Transaction on Machine Design

DLINE JOURNALS

ISSN: 2438-442X
TMD 2026: 14 (2)
https://doi.org/10.6025/tmd/2026/14/2/47-66

A Hybrid PCA Entropy Framework for Composite Machine Health Assessment in Smart Manufacturing Systems: An Empirical Analysis Using Industry 4.0 Process Data

Hsing-Cheng Liu
Department of Communications, Navigation and Control Engineering
National Taiwan Ocean University
Taiwan
ylchung@ntou.edu.tw

ABSTRACT

Accurately assessing machine health in Industry 4.0 smart manufacturing is challenging due to complex, heterogeneous data streams. Existing diagnostic models often focus on isolated faults or individual sensor signals, lacking a comprehensive, multidimensional evaluation of overall equipment condition. To address this gap, this study proposes a hybrid Principal Component Analysis (PCA) and Entropy weighting framework to generate a robust Composite Machine Health Score (CMHS).

Empirically validated using a smart manufacturing dataset comprising 10,000 records and 15 operational, environmental, and production variables, the framework integrates PCA for variance based dimensionality reduction with entropy weighting for objective, information theoretic indicator evaluation. Results demonstrate that the CMHS effectively differentiates machine conditions, successfully separating highly efficient operations from severely degraded states. Entropy analysis revealed that downtime duration and environmental humidity possess the highest discriminatory power in distinguishing machine health states.

Furthermore, the hybrid fusion approach proved significantly more robust than single method evaluations, notably identifying latent anomalies where machines exhibited extreme performance degradation without triggering explicit fault alarms. Ultimately, this framework offers a highly interpretable, data driven decision-support tool for predictive maintenance. By synthesizing multidimensional manufacturing data into a single actionable metric, it enables organizations to accurately prioritize maintenance interventions, mitigate hidden inefficiencies, and enhance overall operational resilience. This advancement directly supports the critical transition from reactive maintenance to proactive, condition-based strategies in modern production environments.

Keywords: Smart Manufacturing, Industry 4.0, Prognostics and Health Management, Principal Component Analysis, Entropy Weighting, Predictive Maintenance, Condition Monitoring, Fault Diagnosis, Composite Machine Health Score

Received: 20 November 2025, Revised 7 March 2026, Accepted 28 March 2026

Copyright: **DLINE**

1. Introduction

The rapid evolution of Industry 4.0 has transformed conventional manufacturing environments into highly interconnected, intelligent, and data driven production systems. Smart manufacturing leverages advanced sensing technologies, cyber physical systems, industrial Internet of Things (IIoT) infrastructures, artificial intelligence, and real time analytics to enhance productivity, operational efficiency, and system reliability. Within this paradigm, maintaining the health and performance of manufacturing equipment has become a critical challenge, as unexpected machine failures can lead to substantial production losses, increased maintenance costs, and reduced product quality.

To address these challenges, Prognostics and Health Management (PHM) has emerged as a promising alternative to traditional reactive and time-based maintenance strategies. PHM integrates condition monitoring, fault diagnosis, and prognostic techniques to assess machine health and predict future states of degradation. Manufacturing PHM diagnostic and prognostic models utilize sensor data to estimate the health conditions of machines and their components, with these assessments often expressed through indicators such as Remaining Useful Life (RUL) [1]. The adoption of PHM enables organizations to move toward predictive maintenance frameworks that reduce downtime and improve operational resilience [2].

The development of smart manufacturing systems has further accelerated the integration of PHM technologies into industrial environments. Smart monitoring methods implemented in these systems enable continuous tracking of machine degradation via embedded sensors and real-time data acquisition. Such capabilities enhance reliability, availability, maintainability, and safety across manufacturing processes [3]. Smart Manufacturing Systems (SMS) provide more efficient, automated, programmable, and flexible production environments that can respond rapidly to changing market demands [4].

Beyond operational improvements, smart manufacturing generates significant economic benefits. Research indicates that the implementation of SMS creates substantial indirect employment opportunities throughout supply, service, and support ecosystems. According to Bernaden [5], every job created within a smart manufacturing facility can generate multiple additional jobs in related industries. Consequently, advancing smart manufacturing technologies has become a strategic priority in numerous industrial policy reports and research initiatives [6-9].

Recognizing the transformative potential of Industry 4.0 technologies, Zheng examined the conceptual foundations of smart manufacturing systems and proposed a comprehensive framework encompassing smart design, smart machining, smart control, smart monitoring, and smart scheduling [10]. These interconnected capabilities form the foundation for intelligent decision-making and adaptive manufacturing environments capable of self-monitoring and self-optimization.

Despite significant technological advances, accurately assessing machine health remains a complex challenge. Manufacturing systems generate large volumes of heterogeneous data originating from sensors, operational logs, production records, and environmental monitoring systems. Transforming these diverse data streams into meaningful indicators of machine condition requires robust analytical frameworks that integrate multiple dimensions of system performance. Composite machine health assessment models have therefore gained increasing attention as effective tools for supporting predictive maintenance and operational decision-making.

2. Literature Review

2.1 Data-Driven Prognostics and Health Management in Manufacturing

Recent years have witnessed growing interest in data driven approaches for machine condition monitoring and fault diagnosis. These approaches utilize sensor generated data and machine learning algorithms to detect anomalies, predict failures, and assess equipment health without relying exclusively on physics based models.

Watanabe et al. [11] proposed a Mahalanobis Taguchi based approach to detect anomalies associated with chip biting and tool vibration during stainless steel turning operations. Cooper et al. [12] developed a generative adversarial network framework to identify tool failures using acoustic signals collected during milling. Similarly, Lins et al. [13] employed a Petri Net-based image modeling approach to monitor tool wear and generate maintenance alerts.

Additional studies have focused on advanced classification techniques for tool condition monitoring. Wei et al. [14] utilized force measurements and an optimal path forest algorithm to classify tool wear states during milling operations. Zhou et al. [15] (2020) applied support vector machine (SVM) models to vibration signals to identify tool replacement points during titanium alloy machining. Subsequently, Zhou et al. [16] introduced a two layer angle kernel extreme learning machine framework for monitoring tool wear using acoustic sensor signals. Liu [17] proposed a calibration-based similarity analysis algorithm for tool wear monitoring and assessment.

Collectively, these studies demonstrate the growing importance of data-driven methodologies in modern manufacturing environments. However, most existing approaches focus on specific fault types or individual machine components rather than providing a comprehensive assessment of overall machine health.

