machines

MDPI

Article

Machine Learning-Enhanced State-Aware Health Assessment of Industrial Assets Under Zero-Label Constraints

Dominik Hornacek and Pavol Tanuska *

Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava, Jána Bottu 25, 917-24 Trnava, Slovakia; dominik.hornacek@stuba.sk
* Correspondence: pavol.tanuska@stuba.sk; Tel.: +421-902-390-091

Abstract

Industrial health assessment often faces the challenge of sensor scarcity and a lack of labelled failure datasets, making conventional monitoring difficult to scale. This study addresses these constraints by proposing a state-aware framework that relies exclusively on routinely measured electrical parameters (active power, current, voltage, and power factor). The main challenge lies in distinguishing benign load variations from actual degradation without process-level context. To overcome this, we integrate automated operating-state recognition using XGBoost with per-state regression modelling to estimate the expected active power. A standardized Health Index (HI) is then derived from the residuals to quantify deviations from normal behaviour. Evaluated on a fleet of three-phase injection moulding machines, the framework demonstrates substantial performance improvements: the state-aware approach increased the median coefficient of determination from 0.64 to 0.86 and reduced residual variability by 30% compared to context-agnostic models. These findings show that synergistic system integration provides a stable and interpretable indicator for early degradation detection and fleet-level benchmarking under strict zero-label industrial constraints.

Keywords: state-aware condition monitoring; health index; ESA; data-driven modelling

Check for updates

Academic Editor: Jong-Myon Kim

Received: 19 January 2026
Revised: 11 February 2026
Accepted: 17 February 2026
Published: 23 February 2026

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Industrial production systems increasingly depend on continuous health monitoring to ensure reliability, safety, and energy efficiency [1–3]. Conventional approaches rely on additional vibration, temperature, or acoustic sensors, which require physical access to equipment and increase maintenance costs. In contrast, electrical parameters such as current, voltage, and power factor are readily available from standard industrial metering systems and inherently reflect the electromechanical behaviour of machines. This makes electrical parameter-based monitoring a practical non-intrusive alternative [4].

The transition toward Industry 4.0 has placed significant emphasis on the digital twin concept and predictive maintenance (PdM) as key drivers for operational excellence. In energy-intensive sectors, such as plastic processing, unplanned downtime can lead to substantial financial losses due to wasted raw materials and disrupted production cycles. While vibration and acoustic monitoring remain the gold standards for mechanical diagnostics, their deployment across entire factories is often hindered by the high cost of specialized sensors and the complexity of data synchronization. Consequently, there is a growing industrial demand for "lean" monitoring solutions that leverage pre-existing infrastructure. Monitoring electrical consumption patterns offers a dual advantage: it serves

as a non-intrusive proxy for mechanical health while simultaneously providing insights into energy efficiency, aligning technical maintenance with corporate sustainability goals.

However, traditional Electrical Signature Analysis (ESA) and its derivatives are primarily designed for steady-state operation. Their performance decreases under variable loads, transient regimes, or multi-mode production cycles, which are common in modern industrial environments [5–7]. Time–frequency and wavelet-based extensions improve sensitivity to transient events but require careful parameter tuning and still do not explicitly account for distinct operating regimes [8,9]. Likewise, recent machine learning approaches—although capable of modelling nonlinear relationships in electrical data—typically depend on labelled fault datasets and often ignore the influence of the operating state, resulting in reduced robustness and a high rate of false alarms [1,10].

Industrial machines rarely operate in a single uniform mode; they cycle through idle, heating, and production phases, each with its own characteristic electrical signature [7,10]. Global anomaly detection methods that do not consider these states can misinterpret normal load-driven variability as degradation [1,10]. This highlights the need for a state-aware, data-driven framework that evaluates machine behaviour relative to the appropriate operating context.

Motivated by these limitations, this work investigates whether state-aware modelling based solely on routinely measured electrical parameters can provide a reliable, fully non-intrusive health assessment without requiring explicit fault annotations [1,7,10].

To address this, we propose a unified state-based predictive framework that integrates the following:

- Automatic operating-state recognition using XGBoost classification [10];
- Per-state regression of expected active power [10];
- A standardized, reproducible Health Index (HI) derived from residual deviations.

