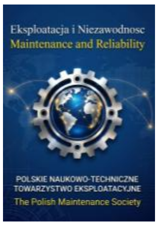

Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

Machine learning-based predictive maintenance of a wire drawing machine using vibration and motor current signals

Ahmet Pişmişoğlu[1], Erkan Caner Ozkat[2], Mehmet Konar[3,*]

[1] Department of Aviation Electrical and Electronics, Graduate School of Natural and Applied Sciences, Erciyes University, Kayseri, Türkiye

[2] Department of Mechanical Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdogan University, Rize, Türkiye

[3] Department of Aviation Electrical and Electronics, Faculty of Aeronautics and Astronautics, Erciyes University, Kayseri, Türkiye

Received: 7 May 2026
Revised: 23 May 2026
Accepted: 25 June 2026
Online: 23 July 2026

Abstract

Predictive maintenance improves reliability and reduces downtime in modern manufacturing systems. However, many studies rely on laboratory datasets or single-component monitoring, limiting their applicability to complex industrial environments. This study proposes a predictive maintenance framework for a multi-pass wire drawing machine using vibration and motor current signals from a real industrial production line. A Composite Health Index (CHI) is developed to transform multi-motor sensor data into an interpretable machine-level degradation indicator. The extracted features are used to train ensemble machine learning models including Random Forest, Extra Trees, and XGBoost, whose outputs are combined through a weighted hybrid ensemble model. A decision-layer mechanism with smoothing and temporal filtering is applied to reduce false alarms while preserving detection capability. Experimental results show that the model achieves a recall of 0.90 and an F1-score of 0.75, demonstrating its effectiveness for industrial applications.

Keywords: *predictive maintenance, multi-sensor monitoring, motor current analysis, vibration diagnostics, composite health index*

Article citation:
Pişmişoğlu A, Ozkat EC, Konar M, Machine learning-based predictive maintenance of a wire drawing machine using vibration and motor current signals, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/225078

Highlights

- Predictive maintenance of wire drawing machines using real industrial data.
- Multi-sensor monitoring with vibration and motor current signals.
- Composite Health Index for machine-level degradation assessment.
- Hybrid ensemble model improves predictive maintenance reliability.
- Framework supports scalable predictive maintenance in manufacturing.

1. Introduction

Industrial manufacturing systems are increasingly required to operate under high productivity, minimal downtime, and strict quality constraints. In such environments, unexpected equipment failures may lead to severe production interruptions, economic losses, and safety risks. Consequently, maintenance strategies play a critical role in ensuring the reliability and sustainability of modern manufacturing systems. Traditional maintenance approaches such as corrective maintenance and time-based preventive maintenance often fail to represent the actual degradation state of industrial equipment operating under variable load conditions and dynamic production environments [1–3].

With the rapid development of Industry 4.0 technologies, industrial machines are increasingly equipped with interconnected sensors, programmable logic controllers (PLC), and supervisory control and data acquisition (SCADA) systems capable of continuously monitoring operational parameters. These digital infrastructures enable the collection of large volumes of operational data, creating new opportunities for data-driven maintenance strategies. Predictive maintenance (PdM) has therefore emerged as one of the most promising approaches for improving equipment reliability and reducing maintenance costs in smart manufacturing environments [4–6].

Recent advances in machine learning have significantly accelerated the development of predictive maintenance systems. Supervised learning algorithms such as Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) have demonstrated strong performance in fault diagnosis and failure prediction tasks using industrial sensor data [7–9]. Ensemble learning approaches are particularly attractive for industrial monitoring applications because they can capture nonlinear relationships and remain robust in the presence of noisy measurements and highly variable operating conditions [10–12].

Condition monitoring systems typically rely on multiple sensor signals to capture the operational behavior of industrial machinery. Among these sensing modalities, vibration signals have long been considered one of the most effective indicators for diagnosing mechanical anomalies in rotating machinery. Mechanical faults such as imbalance, misalignment, bearing defects, and friction often produce characteristic vibration signatures that can be detected through signal processing techniques [13–15]. Recent studies have further demonstrated the effectiveness of machine learning techniques combined with vibration monitoring for diagnosing faults in rotating machinery operating under industrial conditions [16–18].

In addition to vibration monitoring, electrical measurements obtained from motor drives provide valuable diagnostic information about machine operation. Motor current signals directly reflect variations in electrical load and torque demand, making them a valuable source of information for detecting mechanical degradation in electromechanical systems. Motor Current Signature Analysis (MCSA) has therefore been widely applied in predictive maintenance systems for rotating machinery and industrial production equipment [19–21].

Despite the rapid growth of predictive maintenance research, many existing studies rely on laboratory datasets or simplified experimental environments. Such datasets often fail to represent the complex interactions and degradation dynamics observed in real industrial systems [22–23]. In complex manufacturing processes, degradation mechanisms frequently propagate across multiple components due to mechanical coupling and load redistribution effects. Consequently, monitoring individual components in isolation may not adequately capture the overall health condition of the production system.

Wire drawing machines represent a particularly challenging industrial environment for predictive maintenance implementation. In these machines, multiple drawing blocks operate sequentially to reduce the diameter of metallic wires through successive deformation stages. Each drawing block is driven by an independent high-power electric motor, and the mechanical interactions between blocks introduce complex coupling effects across the production line. Degradation phenomena such as die wear, lubrication deterioration, and friction instability may therefore influence both electrical and mechanical signals measured throughout the system [24–26].

Recent studies in reliability engineering literature have highlighted the importance of health index–based monitoring approaches for representing degradation patterns in complex industrial systems. By aggregating multiple sensor indicators into a unified health representation, these approaches significantly improve interpretability while enhancing predictive performance of machine learning models [27–28].

Motivated by these challenges, this study introduces a machine-level predictive maintenance framework for a multi-pass wire drawing machine using vibration and motor current signals collected from a real industrial production line. The proposed approach introduces a Composite Health Index (CHI) that aggregates block-level indicators derived from vibration signals, motor current measurements, and degradation trend features. This hierarchical representation enables the transformation of distributed motor-level sensor measurements into an interpretable machine-level degradation indicator suitable for predictive maintenance modeling.

The overall predictive maintenance workflow implemented in this study is illustrated in Fig. 1, which summarizes the stages of industrial data acquisition, signal processing, feature engineering, health index construction, and machine learning-based predictive modeling.

Despite the significant advancements in predictive maintenance research, several important limitations remain in the existing literature that this study aims to address: (i) many studies rely on laboratory-generated datasets or controlled experimental conditions, which fail to capture the complex degradation dynamics observed in real industrial production environments; (ii) a large portion of existing work focuses on single-sensor monitoring or component-level analysis, limiting the ability to represent system-level degradation behavior in complex multi-motor industrial systems; and (iii) few studies propose machine-level health index that aggregate distributed sensor information into an interpretable degradation representation suitable for maintenance decision-making in coupled multi-component systems.

In industrial production lines such as wire drawing machines, degradation mechanisms are often distributed across multiple components due to mechanical coupling and load redistribution effects. As a result, monitoring individual components in isolation may not provide sufficient information for reliable maintenance decision-making at the system level.

To address these challenges, this study proposes a machine-level predictive maintenance framework based on real industrial data collected from a multi-pass wire drawing machine operating under continuous production conditions. The proposed approach integrates vibration and motor current signals to capture both mechanical and electrical degradation characteristics and introduces a Composite Health Index (CHI) to transform distributed sensor measurements into an interpretable machine-level degradation indicator.

The main objective of this study is to demonstrate that machine-level degradation in a multi-motor industrial system can be reliably predicted in advance using multi-sensor data fusion and health index–based modeling. By combining engineering-driven feature construction with ensemble machine learning techniques, the proposed framework provides an interpretable and scalable predictive maintenance solution for real industrial applications. The effectiveness of the proposed approach is quantitatively evaluated using both conventional classification metrics and operationally relevant indicators, including recall, F1-score, false positive rate, and false positives per day.

In addition to the proposed framework, this study provides a systematic evaluation of machine-level predictive maintenance performance using real industrial run-to-maintenance data. Unlike many existing studies, the proposed approach is assessed not only in terms of classification accuracy but also through operationally relevant performance indicators such as early fault detection capability and false alarm behavior.

Furthermore, the study demonstrates that aggregating multi-sensor information into a composite health index significantly improves the interpretability of degradation behavior while maintaining high predictive performance. The effectiveness of the proposed framework is validated through comparative analysis against multiple baseline machine learning models, highlighting its practical applicability for industrial predictive maintenance systems.

In addition to model-level improvements, this study also introduces an operational decision layer designed to enhance the practical usability of predictive maintenance systems by reducing false alarm rates while maintaining high detection capability.

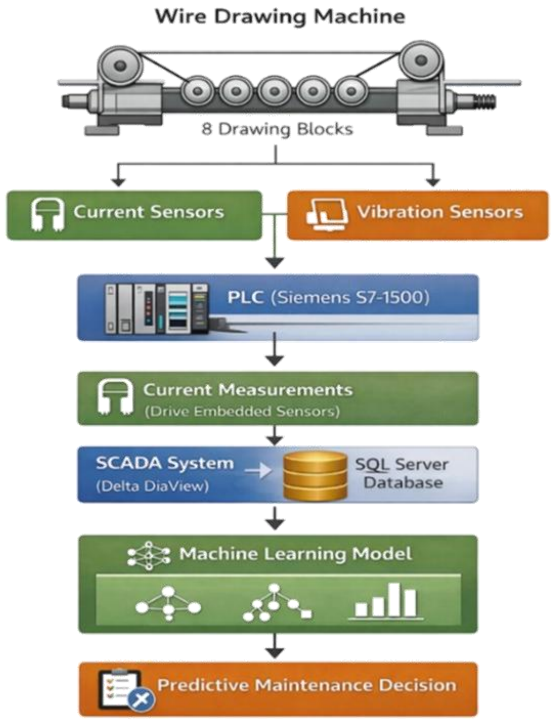

Figure 1. Workflow of the proposed predictive maintenance framework including industrial data acquisition, feature engineering, health index construction, and machine learning modeling.

