementation and online adaptation. In particular, integrating the predictive capabilities of the digital twin into advanced control strategies could further enhance system stability and operational performance. Moreover, the inclusion of additional data sources, such as environmental conditions or grid disturbances, could improve the overall robustness and situational awareness of the system. Finally, the transferability of the framework to other types of hydropower plants and energy systems will be investigated to assess its scalability and general applicability.

Acknowledgements

I would first like to thank God for His guidance, support, and strength throughout the completion of this research. I would also like to express my sincere gratitude to my supervisors, Pr. Ruben Torres Martínez and Pr. Colince Welba, for their continuous support, scientific guidance, and valuable advice during the development of this study.

My appreciation also goes to the professionals and collaborators who kindly provided assistance and whose contributions helped facilitate the successful completion of this work. Their support and expertise were greatly appreciated throughout the research process.

Finally, the authors acknowledge the Journal of Computer and Communications (JCC, SCIRP) for providing the opportunity to disseminate this research and contribute to the advancement of scientific knowledge in the field of intelligent monitoring and predictive maintenance.

Declaration on the Use of Artificial Intelligence

During the preparation of this manuscript, the authors used artificial intelligence (AI) tools, including ChatGPT (OpenAI), solely to assist with language editing, grammar correction, and improvement of the manuscript's clarity and readability. The AI tool was not used to generate the scientific hypotheses, research methodology, mathematical models, experimental design, data analysis, simulation results, interpretation of findings, or conclusions.

All scientific content, including the study design, analyses, validation of results, and final manuscript, was critically reviewed, verified, and approved by the authors. The authors accept full responsibility for the accuracy, originality, and integrity of the published work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Anderson, P.M. and Fouad, A.A. (2002) Power System Control and Stability. 2nd Edition, Wiley-IEEE Press.

[2] Balogun, A.-L., *et al.* (2024) Hydropower Plant Monitoring Architectures and Digitalization Frameworks. *Energy Informatics*, **7**, Article No. 15.

[3] Kundur, P. (1994) Power System Stability and Control. McGraw-Hill Education.

[4] Boschert, S. and Rosen, R. (2016) Digital Twin—The Simulation Aspect. In: Hehenberger, P. and Bradley, D., Eds., *Mechatronic Futures*, Springer International Publishing, 59-74. https://doi.org/10.1007/978-3-319-32156-1_5

[5] Dorf, R.C. and Svoboda, J.A. (2018) Introduction to Electric Circuits. 9th Edition, Wiley.

[6] Fang, X., Misra, S., Xue, G. and Yang, D. (2012) Smart Grid—The New and Improved Power Grid: A Survey. *IEEE Communications Surveys & Tutorials*, **14**, 944-980. https://doi.org/10.1109/surv.2011.101911.00087

[7] Tao, F., Zhang, H., Liu, A. and Nee, A.Y.C. (2019) Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics*, **15**, 2405-2415. https://doi.org/10.1109/tii.2018.2873186

[8] Zhang, W., Yang, D. and Wang, H. (2019) Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal*, **13**, 2213-2227. https://doi.org/10.1109/jsyst.2019.2905565

[9] Grieves, M. (2014) Digital Twin: Manufacturing Excellence through Virtual Factory Replication. White Paper, Florida Institute of Technology.

[10] Fuller, A., Fan, Z., Day, C. and Barlow, C. (2020) Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access*, **8**, 108952-108971. https://doi.org/10.1109/access.2020.2998358

[11] White, F.M. (2011) Fluid Mechanics. 7th Edition, McGraw-Hill Education.

[12] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R.X. (2019) Deep Learning and Its Applications to Machine Health Monitoring. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 3452-3468.

[13] Zhou, K., Fu, C. and Yang, S. (2016) Big Data Driven Smart Energy Management: From Big Data to Big Insights. *Renewable and Sustainable Energy Reviews*, **56**, 215-225. https://doi.org/10.1016/j.rser.2015.11.050

[14] Zhou, Y., *et al.* (2021) Data-Driven Models in Hydropower Monitoring and Fault Diagnosis. *IEEE Access*, **9**, 59324-59340.

[15] Fluid Mechanics White, F.M. (2011) Fluid Mechanics. 7th Edition, McGraw-Hill Education.

[16] Gungor, V.C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., *et al.* (2011) Smart Grid Technologies: Communication Technologies and Standards. *IEEE Transactions on Industrial Informatics*, **7**, 529-539. https://doi.org/10.1109/tii.2011.2166794

[17] Errandonea, I., Beltrán, S. and Arrizabalaga, S. (2020) Digital Twin for Maintenance: A Literature Review. *Computers in Industry*, **123**, Article ID: 103316. https://doi.org/10.1016/j.compind.2020.103316

[18] Jardine, A.K.S., Lin, D. and Banjevic, D. (2006) A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing*, **20**, 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012

[19] Gao, R.X., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M., *et al.* (2015) Cloud-enabled Prognosis for Manufacturing. *CIRP Annals*, **64**, 749-772. https://doi.org/10.1016/j.cirp.2015.05.011

[20] Jones, D., Snider, C., Nassehi, A., Yon, J. and Hicks, B. (2020) Characterising the Dig-

ital Twin: A Systematic Literature Review. *CIRP Journal of Manufacturing Science and Technology*, **29**, 36-52. https://doi.org/10.1016/j.cirpj.2020.02.002

[21] Tao, F., Qi, Q., Liu, A. and Kusiak, A. (2019) Data-Driven Smart Manufacturing. *Journal of Manufacturing Systems*, **48**, 157-169. https://doi.org/10.1016/j.jmsy.2018.01.006

[22] Cassandras, C.G. and Lafortune, S. (2008) Introduction to Discrete Event Systems. 2nd Edition, Springer.

[23] Costa, A., *et al*. (2017). Stochastic Processes in Engineering Systems. Springer.

[24] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[25] Susto, G.A., Schirru, A., Pampuri, S., McLoone, S. and Beghi, A. (2015) Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. *IEEE Transactions on Industrial Informatics*, **11**, 812-820. https://doi.org/10.1109/tii.2014.2349359

[26] Lee, J., Davari, H., Singh, J. and Pandhare, V. (2020) Industrial AI and Digital Twins for Smart Manufacturing. *Manufacturing Letters*, **24**, 34-36.

[27] Pai, M.A. (1989) Energy Function Analysis for Power System Stability. Kluwer Academic Publishers.

[28] Djacba, D.T., Martinez, R.T. and Welba, C. (2026) Towards an Advanced Digital Twin for the Dynamics of a Hydraulic Turboalternator Group under Variable Load Constraints. *Journal of Computer and Communications*, **14**, 80-102. https://doi.org/10.4236/jcc.2026.144004