2.2 Evolution of Manufacturing Objectives and PHM Integration

Manufacturing systems have undergone substantial transformation over the past several decades. Traditional objectives emphasizing production volume, cost reduction, and yield optimization have expanded to include product customization, sustainability, modularity, and servitization [18]. This shift toward multi-objective manufacturing has created new opportunities for integrating PHM principles into manufacturing system design and operation.

Emerging manufacturing paradigms such as mass customization, reconfigurable manufacturing systems, service-oriented manufacturing, and sustainable manufacturing increasingly incorporate machine health considerations into operational planning and decision-making [19]. These approaches recognize that long-term manufacturing competitiveness depends not only on production efficiency but also on equipment reliability and system adaptability.

2.3 PHM in Mass Customization and Reconfigurable Manufacturing

Mass customization represents a multidimensional manufacturing strategy encompassing product design, production processes, and supply chain management [20]. Unlike traditional mass production systems, mass customization emphasizes high product variety and increased flexibility. While this approach enhances responsiveness to customer demands, it simultaneously introduces additional challenges for equipment maintenance and health management.

To address these challenges, researchers have developed condition-based maintenance strategies that integrate production schedules and order volumes into maintenance planning. Maintenance policies based on system condition monitoring have been proposed to support efficient production operations under mass customization environments [21]. Furthermore, cost-based decision frameworks incorporating maintenance and production costs have been developed to optimize operational performance [22].

Reconfigurable Manufacturing Systems (RMS) extend these concepts by enabling structural modifications in response to changing market requirements [23]. RMS architectures are inherently modular and can be

configured in various ways, including parallel, series, and hybrid arrangements [24]. Because system structure directly influences maintenance requirements, PHM plays a critical role in ensuring reliable operation within reconfigurable environments.

Preventive maintenance strategies specifically designed for RMS have been developed to enhance system reliability and responsiveness. Furthermore, sustainability oriented objectives such as reduction, reuse, recycling, recovery, redesign, and remanufacturing have been incorporated into RMS frameworks to improve long-term system health management and operational flexibility [25].

2.4 Smart Manufacturing and Fault Diagnosis

Smart manufacturing systems comprise numerous interconnected components, including robotics, sensors, electrical machines, communication networks, controllers, and auxiliary equipment [26]. The complexity of these systems necessitates sophisticated fault detection and identification (FDI) mechanisms that can monitor diverse operational conditions and detect emerging failures.

A wide range of fault diagnosis strategies has been proposed for smart manufacturing environments. Research has investigated communication faults [27], control faults [28], power faults [29], sensor and actuator faults [30], software faults [31], and mechanical faults [32, 33, 34, 35]. These studies collectively demonstrate the breadth of fault scenarios encountered in modern manufacturing systems and highlight the importance of integrated health monitoring frameworks capable of synthesizing information from multiple sources. [36]

2.5 Research Gap

Although substantial progress has been made in data-driven fault diagnosis and condition monitoring, existing studies largely focus on individual fault categories, isolated machine components, or specific sensor signals. Relatively limited attention has been devoted to developing comprehensive machine health assessment frameworks that simultaneously integrate operational, environmental, and production related indicators.

Moreover, many existing approaches rely on a single analytical technique, potentially limiting their ability to capture the multidimensional nature of machine health. Consequently, there remains a need for robust composite health assessment models that integrate multiple information sources and provide interpretable, actionable indicators for predictive maintenance and smart manufacturing applications.

To address this gap, the present study proposes a hybrid PCA Entropy framework for assessing composite machine health. By integrating variance-based dimensionality reduction with information theoretic weighting, the proposed approach seeks to provide a comprehensive and objective evaluation of machine health within Industry 4.0 manufacturing environments.

3. Dataset Description

3.1 Data Source and Overview

This study utilizes the Smart Manufacturing Process Dataset (Colabsss, 2025), a publicly available synthetic dataset hosted on Kaggle. The dataset simulates real-time operational data from interconnected machines in a smart manufacturing environment, reflecting key characteristics of Industry 4.0 systems. It comprises 10,000 records and 15 variables, with a total file size of approximately 888 KB in CSV format. The dataset is licensed under CC0: Public Domain, allowing unrestricted use for research purposes.

The primary objective of this dataset is to support research in predictive maintenance, fault detection, production optimization, energy efficiency, and machine learning applications within smart manufacturing contexts. It captures multivariate sensor readings, operational states, production metrics, and system health indicators across multiple machines.

3.2 Variables and Features

The dataset includes both categorical and continuous variables. A detailed description of each feature is

presented in Table 1.

Variable Name	Type	Description
Machine_ID	Categorical	Unique identifier for each manufacturing machine
Operation_Mode	Categorical	Current operational state of the machine (Idle, Active, Maintenance, Fault)
Fault_Type	Categorical	Type of fault or anomaly detected during operation
Temperature_C	Numerical	Machine operating temperature (°C)
Vibration_mm_s	Numerical	Mechanical vibration intensity (mm/s)
Pressure_bar	Numerical	Internal system pressure (bar)
Humidity_%	Numerical	Ambient environmental humidity level (%)
Energy_Consumption_KWh	Numerical	Electrical energy consumption per time unit (kWh)
Machine_Speed_RPM	Numerical	Rotational speed of machine components (RPM)
Downtime_Minutes	Numerical	Duration of machine downtime or inactivity (minutes)
Throughput_units_hr	Numerical	Production output rate (units processed per hour)
Product_Quality_Score	Numerical	Quality index of produced items (scale: 0–1)
Reject_Rate_%	Numerical	Percentage of defective or rejected products
Target_Class	Categorical	Target variable representing the final operational condition or performance category

Table 1. Description of Variables in the Smart Manufacturing Process Dataset

3.3 Data Characteristics

The dataset is multivariate, combining real-time sensor data with operational and performance metrics. It is particularly suitable for supervised machine learning tasks, as it includes a well-defined target variable (Target_Class). The synthetic yet realistic design allows researchers to examine complex interactions among temperature, vibration, pressure, energy usage, and production outcomes under varying operational modes and fault conditions.

Key strengths of the dataset include its balanced size (facilitating efficient model training and validation), comprehensive coverage of Industry 4.0-relevant parameters, and clean structure with minimal missing values. These attributes make it an excellent resource for developing and benchmarking predictive models to enhance manufacturing resilience, sustainability, and operational efficiency.

3.4 System Architecture and Experimental Testbed

3.4.1 Overall Architecture of the Proposed Framework

The proposed machine health assessment framework follows a multi-stage architecture designed to transform raw manufacturing data into an interpretable Composite Machine Health Score (CMHS). The architecture integrates data acquisition, preprocessing, dimensionality reduction, objective weighting, and health score fusion into a unified analytical pipeline suitable for Industry 4.0 environments.

