

Rule-Based Pump Health Monitoring with Human-in-the-Loop Decision Gating

***Multika Untung Bahagiya**

Departement of Electrical Engineering, Ronggolawe University of Technology, Jl. Kampus Ronggolawe Blok B No. 1, Mentul, Cepu District, Central Java 58315, Indonesia

[*bahagiya.muti@gmail.com](mailto:bahagiya.muti@gmail.com)

Bagus Tri Kuncoro

Electronic Engineering Education, Faculty of Engineering, State University of Jakarta, Jl. Rawamangun Muka RT/RW 11/14, Rawamangun, Pulo Gadung District, East Jakarta, Jakarta 13220, Indonesia

bagus.tri.kuncoro@unj.ac.id

Indra Mukti Purboyo

Departement of Electrical Engineering, Ronggolawe University of Technology, Jl. Kampus Ronggolawe Blok B No. 1, Mentul, Cepu District, Central Java 58315, Indonesia

indrapurboyo91@gmail.com

Article history: Received June 27, 2026| Revised August 01, 2026| Accepted August 08, 2026

Abstract – This study proposes a rule-based Human-in-the-Loop (HITL) framework for condition-based maintenance in sensor-limited industrial environments. The framework integrates a Numeric Health Score (NHS) based on five operating parameters—vibration, RPM slip, motor temperature, suction pressure, and PDI/PDIT—with an Ordinal Utility Model (OUM) that quantifies qualitative field-operator assessments. The scores are fused using $\alpha = 0.65$ to generate a Health Index (HI), classified into HEALTHY ($HI \geq 75$), WATCH ($55 \leq HI < 75$), DEGRADED ($35 \leq HI < 55$), and CRITICAL ($HI < 35$). Model robustness was evaluated using two Monte Carlo experiments ($N = 10,000$), yielding a Spearman rank correlation of $\rho = 0.9853$ ($p < 0.001$). Bootstrap resampling ($N = 10,000$ per pump) was used to estimate 95% confidence intervals. A three-signal HITL mechanism based on bootstrap confidence level, Classification Stability Rate (CSR), and boundary proximity identifies pumps requiring verification before maintenance decisions. Application to 16 centrifugal pumps over 30 days resulted in 11 HEALTHY (68.75%), 4 WATCH (25.0%), and 1 DEGRADED (6.25%), with no CRITICAL pumps. Two pumps (12.5%) required HITL review, and one was downgraded from WATCH to DEGRADED after its confidence interval crossed the classification boundary. The framework supports maintenance prioritization by directing operator or engineer verification toward pumps with greater uncertainty or deterioration risk, while reducing unnecessary inspection and sensor-deployment requirements. It provides an auditable and reproducible approach for condition-based maintenance and progressive maintenance digitization.

Keywords: Centrifugal Pump, Health Index, Numeric Health Score, Ordinal Utility Model, Human-in-the-Loop.



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

I. INTRODUCTION

Centrifugal pumps are among the most critical pieces of equipment in the oil and gas industry, particularly at fuel terminals. Pump failure can result in operational disruptions, economic losses, and even safety risks.

Therefore, condition monitoring is a crucial activity in condition-based maintenance [1], [2]. The development of machine learning and deep learning technology has encouraged various research, especially related to machine damage diagnosis, including centrifugal pumps [3], [4]. However, such approaches generally assume the availability of integrated and continuous real-time sensor data, as in an integrated SCADA system [5], [6].

Extensive research has been conducted on centrifugal pump condition monitoring using vibration, pressure, and motor current signals. Pressure measurements on the suction and discharge sides of the pump can be used to effectively detect changes in operational conditions [7], [8]. Motor current signature analysis (MCSA) can also be used to identify electrical and mechanical anomalies without additional sensors [9], [10]. On the other hand, the integration of Variable Frequency Drive (VFD) data with the use of deep learning is starting to be applied to predict remaining service life [11], [12]. Although the methods described above have followed current technological developments, they still rely on online sensor data which is not always available in all industrial facilities.

At many fuel terminals in Indonesia, pump operational data collection is still done manually by operators. Data such as RPM (using a tachometer), vibration (vibration meter), temperature (thermogun), and pressure (pressure gauge) are recorded periodically in a logbook or equipment historical data form. This condition often presents challenges related to decision-making on equipment, where available data is sparse, has variability due to human factors, and often may be incomplete. International standards such as ISO 20816-3:2022 provide guidelines for the measurement and evaluation of vibration for industrial machines with capacities above 15 kW and operating speeds of 120 r/min and 30,000 r/min [13], [14]. API 610 specifies the technical requirements for

centrifugal pumps in the oil and gas industry [15], [16], and IEC 60034-1 regulates the performance specifications for rotating electrical machines (including those related to temperature or thermal requirements) [17], [18]. However, the application of these standards in manual data measurement-based environments has rarely been explored in a systematic and integrated manner [19], [20].

In sensor-limited maintenance environments, a standard-based rule system offers a practical alternative to data-intensive machine learning, fuzzy logic, and conventional expert systems. Machine-learning approaches generally require sufficient historical and labeled data, which may not be consistently available when pump measurements are collected manually. Fuzzy logic can accommodate uncertainty and gradual transitions between equipment conditions; however, its implementation requires the definition of membership functions and rule bases that may introduce additional subjective assumptions. Conventional expert systems can incorporate expert knowledge, but their effectiveness depends on the completeness and consistency of the knowledge encoded in the system. In contrast, a standard-based rule system directly translates measured operating conditions into explicit health categories using established engineering thresholds. In this study, ISO 20816-3:2022, API 610, IEC 60034-1:2017, and API 682 are used as the basis for defining parameter thresholds. This approach provides transparent, traceable, and auditable decision criteria and does not require a large training dataset. Therefore, it is appropriate for sparse and manually collected operational data while still allowing qualitative operator knowledge to be incorporated through the Ordinal Utility Model (OUM).

To bridge this gap, this study proposes a rule-based CBM framework specifically designed for sensor-limited environments. The approach integrates two complementary data sources: (1) a Numeric Health Score (NHS) calculated from five operating parameters—vibration, RPM slip, motor temperature, suction pressure, and PDI/PDIT—with Analytic Hierarchy Process (AHP)-based weighting; and (2) an Ordinal Utility Model (OUM) that extracts and quantifies qualitative information from operator records or routine inspection results. Previous studies have emphasized that direct observation and field inspection by operators remain important components in interpreting machine conditions [21], [22]. In this study, these observations are incorporated into a structured and auditable decision framework to complement quantitative measurements and produce a Final Health Class (FHC).