By relying exclusively on standard electrical quantities—active power, current, voltage, and power factor—the framework ensures low deployment cost, high scalability, and compatibility with existing metering infrastructure [1,10]. The approach is consistent with ISO 17359 and ISO 13379 guidelines for data-driven condition monitoring and is applicable across diverse industrial assets [11–13].

The novelty of this work does not lie in the individual components, such as state recognition using XGBoost or fleet-level aggregation, which are well established in the literature. Rather, the novelty arises from the integration of these components into a unified framework that operates under strict industrial constraints. This system is designed to work in zero-label environments, where fault data are not explicitly available. By synergistically combining operating-state recognition with per-state regression and Health Index computation, the framework can continuously monitor machine health without relying on fault-specific labels or additional sensors. The integration of these components provides a scalable, cost-effective solution that meets the real-world needs of Industry 4.0 applications.

2. Literature Review

To maintain focus and relevance, the review in this section concentrates on diagnostic methods that operate on electrical parameters and are directly comparable to the state-aware, regression-based framework developed in this work. Recent studies [1,10] further emphasize the growing shift toward electrical signature-based monitoring due to its non-intrusiveness and compatibility with existing metering systems. Broader fault detection techniques that rely on vibration, acoustic analysis, or high-frequency waveform sensing are therefore mentioned only insofar as they relate to electrical signature-based monitoring.

Electrical parameter-based diagnostics have evolved from traditional frequency-domain analysis to hybrid, data-driven models that combine physical interpretability

with machine learning flexibility [5,7]. This evolution reflects the growing need for scalable, non-intrusive, and intelligent condition monitoring in industrial systems [1,10]. The following subsections summarize the main methodological directions—frequency-domain ESA, time–frequency analysis, data-driven modelling, and state-aware hybrid frameworks—and outline current gaps that motivate the present work.

2.1. Frequency-Domain Electrical Signature Analysis

Early developments in Electrical Signature Analysis (ESA) relied on frequency-domain methods that examine the spectral content of stator currents and voltages to detect characteristic fault signatures. Glowacz et al. [4] first demonstrated that broken rotor bars produce sidebands in the current spectrum, while Benbouzid [5] and Bhosinak et al. [6] extended this principle to eccentricity and bearing defects. These approaches remain valued for their clear physical interpretability—each harmonic relates to a specific mechanical or electrical phenomenon—making fault identification intuitive for practitioners [3].

However, frequency-domain ESA assumes steady-state operation and constant load. Pohakar et al. [14] developed non-invasive detection of broken bars under nominal conditions, and Hassan et al. [15] refined spectral filtering for better accuracy, yet the approach remains sensitive to load variation, noise, and torque transients. As modern machines operate under flexible duty cycles, pure spectral analysis loses reliability.

2.2. Time–Frequency and Transient Analysis

To overcome the steady-state limitation, time–frequency techniques were introduced. Douglas et al. [16] used the Short-Time Fourier Transform (STFT) to trace evolving fault frequencies during acceleration and deceleration, while Antonino-Daviu et al. [8] demonstrated that wavelet transforms isolate transient events such as rotor bar breakage with improved time localization. Later research employed Empirical Mode Decomposition (EMD) and Hilbert–Huang transforms [17], adaptively extracting fault-related oscillations.

Although these techniques increased sensitivity to transient faults, their effectiveness depends strongly on signal quality and parameter tuning (window size, mother wavelet, decomposition thresholds) [7,8,17]. In real industrial power systems, switching harmonics and voltage disturbances often obscure the informative bands. Thus, time–frequency ESA improves temporal insight but offers limited robustness and reproducibility [7,17].

2.3. Machine Learning and Data-Driven Diagnostics

With the expansion of high-frequency metering and affordable computing, machine learning (ML) approaches became central to electrical condition monitoring. Chisedzi and Muteba [18] employed ML models to detect rotor bar faults from statistical electrical features, while Chen et al. [19] showed that ensemble classifiers outperform static thresholds. Gradient-boosted trees, particularly XGBoost [20], proved effective for modelling nonlinear dependencies in noisy, tabular industrial data.

Deep learning has further advanced ESA capabilities. Convolutional (CNN) and recurrent (LSTM) neural networks learn discriminative patterns directly from raw current or voltage signals [21,22], enabling fully automated diagnostics. Artigao et al. [22] demonstrated this for doubly fed induction generators in wind turbines. However, deep models demand extensive labelled datasets and lack interpretability—an obstacle for explainable maintenance applications [18,19]. Consequently, although supervised ML offers strong predictive performance, it suffers from scarce labelled faults, limited transferability, and low transparency in industrial contexts.