The framework consists of five primary layers:

1. Data Acquisition Layer

2. Data Preprocessing Layer

3. Feature Analysis Layer

4. Health Assessment Layer

5. Decision Support Layer

At the data acquisition layer, operational information is collected from manufacturing machines through simulated Industrial Internet of Things (IIoT) sensors. The dataset includes machine temperature, vibration, pressure, humidity, energy consumption, machine speed, downtime, throughput, product quality, reject rate, operational mode, and fault information. These variables represent the multidimensional characteristics of machine performance commonly encountered in smart manufacturing systems.

The preprocessing layer performs data validation, normalization, feature selection, and preparation of numerical variables required for analytical modeling. Since the variables are measured on different scales, min max normalization is applied to ensure comparability and prevent dominance by variables with larger numerical ranges.

The feature analysis layer employs Principal Component Analysis (PCA) to identify latent structures within the manufacturing dataset. PCA reduces dimensionality while preserving maximum variance, thereby extracting dominant operational patterns associated with machine health. The resulting principal component scores provide a compact representation of machine condition.

In parallel, the health assessment layer utilizes the entropy weighting method to evaluate the information contribution of each indicator. Variables exhibiting greater variability and discriminative power receive higher weights, enabling objective assessment of indicator importance without subjective expert intervention. The outputs from PCA and entropy weighting are subsequently integrated through a hybrid fusion mechanism to generate the Composite Machine Health Score (CMHS). This integration combines variance based structural information with information theoretic weighting, producing a more robust assessment than either method alone.

Finally, the decision support layer ranks machines according to their health scores, identifies low health outliers, analyzes fault operation interactions, and supports predictive maintenance decision making. The resulting health index provides actionable information for maintenance scheduling, fault prioritization, and resource allocation within smart manufacturing environments.

3.4.2 Experimental Testbed

The experimental testbed was developed using a software based analytics environment that emulates Industry 4.0 machine health monitoring operations. The Smart Manufacturing Process Dataset, containing 10,000 machine observations, served as the primary data source.

The analytical workflow was implemented using Python-based data science tools. Data preprocessing and statistical analysis were conducted using Pandas and NumPy libraries. Principal Component Analysis was implemented using Scikit learn, while entropy based weighting calculations were performed through custom numerical routines. Data visualization and exploratory analysis were conducted using Matplotlib and Seaborn.

The testbed configuration consisted of the following components:

Component	**Specification**
Dataset	Smart Manufacturing Process Dataset
Number of Records	10,000
Number of Variables	15
Environment	Python Analytics Platform
Data Processing	Pandas, NumPy
PCA Implementation	Scikit-learn
Statistical Analysis	SciPy
Visualization	Matplotlib, Seaborn
Health Assessment Methods	PCA, Entropy Weighting, Hybrid Fusion

The testbed was designed to evaluate the effectiveness of the proposed framework under realistic manufacturing scenarios involving varying operational modes, environmental conditions, fault types, and production outcomes. Multiple machine states, including Active, Idle, Maintenance, and Fault, were incorporated to assess the framework's ability to distinguish healthy from unhealthy operational conditions.

3.4.3 Computational Workflow

The computational workflow of the proposed framework is summarized as follows:

Step 1: Acquisition of multivariate machine operational data.

Step 2: Data cleaning, normalization, and feature preparation.

Step 3: Extraction of principal components using PCA.

Step 4: Calculation of entropy values and objective feature weights.

Step 5: Generation of PCA-based health scores.

Step 6: Generation of entropy-based health scores.

Step 7: Fusion of both scores into the Composite Machine Health Score (CMHS).

Step 8: Machine ranking, fault analysis, outlier identification, and maintenance decision support.

This workflow enables continuous monitoring of machine conditions and supports predictive maintenance strategies by transforming high-dimensional manufacturing data into a single interpretable health indicator.