This study addresses two research gaps. First, there is a need to integrate numerical operating data and qualitative operator assessments through a transparent and auditable rule-based framework, particularly in industrial environments where data are sparse and

manually collected. Second, there is a need for a statistical probability-based Human-in-the-Loop (HITL) mechanism that explicitly incorporates uncertainty into equipment health classification and maintenance decision-making [23]. In this framework, measurement uncertainty is evaluated using Monte Carlo simulation and bootstrap resampling with $N = 10,000$ iterations [24]. The resulting uncertainty information is used as a confidence gate, where stable classifications can be processed automatically, while unstable classifications or conditions requiring further verification are referred to operators. This approach provides a structured mechanism for maintaining human oversight in sparse-data condition monitoring and supports more reliable and explainable maintenance decisions.

II. METHOD

This research develops a Condition-Based Maintenance (CBM) framework with a standard-based rule system specifically designed for industrial environments with limited real-time sensors. The methodology integrates manual data acquisition, Multi-Criteria Decision Making-based parameter weighting, Numeric Health Score (NHS) scoring, Ordinal Utility Model (OUM) fusion, bootstrap confidence estimation using Monte Carlo, and a Final Health Class (FHC) mechanism through HITL gating. Data processing is performed computationally and implemented using Python (with several supporting libraries such as pandas, numpy, scipy, matplotlib) with a documented and reproducible processing flow. Meanwhile, the pump operational data used comes from daily manual operator records at the fuel terminal over a 30-day period.

A. Numeric Health Score (NHS) and Weight Determination with Analytic Hierarchy Process (AHP)

The NHS is a quantitative health score derived from five operational parameters evaluated against applicable engineering standards. It represents objective, measurement-based evidence of pump condition, whereas the OUM represents contextual and qualitative evidence obtained from operator observations and field assessments. The distinction between these two evidence types forms the conceptual basis for their subsequent fusion, in which measurement-based evidence is given greater influence while operator knowledge is retained as complementary information.

1. Score Per Parameter

Each parameter is converted into a discrete numeric score of 100, 75, 50, or 25 using a standard-based stepwise threshold function. The four scores represent the Normal, Watch, Warning, and Alert conditions, respectively. This discrete mapping was selected because the primary objective of the NHS is to translate engineering standard limits into operationally interpretable maintenance categories rather than to

represent a continuous physical degradation function. However, the stepwise approach may introduce discontinuities near threshold boundaries, where a small change in the measured value can result in a change of 25 points in the parameter score. Therefore, the effect of the stepwise scoring scheme on boundary sensitivity is further examined through a piecewise-linear sensitivity analysis.

Table 1. Threshold Scoring per NHS Parameter

Para meter	Reference Standard	Score 100 (Nor mal)	Score 75 (Watc h)	Score 50 (War ning)	Score 25 (Alert)
Vibra tion (mm/ s)	ISO 20816- 3:2022	< 2,3	2,3 – 4,5	4,5 – 7,1	> 7,1
RPM Slip (% absol ute)	API 610 12th Ed.	< 3%	3 – 8%	8 – 12%	> 12%
Temp eratur e	IEC 60034- 1:2017	< 55	55 – 70	70 – 80	> 80
Sucti on (Kg/c m ²)	API 610	≥ 0,3	0,1 – 0,3	0,0 – 0,1	< 0
PDI/P DIT (bar)	API 682	< 0,030	0,030 – 0,080	0,080 – 0,150	> 0,150

The threshold values in Table 1 are applied using a stepwise scoring approach, in which each measured value is assigned to one of four predefined health scores: 100, 75, 50, or 25. This approach provides a clear and consistent classification based on the applicable technical standards. However, the stepwise function creates a discontinuity at each threshold. For example, a small change in vibration around a threshold may cause the parameter score to change from 75 to 50, even though the actual difference in the measured vibration value is relatively small. Therefore, observations located close to the threshold boundaries may be more sensitive to small measurement variations.

To assess this potential sensitivity, a piecewise-linear scoring function is used as a supplementary sensitivity analysis. In this approach, values between two consecutive thresholds receive an interpolated score rather than an immediate transition from one discrete score to another. The linear score is calculated using:

$$S_{linier}(x) = S_i + \left(\frac{S_{i+1} - S_i}{T_{i+1} - T_i} \right) (x - T_i)$$

Where (x) is the observed measurement, T_i and T_{i+1} are two consecutive threshold values, and S_i and S_{i+1}

are their corresponding scores. The primary model retains the stepwise scoring because its objective is to classify pump conditions according to standard-based maintenance categories, while the piecewise-linear formulation is used only as a sensitivity analysis.

$$\Delta HI = |HI_{step} - HI_{linier}|$$

Where (HI_{step}) is the Health Index calculated using the primary stepwise scoring function and (HI_{linier}) is the Health Index calculated using the piecewise-linear scoring function. The absolute difference (ΔHI) represents the sensitivity of the final Health Index to the choice of scoring function. A larger (ΔHI) indicates that the assessment is more affected by the discontinuity of the stepwise scoring scheme.

$$BP = \min_{b \in B} |HI - b|, B = \{35, 55, 75\}$$

Where $(B = \{35, 55, 75\})$ represents the boundaries between the CRITICAL, DEGRADED, WATCH, and HEALTHY classes. (BP) represents the minimum distance between the calculated Health Index and the nearest health-class boundary. A smaller (BP) indicates that the pump is closer to a classification boundary and is therefore more sensitive to small changes in measurement values, scoring functions, or fusion parameters. Conversely, a larger (BP) indicates greater separation from the nearest class boundary and a more stable classification.

The combination of (ΔHI) and (BP) is used to evaluate whether the discontinuity of the stepwise scoring function materially affects the final health classification. In particular, observations with a small (BP) and a relatively large (ΔHI) are considered more sensitive and therefore more likely to require further examination through the Human-in-the-Loop (HITL) mechanism.

The stepwise model is retained as the primary classification model, whereas the piecewise-linear model serves as a robustness check. A classification is considered potentially sensitive when a relatively large (ΔHI) occurs together with a small (BP) , because the difference between the two scoring schemes may alter the assigned health class.