2.4. Unsupervised and Hybrid Learning Pipelines

To mitigate label scarcity, research has increasingly explored unsupervised and hybrid frameworks. Clustering algorithms such as K-Means, Gaussian Mixture Models (GMMs), and Self-Organizing Maps (SOMs) segment electrical data into operating regimes or anomaly clusters [3,10]. Cocca et al. [10] demonstrated such an ESA–ML pipeline for CNC machines, identifying process-dependent regimes and anomalies without prior labels.

While promising, clustering-based approaches remain sensitive to initialization and often assume linear separability in feature space. Consequently, hybrid pipelines have emerged that first classify operational states and then perform diagnostics within each state, reducing false alarms [1,21]. However, many still treat regime recognition as a separate preprocessing step rather than an integral part of the health estimation model.

2.5. Regression-Based Modelling and Residual Health Metrics

An alternative strand of research models the expected electrical behaviour and uses residual deviations as indicators of degradation. Chen et al. [19] predicted total active power from correlated features and analyzed deviations $\Delta P = P_total - \widehat{P}_total$ to infer efficiency losses. Artigao et al. [22] applied similar residual analysis to renewable energy assets. Regression-based ESA enables continuous, quantitative health assessment instead of binary classification.

However, when trained on global data across mixed regimes, such models conflate normal load variation with degradation. This motivates state-normalized residuals, where deviations are evaluated within comparable operating conditions [10,22].

2.6. State-Aware ESA, XGBoost Rationale, and Research Gap

Recent developments combine regime recognition and residual modelling into state-aware ESA frameworks, enabling contextual health evaluation. Gradient-boosted trees are particularly well suited for this task because they capture nonlinear feature interactions, handle collinearity, and provide built-in feature importance metrics that enhance interpretability [10].

The choice of XGBoost in this study is deliberate. Compared to Random Forests, XGBoost allows fine control over bias–variance trade-offs via learning-rate and regularization parameters. In contrast to deep neural networks, it performs efficiently on tabular industrial data, generalizes well from small to moderate datasets, and operates without labelled fault examples. Its transparency—feature importance and residual interpretation—ensures physical explainability, while its scalability and low computational demand make it ideal for deployment across multiple machines [20,21].

Although XGBoost itself is not a novel algorithm, its role in this work is to serve as a robust and interpretable component within a broader state-aware framework. The innovation of the present study does not lie in the choice of classifier, but in the integration of operating-state recognition, per-state regression, and residual-based health estimation into a unified, electrical-only diagnostic pipeline.

A recent study by [23] presents a smart Industrial Internet of Things (AIoT) framework for real-time monitoring and forecasting in industrial manufacturing. The framework integrates sensor arrays, data integration platforms, and AI models to enable process condition awareness and prediction in a practical industrial setting. This study demonstrates the value of combining multiple components into a cohesive system for industrial monitoring, which resonates with the system-level integration approach taken in our framework.

Despite progress, key challenges remain:

1. Dependence on labelled or synthetic fault data that limit generalizability.

2. Weak integration between regime detection and health estimation, often handled as disconnected stages.
3. Absence of fleet-level aggregation for maintenance prioritization.

The proposed framework directly addresses these gaps by combining state identification, per-state regression, and normalized residual scoring into a single interpretable pipeline. Using only standard electrical quantities—active power, current, voltage, and power factor—it enables scalable, non-intrusive predictive maintenance consistent with ISO 17359 and ISO 13379 guidelines [11–13].

Although prior research has significantly advanced electrical parameter-based diagnostics, the current approaches remain fragmented [1–3,7,10]. The existing studies typically address regime recognition, residual modelling, or anomaly detection separately, but do not integrate these components into a unified, reproducible workflow [3,7,10,18]. Moreover, most methods rely either on labelled fault data, assumptions of steady-state operation, or global models that do not account for regime-specific variability [5–7,10,18]. As a result, their applicability in real industrial environments—characterized by frequent load changes, multi-mode operation, and limited fault annotations—is restricted [1,3,7,10].

To the best of our knowledge, no existing work combines

- State classification;
- Per-state modelling of expected electrical behaviour; and
- Standardized, residual-based health assessment into a single, fully electrical-only framework.