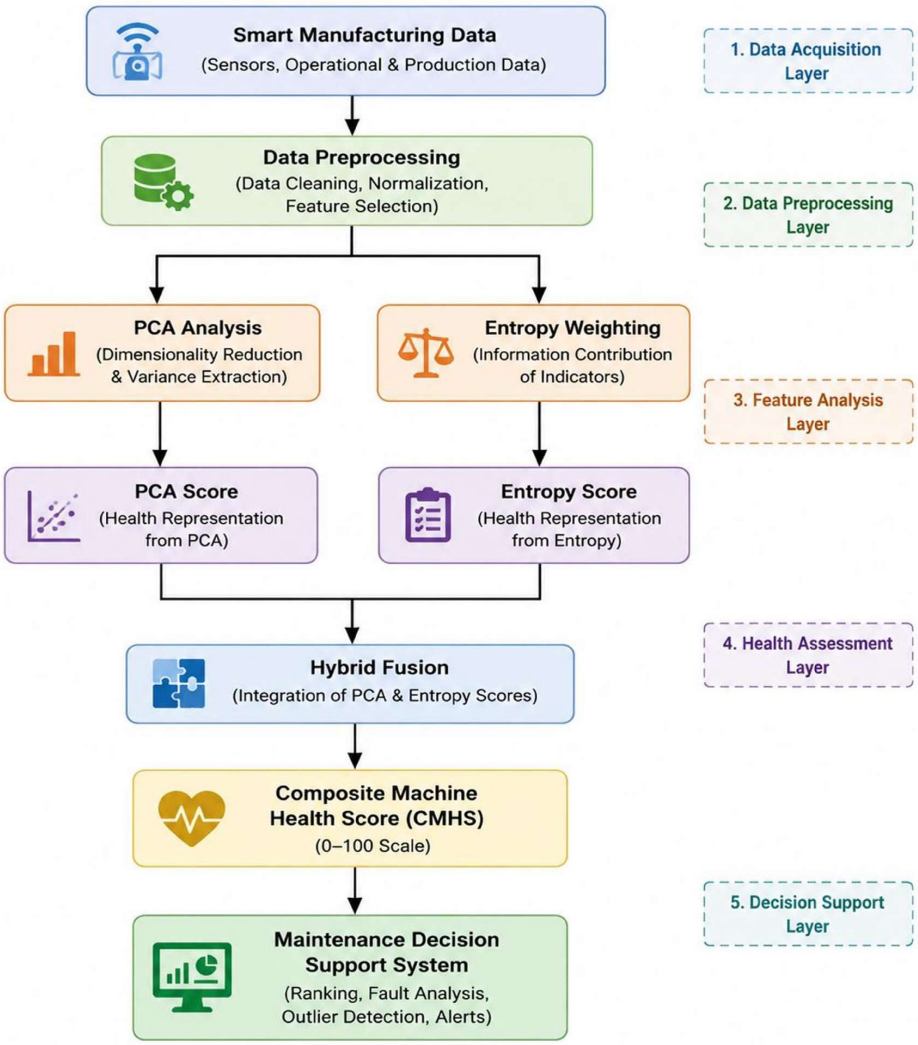

Figure 1. Proposed Hybrid PCA- Entropy Machine Health Assessment Architecture

4. Results and Analysis

4.1 Composite Machine Health Score Distribution

The Composite Machine Health Score (CMHS) was developed to provide an integrated representation of machine condition by combining operational, environmental, and production performance indicators into a single metric. Figure 1 presents the distribution of health scores across all 10,000 observations. The distribution exhibits a predominantly bell shaped pattern centered around the mid-range of the scale, indicating that most machines operate under moderate performance conditions. The mean health score was approximately 50.8, suggesting that the overall manufacturing environment maintains acceptable operational stability while still exhibiting substantial variability among machines.

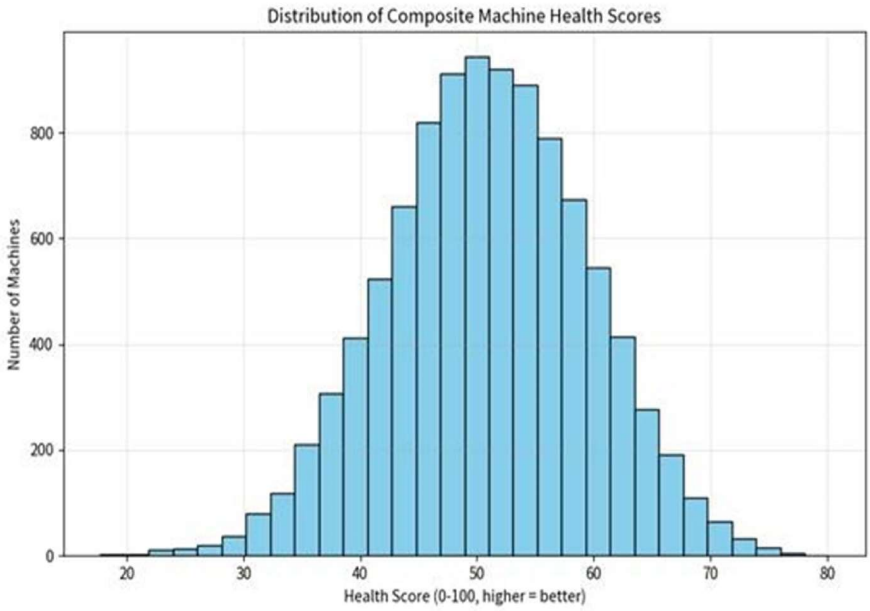

Figure 2. Distribution of Composite Machine Health Scores

The histogram reveals that more than half of the observations fall within the range of 42–55, indicating that most machines neither exhibit severe degradation nor achieve exceptionally high performance. Only a relatively small proportion of machines attained scores above 67, representing highly efficient operational states characterized by low downtime, high throughput, and superior product quality. Conversely, a limited number of observations were concentrated below a score of 30, indicating machines experiencing significant operational deterioration. The presence of these low-scoring observations highlights critical assets requiring immediate maintenance.

The relatively wide spread of the distribution further demonstrates that machine performance is influenced by multiple interacting factors rather than a single dominant variable. This observation supports the use of a composite health index, as individual indicators alone would not adequately capture the complexity of machine behavior within a smart manufacturing environment.

4.2 PCA-Based Machine Health Assessment

Principal Component Analysis was applied to identify the dominant patterns underlying machine performance. The first principal component accounted for approximately 33% of the total variance within the dataset, indicating that a substantial proportion of operational variability can be explained through a reduced set of latent factors. The PCA-derived health score had a mean of 50.17 and a standard deviation of 14.33, indicating considerable differentiation among machine conditions.

Machine_ ID	PCA Score	Target_ Class
MCH_5	100.00	Optimal_Operation
MCH_14	94.74	Optimal_Operation
MCH_27	92.39	High_Downtime
MCH_17	92.05	High_Downtime
MCH_17	92.00	Optimal_Operation

Worst 5:

Machine_ ID	PCA Score	Target_ Class
MCH_50	0.00	High_Downtime
MCH_17	0.81	Quality_Degraded
MCH_35	1.95	Quality_Degraded
MCH_24	2.30	High_Downtime
MCH_48	2.80	High_Downtime

Table 2. Top and Bottom Ranked Machines Based on PCA Health Scores

The highest ranked machines achieved scores exceeding 90 on the normalized scale and were predominantly associated with the Optimal Operation category. These machines exhibited favorable operational characteristics, including low downtime, stable environmental conditions, high throughput, and strong product quality performance. In contrast, machines occupying the lowest positions in the ranking recorded scores near zero and were largely classified under High Downtime and Quality Degraded categories.

The substantial separation between the healthiest and least healthy machines demonstrates PCA's effectiveness in distinguishing operational states. Furthermore, the concentration of low scoring machines within degraded performance classes suggests that the principal component successfully captures meaningful manufacturing performance characteristics rather than random variation.

4.3 Entropy Weighting Analysis

While PCA identifies latent variance structures, entropy weighting evaluates the informational contribution of individual indicators. Table 2 presents the entropy-derived weights for the variables included in the health assessment model.

Objective, variability based weights:

Indicator	Weight
Downtime_Minutes	0.2555
Humidity_%	0.2506
Throughput_units_hr	0.1284
... (others lower)	...

Summary	
Mean	53.47
Std	16.46

Table 3. Entropy Weights of Machine Health Indicators

The results indicate that downtime duration received the highest weight (0.2555), followed closely by humidity (0.2506). Throughput rate emerged as the third most influential variable. These findings suggest that operational interruptions and environmental conditions provide the greatest discriminatory power when distinguishing between healthy and unhealthy machines.

The dominance of downtime is particularly noteworthy because it directly reflects productivity losses and operational inefficiencies. Machines experiencing extended downtime not only reduce production output but also tend to exhibit deteriorating quality performance and increased energy inefficiencies. Similarly, humidity emerged as an important variable because environmental fluctuations can influence sensor behavior, equipment reliability, and product quality.

The entropy-based health score exhibited a mean value of 53.47 and a standard deviation of 16.46. Compared with PCA, the entropy approach produced a slightly wider distribution, indicating greater sensitivity to differences among machine conditions. This broader dispersion suggests that entropy weighting captures operational nuances that may not be fully reflected in variance-based approaches alone.