Here is a visualization of how threshold scoring works visually:



Figure 1. Heatmap Visualization of Sub-Score Per Pump

2. Determining Weights Using The Analytic Hierarchy Process (AHP) Method

The Analytic Hierarchy Process (AHP) method developed by Saaty was used to determine the weight of each parameter. The pairwise comparison matrix was compiled based on the severity potential and early warning capability of each parameter, as presented in Table 2 below:

Table 2. Pairwise Comparison Matrix (PCM)

Parameter	Vibration	RPM Slip	Motor Temp	Suction	PDI
Vibration	1	1	1	3	3
RPM Slip	1	1	1	3	3
Motor Temp	1	1	1	3	3
Suction	1/3	1/3	1/3	1	1
PDI/PDI	1/3	1/3	1/3	1	1

The normalization of columns and priority vectors (eigenvectors) is obtained from the following calculation:

Each element a_{ij} is divided by the number of columns $\sum_{i=1}^n a_{ij}$

Where:

- a_{ij} : the pairwise comparison value between the i-th parameter and the j-th parameter using the Saaty fundamental scale (1–9). Where $a_{ij} > 1$ indicates that parameter i is more important than parameter j, while $a_{ij} < 1$ indicates the opposite.
- a_{norm} : the column normalization of the pairwise comparison matrix in Table 2.
- n : parameters used in the study

The number of each PCM column above is: Vibration column = 3.667; RPM Slip column = 3.667; Motor Temp column = 3.667; Suction column = 11,000; PDI column = 11,000. After obtaining the PCM column normalization above, the next step is to calculate the priority vector (eigenvector) as the row average of the normalization matrix, as shown in Table 3 below:

Table 3. AHP Weights – Eigenvector Results of PCM Normalization

Parameter	Weight (w)	Percent age	Justification
Vibration Pump DE	0,2727	27,27%	Mechanical indicators
RPM Slip	0,2727	27,27%	Electrical and mechanical efficiency indicators
Motor Temperature	0,2727	27,27%	Motor insulation and cooling conditions

Suction Pressure	0,0909	9,09%	Situational, depending on process conditions
PDIT	0,0909	9,09%	Situational as a comparison of the pressure measured on the pressure gauge with the transmitter
Total	1,0000	100%	

After obtaining the above results, a consistency test (Consistency Ratio (CR)) was carried out with the following formulation:

$$\lambda_{max} = \frac{1}{n} \sum_{i=1}^n \frac{(A \cdot \omega)_i}{\omega_i} = 5 \dots\dots\dots(1)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} = \frac{5 - 1}{5 - 1} = 0 \dots\dots\dots(2)$$

For n=5, the standard Saaty Random Index (RI) value is 1.12.

Where:

- λ_{max} : the largest eigenvalue of the pairwise comparison matrix A
- ω : represents the relative weight of each parameter in the condition monitoring system.
- Consistency Index (CI): a quantitative metric that measures the degree of logical inconsistency in the pairwise comparison matrix.

The CR value is as follows:

$$CR = \frac{CI}{RI} = \frac{0}{1.12} = 0 \dots\dots\dots(3)$$

Because $CR < 0.10$ the matrix is declared consistent and the weights can be used.

B. Ordinal Utility Model (OUM) and NHS–OUM FUSION

OUM quantifies the qualitative assessment of operators in the field that cannot be captured by sensor data. The four OUM components are obtained from three different data sources as shown in Table 4 below:

Table 4. OUM Components and Ordinal Mapping to Numeric Scores

Components	Data Source	Weight	Mapping Ordinal → Score
oum_flag	FLAG	30%	OK→10
oum_rpm	_RPM		0
oum_slip	_SLIP (daily data)		WARN →50
			ALERT →25

oum_c onditio n	Overall		HEALT
	l		HY→10
	Condit		0
	ion		DEGRA
	(daily	30%	DED→5
	data)		0
			CRITIC
			AL→25
oum_s everity	Self-		LOW→1
	Analys		00
	is (per		WATCH
	pump)		→65
		25%	MEDIU
			M→50
			HIGH→
			35
			CRITIC
			AL→25
oum_ manpo	Inspect		Normal
	ion		→100
	Refere		Standby
	nce		→85
	(per	15%	Monitor
	pump)		→65
			Perhatian
			→50
			Tindakan
			→25
oum_c onditio n	FLAG		OK→10
	_RPM		0
	_SLIP		WARN
	(daily	30%	→50
	data)		ALERT
			→25

$$OUM_{Score} = \frac{\sum_{k=1}^4 (s_k \times w_k)}{\sum_{k=1}^4 w_k} \dots\dots\dots(4)$$

Where:

- 1) s_k is the numerical score of the k th component resulting from ordinal mapping
- 2) w_k is the weight of the k th component (in percentage, with $\sum w_k = 100\%$)

The final Health Index (HI) is obtained through linear fusion of NHS and OUM with parameter $\alpha = 0.65$:

$$HI = \alpha \times NHS_{final} + (1 - \alpha) \times OUM_{Score},$$

$$\alpha = 0,65 \dots\dots\dots(5)$$

Where:

NHS_{final} is the numerical health score of the operating parameters (vibration, RPM slip, motor temperature, suction pressure, and PDI/PDIT) that has gone through the AHP and aggregation process,

OUM_{Score} is the score from Equation 4, α is the weighting coefficient that determines the relative dominance of NHS over OUM.

The fusion coefficient ($\alpha = 0.65$) is defined as the baseline configuration to give greater influence to the quantitative NHS while retaining OUM as

complementary qualitative evidence. This weighting rationale is based on the different evidential characteristics of the two components: NHS is derived from five operational parameters evaluated against engineering standard thresholds, whereas OUM represents contextual information obtained from operator observations that may not be captured by the available measurements. Therefore, the 65:35 configuration is intended to prioritize objective standard-based evidence without eliminating the contribution of field-based expert knowledge.

Importantly, ($\alpha = 0.65$) is not assumed to be a universally optimal value. It is treated as the baseline configuration and its robustness is examined through a Monte Carlo sensitivity analysis in which α is varied from 0.50 to 0.80. For each simulated α value, the resulting Health Index, classification stability rate, boundary-crossing frequency, and HITL activation are evaluated and compared with the nominal $\alpha = 0.65$ configuration. The 65:35 configuration is considered robust when its classification performance remains stable and does not show systematic deterioration relative to neighboring α values. This analysis therefore provides an empirical basis for retaining $\alpha = 0.65$ rather than treating the 65:35 ratio as an arbitrary fixed assumption. HI is classified into four classes: HEALTHY ($HI \geq 75$), WATCH ($55 \leq HI < 75$), DEGRADED ($35 \leq HI < 55$), and CRITICAL ($HI < 35$).