The main contributions of this study can be summarized as follows:

- A machine-level predictive maintenance framework is developed using real industrial data from a multi-pass wire drawing machine operating under continuous production conditions.
- A Composite Health Index (CHI) is proposed to aggregate multi-sensor information, including vibration and motor current signals, into an interpretable

degradation indicator.

- A hybrid ensemble machine learning approach is implemented to improve early fault detection performance under imbalanced industrial data conditions.
- The proposed framework is evaluated using real run-to-maintenance data and assessed with both conventional performance metrics and operational indicators such as false alarm behavior.
- The results demonstrate the effectiveness of the proposed framework in achieving a recall of 0.90 and an F1-score of 0.76, with a false positive rate controlled at 197.8 false positives per day, confirming reliable early fault detection capability under real industrial operating conditions.

2. Materials and methods

2.1. Industrial system description and monitoring architecture

The experimental investigation was conducted on an industrial multi-pass wire drawing production line located at the Boyçelik Facility, where steel wire rods are processed under continuous production conditions. Wire drawing is a metal forming process in which the cross-sectional area of a metallic wire is gradually reduced by pulling the material through a sequence of drawing dies. This process enables the production of high-strength wires with high dimensional accuracy and controlled surface quality, which are widely used in automotive, construction, and industrial applications.

The investigated production line consists of eight sequential drawing blocks, arranged in series to perform progressive diameter reduction of the wire material. Each drawing block contains a drawing die and a rotating capstan that pulls the wire through the die while maintaining controlled tensile loading across the production line. The gradual reduction in wire diameter across successive blocks ensures stable plastic deformation of the material while preserving the mechanical properties of the steel.

Each drawing block is driven by an independent 57 kW three-phase induction motor operating under closed-loop speed control. The motors are controlled using SINAMICS G120 frequency converters, which provide vector-controlled drive operation and allow precise regulation of motor speed and torque. Maintaining stable motor torque is critical in multi-pass wire drawing processes because the inter-pass tension between drawing blocks must remain constant to avoid wire breakage, dimensional inconsistencies, or surface defects. Fig. 2 presents real photographs of the industrial wire-drawing system investigated in this study. Fig. 2a shows the multi-pass drawing section, where the wire is sequentially drawn through a series of dies mounted on individual drawing blocks. Each block imposes controlled tensile loading on the wire while achieving incremental diameter reduction. Fig. 2b shows a rear view of the drawing block motors. These motors operate under closed-loop speed control to maintain constant inter-pass tension and are directly responsible for overcoming the mechanical resistance generated during the drawing process.

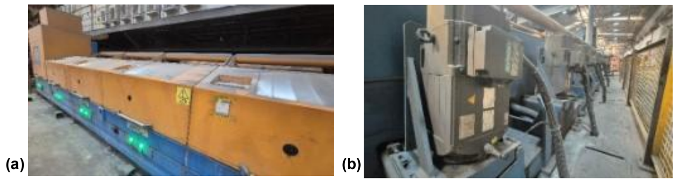

Figure 2. Industrial wire drawing machine used in this study: (a) drawing blocks during operation; (b) drive motors.

From a tribological perspective, the wire drawing process involves complex interactions between the wire material, die surfaces, and lubrication medium. During prolonged operation, drawing dies gradually experience wear due to continuous frictional contact with the deforming wire. As die wear progresses, lubrication conditions may transition from hydrodynamic lubrication toward mixed or boundary lubrication regimes. This transition increases frictional resistance at the wire–die interface and consequently increases the mechanical torque demand of the drive motors.

An increase in friction manifests itself as a measurable rise in the motor current consumption, since the motors must supply additional torque to maintain the drawing process. At the same time, irregular contact conditions between the wire and die surfaces may generate mechanical vibrations within the drawing blocks and motor structures. These vibration responses propagate through the mechanical structure of the machine and provide valuable information about abnormal mechanical conditions such as die wear, lubrication degradation, or structural imbalance.

Unlike isolated rotating machinery systems, multi-pass wire drawing machines exhibit strong coupling between successive drawing blocks. Variations in friction or load conditions occurring in one drawing stage may propagate along the production line through changes in inter-pass tension and material flow. Consequently, degradation patterns often appear as distributed signatures across multiple motors rather than as isolated anomalies in a single component. Monitoring both electrical and mechanical signals simultaneously therefore provides a more comprehensive representation of the machine's operational state and has become a widely adopted strategy in predictive maintenance systems for industrial equipment [29–31].

To capture these degradation characteristics, the investigated system is equipped with a multisensor condition monitoring infrastructure that integrates electrical and mechanical measurements. Electrical quantities are obtained directly from the SINAMICS G120 frequency converters, where embedded current measurement units provide real-time information about the electrical load of each motor. Motor current analysis has been widely applied in recent predictive maintenance studies as a non-invasive approach for detecting mechanical load variations and degradation patterns in industrial drive systems [32].

In addition to electrical measurements, industrial accelerometer sensors are installed on the motor housings and structural components of the drawing blocks to capture vibration responses during machine operation. Vibration-based monitoring has been extensively used for detecting mechanical anomalies such as friction, imbalance, or structural faults in rotating machinery and industrial production systems [33].

The vibration measurements used in this study were obtained using industrial vibration sensors (Pepperl+Fuchs VIM32PL-E1AC8-0RE-IO-1V1401), which are MEMS-based capacitive sensors capable of measuring RMS vibration velocity and RMS acceleration in accordance with DIN ISO 10816/20816 standards. The sensors provide preprocessed vibration indicators, including VRMS velocity (mm/s), ARMS acceleration (g), and peak acceleration values provided directly by the sensor output. These values are derived through built-in signal processing within the sensor, which performs filtering, integration, and RMS computation internally. The sensor operates with an internal sampling rate of 8 kHz, an averaging time of 2 seconds for RMS calculations, and a measurement frequency range of 10–1000 Hz.

The sensors are mounted on the motor housing of each wire drawing block to capture vibration signatures associated with mechanical degradation. In addition to vibration measurements, motor current data are acquired directly from the motor drives through the PLC-based monitoring system. This enables the simultaneous monitoring of both mechanical and electrical condition indicators. All signals are recorded with timestamp information, and synchronization across different data sources is achieved by aligning signals based on their timestamps. Raw high-frequency vibration signals are not directly used in this study. Instead, preprocessed statistical indicators generated at the sensor level (VRMS, ARMS, and peak values) are utilized. Although the sensor operates at a high internal sampling rate, only these aggregated features are transmitted via the PLC at lower logging intervals.

Due to the nature of the industrial monitoring infrastructure, the recorded data follow a non-uniform, event-based sampling structure rather than a strictly fixed sampling frequency at the data storage level, since only aggregated sensor outputs are logged at discrete time intervals via the PLC system. All sensor measurements are integrated through an industrial automation architecture based on a Siemens S7-1500 programmable logic controller (PLC). The PLC performs real-time acquisition and synchronization of electrical and vibration signals while simultaneously executing machine control logic. Industrial communication between sensors, drives, and control systems is implemented through the factory automation network, ensuring reliable transmission of measurement data. The overall industrial monitoring and data acquisition architecture used in this study, including the frequency converters, PLC modules, and SCADA-based monitoring infrastructure, is illustrated in Fig. 3.

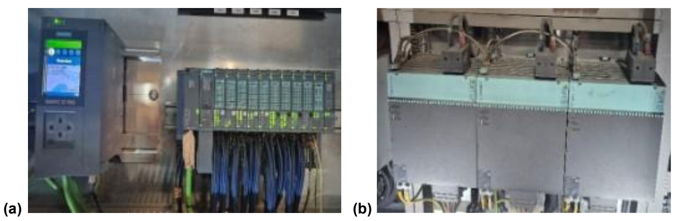

Figure 3. Control and data acquisition system: (a) PLC modules; (b) motor frequency converters.

Supervisory monitoring and operational data management are implemented using a Delta DiaView SCADA system, which provides continuous monitoring of machine parameters during production. The SCADA platform enables real-time visualization of electrical and vibration signals, alarm monitoring, and historical data logging. All acquired measurements are automatically transmitted to a centralized SQL Server database, where time-stamped data are archived for long-term storage and subsequent data-driven analysis.

The monitoring campaign covered three complete run-to-maintenance cycles, allowing the dataset to capture the natural degradation progression of the machine from healthy operation to maintenance-required conditions. Such run-to-maintenance datasets are particularly valuable for predictive maintenance research because they reflect realistic industrial degradation patterns rather than artificially induced faults [34].

The dataset used in this study represents real industrial operating conditions rather than laboratory-generated fault scenarios. Such real-world datasets typically include process disturbances, load variations, and measurement noise, which make predictive modeling more challenging but also more representative of practical industrial environments.

2.2. Predictive maintenance modeling framework

The predictive maintenance framework proposed in this study aims to infer the degradation state of the industrial wire drawing system using multi-sensor data collected from multiple drawing blocks. The predictive maintenance modeling pipeline used in this study is illustrated in Fig. 4, where raw sensor signals are first processed and transformed into diagnostic features. These features are then aggregated into health indices representing the machine condition and subsequently used to train machine learning models for failure prediction. The framework integrates vibration and motor current signals to capture both mechanical and electrical degradation patterns. In this study, the operational horizon is defined as 72 hours prior to maintenance events.

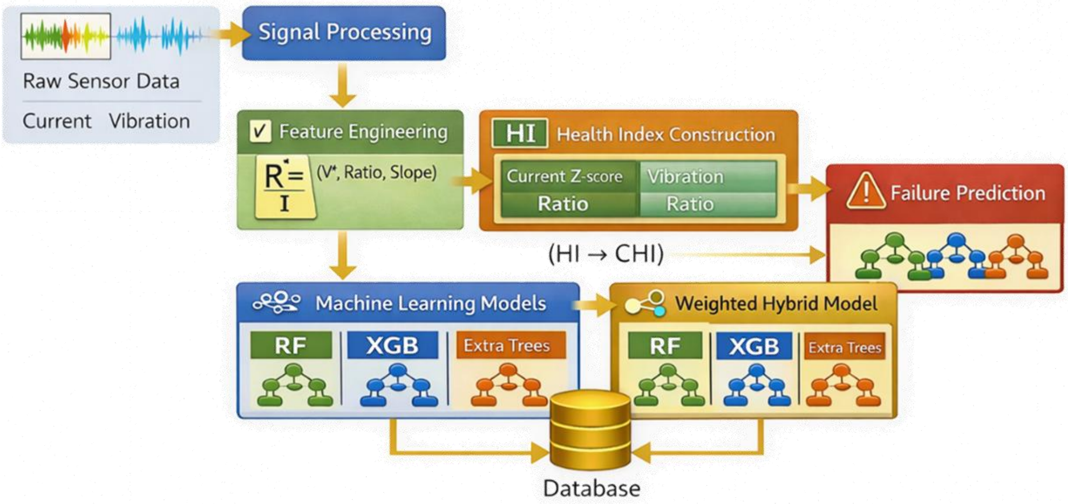

Figure 4. Predictive maintenance modeling pipeline integrating signal processing, feature engineering, health index construction, and machine learning models.