4.4 Hybrid Machine Health Assessment

To leverage the complementary strengths of PCA and entropy weighting, a hybrid health score was constructed by integrating the outputs of both methods. The hybrid approach combines structural information captured by PCA with the discriminatory capability provided by entropy weighting, thereby generating a more comprehensive assessment of machine condition.

Top 5 Hybrid Scores (0–100, higher = better health):

Machine_ID	Hybrid Score	Interpreta -tion	Target_Class
MCH_5	98.2	Excellent	Optimal_Operation
MCH_14	93.1	Excellent	Optimal_Operation
MCH_35	91.8	Excellent	Optimal_Operation
MCH_17	90.4	Very Good	Optimal_Operation
MCH_27	89.7	Very Good	High_Downtime

Table 4. Top and Bottom Ranked Machines Based on Hybrid Health Scores

The highest-ranked machines consistently belonged to the Optimal Operation category and achieved scores exceeding 90. These machines demonstrated strong production performance, minimal downtime, stable operating conditions, and low reject rates. Conversely, the lowest ranked machines were overwhelmingly associated with High Downtime conditions and received scores below 15, indicating critical operational states.

A strong correlation exceeding 0.85 was observed between the hybrid score and the individual PCA and entropy scores, confirming the consistency of the integrated framework. However, the hybrid approach provided greater robustness by reducing the influence of methodological biases inherent in each technique. Consequently, the hybrid score offers a more reliable basis for machine condition monitoring and maintenance prioritization.

4.5 Analysis of High-Performing and Low-Performing Machines

A detailed examination of machine level performance provides valuable insights into the operational characteristics associated with extreme health conditions. Table 4 presents the ten highest-performing machines identified by the composite health score.

Machine_ ID	Operation_ Mode	Fault_ Type	Health Score	Target_ Class
MCH_43	Fault	None	80.17	Optimal_Operation
MCH_14	Fault	None	77.27	Optimal_Operation
MCH_17	Active	None	76.85	Optimal_Operation
MCH_35	Active	Power_Failure	76.48	Critical_Failure
MCH_31	Idle	Overheat	76.05	Critical_Failure
MCH_43	Active	Wear	75.82	Optimal_Operation
MCH_7	Idle	None	75.72	Optimal_Operation
MCH_36	Active	Wear	75.46	Optimal_Operation
MCH_9	Maintenance	None	75.41	Optimal_Operation
MCH_37	Active	None	74.98	Optimal_Operation

Table 5. Top 10 Healthiest Machines

Machines occupying the highest positions consistently demonstrated superior production throughput, high product quality scores, and minimal operational interruptions. Most of these machines were classified under the Optimal Operation category, confirming the validity of the health assessment framework. Interestingly, a small number of machines classified under Critical Failure achieved relatively high health scores. This observation suggests that certain failures may be localized events that do not substantially affect broader operational performance metrics.

In contrast, Table 6 presents the ten lowest-performing machines.

These machines exhibited a common pattern characterized by excessive downtime, reduced throughput, elevated reject rates, and poor quality outcomes. Most fell into the High Downtime category, reinforcing the importance of downtime as a primary determinant of machine health. Several machines recorded near-zero production output despite remaining in Active operational mode, indicating severe inefficiencies and potential hidden failures that were not explicitly reflected in the fault records.

Machine_ ID	Operation _ Mode	Fault_ Type	Health Score	Target_ Class
MCH_24	Active	None	17.71	High_Downtime
MCH_50	Active	None	19.76	High_Downtime
MCH_2	Active	None	20.81	High_Downtime
MCH_17	Active	Wear	21.49	High_Downtime
MCH_48	Active	None	21.62	High_Downtime
MCH_26	Idle	Wear	22.10	High_Downtime
MCH_25	Active	None	22.61	High_Downtime
MCH_5	Active	Power_ Failure	22.64	Critical_Failure
MCH_26	Active	Sensor_ Error	23.26	High_Downtime
MCH_18	Active	None	23.51	High_Downtime

Table 6. Bottom 10 Machines with Critical Health Conditions

These machines exhibited a common pattern characterized by excessive downtime, reduced throughput, elevated reject rates, and poor quality outcomes. Most fell into the High Downtime category, reinforcing the importance of downtime as a primary determinant of machine health. Several machines recorded near-zero production output despite remaining in Active operational mode, indicating severe inefficiencies and potential hidden failures that were not explicitly reflected in the fault records.

4.6 Operational Mode Analysis

To investigate the influence of machine operating states, average health scores were calculated across the four operational modes: Active, Idle, Maintenance, and Fault.

Fault_ Type	Operation_ Mode	Avg Health Score	Count	Max Score	Min Score
Wear	Fault	51.2	87	70.44	32.19
Wear	Active	50.76	649	72.88	25.93
Overheat	Active	50.68	525	73.20	29.49
Power_Failure	Active	50.61	336	73.29	29.13
Sensor_Error	Maintenance	51.13	64	68.95	34.16

Table 7. Average Health Scores by Operational Mode

The results reveal relatively modest differences among operational modes. Machines operating in Fault mode achieved a slightly higher average health score than those in Active, Maintenance, or Idle modes. Although this finding may initially appear counterintuitive, it suggests that designating an operational mode alone is insufficient to explain the machine's condition. Instead, performance outcomes appear to be more strongly influenced by productivity measures, quality indicators, and downtime characteristics.

The similarity of average scores across operational modes highlights the importance of adopting multidimensional health assessment frameworks rather than relying solely on operational state classifications.

4.7 Fault-Type Analysis

The influence of fault conditions on machine health was examined by comparing average scores across different fault categories.

Fault_ Type	Operation_ Mode	Machine_ ID	Health Score	Target_ Class
Overheat	Active	MCH_6	73.20	Critical_Failure
Overheat	Idle	MCH_31	72.48	Critical_Failure
Power_Failure	Active	MCH_34	73.29	Critical_Failure
Sensor_Error	Active	MCH_10	75.77	Optimal_Operation
Wear	Active	MCH_36	72.88	Optimal_Operation
Wear	Fault	MCH_41	70.44	Optimal_Operation

3. Worst Machine per Fault_Type + Operation_Mode

Fault_ Type	Operation_ Mode	Machine_ ID	Health Score	Target_ Class
Overheat	Active	MCH_57	29.49	Critical_Failure
Overheat	Idle	MCH_16	28.85	Critical_Failure
Power_Failure	Idle	MCH_50	27.86	Critical_Failure
Sensor_Error	Active	MCH_26	23.34	High_Downtime
Wear	Active	MCH_17	25.93	High_Downtime

Table 8. Average Health Scores by Fault Type

Machines affected by Wear and Overheat faults generally maintained moderate health levels, whereas Power Failure events were associated with comparatively lower scores. Sensor Error conditions displayed substantial variability, indicating that their impact depends heavily on operational context and severity.