C. Monte Carlo Validation and Bootstrap Confidence Estimation

Statistical validation was performed using Monte Carlo simulation and bootstrap resampling to evaluate the robustness of the AHP-derived weights, the NHS–OUM fusion coefficient, and the uncertainty of the resulting Health Index. A total of 10,000 Monte Carlo iterations were performed for each experiment.

Experiment A: Robustness of AHP Weights. The stability of the AHP-derived parameter weights was evaluated by generating perturbed weight vectors using a Dirichlet distribution:

$$w_{sim} \sim Dirichlet(\alpha = W_{nominal} \times \kappa), \kappa = 10 \dots\dots\dots(6)$$

Where:

- 1) $W_{nominal}$ is the nominal weight vector.
- 2) κ is the concentration parameter controlling the magnitude of weight perturbation, and,
- 3) w_{sim} is the simulated weight vector. For each iteration, the NHS and subsequent Health Index were recalculated using the simulated weights. This procedure evaluates whether small plausible variations in the AHP weights materially affect pump ranking and health classification.

Experiment B: Robustness of the NHS–OUM Fusion Coefficient. The robustness of the proposed fusion coefficient was evaluated by varying (α)

over the interval 0.50–0.80:

$$\alpha_{sim} \sim Uniform(0,50; 0,80), N = 10.000 \text{ iterasi} \dots\dots\dots(7)$$

For each iteration, a value of (α_{sim}) was sampled and used to recalculate the Health Index according to:

$$HI_{sim} = \alpha_{sim}NHS_{final} + (1 - \alpha_{sim})OUM_{score}$$

The resulting Health Index, health-class assignment, classification stability rate (CSR), boundary-crossing frequency, and HITL activation were compared with the nominal configuration of ($\alpha = 0.65$). The purpose of this analysis is not to assume that ($\alpha = 0.65$) is universally optimal, but to determine whether the proposed 65:35 configuration remains stable under plausible variations in the relative contribution of quantitative and qualitative evidence.

The Classification Stability Rate (CSR) is calculated as:

$$CSR = \frac{N_{same \text{ class}}}{N_{total}} \times 100\%$$

Where ($N_{same \text{ class}}$) is the number of Monte Carlo iterations that produce the same health class as the nominal configuration and ($N_{total} = 10.000$). A higher CSR indicates greater robustness of the classification to parameter perturbation. A CSR below 80% is considered an indication that the classification is insufficiently stable for fully automated decision-making and therefore activates the HITL mechanism.

The coefficient of variation (CV) is used to quantify the relative variability of the simulated Health Index:

$$CV = \frac{\sigma HI}{\mu HI} \times 100\%$$

Where (σHI) is the standard deviation and (μHI) is the mean of the simulated Health Index. A lower CV indicates greater numerical stability of the model.

The Spearman rank correlation coefficient (ρ) is used to evaluate the consistency between the nominal pump ranking and the median ranking obtained from the Monte Carlo simulations. A value of ($\rho > 0.90$) is used as the criterion for strong rank-order consistency.

Sensitivity to individual parameters is further examined using a One-at-a-Time (OAT) analysis. The normalized sensitivity index is calculated as:

$$SI_i = \frac{|\Delta \overline{HI}_i|}{\sum_{j=1}^p |\Delta \overline{HI}_j|} \dots\dots\dots(8)$$

Where:

- 1) SI_i is normalized sensitivity index of the i -th parameter;
- 2) $|\Delta \overline{HI}_i|$ is the average change in the Health Index resulting from variation in parameter;
- 3) p is total number of parameters varied in the OAT analysis;
- 4) $\sum_{j=1}^p |\Delta \overline{HI}_j|$ is total absolute average change in the Health Index across all varied parameters.

Bootstrap resampling was subsequently performed to quantify the uncertainty associated with the observed daily Health Index values. For each pump with (n) daily observations, bootstrap samples were generated with replacement and repeated 10,000 times. The bootstrap distribution of the mean Health Index was used to estimate the 95% confidence interval:

$$CI_{95\%} = [\text{percentile}(bs_{means}, 2,5); \text{percentile}(bs_{means}, 97,5)] \dots\dots\dots(9)$$

The Confidence Level is defined as:

- 1) HIGH if the bootstrap Coefficient of Variation (CV) is <5% and the Coefficient Index (CI) width is <10;
- 2) MEDIUM if the CV is <15% and the CI width is <25; and,
- 3) LOW when either criterion for the MEDIUM category is not satisfied. The confidence-interval width is calculated as:

$$CI_{width} = CI_{upper} - CI_{lower}$$

Where:

- 1) CI_{width} is width of the 95% bootstrap confidence interval;
- 2) CI_{upper} is upper limit of the 95% confidence interval;
- 3) CI_{lower} , lower limit of the 95% confidence interval.

$$CI_{95\%} = [CI_{lo}, CI_{hi}]$$

dan

$$CI_{cross} = \begin{cases} 1, \exists b \in \{35, 55, 75\} : CI_{lo} < b < CI_{hi} \\ 0, \text{otherwise} \end{cases}$$

CI_{cross} menunjukkan bahwa ketidakpastian bootstrap melintasi batas kelas sehingga klasifikasi otomatis tidak dianggap aman dan **HITL diaktifkan**.

The combination of Monte Carlo stability, bootstrap uncertainty, and boundary proximity provides the basis for determining whether a pump can be assigned a health class automatically or requires additional human verification through the Human-in-the-Loop (HITL) mechanism.

The following diagram proves that the selected AHP weights produce sensitivity to several previously defined parameters:

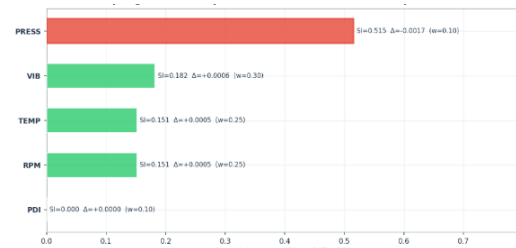


Figure 2. Tornado Diagram Visualization

D. Human-In-The-Loop (HITL) Gating and Final Health Class (FHC)

The Human-in-the-Loop (HITL) mechanism is the final decision layer before the Final Health Class (FHC) is assigned. Its purpose is to automatically identify pumps with high enough classification ambiguity to require verification by an operator or engineer before a maintenance decision is made.

Table 5. HITL Parameters

Parameter	Active Condition	Justification
Confidence Level	= LOW (CV $\geq$ 15% atau $CI_{width} \geq 25$)	High uncertainty prevents reliable automated classification.
CSR (Monte Carlo)	< 80%	Health class changes in more than 20% of Monte Carlo iterations.
Boundary Proximity	< 8 points	HI is close to a class boundary and may be sensitive to small variations.