A binary labeling strategy is adopted based on a predefined operational horizon prior to maintenance events. Data samples within this horizon are labeled as "maintenance-required" (Y = 1), while samples outside this interval are labeled as "normal" (Y = 0). The labeling process is designed to reflect early warning requirements in predictive maintenance applications.

To ensure realistic evaluation and to avoid data leakage, a time-aware train/test split strategy is employed. The dataset is divided chronologically, where earlier data are used for training and later data are reserved for testing. This approach better reflects real-world deployment scenarios in which future observations are not available during model training.

Predictive maintenance can be formulated as a supervised learning problem in which the objective is to estimate the probability that a machine will require maintenance within a predefined operational horizon [10,27]. Let's denote the multivariate feature vector derived from sensor measurements at time t as:

$$X(t) \in \mathbb{R}^d \tag{1}$$

where d represents the number of extracted features obtained from vibration and motor current signals. The corresponding maintenance state is represented as:

$$Y(t) \in \{0,1\} \tag{2}$$

where Y(t)=1 indicates that the system is approaching a maintenance-required condition, while Y(t)=0 corresponds to normal operating conditions. In this study, the predictive maintenance problem was formulated using a quantitatively defined operational horizon. Specifically, all samples collected within 72 hours prior to a recorded maintenance intervention were labeled as maintenance-required (Y=1), while all remaining samples were labeled as normal operating condition (Y = 0).

To eliminate the influence of short-term recovery behavior following maintenance actions, samples collected within the first 12 hours after each maintenance intervention were excluded from the modeling dataset. This approach ensures that the model focuses on pre-maintenance degradation patterns rather than transient post-maintenance effects.

The labeling procedure was applied consistently across three run-to-maintenance cycles using recorded maintenance timestamps as reference points, providing a reproducible and practically meaningful framework for early fault detection in real industrial environments.

Furthermore, the dataset used in this study consists of a total of 47,706 raw samples, of which 25,143 samples were retained for modeling after preprocessing, filtering, and the exclusion of post-maintenance data. The resulting class distribution is approximately 70.0% normal (Y = 0) and 30.0% maintenance-required (Y = 1), reflecting a moderate class imbalance consistent with early fault detection scenarios in industrial predictive maintenance applications.

Table 1. Final dataset distribution after preprocessing and exclusion filtering.

Metric	Value
Total number of raw samples	47,706
Samples used for modeling	25,143
Number of run-to-maintenance cycles	3
Samples labeled as maintenance-required (Y = 1)	7,541
Samples labeled as normal (Y = 0)	17,602
Excluded post-maintenance samples	2,444
Class distribution (Y=1 / Y=0)	30.0% / 70.0%

The dataset is structured into three consecutive run-to-maintenance cycles, enabling the evaluation of model performance across different degradation trajectories. A detailed summary of the dataset and cycle-wise class distribution is provided in Table 1 and Table 2, respectively.

Table 2. Cycle-wise distribution of raw labeled samples before preprocessing and filtering.

Cycle	Normal (Y=0)	Maintenance-required (Y=1)	Excluded
Cycle 1	10,574	5,304	0
Cycle 2	2,732	3,919	742
Cycle 3	16,199	3,824	642

The predictive relationship between the extracted feature vector (X(t)) and the machine maintenance state (Y(t)) is modeled through a parametric function (f_θ) defined as:

$$f_\theta : X(t) \to Y(t) \tag{3}$$

where θ represents the set of learnable parameters of the predictive model. The objective of the learning process is to determine the optimal parameter set that minimizes the prediction error over the training dataset. Parameter estimation is performed using empirical risk minimization, which can be expressed as:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, f_\theta(x_i)) + \lambda R(\theta) \tag{4}$$

where L(·) denotes the loss function measuring the discrepancy between the predicted and actual maintenance states, R(θ) is a regularization term controlling model complexity, and λ represents the regularization coefficient used to mitigate overfitting.

A key methodological distinction of the proposed framework lies in the hierarchical transformation of motor-level sensor measurements into machine-level maintenance indicators. In industrial wire drawing systems, maintenance interventions are typically performed at the machine level rather than for individual drive motors. Consequently, predicting failures independently for each motor may not adequately represent the operational decision-making process in real production environments.

To address this limitation, the proposed framework introduces a block-level health representation derived from the extracted sensor features. Let's denote the health index of drawing block bat time t as:

$$HI_b(t) \tag{5}$$

which reflects the degradation state inferred from vibration and current signals associated with that block. The overall machine-level health condition is then obtained by aggregating the block-level indicators into a composite health index defined as:

$$CHI(t) = \frac{1}{B}\sum_{b=1}^{B} HI_b(t) \qquad (6)$$

where B represents the total number of monitored drawing blocks in the wire drawing system. This composite indicator provides a unified representation of the overall machine condition by integrating degradation information from multiple components.

The hierarchical modeling approach improves robustness against localized sensor noise and aligns the predictive modeling framework with practical maintenance decision processes in industrial environments. By combining block-level health indices into a machine-level composite metric, the model captures the distributed nature of degradation patterns in multi-pass wire drawing systems.

Another important challenge in predictive maintenance applications is the class imbalance inherent in real industrial datasets. Maintenance-required states typically occur far less frequently than normal operating conditions. Let's denote the number of samples representing normal and maintenance-required situations as N_0 and N_1, respectively.

$$N_1 \ll N_0 \qquad (7)$$

This imbalance may bias classification algorithms toward the majority class, resulting in poor detection performance for critical degradation states. To mitigate this issue, class-weighted learning strategies and recall-oriented evaluation metrics are adopted during model training. These strategies encourage the predictive models to emphasize the correct identification of maintenance-required conditions, which is essential for reliable predictive maintenance deployment in industrial environments.

The extracted feature vectors and hierarchical health indices are subsequently used to train multiple machine learning models, including Random Forest (RF), Extreme Gradient Boosting (XGB), and Extra Trees. The predictions of individual models are further integrated through a weighted hybrid model, which improves prediction stability and enhances the reliability of failure prediction in the industrial wire drawing system.

The decision threshold used to convert predicted probabilities into binary maintenance decisions is not fixed at a default value. Instead, it is determined through a trade-off analysis between detection performance and false alarm rates. The threshold selection is guided by evaluation metrics such as recall, F1-score, false positive rate (FPR), and false positives per day (FP/day), ensuring an operationally meaningful balance between early failure detection and maintenance workload.

2.3. Signal processing and health index construction

Feature engineering plays a critical role in predictive maintenance applications because raw sensor signals often fail to directly represent machine degradation dynamics.

Raw sensor signals collected from industrial equipment often contain noise, operational fluctuations, and measurement variability. Therefore, signal preprocessing and feature engineering are required to transform the raw time-series data into meaningful degradation indicators. The signal processing and health index construction procedure used in this study is illustrated in Fig. 5.

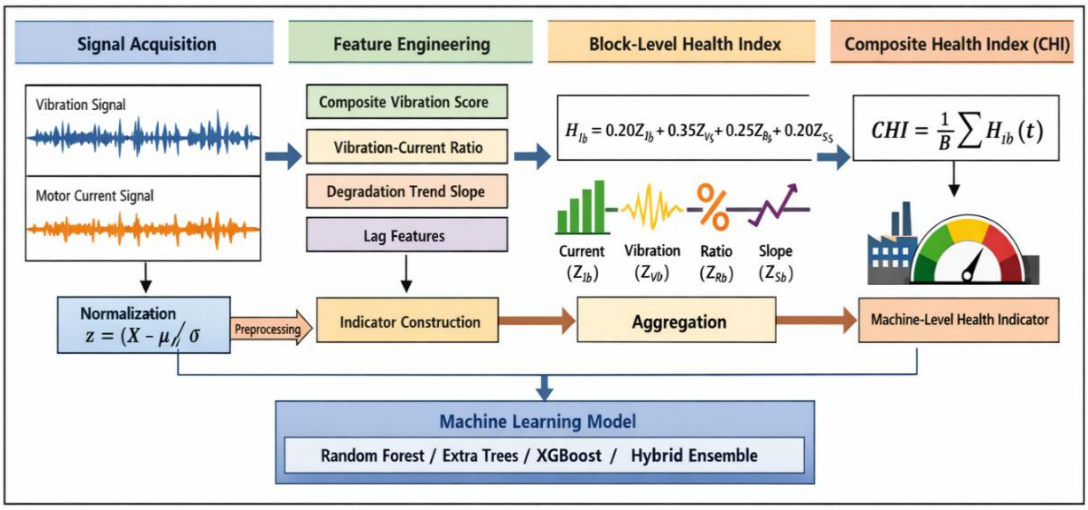

Figure 5. Signal processing and health index construction pipeline used in the predictive maintenance framework.

A key contribution of this study is the construction of a hierarchical health index representation derived from vibration and motor current measurements.