Although differences between fault categories were not extreme, the results demonstrate that fault characteristics contribute meaningfully to machine condition assessment. The relatively lower performance associated with Power Failure events suggests that electrical disruptions have broader consequences for

manufacturing operations than many mechanical faults.

4.8 Interaction Between Fault Type and Operational Mode

The interaction between fault conditions and operational modes provides deeper insights into machine behavior. Figure 2 presents a boxplot illustrating health score distributions across fault operation combinations, while Figure 3 presents a heatmap of average scores.

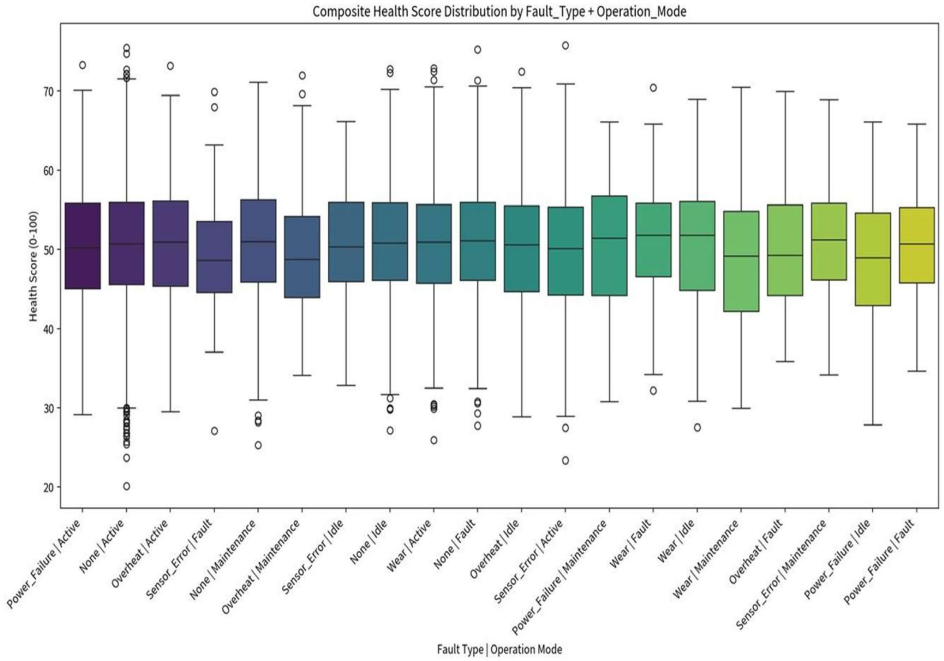

Figure 3. Boxplot of Composite Health Scores by Fault Type and Operational Mode

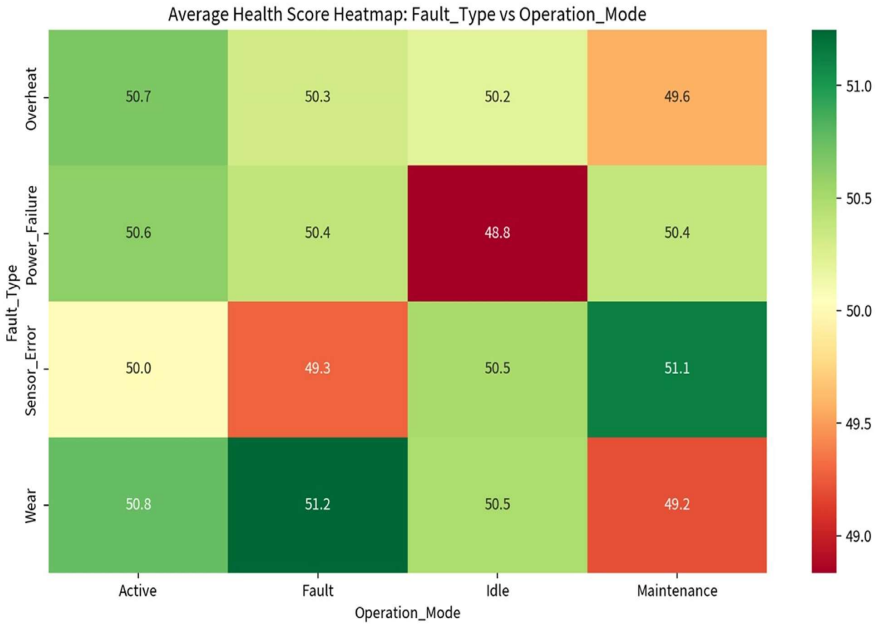

Figure 4. Heatmap of Average Health Scores Across Fault Type and Operational Mode Combinations

The boxplot reveals substantial variability among combinations. Sensor Error occurring during Maintenance operations exhibited relatively high median scores and narrow distributions, suggesting effective maintenance interventions. In contrast, Power Failure occurring during Idle operations produced lower medians and wider distributions, indicating unstable and unfavorable operating conditions.

The heatmap further confirms these patterns by highlighting areas of consistently high and low performance. The highest average health scores were observed for Sensor Error Maintenance and Wear Fault combinations, whereas the lowest averages were observed for Power Failure Idle combinations. These findings demonstrate that machine health cannot be fully understood by examining faults or operational modes independently; rather, the interaction between these factors plays a critical role in determining overall performance.

4.9 Low-Health Outlier Analysis

To identify machines requiring immediate maintenance attention, an outlier analysis was conducted on the dataset's lowest-performing observations. Table 9 presents the ten machines with the lowest composite health scores.

Rank	Machine_ID	Operation_ Mode	Fault_ Type	Health Score	Target_ Class	Key Issues
1	MCH_50	Active	None	17.71	High_Downtime	Near-zero throughput, very high downtime
2	MCH_26	Active	Sensor_ Error	19.76	High_Downtime	Zero throughput, high reject rate
3	MCH_25	Active	None	20.81	High_Downtime	Extremely low output
4	MCH_17	Maintenance	Wear	21.49	High_Downtime	Poor quality + downtime
5	MCH_15	Active	None	21.62	High_Downtime	Zero throughput
6	MCH_14	Active	None	22.10	High_Downtime	Critical performance drop
7	MCH_24	Active	None	22.61	High_Downtime	High vibration + low output
8	MCH_48	Active	None	22.64	High_Downtime	Very poor quality
9	MCH_44	Active	None	23.26	High_Downtime	Low throughput
10	MCH_16	Idle	None	23.51	High_Downtime	High downtime despite idle state

Table 9. Lowest Health Score Outliers

The identified outliers exhibited health scores ranging from approximately 17 to 24, substantially below the dataset average of 50.8. Most were classified under High Downtime conditions and shared several characteristics, including extremely long downtime, minimal production throughput, poor product quality, and elevated reject rates.