The HITL mechanism is activated when at least one predefined uncertainty or classification-ambiguity condition is satisfied. A pump is flagged for HITL review when the bootstrap confidence level is LOW, the Monte Carlo CSR is below 80%, or the boundary proximity is less than 8 points. These criteria are designed as operational safeguards to prevent automated maintenance decisions when the numerical assessment is unstable, the classification changes frequently under parameter perturbation, or the Health Index is close to a class boundary. Meanwhile, to determine the Final Health Class (FHC), follow the following decision logic in Table 6 below:

Table 6. Final Health Class (FHC)

Decision Path	Condition	Final Health Class
AUTO_CONFIRMED	HITLRequired = False	Nominal HIClass
HITL_CONFIRMED_SINGLE_CLASS	HITL required = True; CI lower and upper limits remain within the same health class	Nominal HI Class; HITL review required
HITL_DOWNGRADED	HITL required = True; bootstrap CI crosses one or more class boundaries	Lower health class is assigned conservatively

III. RESULTS AND DISCUSSION

The research discussion on the monitoring or monitoring system of 16 centrifugal pump units at the

Fuel Terminal was conducted using the following integrated approaches: Numeric Health Score (NHS), Ordinal Utility Model (OUM), Monte Carlo validation, and Human-in-the-Loop (HITL) gating mechanism. The data used is operational data within 30 days or 1 month. The data results obtained are used as a reference to identify the health condition of each pump, determine maintenance priorities, and validate the model consistency statistically.

A. Numeric Health Score (NHS) Per Pump

NHS calculations were performed on five operational parameters: pump vibration (ISO 20816-3:2022), RPM slip (API 610), motor temperature (IEC 60034-1), suction pressure (API 610), and PDI/PDIT (API 682). The sub-score results for each parameter and the final NHS for 16 pumps over 30 days are shown in Table 7.

Table 7. NHS Per Pump (Monthly Average)

Pump	Vib DE	RPM Slip	Motor Temp	Suction	NHS final	OUM Score	HI
P.TGL-110A	100	100	100	100	100,00	96,74	97,13
P.TGL-110B	100	100	100	100	100,00	95,74	96,03
P.TGL-120A	100	100	100	87,5	98,61	95,73	95,84
P.TGL-120B	100	100	100	100	100,00	95,73	95,87
P.TGL-120C	100	100	100	87,5	98,61	95,62	95,84
P.TGL-120D	100	100	100	87,5	98,61	95,12	95,36
P.TGL-120E	100	100	100	100	100,00	95,17	95,43
P.TGL-210A	100	100	100	100	95,50	57,98	59,84
P.TGL-210B	100	100	100	100	97,60	57,98	61,90
P.TGL-220A	100	50	100	100	65,00	66,65	66,06
P.TGL-220B	100	50	100	100	65,00	65,75	65,75
P.TGL-220C	100	50	100	100	65,00	66,20	66,46
P.TGL-320A	100	50	100	100	65,00	94,83	88,24
P.TGL-320B	100	70	100	100	79,00	94,83	93,16
P.TGL-320C	100	100	100	87,5	98,61	95,28	95,46
P.TGL-320D	100	100	100	100	100,00	94,94	95,36

In the table above, the most crucial finding is the contrast between NHS and OUM on P.TGL-210A and P.TGL-210B. The final NHS of these two pumps reached 95.50 and 97.60 because all sensor parameters were in the normal zone. However, the OUM Score on both pumps was only 57.98 due to the Severity = HIGH assessment by the field operator. After fusion ($\alpha = 0.65$), the Health Index (HI) dropped to 59.84 and 61.90. The HI-NHS delta of -35.66 and -35.70 points was the largest in the dataset, confirming that the operator's tacit knowledge assessment successfully

corrected the optimism of the measurement or sensor data significantly.

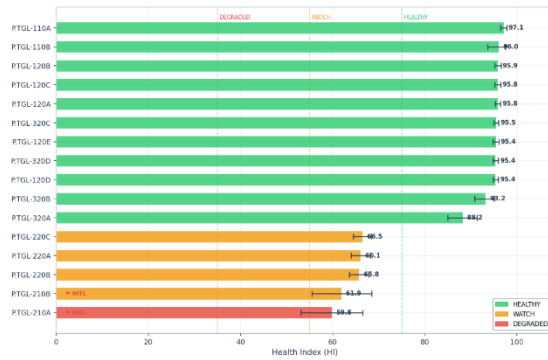


Figure 3. Ranking Health Index 16 Pompa

B. Health Index (HI) and Final Health Class (FHC) Ranking

The pumps are ranked in ascending order of HI, with Rank 1 representing the lowest Health Index and consequently the highest maintenance priority. The FHC determination is performed through a HITL gating mechanism based on three signals: the Confidence Level from the bootstrap, the Classification Stability Rate (CSR) from Monte Carlo, and boundary proximity. The HI and FHC rankings are shown in Table 8:

Table 8. Ranking Health Index and Final Health Class 16 Pompa FT Tegal

Rank	Pump	HI Mean	HI Nominal	Final Class	Decision Path	Conf.	HITL	Trend
1	P.TGL-210A	59,84	WATCH	DEGRADED	HITL_DOWN	ME DIU M	YES	DECLINING
2	P.TGL-210B	61,90	WATCH	WATCH	HITL_CONFIRM	ME DIU M	YES	DECLINING
3	P.TGL-220B	65,75	WATCH	WATCH	AUTO	HIGH	no	STABLE
4	P.TGL-220A	66,06	WATCH	WATCH	AUTO	HIGH	no	STABLE
5	P.TGL-220C	66,46	WATCH	WATCH	AUTO	HIGH	no	STABLE
6	P.TGL-320A	88,24	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
7	P.TGL-320B	93,16	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
8	P.TGL-120D	95,36	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
9	P.TGL-320D	95,36	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
10	P.TGL-120E	95,43	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
11	P.TGL-320C	95,46	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
12	P.TGL-120A	95,84	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
13	P.TGL-120C	95,84	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
14	P.TGL-120B	95,87	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
15	P.TGL-110B	96,03	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE
16	P.TGL-110A	97,13	HEALTHY	HEALTHY	AUTO	HIGH	no	STABLE

Based on Table 8, the Final Health Class distribution shows that 11 pumps (68.75%) were classified as HEALTHY, 4 pumps (25.00%) as WATCH, and 1 pump (6.25%) as DEGRADED. No pump reached the CRITICAL class during the 30-day observation period. Fourteen pumps (87.50%) were assigned through the AUTO_CONFIRMED path, indicating

that their classification remained sufficiently stable for automated assessment. Two pumps, P.TGL-210A and P.TGL-210B, required HITL review because uncertainty and classification-ambiguity signals were activated. Both pumps recorded CSR = 0% relative to their nominal class across the 10,000 Monte Carlo iterations, with boundary proximity values of 4.8 and 6.9 points, respectively, from the WATCH–DEGRADED boundary.