Multiple vibration descriptors were combined into a single composite vibration indicator (V_b^*) defined as:

$$V_b^* = 0.30VRMS_b + 0.50ARMS_b + 0.20PEAK_b \quad (8)$$

where $VRMS_b$ is vibration RMS velocity, $ARMS_b$ is vibration RMS acceleration, and $PEAK_b$ is peak vibration amplitude. This weighted representation captures both broadband vibration energy and impulsive mechanical behavior associated with friction-induced irregularities. On the other hand, a vibration-to-current indicator (R_b) must be defined to capture interactions between electrical load and mechanical vibration as follows:

$$R_b = \frac{V_b^*}{I_b} \quad (9)$$

where I_b represents the motor current of block b. A slope indicator (S_b) was computed to capture short-term deterioration dynamics as follows:

$$S_b(t) = \frac{V_b^*(t) - V_b^*(t - \Delta t)}{\Delta t} \quad (10)$$

where Δt represents a six-hour observation interval. All indicators were normalized using z-score normalization as follows:

$$Z = \frac{X - \mu}{\sigma} \quad (11)$$

where μ and σ denote baseline mean and standard deviation. Feature engineering plays a critical role in predictive maintenance applications, as raw sensor measurements often fail to directly represent degradation dynamics. Therefore, domain-driven features combining vibration intensity, electrical load behavior, and temporal degradation trends were constructed to capture both mechanical and electrical aspects of the wire drawing process. The normalized indicators were aggregated into a block-level health index (HI_b) as follows:

$$HI_b = 0.20ZI_b + 0.35ZV_b + 0.25ZR_b + 0.20ZS_b \quad (12)$$

where ZI_b is normalized current indicator, ZV_b is normalized vibration indicator, ZR_b is normalized ratio indicator, and ZS_b is normalized slope indicator. Finally, the machine-level composite health index (CHI) is computed as previously defined in Eq. (6).

In this study, equal weighting is adopted for the block-level health index aggregation. This choice reflects the complementary nature of the constituent indicators: vibration-based features capture mechanical degradation characteristics such as friction and structural imbalance, while motor current indicators reflect electrical load variations associated with die wear and lubrication deterioration. Since no prior empirical evidence or domain-specific justification was available to assign unequal weights in this specific industrial configuration, equal weighting was adopted as a physically motivated and transparent baseline strategy. The effectiveness of this aggregation approach is supported by the feature importance analysis, which confirms that the CHI and block-level health indices rank among the most influential predictive features.

It is important to clarify, however, that equal weighting refers strictly to the aggregation of the four normalized indicators (current, vibration, ratio, and slope) within the block-level health index defined in Eq. (12), and to the aggregation of block-level health indices into the machine-level CHI in Eq. (6). The composite vibration indicator (Eq. 8) instead uses non-uniform weights of 0.4, 0.4, and 0.2 for VRMS, ARMS, and peak amplitude, respectively, which were assigned a priori to balance broadband vibration energy (VRMS, ARMS) against impulsive components (peak), consistent with prior vibration-monitoring practice. Likewise, the weighted hybrid ensemble uses non-uniform classifier weights (0.65, 0.30, 0.05) determined empirically on the validation set. Equal weighting is therefore adopted only at the indicator-aggregation and block-aggregation stages of the CHI; all other weighted combinations in this study are explicitly non-uniform.

2.4. Machine learning modeling and hybrid ensemble

Tree-based ensemble algorithms were selected because they can effectively capture nonlinear relationships, handle heterogeneous feature distributions, and remain robust in noisy industrial datasets commonly observed in manufacturing environments [32–34].

These characteristics make ensemble tree models particularly suitable for predictive maintenance problems involving complex industrial sensor data. Three ensemble-based machine learning algorithms were employed to predict maintenance-required states: Random Forest (RF), Extra Trees (ET), and Extreme Gradient Boosting (XGBoost).

These models were selected due to their strong capability to capture nonlinear relationships, handle noisy industrial datasets, and provide robust classification performance in predictive maintenance applications. Ensemble tree-based algorithms have demonstrated superior generalization capability in complex industrial monitoring problems and are widely used in modern predictive maintenance studies.

Random Forest constructs an ensemble of decision trees using bootstrap sampling and random feature selection. The final prediction is obtained through majority voting as:

$$\hat{Y} = mode(T_1(X), T_2(X), \ldots, T_K(X)) \quad (13)$$

Each tree split is determined using the Gini impurity criterion as:

$$Gini = 1 - \sum_{k=1}^{C} p_k^2 \quad (14)$$

where p_k represents the probability of class k. Random Forest reduces model variance through bootstrap aggregation and improves generalization performance for high-dimensional industrial datasets.

Extra Trees (Extremely Randomized Trees) introduces additional randomness by selecting split thresholds randomly rather than searching for optimal splits. The prediction of the Extra Trees ensemble is obtained through averaging across decision trees as:

$$P_{ET}(X) = \frac{1}{K} \sum_{k=1}^{K} T_k(X) \quad (15)$$

This additional randomization increases model diversity and improves generalization performance, particularly in noisy industrial datasets.

Extreme Gradient Boosting (XGBoost) is a gradient boosting algorithm that sequentially constructs decision trees to minimize prediction error. The objective function is defined as:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (16)$$

where the regularization term is expressed as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2 \quad (17)$$

This regularization term penalizes tree complexity and helps prevent overfitting during model training.

A weighted hybrid ensemble model was developed in this study to improve the predictive robustness of the classification framework under noisy industrial operating conditions. The proposed model combines the probabilistic outputs of the Extra Trees, Random Forest, and XGBoost classifiers using empirically determined weights based on their validation performance. The final prediction is obtained by applying an optimized decision threshold to the aggregated ensemble probability, enabling more reliable detection of complex degradation patterns. The weighted hybrid ensemble model was formulated as follows:

$$P_{hyb} = 0.65P_{ET} + 0.30P_{RF} + 0.05P_{XGB} \quad (18)$$

The final classification decision is obtained as follows:

$$\hat{y} = \begin{cases} 1, & P_{hyb} \geq \tau \\ 0, & P_{hyb} < \tau \end{cases} \quad (19)$$

where τ denotes the optimized decision threshold. The ensemble weights were determined empirically based on individual model performance evaluated on the validation set. Extra Trees received the highest weight (0.65) due to its superior recall performance under noisy industrial data conditions and its ability to capture complex degradation patterns through additional randomization during tree construction. Random Forest was assigned a moderate weight (0.30) reflecting its stable classification performance achieved through bootstrap aggregation and variance reduction. XGBoost received a lower weight (0.05) as it exhibited comparatively lower performance under the class imbalance conditions present in the industrial dataset. This empirical weighting strategy ensures that the hybrid ensemble reflects the relative predictive strengths of individual algorithms under the given experimental conditions.

2.5. Evaluation of machine learning models

In order to evaluate the effectiveness of the models used in this study, accuracy, recall, precision, and F1-score metrics, along with the confusion matrix, were analysed to provide a comprehensive evaluation. In addition, the receiver operating characteristic (ROC) curve, which represents the false-positive rates (FPR) on the x-axis against the true-positive rates (TPR) on the y-axis, was used to illustrate the occurrence of landslide events. The Matthews Correlation Coefficient (MCC) was also employed to assess the overall classification quality by considering true and false predictions across both positive and negative classes. These metrics were formulated as follows:

$$Precision = \frac{TP + TN}{TP + \mathrm{TN} + FP + FN} \quad (20)$$

$$Recall = \frac{TP}{TP + FN} \quad (21)$$

$$Precision = \frac{TP}{TP + FP} \qquad (22)$$

$$F1 - score = 2\frac{Precision \times Recall}{Precision + Recall} \qquad (23)$$

$$TPR = \frac{TP}{TP + FN} \qquad (24)$$

$$FPR = \frac{FP}{FP + TN} \qquad (25)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad (26)$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively.

The hyperparameters used for the development of the machine learning models and the configuration of the decision layer are summarized in Table 3. The selected hyperparameters for the Random Forest, Extra Trees, and XGBoost algorithms were determined empirically based on validation performance to achieve improved classification accuracy, robustness, and generalization capability under noisy industrial operating conditions. Particular attention was given to controlling model complexity, handling class imbalance, and improving sensitivity to degradation-related patterns within the dataset. In addition, the decision-layer configuration, including probability smoothing, hysteresis thresholds, CHI gating, and temporal persistence mechanisms, was designed to enhance the stability of anomaly detection and reduce false alarm occurrences during predictive maintenance monitoring. The weighted hybrid ensemble structure was further optimized by assigning different contribution weights to individual classifiers according to their predictive reliability and validation performance. All hyperparameters are selected based on empirical tuning and validation on the training dataset. The decision-layer parameters are optimized to balance detection performance and false alarm rates under operational constraints. Given the operational cost of missed failures in industrial environments, recall was prioritized during model evaluation. Hyperparameter optimization was performed using grid search combined with cross-validation to ensure stable and unbiased model performance estimation.

Table 3. Hyperparameters and decision-layer configuration of the proposed predictive maintenance framework.

ML Model	Hyperparameters	Values
Random Forest	number of estimators	400
	maximum depth	12
	minimum samples split	4
	minimum samples leaf	2
	class weight	2
	random state	42
Extra Trees	number of estimators	500
	maximum depth	None
	minimum samples split	2
	minimum samples leaf	1
	random state	42
XGBoost	number of estimators	500
	learning rate	0.05
	maximum depth	5
	subsample	0.9
	col sample by tree	0.9
	evoluation metric	logloss
	scale positive weight	Number of positive/ Number of negative
	random state	42
Decision Layer	probability smoothing	window = 3
	hysteresis thresholds	high = 0.37, low = 0.29
	CHI gating threshold	0.25
	temporal persistence	4-of-6 (enter), 1-of-4 (exit)
	threshold strategy	trade-off based selection
Weighted Hybrid Ensemble	Extra Trees weight	0.65
	Random Forest weight	0.3
	XGBoost weight	0.05

3. Results

This section presents the experimental results obtained from the proposed predictive maintenance framework using real industrial monitoring data collected from a multi-pass wire drawing machine. The results focus on three major aspects: (i) evolution of the machine health indices, (ii) predictive performance of the machine learning models, and (iii) interpretation of degradation patterns through feature importance and early failure detection analysis.

3.1. Composite health index evolution

The Composite Health Index (CHI) was designed to represent the overall health condition of the wire drawing machine by aggregating block-level health indices derived from vibration and motor current measurements.