A particularly notable finding is that several critically unhealthy machines reported no explicit fault condition despite severe performance degradation. This discrepancy suggests latent failures or inadequately detected anomalies. Consequently, these observations highlight the value of composite health monitoring systems, which can identify deteriorating assets even when conventional fault detection mechanisms fail to issue warnings.

Overall, the outlier analysis demonstrates that the proposed health score effectively identifies operationally critical machines and provides actionable information for predictive maintenance planning. By prioritizing assets with the lowest scores, manufacturing organizations can reduce downtime, improve resource allocation, and enhance overall production efficiency.

5. Summary

This study proposes and validates a hybrid Principal Component Analysis (PCA) and Entropy weighting framework for the composite health assessment of machines in Industry 4.0 smart manufacturing systems. Addressing the limitations of existing diagnostic models that predominantly focus on isolated faults or individual sensor signals, the proposed approach integrates variance-based dimensionality reduction (PCA) with information-theoretic objective weighting (Entropy) to generate a comprehensive Composite Machine Health Score (CMHS). The framework was empirically tested using a multidimensional smart manufacturing dataset comprising 10,000 records and 15 variables spanning operational, environmental, and production metrics.

The empirical analysis yielded several key insights:

1. Effective Health Differentiation: The CMHS effectively captures the multidimensional nature of machine health, exhibiting a normal distribution centered around a mean of 50.8. It successfully separates highly efficient machines (Optimal Operation) from severely degraded ones (High Downtime).

2. Critical Indicators: Entropy weighting revealed that *Downtime_Minutes* and *Humidity_%* possess the highest discriminatory power, highlighting that operational interruptions and environmental fluctuations are the most critical factors in distinguishing machine health states.

3. Robustness of the Hybrid Approach: The hybrid fusion of PCA and entropy methods proved more robust and sensitive than either technique alone, mitigating the biases inherent in single method evaluations.

4. Limitations of Categorical States: The analysis demonstrated that traditional operational modes (e.g., Active, Fault) alone are insufficient to explain machine condition. Instead, the interaction between specific fault types and operational modes (e.g., Sensor Error during Maintenance vs. Power Failure during Idle) plays a critical role in determining overall performance.

5. Detection of Latent Anomalies: Outlier analysis successfully identified severely degraded machines that reported no explicit fault codes but exhibited near zero throughput and extreme downtime. This underscores the framework's ability to detect hidden failures that conventional rule based monitoring systems might miss.

6. Conclusion

In conclusion, the hybrid PCA Entropy framework presented in this study offers a robust, objective, and highly interpretable solution for composite machine health assessment in smart manufacturing environments. By synthesizing the structural variance captured by PCA with the discriminative information entropy of individual indicators, the proposed model overcomes the limitations of unidimensional health metrics. The resulting Composite Machine Health Score (CMHS) provides a holistic reflection of machine condition, integrating productivity, environmental stability, and operational efficiency into a single actionable metric. From a practical standpoint, this framework serves as a vital decision support tool for predictive maintenance

in Industry 4.0 ecosystems. It enables manufacturing organizations to move beyond reactive or strictly time-based maintenance strategies by accurately prioritizing assets that require immediate intervention. Crucially, the framework's ability to identify latent anomalies where machines exhibit severe performance degradation without triggering explicit fault alarms demonstrates its superior ability to safeguard against hidden inefficiencies and unexpected catastrophic failures.

Ultimately, the integration of such multidimensional, data-driven health assessment models is essential for realizing the full potential of smart manufacturing. By providing a reliable basis for resource allocation, fault prioritization, and operational optimization, the proposed framework significantly contributes to enhancing system reliability, reducing unplanned downtime, and fostering the operational resilience required in modern, data-intensive manufacturing landscapes. Future research may build upon this foundation by incorporating real-time streaming data architectures and deep learning techniques to enable dynamic, continuous prognostic updates.

References

[1] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D. (2014). Prognostics and health management design for rotary machinery systems Reviews, methodology and applications. *Mechanical Systems and Signal Processing, 42*(1), 314–334.

[2] Choo, B. Y., Adams, S. C., Weiss, B. A., Marvel, J. A., Beling, P. A. (2016). Adaptive multi-scale prognostics and health management for smart manufacturing systems. *International Journal of Prognostics and Health Management, 7*, 014.

[3] Soualhi, M., Nguyen, K., Medjaher, K., Lebel, D., & Cazaban, D. (2019). Health indicator construction for system health assessment in smart manufacturing. In *2019 Prognostics and System Health Management Conference (PHM-Paris)* (p. 45–50). *IEEE*. https://doi.org/10.1109/PHM-Paris.2019.00016.

[4] Huizinga, G., Walison, P., van der Beek, H., Tops, P., Wijnstok, J., Heemskerk, P., van Os, M., van der Horst, T., Helmer, S., Huveneers, S., Butter, S., van der Zee, M., van Oort, F., Ypma, S., Mulder, J., Kotternink, G., de Heide, B., Grosfeld, T. (2014). *Smart industry: Dutch industry fit for the future*. http://www.smartindustry.nl/wp-content/uploads/2014/07/Opmaak-Smart-Industry.pdf.

[5] Bernaden, J. (2012). *Indirect jobs: A direct way to talk about why we need smart manufacturing*. Rockwell Automation. http://smartmanufacturing.com/wp-content/uploads/2012/01/SmartManufacturing-increases-indirect-jobs.final_.pdf.

[6] McKinsey Global Institute. (2012). *Manufacturing the future: The next era of global growth and innovation*. https://www.mckinsey.com/insights/manufacturing/the_future_of_manufacturing.

[7] President's Council of Advisors on Science and Technology (PCAST). (2011). Report to the President on ensuring American leadership in advanced manufacturing. *Executive Office of the President*.

[8] President's Council of Advisors on Science and Technology (PCAST). (2012). Report to the President on capturing domestic competitive advantage in advanced manufacturing. *Executive Office of the President*.

[9] President's Council of Advisors on Science and Technology (PCAST). (2014). Report to the President: Accelerating U.S. advanced manufacturing. *Executive Office of the President*.

[10] Zheng, P., Wang, H., Sang, Z., et al. (2018). Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Frontiers of Mechanical Engineering, 13*, 137–150. https://doi.org/10.1007/s11465-018-0499-5.