P.TGL-210A was the only pump whose class changed through the HITL_DOWNGRADED path, from nominal WATCH to final DEGRADED. Its 95% bootstrap confidence interval of 53.17–66.51 crossed the WATCH–DEGRADED boundary at 55.0 points. Therefore, the lower health class was assigned conservatively according to the predefined FHC decision rule. In contrast, P.TGL-210B had a 95% bootstrap confidence interval of 55.47–68.47, which remained entirely within the WATCH class. Consequently, its nominal WATCH classification was retained through the HITL_CONFIRMED_SINGLE_CLASS path. These results demonstrate that the HITL mechanism selectively identifies cases in which uncertainty or proximity to a classification boundary may compromise the reliability of automated health classification.

The maintenance implications of the results are particularly evident for P.TGL-210A and P.TGL-210B. Although the quantitative NHS values of these pumps remained high at 95.50 and 97.60, respectively, both pumps obtained an OUM Score of 57.98 because the field operator assigned a HIGH severity assessment. This discrepancy indicates that the maintenance response should not be based solely on the numerical sensor-derived condition. For P.TGL-210A, the recommended first response is priority field re-inspection to verify the HIGH-severity operator assessment, repeat the relevant operational measurements, and review recent operating conditions and loading history. If the degraded condition is confirmed, corrective maintenance should subsequently be initiated based on the identified cause. For P.TGL-210B, the recommended response is targeted operator/engineer verification and repeated inspection because the pump remains within the WATCH class but has sufficient uncertainty to activate HITL. Thus, the framework converts uncertainty and operator knowledge into specific verification actions rather than directly prescribing component replacement without field confirmation.

For pumps classified as HEALTHY and AUTO_CONFIRMED, routine operation, daily condition recording, and scheduled preventive inspection remain appropriate. Pumps classified as WATCH without HITL activation should receive increased monitoring frequency and targeted inspection of the parameters contributing to the reduced Health Index. WATCH pumps that activate HITL require operator or engineer verification and

repeated measurement before corrective maintenance decisions are made. Pumps classified as DEGRADED require priority field inspection and confirmation of the underlying degradation before corrective maintenance, operating-condition adjustment, or component replacement is undertaken. Although no CRITICAL pump was observed during the study period, a CRITICAL classification would require immediate engineering assessment and, where equipment integrity or operational safety is at risk, restriction or suspension of operation followed by corrective maintenance before return to service. Accordingly, the HITL mechanism functions as both a classification safeguard and a maintenance-escalation trigger.

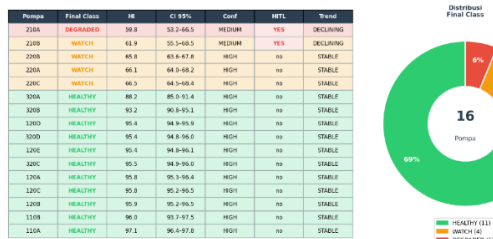


Figure 4. HITL Summary : Final Health Class Per Pompa

C. Monte Carlo and Bootstrap Validation Results

Tabel 9. Monte Carlo and Bootstrap Validation Results

Validation Metric	Result	Criterion	Interpretation
Monte Carlo iterations	10.000	-	Simulation basis
CV	-	<5%	Stable/Unstable
Spearman ρ	-	>0.90	Consistent/Not consistent
CSR P.TGL-210A	0%	$\geq 80\%$	Unstable; HITL activated
CSR P.TGL-210B	0%	$\geq 80\%$	Unstable; HITL activated
Bootstrap CI P.TGL-210A	53.17–66.51	-	Crosses WATCH–DEGRADED boundary
Bootstrap CI P.TGL-210B	55.47–68.47	-	Within WATCH

Table 9 summarizes the statistical validation results used to evaluate the robustness of the proposed Health Index classification. The Monte Carlo analysis was performed using 10,000 iterations to examine the effect of plausible variations in the model parameters

and fusion coefficient on the resulting Health Index and health-class assignment. The use of 10,000 iterations provides the basis for estimating classification stability and identifying cases in which small parameter variations may produce different health classes.

The coefficient of variation (CV) is used to evaluate the numerical stability of the simulated Health Index. A CV below 5% indicates that the variation of the simulated Health Index is relatively small compared with its mean, suggesting that the numerical output is stable. Conversely, a CV equal to or above 5% indicates greater variability and should be interpreted as evidence that the Health Index is more sensitive to the simulated parameter variations. Therefore, the CV result should be considered together with CSR and boundary proximity rather than as an isolated indicator of model reliability.

The Spearman rank correlation coefficient (ρ) evaluates whether the relative ranking of pumps remains consistent between the nominal model and the Monte Carlo results. A value of $\rho > 0.90$ indicates strong rank-order consistency, meaning that the relative maintenance priority of the pumps is not substantially altered by the simulated parameter variations. A lower value would indicate that the ranking is sensitive to parameter perturbations and that maintenance prioritization should be interpreted with greater caution. The reported CV and Spearman results should therefore be used to assess the overall numerical and ranking robustness of the model.

The most critical finding concerns the Classification Stability Rate (CSR) of P.TGL-210A and P.TGL-210B. Both pumps obtained a CSR of 0%, substantially below the predefined threshold of 80%. This means that, under the 10,000 Monte Carlo simulations, neither pump retained its nominal health class in the simulated classification results. The result indicates a high degree of classification instability and directly supports activation of the HITL mechanism. In practical terms, the automated classification for these two pumps should not be considered sufficiently reliable without additional operator or engineer verification.

The bootstrap confidence intervals provide additional evidence regarding the source of this classification ambiguity. P.TGL-210A has a 95% bootstrap confidence interval of 53.17–66.51. Because the interval crosses the WATCH–DEGRADED boundary at 55 points, the possible Health Index values span two health classes. This indicates that the classification of P.TGL-210A is statistically uncertain and provides the basis for the HITL_DOWNGRADED decision. In accordance with the predefined conservative FHC rule, the lower health class, DEGRADED, is assigned when the confidence interval crosses the class boundary.