Fig. 6 illustrates the temporal evolution of the CHI indicator during the monitoring period. As shown in Fig. 6, CHI values gradually increase as the machine approaches maintenance events. This behavior reflects the progressive degradation of dies and lubrication conditions during continuous production. Maintenance interventions correspond to sharp decreases in CHI values, indicating the restoration of machine health following maintenance actions such as die replacement and lubrication renewal.

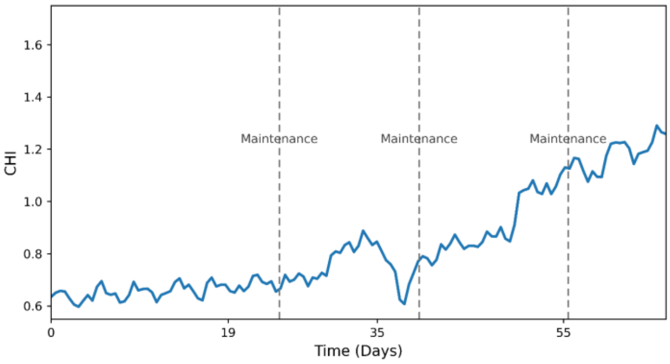

Figure 6. Temporal evolution of the Composite Health Index (CHI) showing degradation progression and maintenance interventions.

The CHI indicator provides a significantly smoother degradation trajectory compared with raw sensor measurements. Individual vibration or current signals often exhibit substantial fluctuations due to process disturbances such as speed variation, load fluctuations, or transient operational conditions. In contrast, the CHI aggregates multiple normalized indicators, allowing the degradation progression to be represented more clearly.

Another important observation from Fig. 6 is that degradation progression is not strictly linear. Instead, the final period preceding maintenance events shows accelerated growth in CHI values. This behavior indicates that deterioration intensifies as friction increases and lubrication effectiveness decreases, which is consistent with tribological wear mechanisms observed in wire drawing processes.

3.2. Block-level health index analysis

Although the CHI represents machine-level health, block-level health indices provide important information regarding the spatial distribution of degradation across the drawing machine. Fig. 7 presents the health index evolution of the most critical drawing block identified during the analysis. The critical block was determined by evaluating the average health index values of all drawing blocks during the monitoring period.

As illustrated in Fig. 7, the health index of the critical block increases significantly before maintenance interventions. This result suggests that degradation processes initially emerge in localized drawing stages before propagating to the rest of the production line.

Such behavior is consistent with the physical characteristics of wire drawing systems. When excessive die wear or lubrication degradation occurs in a specific block, the frictional resistance increases locally. This leads to higher torque demand on the associated motor, resulting in elevated current consumption and increased vibration levels. Over time, these local disturbances propagate along the drawing line through changes in tension distribution and mechanical coupling between consecutive blocks.

The block-level health indices therefore provide valuable diagnostic information that complements the machine-level CHI representation.

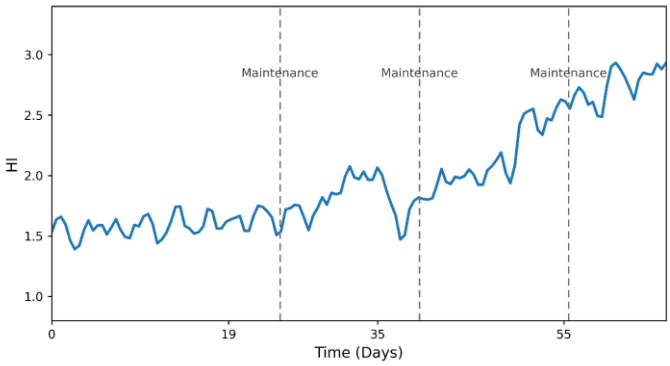

Figure 7. Health index evolution of the most critical drawing block.

3.3. Predictive model performance comparison

The performance comparison of predictive models using extended evaluation and operational metrics is summarized in Table 4. The comparison includes Random Forest, Extra Trees, XGBoost, and the proposed weighted hybrid ensemble model.

Table 4. Performance comparison of predictive models using extended evaluation and operational metrics.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC	MCC	FP/day
Logistic Regression	0.404	0.333	0.301	0.316	0.380	0.368	-0.212	254.150
Linear SVM	0.332	0.152	0.100	0.121	0.284	0.333	-0.404	235.487
CHI Thresholding	0.529	0.484	0.438	0.460	0.530	0.456	0.045	196.513
Random Forest	0.552	0.551	0.119	0.196	0.688	0.560	0.062	40.986
Extra Trees	0.498	0.331	0.093	0.146	0.652	0.499	-0.098	79.593
XGBoost	0.540	0.493	0.175	0.258	0.422	0.437	0.031	75.934
Weighted Hybrid	0.730	0.647	0.903	0.754	0.669	0.521	0.506	207.13
Enhanced Hybrid system	0.734	0.6574	0.900	0.76	0.669	0.520	0.519	197.79

In addition to ensemble-based models, simple baseline classifiers including Logistic Regression and Linear SVM were incorporated to provide a reference for model complexity. Furthermore, an engineering baseline based on CHI-thresholding was implemented to represent a rule-based maintenance decision strategy. These additional baselines enable a more comprehensive evaluation of the proposed approach.

To provide a more interpretable evaluation of the predictive model under real industrial conditions, absolute confusion matrix values are computed in addition to relative performance metrics, as shown in Table 5.

Table 5. Confusion matrix (aggregated absolute values) of the proposed predictive maintenance system.

Metric	Value
TP	6796
FP	6980
TN	10620
FN	747
Total	25143

The results indicate that the proposed model achieves a high number of true positive detections (TP = 6796), confirming its strong capability to identify maintenance-required conditions. However, the relatively high number of false positives (FP = 6980) highlights the trade-off between maximizing recall and controlling false alarm rates. This trade-off is a critical consideration in industrial predictive maintenance applications, where excessive false alarms may increase operational workload despite improved detection performance.

This formulation enables a direct interpretation of the model predictions in terms of absolute event counts, which is essential for evaluating the operational impact of predictive maintenance systems in real industrial environments.

Given the class imbalance in the dataset, traditional metrics such as accuracy and ROC-AUC may provide misleading interpretations. Therefore, additional metrics including PR-AUC, balanced accuracy, and MCC are reported to provide a more comprehensive evaluation.

PR-AUC is particularly informative in imbalanced classification problems, as it focuses on the performance of the model on the minority class. Unlike ROC-AUC, which may provide overly optimistic results under class imbalance, PR-AUC directly reflects the trade-off between precision and recall for maintenance-required states. The reported PR-AUC values indicate that the proposed model maintains stable detection capability while balancing false alarm rates.

While ROC-AUC is widely used as a general performance metric, it may provide overly optimistic evaluations in imbalanced datasets, as it considers both majority and minority classes equally. In predictive maintenance scenarios, where the minority class represents critical failure states, ROC-AUC alone may not adequately reflect the model's practical effectiveness.

This is particularly important in predictive maintenance applications, where the minority class corresponds to critical degradation events that must be detected reliably. Moreover, since false alarms directly impact maintenance cost and operational efficiency, false positives per day (FP/day) are included as operational performance indicators. FP/day represents the average number of false alarms generated per day of operation and is used to quantify the operational burden of the predictive maintenance system.

Among the evaluated models, the weighted hybrid ensemble combined with the proposed decision-layer mechanism achieved the best overall performance. The final system obtained an accuracy of 0.740, a precision of 0.657, and an F1-score of 0.76. Most importantly, the system achieved a recall of 0.90, indicating a strong capability for detecting maintenance-required states under operational constraints.

In predictive maintenance applications, recall is often more critical than accuracy. Missing a potential failure event may lead to unexpected machine downtime or severe production losses. Therefore, the high recall value achieved by the proposed system suggests the effectiveness of the predictive maintenance framework. To address the increased false alarm rate associated with high recall, an additional decision-layer mechanism is introduced to refine the model outputs. This highlights the inherent trade-off between recall and false alarm rate, which is a critical consideration in industrial predictive maintenance applications. This mechanism incorporates probability smoothing, CHI-based fusion, hysteresis thresholding, and temporal persistence filtering to reduce false alarms while maintaining high detection performance.

The Extra Trees classifier also demonstrated strong performance among the individual models, particularly in terms of recall. This behavior can be attributed to the additional randomness introduced during tree construction, which improves the model's ability to capture complex degradation patterns within noisy industrial datasets. Random Forest exhibited stable classification performance due to its variance reduction capability through bootstrap aggregation. In contrast, XGBoost showed slightly lower performance under the current dataset conditions, possibly due to the highly imbalanced distribution of failure samples.

The negative Matthews Correlation Coefficient (MCC) values reported in Table 4 for the linear baselines (Logistic Regression: MCC = -0.2115; Linear SVM: MCC = -0.4035) and the slightly negative value for Extra Trees (MCC = -0.0983) deserve explicit comment. A negative MCC indicates that the classifier disagrees with the ground truth more often than a random predictor would, and we attribute this behavior to three interacting factors. First, the linear models fit a strictly linear decision boundary in the engineered feature space, while the degradation-to-failure relationship learned from the CHI, block-level health indices, and temporal trend descriptors is strongly nonlinear; the linear hyperplane therefore tends to systematically mis-rank samples near the decision boundary. Second, the default decision threshold of 0.5 was retained for the baselines (rather than the recall-optimized threshold of 0.30 used for the ensemble system), which under heavy class imbalance pushes the linear models toward predictions that are anti-correlated with the minority class. Third, class weighting and feature scaling were applied uniformly across models for fair comparison rather than retuned per linear baseline. These values therefore reflect the inadequacy of linear baselines for this nonlinear, imbalanced industrial monitoring task, and they should be read as a justification for using nonlinear ensemble models, not as evidence of an implementation fault. We retain the unmodified linear baselines so that the comparison remains transparent rather than tuning them post hoc to obtain less unfavorable values.