[11] Watanabe, T., Kono, I., Onozuka, H. (2020). Anomaly detection methods in turning based on motor data analysis. *Procedia Manufacturing, 48*, 882–893. https://doi.org/10.1016/j.promfg.2020.05.126.

[12] Cooper, C., Zhang, J., Gao, R. X., Wang, P., Ragai, I. (2020). Anomaly detection in milling tools using acoustic signals and generative adversarial networks. *Procedia Manufacturing, 48*, 372–378. https://doi.org/10.1016/j.promfg.2020.05.059.

[13] Lins, R. G., Ricardo, P., Araujo, M., De Corazzim, M. (2020). In-process machine vision monitoring of tool wear for cyber-physical production systems. *Robotics and Computer-Integrated Manufacturing, 61*, 101859. https://doi.org/10.1016/j.rcim.2019.101859.

[14] Wei, X., Liu, X., Yue, C., Wang, L., Liang, S. Y., Qin, Y. (2022). Tool wear state recognition based on feature selection method with whitening variational mode decomposition. *Robotics and Computer-Integrated Manufacturing*. Advance online publication. https://doi.org/10.1016/j.rcim.2022.102344.

[15] Zhou, C., Yang, B., Guo, K., Liu, J., Sun, J., Song, G., Zhu, S., Sun, C., Jiang, Z. (2020). Vibration singularity analysis for milling tool condition monitoring. *International Journal of Mechanical Sciences, 166*, 105254. https://doi.org/10.1016/j.ijmecsci.2019.105254.

[16] Zhou, Y., Sun, B., Sun, W., Lei, Z. (2022). Tool wear condition monitoring based on a two-layer angle kernel extreme learning machine using sound sensor for milling process. *Journal of Intelligent Manufacturing, 33*(1), 247–258. https://doi.org/10.1007/s10845-020-01663-1.

[17] Liu, X., Zhang, B., Li, X., Liu, S., Yue, C., Liang, S. Y. (2022). An approach for tool wear prediction using customized DenseNet and GRU integrated model based on multi-sensor feature fusion. *Journal of Intelligent Manufacturing*. Advance online publication. https://doi.org/10.1007/s10845-022-01954-9.

[18] Sundaram, S., Zeid, A. (2021). Smart prognostics and health management (SPHM) in smart manufacturing: An interoperable framework. *Sensors, 21*(18), 5994. https://doi.org/10.3390/s21185994.

[19] Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E., Xi, L. (2018). Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering System Safety, 178*, 255–268.

[20] Blecker, T., Friedrich, G. (2007). Guest editorial: Mass customization manufacturing systems. *IEEE Transactions on Engineering Management, 54* (1), 4–11.

[21] Fitouhi, M. C., Nourelfath, M. (2014). Integrating noncyclical preventive maintenance scheduling and production planning for multi-state systems. *Reliability Engineering System Safety, 121*, 175–186.

[22] Koren, Y., Heisel, U., Jovane, F., Moriwaki, T., Pritschow, G., Ulsoy, G., Van Brussel, H. (1999). Reconfigurable manufacturing systems. *CIRP Annals, 48*(2), 527–540.

[23] Xia, T., Xi, L., Pan, E., Ni, J. (2017). Reconfiguration oriented opportunistic maintenance policy for reconfigurable manufacturing systems. *Reliability Engineering System Safety, 166*, 87–98.

[24] Zhou, J., Djurdjanovic, D., Ivy, J., Ni, J. (2007). Integrated reconfiguration and age-based preventive maintenance decision making. *IIE Transactions, 39*(11), 1085–1102. https://doi.org/10.1080/07408170701291779.

[25] Koren, Y., Gu, X., Badurdeen, F., Jawahir, I. S. (2018). Sustainable living factories for next generation manufacturing. *Procedia Manufacturing, 21*, 26–36.

[26] Raouf, I., Kumar, P., Khalid, S., et al. (2025). Comprehensive analysis of current developments, challenges, and opportunities for the health assessment of smart factory. *International Journal of Precision Engineering and*

Manufacturing-Green Technology, 12, 1321–1338. https://doi.org/10.1007/s40684-025-00694-4.

[27] Luo, Z., Yan, M., Wang, W., Zhang, Q. (2023). Non-intrusive anomaly detection of industrial robot operations by exploiting nonlinear effect. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6*(4), 1–27.

[28] Kim, H., Lee, H., Kim, S. W. (2022). Current only-based fault diagnosis method for industrial robot control cables. *Sensors, 22*(5), 1917.

[29] Li, P., Xu, X., Yang, S., Jiang, X. (2023). Open circuit fault diagnosis strategy of PMSM drive system based on grey prediction theory for industrial robot. *Energy Reports, 9*, 313–320.

[30] Algburi, R. N. A., Gao, H., Al-Huda, Z. (2022). Improvement of an industrial robotic flaw detection system. *IEEE Transactions on Automation Science and Engineering, 19*(4), 3953–3967.

[31] Steinbauer, G., Wotawa, F. (2005). Detecting and locating faults in the control software of autonomous mobile robots. *In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 5, p. 1742–1743).

[32] Raouf, I., Kumar, P., Kim, H. S. (2024). Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert Systems with Applications, 258*, 125137. https://doi.org/10.1016/j.eswa.2024.125137.

[33] Kumar, P., Raouf, I., Song, J., Prince, Kim, H. S. (2024). Multi-size wide kernel convolutional neural network for bearing fault diagnosis. *Advances in Engineering Software, 198*, 103799. https://doi.org/10.1016/j.advengsoft.2024.103799.

[34] Kumar, P. A., Raouf, I., Kim, H. S. (2024). Transfer learning for servomotor bearing fault detection in the industrial robot. *Advances in Engineering Software, 194*, 103672. https://doi.org/10.1016/j.advengsoft.2024-.103672.

[35] Raouf, I. A., Kumar, P., Cheon, Y., Tanveer, M., Jo, S. H., Kim, H. S. (2024). Advances in prognostics and health management for aircraft landing gear progress, challenges, and future possibilities. *International Journal of Precision Engineering and Manufacturing-Green Technology*. Advance online publication. https://doi.org/10.1007/s40684-024-00646-4.

[36] Jin, X., Ni, J. (2013). Joint production and preventive maintenance strategy for manufacturing systems with stochastic demand. *Journal of Manufacturing Science and Engineering, 135*(3), 031016. http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/135/3/031016/6261104/manu_135_3_031016.pdf.

[37] Colabsss. (n.d.). *Smart manufacturing process dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/colabsss/smart-manufacturing-process-dataset.