In contrast, P.TGL-210B has a 95% bootstrap confidence interval of 55.47–68.47. Although this

pump also activates HITL because of its classification instability and boundary proximity, the entire confidence interval remains above the 55-point boundary and therefore remains within the WATCH class. Consequently, the nominal WATCH classification is retained through the HITL_CONFIRMED_SINGLE_CLASS pathway. The distinction between P.TGL-210A and P.TGL-210B demonstrates the importance of combining Monte Carlo classification stability with bootstrap confidence intervals: two pumps may both require human verification, but the final maintenance decision can differ depending on whether the uncertainty interval crosses a health-class boundary.

Overall, the validation results demonstrate that the proposed framework does not rely solely on the nominal Health Index value. Monte Carlo simulation evaluates the stability of the classification under parameter perturbations, while bootstrap resampling evaluates uncertainty in the observed Health Index. Boundary proximity and confidence-interval crossing then provide an explicit link between statistical uncertainty and the HITL decision. This combination enables the framework to distinguish between a pump that requires verification but can retain its nominal class, such as P.TGL-210B, and a pump whose uncertainty warrants conservative downgrading and higher maintenance priority, such as P.TGL-210A.

Table 10. Maintenance Response Recommendations Based on FHC and HITL Status

Health/FHC Status	HITL Condition	Recommended Operator/Engineer Response	Maintenance Priority
HEALTHY	HITL = No	Continue routine operation and scheduled preventive inspection; maintain daily operational recording.	Routine
WATCH	HITL = No	Increase monitoring frequency and perform targeted inspection of the parameters contributing to the reduced HI.	Moderate
WATCH	HITL = Yes; CI remains within	Conduct operator/engineer verification, repeat relevant	Moderate –High

	WATC H	measurements, and review recent operating trends before making corrective-maintenance decisions.	
DEGRADED	CI crosses	Conduct priority field re-inspection, verify the source of degradation or operator-reported severity, evaluate operating/load conditions, and initiate corrective maintenance if degradation is confirmed.	High

For the observed dataset, P.TGL-210A receives the highest maintenance priority because its nominal WATCH classification was downgraded to DEGRADED after the bootstrap confidence interval crossed the WATCH–DEGRADED boundary. The appropriate initial response is priority field re-inspection and verification of the HIGH-severity operator assessment, followed by repeated measurement and investigation of the operating conditions. Corrective maintenance or component replacement should only be initiated when the degraded condition and its underlying cause are confirmed. P.TGL-210B remains in the WATCH class because its bootstrap confidence interval remains entirely within the WATCH range. Nevertheless, its HITL activation indicates that additional operator/engineer verification and targeted re-inspection are required before maintenance decisions are finalized. The remaining HEALTHY and AUTO_CONFIRMED pumps can remain under routine monitoring and scheduled preventive maintenance.

IV. CONCLUSION

This study developed and evaluated a condition-based maintenance framework that integrates standard-based Numeric Health Score (NHS), operator-based Ordinal Utility Model (OUM), Monte Carlo robustness analysis, bootstrap uncertainty estimation, and Human-in-the-Loop (HITL) gating for centrifugal pump condition assessment in an industrial fuel-terminal environment. The results demonstrate that combining quantitative operational measurements with structured operator

assessments provides a more comprehensive basis for health classification than relying on measurement-based scoring alone. In particular, the OUM component allows contextual field information that may not be reflected in the measured operational parameters to contribute to the final Health Index.

The proposed framework also provides a decision mechanism that links statistical uncertainty to maintenance action. Instead of treating the Health Index classification as an absolute automated decision, the HITL mechanism identifies cases in which classification stability, confidence, or proximity to a health-class boundary indicates insufficient certainty. The cases of P.TGL-210A and P.TGL-210B illustrate this distinction. Both pumps required human verification, but only P.TGL-210A was conservatively downgraded because its bootstrap confidence interval crossed the WATCH–DEGRADED boundary. This demonstrates that the framework can differentiate between classification uncertainty that requires verification and uncertainty that warrants escalation of maintenance priority.

From an operational perspective, the framework provides a direct connection between health assessment and maintenance response. Stable HEALTHY pumps can remain under routine monitoring and preventive maintenance, whereas WATCH pumps can be subjected to increased monitoring and targeted re-inspection. Pumps that trigger HITL require operator or engineer verification, while a DEGRADED classification resulting from boundary-crossing uncertainty can be used to prioritize field inspection and corrective maintenance. Therefore, the proposed approach supports condition-based maintenance by reducing unnecessary interventions for stable equipment while directing maintenance resources toward pumps with greater evidence of deterioration or classification ambiguity.

Nevertheless, several limitations should be considered. First, the analysis was based on 30 days of operational records from a single fuel-terminal environment and therefore may not fully represent long-term equipment degradation patterns or operating conditions at other facilities. Second, several OUM inputs were derived from manual operator records and qualitative assessments. Consequently, the reliability of the OUM and the resulting Health Index is dependent on the consistency, completeness, and objectivity of operator observations. Differences in operator experience or recording practices may introduce subjectivity that cannot be completely eliminated through the current scoring mechanism. Third, the selected AHP pairwise comparisons, the fusion coefficient of $\alpha = 0.65$, and the HITL thresholds represent the baseline configuration of this study and require further validation across different equipment populations and operating environments. Finally, the absence of a longer historical record limits the ability of the framework to establish long-term predictive relationships between Health Index deterioration and actual equipment failure or maintenance events.

Future research should therefore extend the observation period and include multiple fuel terminals

or industrial facilities to evaluate the transferability and generalizability of the framework. Longitudinal data should be used to investigate Health Index trajectories and their relationship with actual failure, repair, component replacement, and maintenance records. The consistency of operator-based OUM assessments should also be evaluated through inter-operator agreement analysis and standardized digital inspection forms to reduce subjectivity in manual data collection. In addition, future work should investigate the integration of additional real-time or semi-automated sensors where available, while retaining the proposed manual-data framework for environments with limited sensing infrastructure. Further validation of the AHP weights, fusion coefficient, and HITL thresholds using larger datasets and independent operational sites is also recommended. These extensions would strengthen the evidence for deploying the framework as a scalable decision-support tool for condition-based maintenance in oil and gas and other process industries.

V. ACKNOWLEDGMENTS

Author extends their gratitude to Ronggolawe University of Technology for providing the opportunity to collaborate with the State University of Jakarta, which contributed to the writing of this article.