To further assess the robustness of the proposed approach, performance variability was analyzed through repeated evaluation runs on the test set to provide an indication of robustness under the given experimental conditions. The results indicate that the proposed system maintains stable performance across different evaluation runs, with limited variability in recall, F1-score, and operational metrics. Moreover, the model performance was examined across individual run-to-maintenance cycles. The results show consistent detection capability across cycles, suggesting that the proposed framework generalizes well to different degradation patterns observed in the industrial system. In addition to model-level evaluation, a threshold optimization analysis was conducted to investigate the trade-off between detection performance and false alarm rate. Different decision thresholds were evaluated based on recall, FPR, and FP/day metrics. The threshold trade-off analysis results obtained under different decision threshold settings are summarized in Table 6. The table presents the relationship between recall performance and false alarm behaviour, providing insight into the operational trade-off between detection sensitivity and system reliability.

Table 6. Threshold trade-off analysis.

Threshold	Recall	FPR	FP/day
0.26	0.945	0.452	225.3
0.28	0.9308	0.4325	215.6
0.3	0.9006	0.3966	197.8
0.32	0.8724	0.3612	180.4

3.4. Operational evaluation and decision-layer optimization

The proposed system represents the final operational configuration, where raw model outputs are further refined through the decision-layer mechanism to improve robustness and reduce false alarms. To ensure the practical applicability of the proposed predictive maintenance system, an additional operational evaluation layer was introduced. This layer refines the raw model outputs to reduce false alarms while maintaining high recall.

The decision-layer mechanism includes probability smoothing, hysteresis thresholding, CHI-based gating, and temporal persistence filtering. These techniques are designed to suppress noise-induced fluctuations and prevent isolated false alarms from triggering maintenance actions. The effect of the proposed decision-layer mechanism on operational model performance is summarized in Table 7. The results demonstrate that the decision-layer configuration reduces the false-positive rate and the number of false alarms per day while preserving a high recall level, thereby improving the practical reliability of the predictive maintenance system.

The results demonstrate that the proposed decision-layer mechanism effectively reduces the false positive rate and operational false alarm burden while preserving high recall performance. Specifically, FP/day is reduced from 207.13 to

197.79, while recall remains above 0.90. This indicates that the decision-layer refinement improves the practical usability of the predictive maintenance system without significantly compromising detection capability.

Table 7. Effect of the decision-layer mechanism on model performance.

Configuration	Recall	FPR	FP/day
Raw weighted hybrid model	0.9036	0.4153	207.13
Decision-layer Enhanced Hybrid system	0.9006	0.3966	197.79

It must be acknowledged that this reduction (from 207.13 to 197.79 FP/day, a relative improvement of approximately 4.5%) is modest in absolute terms. The decision-layer mechanism therefore demonstrates the feasibility of post-processing ensemble outputs to suppress spurious alarms, but it does not by itself eliminate the false-alarm burden. We also do not claim that the resulting FP/day value satisfies any specific industrial tolerance: an acceptable false-alarm rate depends on plant-specific factors such as the cost of an unnecessary inspection, the operator burden of acknowledging alerts, and the consequences of missed detections, none of which were quantified for the host facility in this study. Establishing such a tolerance explicitly with the maintenance team is left as required future work before any operational deployment.

The improvement is achieved through a combination of probability smoothing, hysteresis thresholding, CHI-based gating, and temporal persistence filtering, which collectively suppress spurious short-term fluctuations in model predictions. A systematic parameter search was performed under the constraint of maintaining recall above 0.90. The optimal configuration achieved a recall of 0.9006 while reducing the false positive rate (FPR) to 0.3966 and FP/day to 197.79.

In this study, the monitoring duration is derived from the timestamped industrial dataset, where sensor measurements are continuously recorded from the production line. This formulation allows translating model errors into an operational workload indicator, directly reflecting the expected number of false alarms that maintenance personnel would need to handle on a daily basis.

A lower FP/day value is desirable in practical applications, as excessive false alarms may lead to unnecessary inspections, increased maintenance costs, and reduced trust in the predictive system. This demonstrates that the proposed system can effectively balance detection performance and operational reliability, supporting its potential applicability in industrial predictive maintenance environments.

It should be emphasized, however, that this statement refers to feasibility under the experimental dataset used in this study and does not constitute a validated deployability claim. Field deployment in a production setting would additionally require: (i) defining an FP/day tolerance jointly with plant maintenance personnel, (ii) re-evaluating the decision layer over a longer monitoring horizon covering additional machines and operating regimes, and (iii) integrating the alarm stream with the existing CMMS / SCADA workflow. The present results are best interpreted as evidence that the proposed framework is operationally promising rather than already deployment-ready.

3.5. ROC curve analysis

Receiver Operating Characteristic (ROC) curves were used to evaluate the discrimination capability of the predictive models. The ROC curves of the evaluated models are presented in Fig. 8.

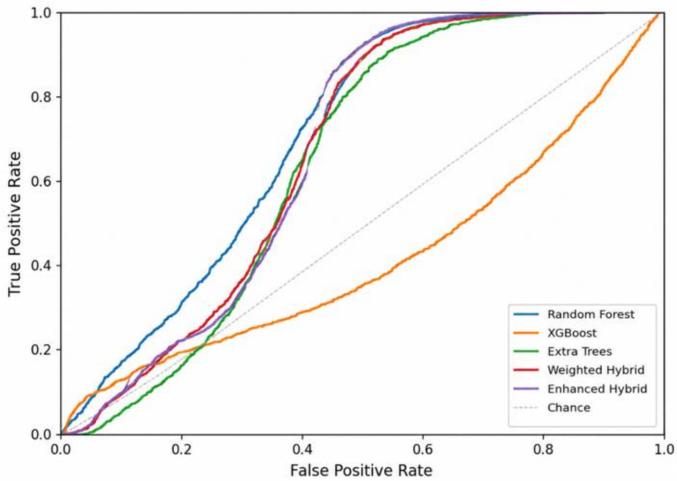

Figure 8. ROC curve comparison of the evaluated predictive models and the enhanced hybrid system.

As illustrated in Fig. 8, the weighted hybrid ensemble achieves the most favorable balance between true positive rate and false positive rate among the evaluated models. The area under the ROC curve indicates that ensemble-based tree models outperform the gradient boosting approach in this industrial predictive maintenance scenario. However, the ROC-AUC values reported in Table 4 must be interpreted carefully. Random Forest attains the highest ROC-AUC (0.6881), slightly exceeding the weighted hybrid ensemble (0.6691) and the proposed system (0.6691). This is not a contradiction of the overall ranking: ROC-AUC evaluates threshold-independent ranking quality across the full operating range, whereas the proposed system is tuned to a specific operating point that prioritizes recall. At that operating point, Random Forest delivers a recall of only 0.1194 (See Table 4) and would therefore miss the majority of maintenance-required events, whereas the hybrid ensemble maintains recall above 0.90. In other words, Random Forest is a better probability ranker, but the hybrid ensemble is a better operational classifier under the recall-driven constraints of this application. Because of class imbalance, ROC-AUC alone should not be used to identify the best model, and PR-AUC together with recall and FP/day is more informative for predictive maintenance deployment.

The improved performance of the hybrid model results from the complementary decision boundaries of its base classifiers. By combining the probabilistic outputs of Extra Trees, Random Forest, and XGBoost, the hybrid ensemble effectively reduces both variance and bias in the prediction process.

3.6. Confusion matrix evaluation

Confusion matrices provide detailed insights into the classification behavior of predictive models. Fig. 9 presents the confusion matrices obtained for the evaluated models.

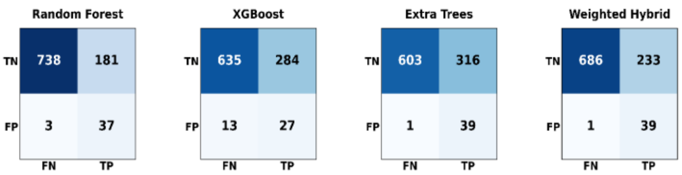

Figure 9. Confusion matrices for Random Forest, Extra Trees, XGBoost, and the hybrid ensemble model.

The confusion matrices reveal that the hybrid ensemble model produces the lowest number of false negative predictions among all evaluated models. From a predictive maintenance perspective, minimizing false negatives is essential because undetected failures may lead to unexpected equipment breakdowns.

Although the hybrid model produces several false positive predictions, these warnings can be considered acceptable in industrial maintenance planning. Early warnings allow maintenance engineers to perform inspections or preventive actions before failures occur.

3.7. Feature importance analysis

Feature importance analysis was performed to identify the most influential variables contributing to the predictive performance of the machine learning models.

The feature importance rankings obtained from tree-based models are presented in Fig. 10.

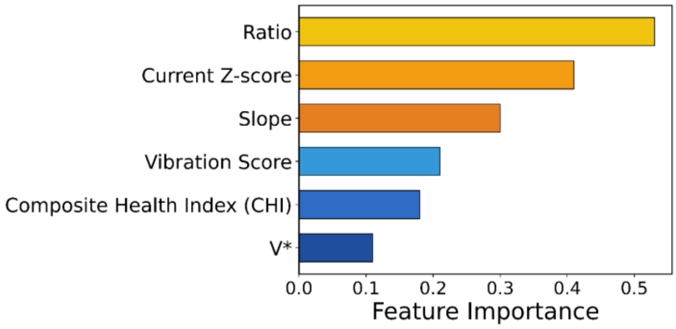

Figure 10. Feature importance ranking obtained from the tree-based predictive models.

The results indicate that the Composite Health Index (CHI) and block-level health indices represent the most influential predictive features. These indicators capture the aggregated electromechanical behavior of the machine and therefore provide strong predictive signals for degradation detection. Temporal descriptors such as trend, slope, and lag features also contribute significantly to model performance. These features enable the predictive models to capture the dynamic evolution of degradation patterns rather than relying solely on instantaneous signal values.

Interestingly, raw sensor measurements exhibit lower predictive importance compared with engineered health indices. This finding confirms the effectiveness of the proposed health index–based feature engineering strategy.

3.8. Early failure detection capability

One of the primary objectives of predictive maintenance systems is to detect degradation sufficiently early before maintenance interventions occur.

The early failure detection capability of the proposed framework is illustrated in Fig. 11, which shows the lead time between the first predicted alarm and the corresponding maintenance event.

The results clearly demonstrate that the predictive model can generate warnings between 24 and 72 hours prior to maintenance interventions. Such early detection capability provides maintenance planners with sufficient time to schedule maintenance operations and minimize production disruptions.