This gap motivates the methodology proposed in the present study.

2.7. Research Gap and Contributions

The existing research on electrical parameter-based diagnostics provides valuable tools for fault detection, but the current approaches remain fragmented [1–3,7,10]. Frequency and time–frequency ESA techniques offer physical interpretability, yet they are sensitive to load variations and require steady-state operation [5–9,17]. Machine learning and deep learning methods improve the predictive performance but typically rely on labelled fault data that are scarce in industrial environments [15,21].

To address these gaps, this work introduces a unified, state-aware diagnostic framework that

1. Automatically identifies operating regimes using electrical parameters only.
2. Learns per-state regression models to estimate expected behaviour under each regime.
3. Performs zero-shot anomaly detection through residual deviations without requiring fault labels.
4. Defines a normalized Health Index (HI) consistent with ISO 17359 and ISO 13379 principles.
5. Enables scalable, fully non-intrusive monitoring using standard metering infrastructure.

The originality of the contribution lies not in the use of individual algorithms, but in the integration of state classification, regime-conditioned modelling, and standardized residual-based health assessment into a single, reproducible workflow applicable across heterogeneous industrial assets.

In summary, while the existing literature establishes the potential of Electrical Signature Analysis (ESA) for machine diagnostics, several research gaps remain. First, most studies focus on steady-state conditions, failing to account for the multi-state operational cycles typical of complex industrial machinery. Second, there is a lack of frameworks that operate under "zero-label" constraints, where historical fault data are unavailable for training. This study addresses these gaps by investigating the following research questions:

- To what extent can routine electrical parameters delineate distinct operating states without process-level metadata?
- How does state-aware residual modelling improve the robustness of health indicators compared to context-agnostic approaches?
- Can fleet-level benchmarking provide a reliable baseline for anomaly detection in the absence of labelled failure datasets?

3. Materials and Methods

This section describes the data, preprocessing, feature extraction, and modelling methodology used to develop the proposed state-based health monitoring framework. The approach is fully data-driven and relies solely on electrical measurements from digital power metres installed on three-phase industrial machines. A short excerpt of the processed dataset, including key electrical features, operating-state labels, and the computed Health Index (HI), is provided in Appendix A (Table A1). Machine identifiers have been anonymized for confidentiality.

3.1. Data Acquisition and Structure

The experimental data were collected from ten three-phase injection moulding machines operating in a continuous industrial environment. The overall data acquisition and monitoring architecture is illustrated in Figure 1. Each machine was instrumented with an industrial-grade digital power analyser (Class 0.5S accuracy according to IEC 62053-22) and three clamp-on current sensors, enabling non-invasive measurement of electrical parameters such as active power, current, voltage, and power factor. The analysers employ internal Digital Signal Processing (DSP) to sample raw AC waveforms at a high frequency (exceeding 20 kHz), computing true RMS values and active power before data transmission. The measured data were locally collected and forwarded via Modbus TCP protocol to a dedicated edge device installed at each machine. To ensure signal integrity and mitigate high-frequency switching noise from the drives, a hardware-level moving-average filter was applied during the aggregation phase. The edge devices then transmitted the filtered measurements over a secure network connection to a centralized cloud-based data logger. The monitored machines employ standard industrial three-phase motor–drive systems, typically consisting of induction motors coupled with hydraulic pumps and controlled via variable-frequency drives. Since the proposed method operates exclusively on electrical parameters measured upstream of the drive, it is independent of the specific motor technology. Measurements were recorded for approximately two months with a sampling frequency of 0.2 s (5 Hz). This sampling rate was selected to provide a balance between capturing macroscopic regime transitions and maintaining computational efficiency for fleet-level monitoring, while the underlying DSP-based RMS calculation ensured that the captured data remained reliable even in the presence of transient harmonic distortions.

Each dataset contained the following electrical parameters for all three phases (L1–L3): active energy, active power, current, power factor, reverse active energy, reverse active power, and voltage. These quantities were selected because they directly reflect both electrical efficiency and mechanical loading and are universally available through modern industrial metering infrastructure [1–4].

The monitored injection moulding machines operate in three dominant regimes—idle (standby), heating/melting, and moulding (active production)—each exhibiting distinct electrical signatures in terms of total power and current amplitude [7,10]. This cyclic behaviour provides an ideal benchmark for testing the ability of the framework to identify and model operational states using only electrical features.