REFERENCES

- [1] C. Kalmár and F. Hegedűs, “Condition Monitoring of Centrifugal Pumps Based on Pressure Measurements,” no. 1, pp. 1–10.
- [2] I. Adabara and Y. I. Danjuma, “Trustworthy and Ethical AI for Intrusion Detection in Healthcare IoT (IoMT) Systems: An Agentic Decision Loop Framework,” 2026.
- [3] H. Wang, H. Wang, and X. Tang, “A Review of Deep Learning in Rotating Machinery Fault Diagnosis and Its Prospects for Port Applications,” pp. 1–47, 2025.
- [4] A. Khedkar, S. Sahare, and S. Musale, “AI Driven Predictive Monitoring and Optimization of Decentralized Water Treatment Infrastructure in Resource Constrained Urban Settlements,” vol. 44, no. 1, pp. 104–118.
- [5] H. Adaika, Z. Tir, and M. Sahraoui, “PumpSpectra: An MCSA-Based Platform for Fault Detection in Centrifugal Pump Systems,” pp. 1–20, 2025.
- [6] M. Adil *et al.*, “When Healthcare IoT Applications Meet Agentic AI: Emerging Security and Privacy Challenges,” no. February, 2026.
- [7] R. urunen, T., Miettinen, J., Hämäläinen, A., Karhinen, A., Viitala, “Deep Learning for Centrifugal Pump Condition Monitoring Using Data from Variable Frequency Drive. In: Okada, M. (eds) *Advances in Mechanism and Machine Science. IFToMM WC 2023*,” vol. 147, p. 45705, 2023.

- [8] S. Pardo-pina, J. Germer, R. Suay-cortés, and M. Platero-horcajadas, "Smart Agricultural Technology An agent-Based service architecture for smart greenhouses : Telemetry analytics and decision support with RAG-grounded LLM agents," *Smart Agric. Technol.*, vol. 13, no. December 2025, p. 101872, 2026, doi: 10.1016/j.atech.2026.101872.
- [9] A. Kiswantono, A. F. Saputra, P. S. Elektro, F. Teknik, and U. B. Surabaya, "SISTEM MONITORING DAN PROTEKSI MOTOR MINI CONVEYOR TERHADAP ANOMALI ARUS DAN," vol. 13, no. 2, 2025.
- [10] Q. Zeng *et al.*, "Materials and System Design for Self-Decision Bioelectronic Systems."
- [11] B. M. Dinata *et al.*, "PEMODELAN HYBRID SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGBORS UNTUK PERAMALAN PERMINTAAN SUKU CADANG OTOMOTIF: PENDEKATAN MACHINE LEARNING TERHADAP DATA FLUKTUATIF," vol. 14, no. 1, pp. 178–188, 2026.
- [12] S. Pardo-Pina, J. G. R. S.-C. Platero-Horcajadas, Manuel Platero-Horcajada, Francisco-Javier, and Ferrández-Pastor, "An agent-Based service architecture for smart greenhouses: Telemetry analytics and decision support with RAG-grounded LLM agents," 2026.
- [13] F. R. Mukhnizar, Risal Abu, "ANALYSIS OF OVERHEAT OF PERTAMAX FUEL DISTRIBUTION PUMP FOR MAINTENANCE ACTION AT PT. PERTAMINA PATRA NIAGA-FUEL TERMINAL MEDAN GROUP," vol. 6, no. 1, pp. 1610–1626, 2024.
- [14] S. Sardar *et al.*, "Artificial intelligence for antimicrobial resistance : advancing reproducibility , interpretability , and clinical deployment," no. May, 2026.
- [15] Z. F. Bagas Dwi Pangestu, "PERANCANGAN TANGKI TIMBUN CRUDE OIL DENGAN WORKING VOLUME 5.372 BARREL BERDASARKAN STANDAR API 650 DI FASILITAS PRODUKSI LAPANGAN PT X," vol. 5, no. November, pp. 324–333, 2025.
- [16] E. O. Ibam and Ayobami Emmanuel Mesioye, "Decoupling Cyber Sabotage in Industrial Telemetry from Mechanical Faults Using Gradient-Based Forensic Edge XAI," 2026.
- [17] S. Rahyadi, D. Desmira, P. Vokasional, T. Elektro, and F. Keguruan, "Analisis Preventive Maintenance Motor Condensate Extraction Pump 10 , 5kv Unit 6 di PT PLN Indonesia Power UBP Suralaya," vol. 12, no. 2, pp. 354–362, 2025.
- [18] M. A, Nishat Tasneem Meem, M. Masud, Shamim Ahmed Kamal Uddin Khan, Md. Shirajum Munir, and A. K. Bairagi, "Toward sustainable hydroponic farming: An AI-driven IoT framework for virtual pH sensing and sensor lifespan extension".
- [19] I. Nasri, M. Mansouri, and S. Member, "A Review on Multimodal-Based Fault Detection and Diagnosis in Photovoltaic Systems : Fusion and Attention Mechanisms," *IEEE Access*, vol. 14, no. February, pp. 44382–44402, 2026, doi: 10.1109/ACCESS.2026.3675686.
- [20] P. D. R. R. B. W. C. Kravaris, "Model-based fault diagnosis and fault tolerant control in safety-critical chemical reactors: An experimental study."
- [21] F. Rachmanu, "PEMANTAUAN KONDISI POMPA SENTRIFUGAL P-12A MENGGUNAKAN ANALISIS VIBRASI STUDI KASUS DI PT. X," vol. 2, no. 1, pp. 14–19, 2022.
- [22] A. Naug *et al.*, "LC-Opt: Benchmarking Reinforcement Learning and Agentic AI for End-to-End Liquid Cooling Optimization in Data Centers ' r," no. LC, pp. 1–50, 2025.
- [23] K. Besigomwe, "Human - in - the - Loop Self - Healing Systems: Integrating Human Oversight for Autonomous Failure Detection , Repair and System Optimization," vol. 5, no. 3, pp. 254–267, 2025, doi: 10.47760/cognizance.2025.v05i03.020.
- [24] H. Wan, X. Lu, Y. Chen, K. Devaprasad, and L. Hinkle, "Automating Modelica Module Generation Using Large Language Models : A Case Study on Building Control Description Language".
- [25] M. C. N. Benjamin Baasch, Jorn C Groos, Judith Heusel, "Data-driven sparse coding for onboard condition monitoring of railway tracks."
- [26] N. Baddou, afaf dadda, Bouchra RZINE, and hala hmamed, "A Data-Driven Human-in-the-Loop Framework for Carbon Reduction: Enhancing Fault Diagnosis and Energy Performance Monitoring in Industry 5.0 with Large Language Model and Deep reinforcement learning," 2025.