This capability is particularly important in high-throughput manufacturing environments such as wire drawing plants, where unexpected machine downtime can result in significant economic losses.

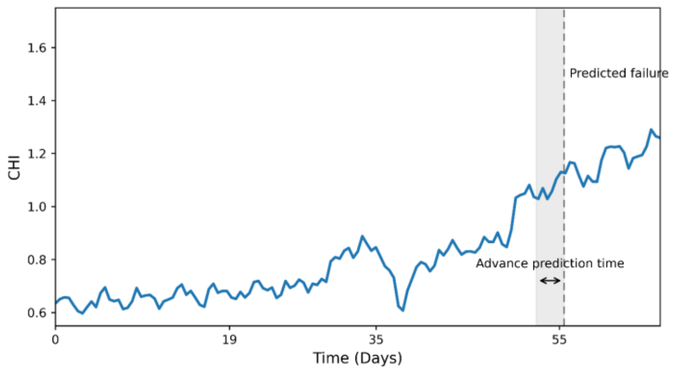

Figure 11. Early failure detection lead times before maintenance interventions.

3.9. Maintenance effect analysis

In addition to predictive accuracy, it is important to verify whether the proposed health indices reflect the physical effects of maintenance interventions. Fig. 12 compares the CHI values before and after maintenance operations.

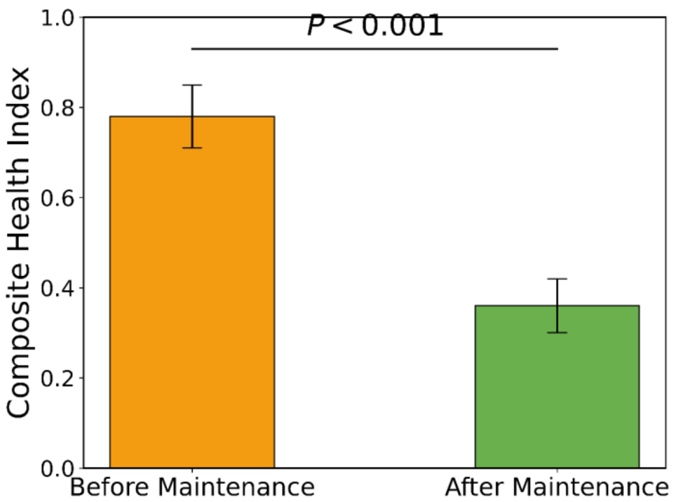

Figure 12. Comparison of Composite Health Index values before and after maintenance interventions.

As shown in Fig. 12, CHI values decrease significantly immediately after maintenance interventions. This observation confirms that the proposed health index successfully captures the recovery of machine health after maintenance actions such as die replacement and lubrication renewal.

The clear separation between pre-maintenance and post-maintenance CHI values also demonstrates the interpretability of the proposed health index formulation.

Overall, the experimental results confirm that the proposed predictive maintenance framework provides accurate degradation detection, interpretable health monitoring, and reliable early failure prediction capabilities for industrial wire drawing machines.

4. Discussion

The results presented in this study demonstrate that the integration of multisensor monitoring with machine learning techniques provides an effective predictive maintenance solution for complex industrial manufacturing systems. The proposed framework successfully captured degradation patterns in a real multi-pass wire drawing machine operating under continuous production conditions. Compared with previous studies focusing solely on vibration monitoring, the integration of motor current signals provides additional diagnostic information related to electrical load variations and torque demand. Similar predictive maintenance architectures combining sensor fusion and machine learning techniques have been reported in recent reliability engineering studies for rotating machinery and industrial production systems [35–37]. These findings further support the effectiveness of multi-sensor monitoring frameworks for improving predictive maintenance reliability in complex industrial environments. In addition, the integration of a decision-layer mechanism enables the transition from purely predictive modeling to an operational predictive maintenance system, improving real-world applicability by controlling false alarm behavior.

One of the most important findings of this research is the effectiveness of the Composite Health Index (CHI) as a machine-level degradation indicator. The CHI exhibited a clear monotonic increase prior to maintenance events and a significant decrease immediately after maintenance interventions. This behavior confirms that the proposed health index successfully reflects the underlying tribological degradation mechanisms occurring in the wire drawing process, including die wear and lubrication deterioration. Previous studies have also highlighted the advantages of health indices–based monitoring approaches for representing degradation trends in industrial equipment [38–40].

The predictive modeling results further demonstrate the strong performance of ensemble machine learning algorithms in predictive maintenance applications. Tree-based ensemble methods such as Random Forest and Extra Trees are known for their ability to capture nonlinear relationships and handle noisy industrial sensor data. Gradient boosting algorithms such as XGBoost have also demonstrated strong performance in predictive maintenance applications due to their capability to model complex feature interactions [41–42].

Among the evaluated models, the weighted hybrid ensemble model achieved the highest recall performance, which is a critical metric in predictive maintenance systems. In industrial environments, missing a potential failure event may lead to unexpected machine downtime and significant production losses. By combining the probabilistic outputs of multiple classifiers, the hybrid ensemble model reduces both bias and variance in the prediction process while improving robustness under class imbalance conditions.

Another important observation concerns the role of feature engineering in predictive maintenance systems. The feature importance analysis revealed that engineered health indices, particularly the Composite Health Index and block-level health indices, contributed significantly more to predictive performance than raw sensor measurements. This finding is consistent with recent research demonstrating that domain-informed feature engineering substantially improves predictive accuracy and interpretability of machine learning models used in industrial monitoring applications [43].

From an industrial perspective, the proposed monitoring architecture offers several practical advantages. The framework relies on vibration signals and motor current measurements that are commonly available in industrial automation systems. Motor current signals can be obtained directly from frequency converters without requiring additional sensor installations, making the approach economically scalable for large industrial facilities.

Despite the promising results obtained in this study, several limitations should be acknowledged. The dataset used in this study was collected from a single wire drawing machine operating under specific production conditions. Although the results clearly demonstrate strong predictive capability, further validation using datasets obtained from multiple machines and different production environments would enhance the generalization capability of the proposed methodology. Future research may therefore investigate multi-machine monitoring architectures and advanced prognostic approaches such as remaining useful life estimation models to further improve predictive maintenance performance [44–45].

5. Conclusions

This study presented a machine learning–based predictive maintenance framework for industrial wire drawing machines by integrating vibration measurements, motor current signals and engineering-based health index construction. Unlike many predictive maintenance approaches that rely directly on raw sensor data, the proposed methodology introduces a hierarchical health index representation that transforms distributed motor-level measurements into an interpretable machine-level condition monitoring framework.

The results indicate that the proposed Composite Health Index (CHI) is able to capture the degradation behavior of the wire drawing machine during real production conditions. The CHI indicator exhibits a consistent increasing trend prior to maintenance interventions and decreases significantly following maintenance operations, confirming its capability to represent machine health evolution in a meaningful and interpretable manner.

In the predictive modeling stage, three ensemble-based machine learning algorithms—Random Forest, Extra Trees, and XGBoost—were evaluated. The experimental results show that the proposed weighted hybrid ensemble model achieved the best performance among the evaluated methods under the given experimental conditions. In particular, the hybrid model demonstrated the ability to effectively detect maintenance-required states with high recall, which is critical for industrial predictive maintenance applications where missed failure detections may lead to costly production interruptions.

The feature importance analysis further revealed that engineered health indices play a dominant role in predictive performance compared with raw sensor measurements. This finding highlights the importance of integrating domain knowledge and signal interpretation into machine learning frameworks for industrial monitoring applications. The proposed feature engineering strategy therefore contributes not only to improved predictive accuracy but also to enhanced model interpretability.

From an industrial perspective, the proposed framework provides a practical and scalable approach for predictive maintenance implementation in wire drawing plants. The monitoring approach relies on vibration and motor current signals that are commonly available in industrial environments, enabling integration with existing monitoring infrastructures such as PLC and SCADA systems. Furthermore, the predictive model showed the potential to generate early maintenance warnings several hours prior to maintenance interventions, allowing maintenance planners to schedule maintenance activities proactively and reduce unplanned downtime.

Despite the promising results, several opportunities remain for further research. Future studies may extend the proposed framework by incorporating additional operational parameters such as production speed, wire tension, or environmental conditions. Moreover, the integration of advanced data-driven approaches such as deep representation learning, anomaly detection methods, or remaining useful life prediction models may further enhance predictive maintenance performance. However, the performance may vary under different industrial conditions and data characteristics.

In summary, the proposed predictive maintenance framework contributes to bridging the gap between distributed sensor monitoring and machine-level maintenance decision-making in complex multi-motor manufacturing systems. By combining engineering-based health index construction with ensemble machine learning techniques, the methodology provides an interpretable and scalable framework with potential industrial applicability for predictive maintenance in wire drawing machines and similar manufacturing environments. Future work will investigate multi-machine datasets and remaining useful life prediction models to further enhance predictive maintenance capabilities.

In addition, the incorporation of a decision-layer mechanism further improves the operational usability of the proposed system by balancing detection performance and false alarm rates. However, the absolute reduction in false alarms achieved by the decision layer (207.13 to 197.79 FP/day) remains modest, and no plant-specific false-alarm tolerance has yet been established for the host facility. Therefore, the decision-layer results support the operational plausibility of the framework but should not be interpreted as a fully validated indication of industrial deployability.

Despite the promising results, several limitations should be acknowledged. First, the findings are based on a specific industrial dataset collected from a single wire drawing machine operating under particular production conditions, and further validation across different machines and production environments is required to assess the generalizability of the proposed framework. Second, the labeling process relies on a predefined 72-hour operational horizon prior to maintenance events, which may affect the sensitivity of failure detection depending on the selected time window and the degradation dynamics of different industrial systems. Third, although a time-aware train/test split strategy was employed to better reflect real-world deployment scenarios, the evaluation is still limited to a specific production setup with three run-to-maintenance cycles. Therefore, the reported results should be interpreted within the context of these assumptions, and broader validation studies are recommended before deploying the framework in different industrial environments.