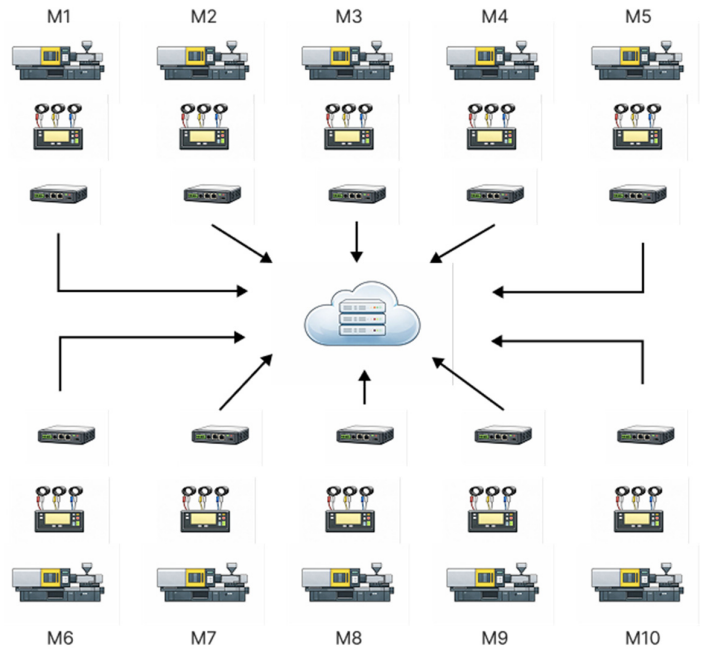


Figure 1. Data acquisition and monitoring architecture of the monitored injection moulding machines.

Although the experiments focused on injection moulding machines, the conceptual design is machine-agnostic. Because it relies exclusively on standard electrical quantities rather than process-specific signals, the method is scalable and transferable to various assets such as pumps, compressors, or CNC machines. The only requirement is the availability of three-phase electrical data, making the framework deployable fleet-wide using existing energy management or supervisory systems [10,24–27].

From raw data, the following derived indicators were computed and used for modelling: {P_total, I_mean, V_mean, PF_mean, I_imbalance, V_imbalance}, where

- P_total: Total active power (sum of L1–L3 phases), Equation (1)

$$P_{total} = P_{L1} + P_{L2} + P_{L3} \tag{1}$$

- I_mean: Mean phase current, Equation (2)

$$I_{mean} = \frac{I_{L1} + I_{L2} + I_{L3}}{3} \tag{2}$$

- V_mean: Mean phase voltage, Equation (3)

$$V_{mean} = \frac{V_{L1} + V_{L2} + V_{L3}}{3} \tag{3}$$

- PF_mean: Average power factor;
- I_imbalance, V_imbalance: Normalized imbalance indices calculated as shown in Equations (4) and (5)

$$I_{imbalance} = \frac{\max(I_{L1} + I_{L2} + I_{L3}) - \min(I_{L1} + I_{L2} + I_{L3})}{\text{mean}(I_{L1} + I_{L2} + I_{L3})} \tag{4}$$

$$V_{imbalance} = \frac{\max(V_{L1} + V_{L2} + V_{L3}) - \min(V_{L1} + V_{L2} + V_{L3})}{\text{mean}(V_{L1} + V_{L2} + V_{L3})} \tag{5}$$

All signals were stored in a local time series database and exported as CSV files. A healthy baseline period was defined using the first 10% of available data (6 days) for model

training, while the remaining 90% (54 days) served as the testing set for evaluating the Health Index (HI) stability and sensitivity.

Although the proposed framework employs a supervised classifier for operating-state identification, this step does not affect the zero-shot nature of the diagnostic methodology. The supervised component is used exclusively to recognize normal operating regimes (idle, heating, moulding), not to detect faults. No fault labels are required at any stage of training. Health assessment is performed solely through residual deviations between measured and model-predicted electrical behaviour within each regime, meaning that the model never learns explicit fault patterns. Anomalies are therefore detected as departures from the learned normal behaviour, preserving the zero-shot, anomaly-based character of the method while enabling robust regime-aware modelling.