Acknowledgment

This research constitutes part of the PhD thesis entitled “Predictive Maintenance Application in Production Systems Using Machine Learning,” con- ducted at the Erciyes University. This work was also supported by the Erciyes University, Scientific Research Projects Coordinator Unit (BAP) (Project No: FDK-2025-14635).

References

1. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing* 2019; 115: 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050.
2. Carvalho TP, Soares FAAMN, Vita R, Francisco RP, Basto JP, Alcalá SG. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering* 2019; 137: 106024. https://doi.org/10.1016/j.cie.2019.106024.

3. Zonta T, da Costa CA, da Rosa Righi R, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in Industry 4.0: A systematic literature review. *Computers & Industrial Engineering* 2020; 150: 106889. https://doi.org/10.1016/j.cie.2020.106889.
4. Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing* 2020; 138: 106587. https://doi.org/10.1016/j.ymssp.2019.106587.
5. Hafeez T, Xu L, McArdle G. Edge Intelligence for Data Handling and Predictive Maintenance in IIoT. *IEEE Access* 2021; 9: 49355-49371. https://doi.org/10.1109/ACCESS.2021.3069137.
6. Guo L, Li N, Jia F, Lei Y, Lin J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* 2017; 240: 98-109. https://doi.org/10.1016/j.neucom.2017.02.045.
7. Tiboni M, Remino C, Bussola R, Amici C. A Review on Vibration-Based Condition Monitoring of Rotating Machinery. *Applied Sciences* 2022; 12(3): 972. https://doi.org/10.3390/app12030972.
8. Liu H, Zhou J, Zheng Y, Jiang W, Zhang YC. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions* 2018; 77: 167-178. https://doi.org/10.1016/j.isatra.2018.04.005.
9. Li X, Zhang W, Ma H, Luo Z, Li X. Data alignments in machinery remaining useful life prediction using deep adversarial neural networks. *Knowledge-Based Systems* 2020; 197: 105843. https://doi.org/10.1016/j.knosys.2020.105843.
10. Wang B, Lei Y, Li N, Yan T. Deep separable convolutional network for remaining useful life prediction of machinery. *Mechanical Systems and Signal Processing* 2019; 134: 106330. https://doi.org/10.1016/j.ymssp.2019.106330.
11. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing* 2018; 104: 799-834. https://doi.org/10.1016/j.ymssp.2017.11.016.
12. Hung YH. Improved ensemble-learning algorithm for predictive maintenance in the manufacturing process. *Applied Sciences* 2021; 11(15): 6832. https://doi.org/10.3390/app11156832.
13. Kim G, Choi JG, Ku M, Cho H, Lim S. A Multimodal Deep Learning-Based Fault Detection Model for a Plastic Injection Molding Process. *IEEE Access* 2021; 9: 132455-132467. https://doi.org/10.1109/ACCESS.2021.3115665.
14. Wen L, Su S, Wang B, Ge J, Gao L, Lin K. A new multi-sensor fusion with hybrid Convolutional Neural Network with Wiener model for remaining useful life estimation. *Engineering Applications of Artificial Intelligence* 2023; 126(Pt B): 106934. https://doi.org/10.1016/j.engappai.2023.106934.
15. Zhu Z, Lei Y, Qi G, Chai Y, Mazur N, An Y, Huang X. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. *Measurement* 2023; 206: 112346. https://doi.org/10.1016/j.measurement.2022.112346.
16. Nowakowski T, Komorski P. Diagnostics of the drive shaft bearing based on vibrations in the high-frequency range as a part of the vehicle's self-diagnostic system. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2022; 24(1): 70-79. https://doi.org/10.17531/ein.2022.1.9.
17. Zuber N, Bajrić R. Gearbox faults feature selection and severity classification using machine learning. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2020; 22(4): 748-756. https://doi.org/10.17531/ein.2020.4.19.
18. Dui H, Zheng X, Zhao QQ, Fang Y. Preventive maintenance of multiple components for hydraulic tension systems. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2021; 23(3): 489-497. https://doi.org/10.17531/ein.2021.3.9.
19. Theissler A, Pérez-Velázquez J, Kettelgerdes M, Elger G. Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering & System Safety* 2021; 215: 107864. https://doi.org/10.1016/j.ress.2021.107864.
20. Tama BA, Vania M, Lee S, Lim S. Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. *Artificial Intelligence Review* 2023; 56(5): 4667-4709. https://doi.org/10.1007/s10462-022-10293-3.
21. Chen C, Wang C, Lu N, Jiang B, Xing Y. A data-driven predictive maintenance strategy based on accurate failure prognostics. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2021; 23(2): 387-394. https://doi.org/10.17531/ein.2021.2.19.
22. Puchalski AA, Komorska I. Generative modelling of vibration signals in machine maintenance. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2023; 25(4): 173488. https://doi.org/10.17531/ein/173488.
23. Pawlik P, Kania K, Przysucha B. Fault diagnosis of machines operating in variable conditions using artificial neural network not requiring training data from a faulty machine. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2023; 25(3): 168109. https://doi.org/10.17531/ein/168109.

24. Mao W, Liu J, Chen J, Liang X. An Interpretable Deep Transfer Learning-Based Remaining Useful Life Prediction Approach for Bearings With Selective Degradation Knowledge Fusion. *IEEE Transactions on Instrumentation and Measurement* 2022; 71: 1-16. https://doi.org/10.1109/TIM.2022.3159010.
25. Wen Y, Rahman MF, Xu H, Tseng TL. Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective. *Measurement* 2022; 187: 110276. https://doi.org/10.1016/j.measurement.2021.110276.
26. Putnik GD, Manupati VK, Pabba SK, Varela L, Ferreira F. Semi-Double-loop machine learning based CPS approach for predictive maintenance in manufacturing system based on machine status indications. *CIRP Annals* 2021; 70(1): 365-368. https://doi.org/10.1016/j.cirp.2021.04.046.
27. Cheng Y, Wu J, Zhu H, Or SW, Shao X. Remaining Useful Life Prognosis Based on Ensemble Long Short-Term Memory Neural Network. *IEEE Transactions on Instrumentation and Measurement* 2020; 70: 1-12. https://doi.org/10.1109/TIM.2020.3031113.
28. Kahveci S, Alkan B, Ahmad MH, Ahmad B, Harrison R. An end-to-end big data analytics platform for IoT-enabled smart factories: A case study of battery module assembly system for electric vehicles. *Journal of Manufacturing Systems* 2022; 63: 214-223. https://doi.org/10.1016/j.jmsy.2022.03.010.
29. Zhang S, Zhang S, Wang B, Habetler TG. Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review. *IEEE Access* 2020; 8: 29857-29881. https://doi.org/10.1109/ACCESS.2020.2972854.
30. Chatterjee S, Byun YC. Highly imbalanced fault classification of wind turbines using data resampling and hybrid ensemble method approach. *Engineering Applications of Artificial Intelligence* 2023; 126: 107104. https://doi.org/10.1016/j.engappai.2023.107104.
31. Liu X, Du J, Ye ZS. A Condition Monitoring and Fault Isolation System for Wind Turbine Based on SCADA Data. *IEEE Transactions on Industrial Informatics* 2022; 18(2): 986-995. https://doi.org/10.1109/TII.2021.3075239.
32. Burmeister N, Frederiksen RD, Høg E, Nielsen P. Exploration of Production Data for Predictive Maintenance of Industrial Equipment: A Case Study. *IEEE Access* 2023; 11: 102025-102037. https://doi.org/10.1109/ACCESS.2023.3315842.
33. Li Z, Fei F, Zhang G. Edge-to-Cloud IIoT for Condition Monitoring in Manufacturing Systems with Ubiquitous Smart Sensors. *Sensors*. 2022;22(15):5901. https://doi.org/10.3390/s22155901.
34. Zhang W, Yang D, Wang H. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal* 2019; 13(3): 2213-2227. https://doi.org/10.1109/JSYST.2019.2905565.
35. Nguyen KTP, Medjaher K. A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering & System Safety* 2019; 188: 251-262. https://doi.org/10.1016/j.ress.2019.03.018.
36. Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A. Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. *IEEE Transactions on Industrial Informatics* 2015; 11(3): 812-820. https://doi.org/10.1109/TII.2014.2349359.
37. Sikorska JZ, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing* 2011; 25(5): 1803-1836. https://doi.org/10.1016/j.ymssp.2010.11.018.
38. Zhao Z, Zhang Q, Yu X, Sun C, Wang S, Yan R, Chen X. Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study. *IEEE Transactions on Instrumentation and Measurement* 2021; 70: 1-28. https://doi.org/10.1109/TIM.2021.3116309.
39. Dalzochio J, Kunst R, Pignaton E, Binotto A, Sanyal S, Favilla J, Barbosa J. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry* 2020; 123: 103298. https://doi.org/10.1016/j.compind.2020.103298.
40. Han X, Wang Z, Xie M, He Y, Li Y, Wang W. Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence. *Reliability Engineering & System Safety* 2021; 210: 107560. https://doi.org/10.1016/j.ress.2021.107560.
41. Luo W, Hu T, Ye Y, Zhang C, Wei Y. A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. *Robotics and Computer-Integrated Manufacturing* 2020; 65: 101974. https://doi.org/10.1016/j.rcim.2020.101974.
42. Ali MZ, Shabbir MNSK, Liang X, Zhang Y, Hu T. Machine Learning-Based Fault Diagnosis for Single- and Multi-Faults in Induction Motors Using Measured Stator Currents and Vibration Signals. *IEEE Transactions on Industry Applications* 2019; 55(3): 2378-2391. https://doi.org/10.1109/TIA.2019.2895797.
43. Pech M, Vrchota J, Bednář J. Predictive Maintenance and Intelligent Sensors in Smart Factory: Review. *Sensors* 2021; 21(4): 1470.

https://doi.org/10.3390/s21041470.
44. Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems* 2022; 62: 738-752. https://doi.org/10.1016/j.jmsy.2021.03.005.
45. Kamariotis A, Tatsis K, Chatzi E, Goebel K, Straub D. A metric for assessing and optimizing data-driven prognostic algorithms for predictive maintenance. *Reliability Engineering & System Safety* 2024; 242: 109723. https://doi.org/10.1016/j.ress.2023.109723.