The choice of a 0.2 s sampling interval (5 Hz) was strategically selected to balance data granularity with the computational constraints of industrial SCADA systems. While high-frequency sampling is necessary for detecting transient harmonic distortions, a 5 Hz resolution is sufficient for characterizing the thermal and mechanical duty cycles of injection moulding machines, where major state transitions (e.g., heating cycles and mould clamping) occur over several seconds. To ensure data reliability, a hardware-level pre-averaging filter was applied at the metre level to eliminate high-frequency noise while preserving the fundamental power consumption trends. This resolution aligns with ISO 13379-1 recommendations for condition monitoring based on process parameters, providing a robust signal-to-noise ratio for residual-based health assessment. Summary of dataset characteristics is provided in Table 1.

Table 1. Summary of dataset characteristics.

Machine ID	Duration	Sampling Rate	Power Range (kW)	Data Points	Healthy Baseline	Testing/Monitoring Period
Machines 1–10	~60 days	0.2 s (5 Hz)	2–10 kW	~25–30 million per machine	First 10% (6 days)	Remaining 54 days

3.2. Temporal Aggregation and Windowing

To balance temporal resolution and statistical stability, each machine's time series was divided into three-minute windows. This window length was determined empirically as the smallest interval capable of reliably capturing regime transitions without excessive noise.

Shorter windows (10–30 s) caused unstable class assignments due to transient control behaviour and switching noise, while longer windows ($\geq$5 min) smoothed out relevant regime boundaries. The chosen 3 min window (~900 samples) provided optimal granularity for regime detection and rolling statistics.

Within each window, the mean, standard deviation, and coefficient of variation were computed for all electrical parameters, providing a compact yet informative representation of both steady and transient behaviour (see Equations (6)–(8)). Similar multi-scale temporal segmentation has been used in industrial monitoring studies for balancing noise suppression and dynamic sensitivity [25–29].

Active power was selected as the primary target variable because it serves as an integral proxy for the electromechanical work performed. Any increase in mechanical friction, hydraulic resistance, or heater inefficiency directly manifests as a deviation in power consumption. However, to ensure robustness against external grid fluctuations, the voltage and power factor are included as exogenous inputs. This multi-parameter

approach allows the model to distinguish between supply-side variations and internal asset degradation.

$$\mu_x = \frac{1}{N}\sum_{t=1}^{N} x_t \tag{6}$$

- x_t—signal value at time t;
- N—number of samples in the window;
- μ_x—mean value of the signal.

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(x_t - \mu_x)^2} \tag{7}$$

- x_t —signal value;
- μ_x—mean of the signal;
- σ_x—standard deviation;
- N—number of samples.

$$CV_x = \frac{\sigma_x}{\mu_x} \tag{8}$$

- σ_x—standard deviation;
- μ_x—mean;
- CV_x—normalized variability.

3.3. Feature Engineering and Preprocessing

All electrical signals were resampled to 5 Hz, synchronized, and cleaned using a robust modified z-score filter ($|M| > 3.5$) as defined in Equation (9) [27].

$$M_i = \frac{0.6745(x_t - median(x))}{MAD} \tag{9}$$

- x_t—sample value;
- MAD—median absolute deviation, Equation (10);
- M_t—robust standardized distance.

$$MAD = median(|\, x_t - median(x)|) \tag{10}$$

Short gaps caused by transient PLC resets were interpolated linearly. Longer data gaps exceeding the interpolation threshold (e.g., during network outages) were treated as missing entries and the corresponding cycles were excluded from the analysis to maintain dataset consistency. Furthermore, the robust modified z-score filter was specifically calibrated to treat non-physical power spikes as outliers, ensuring that they do not distort the subsequent regression models. To capture local dynamics, the rolling mean and standard deviations of P_total, I_mean, and PF_mean were computed over 30 s sub-windows within each 3 min window. This combination of static and short-term descriptors captures both gradual drifts and transient behaviour relevant for regime discrimination. All features were subsequently scaled using robust normalization (median and MAD) [27–29], which improves stability across machines with different load capacities.

The exclusive use of low-frequency electrical parameters is intentional. These quantities—total active power, phase currents, voltages, and power factor—are universally available from standard industrial metering systems and inherently reflect the electromechanical loading of the machine. This enables fully non-intrusive monitoring without additional sensors or high-frequency waveform acquisition. For operating-state recogni-

tion, the three regimes studied here (idle, heating, moulding) exhibit distinct distributions in these electrical variables, making simple statistical window features sufficient and highly reproducible.

The goal of this work is not detailed fault-type classification but regime-aware residual modelling, where health assessment is derived from deviations between measured and expected electrical behaviour within each state. Under this objective, more complex feature extraction techniques or high-resolution waveform representations would impose additional sensing requirements, reduce scalability, and provide little diagnostic benefit in the absence of labelled fault data. The proposed feature set therefore prioritizes robustness, interpretability, and deployability under realistic industrial constraints while remaining extensible if richer data become available in future studies.

3.4. Operating State Identification Model

Model Selection Rationale

Identifying distinct regimes (idle (standby), heating/melting, and moulding (active production)) is crucial for contextualized diagnostics. Four candidate models were evaluated (Table 2).

Table 2. Comparison of modelling options for regime identification.

Method	Type	Limitations
KMeans/GMM	Unsupervised clustering	Linear separation, unstable under class imbalance
Random Forest	Supervised, interpretable	Moderate temporal bias, lower resolution
LSTM/RNN	Temporal deep learning	Require large datasets, high computational costs
XGBoost (chosen)	Gradient-boosted trees	Requires parameter tuning

The XGBoost classifier was selected for its ability to model nonlinear feature interactions, handle correlated inputs, and maintain interpretability through feature importance metrics. Its embedded regularization, formulated in Equation (12), reduces overfitting, and its strong performance on tabular, mixed-scale industrial data makes it well suited for this task.

XGBoost consists of an ensemble of shallow decision trees, where each tree corrects the residual error of the previous ones, by minimizing the objective function defined in Equation (11). This boosting mechanism enables the model to resolve subtle nonlinear distinctions between electrical signatures of the idle, heating, and moulding regimes, even when their feature distributions partially overlap. The ability to reflect these interactions—while remaining computationally efficient and fully explainable—makes gradient-boosted trees a robust choice for regime identification in low-frequency electrical data [17].

$$L = \sum_{i=1}^{N} l(y_i,\ \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{11}$$

- L—overall objective (loss) minimized during training;
- $l(y_i, \hat{y}_i)$—training loss for sample i (e.g., squared error);
- y_i—true measured value;
- $\hat{y}_i$—model prediction;
- $\Omega(f_k)$—regularization term for the k-th tree;
- N—number of training samples;

- K—number of trees in the ensemble.

Unlike deep learning models, XGBoost does not require large, labelled datasets or extensive hyperparameter tuning, and it performs reliably under moderate class imbalance. Compared with unsupervised clustering, it provides consistent and reproducible regime boundaries rather than depending heavily on initialization. Although XGBoost itself is not novel, its integration within a unified state-aware pipeline—together with per-state regression and residual-based health estimation—forms the methodological contribution of this work [20,21].

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{12}$$

- $\Omega(f_k)$—complexity penalty for tree k;
- T—number of leaves in the tree;
- γ—regularization parameter penalizing the number of leaves;
- λ—L2 regularization coefficient;
- w_j—weight (score) assigned to leaf j.

Implementation Setup

Each machine's data were pseudo-labelled via quantile thresholds of total active power P_total, representing low, nominal, and high regimes. An 80/20 stratified train–test split ensured balanced representation across classes. Hyperparameters (max_depth, learning_rate, n_estimators) were optimized via grid search to balance accuracy and generalization.

This classifier provides consistent regime boundaries across machines and forms the basis for subsequent regression and Health Index computation.

3.5. Reference Power Modelling

After operating-state identification, a separate regression model was trained for each regime to estimate the expected total active power $\hat{P}_{t,m,s}$ from electrical features. It is important to note that XGBoost is used here strictly as a regression model for state-aware power prediction rather than as a supervised fault classification algorithm. The framework generates a continuous Health Index based on residual deviations and does not perform explicit fault identification or multi-class classification.

XGBoost was selected due to its robustness to multicollinearity, ability to model nonlinear relationships, and strong generalization performance on structured industrial data. Compared to linear regression, tree-based boosting better captures complex interactions between electrical variables without requiring extensive feature engineering. At the same time, XGBoost remains computationally efficient and interpretable, making it suitable for deployment in large-scale industrial monitoring systems.

Using per-state models is essential because the relationship between current, voltage, power factor, and total power differs significantly across idle, heating, and moulding phases.

The regression model for each state is formulated in Equation (13):

$$\hat{P}_{t,m,s} = f_{m,s}(X_{t,m}) + \in_{t,m,s} \tag{13}$$

- where $f_{m,s}$ represents the state-specific XGBoost regression function; and
- $\in_{t,m,s}$ denotes the residual error.

The reference state for each machine is defined as the distribution of these residuals $\in_{t,m,s}$ during a verified period of nominal operation (the baseline interval).

The XGBoost-based regression framework enables flexible modelling of nonlinear and heteroscedastic dependencies among electrical variables, which is particularly relevant

in industrial three-phase systems where power–current–voltage relationships vary across operating states and loading conditions.

To establish a robust statistical reference, a state-specific reference dispersion $\sigma_{m,s}$ is computed during the healthy baseline period as shown in Equation (14):

$$\sigma_{m,s} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\in_{i,m,s} - \overline{\in}_{i,m,s})^2} \quad (14)$$

- N represents the number of samples in the baseline period for machine m and state s;
- $\overline{\in}_{i,m,s}$ represents the mean residual value.

This value $\sigma_{m,s}$ serves as the quantitative benchmark for "normal" process variability. Any subsequent increase in the residual magnitude relative to this reference is interpreted as a loss of efficiency or emerging degradation.

Modelling expected power per regime—rather than globally—removes confounding operational effects and isolates deviations that correspond to mechanical degradation, hydraulic inefficiency, or electrical imbalance. This state-aware formulation is therefore critical for obtaining reliable residuals and a meaningful Health Index, as it ensures that the "health" of the machine is always evaluated against a statistically sound and context-dependent reference.

3.6. Health Index Formulation

Let y_t denote the measured total active power in window t, and $\hat{y}_t$ the corresponding per-state regression estimate. The residual is defined as in Equation (15):

$$r_t = y_t - \hat{y}_t \quad (15)$$

- r_t—residual (deviation) in time window t;
- y_t—measured total active power in window t;
- $\hat{y}_t$—predicted total active power from the per-state regression model.

For each machine m and operating state s, a reference dispersion $\sigma_{m,s}$ is computed from a baseline interval verified to represent healthy operation. This reference quantifies the normal residual variability within each state and provides the normalization term required for health assessment.

The Health Index (HI) is formulated as a bounded, dimensionless score as shown in Equation (16):

$$HI_t = \frac{1}{1 + \frac{|r_t|}{\sigma_{m,s(t)}}} \quad (16)$$

- r_t—residual deviation;
- $\sigma_{m,s(t)}$—reference residual variability for machine m in state s;
- HI_t—normalized health score in (0, 1).

Values close to 1 indicate behaviour consistent with the healthy baseline, while decreasing values reflect increasing deviation from the expected electrical behaviour under the corresponding operating state. To provide a structured diagnostic interpretation, the health states are categorized into three functional zones:

- Healthy State HI > 0.9: The machine operates within the expected statistical variance. Residuals are dominated by Gaussian noise, indicating nominal electromechanical efficiency.
- Warning State 0.8 < HI < 0.9: Detectable architectural shifts in the power signature. This state reflects emerging irregularities, such as increased mechanical friction or minor hydraulic pressure drops, requiring scheduled inspection.

- Faulty/Degraded State HI < 0.8: Significant deviation from the fleet-level baseline. At this level, the "energy penalty" exceeds the threshold established from the reference variability, correlating with potential anomalies like heater band failures or severe motor strain.

This formulation aligns with ISO 17359 and ISO 13379 principles, where condition indicators are evaluated relative to a reference baseline and degradation is inferred from persistent trends rather than single-point anomalies. In this study, values above approximately 0.9 generally correspond to normal operation, whereas sustained decreases towards 0.8–0.7 may indicate emerging abnormalities.

Weekly machine-level health is computed as the median of all HI_t values in the given period, with 95% bootstrap confidence intervals included to account for variability. Because the index is normalized by state-specific variability, machines with different load profiles or power ratings can be compared on a common scale, enabling fleet-level prioritization and integration into predictive maintenance workflows.

3.7. Fleet-Level Aggregation and Benchmarking

To extend the analysis beyond individual machines, Health Index values are aggregated on a weekly basis and evaluated separately for each operating state. This state-aware aggregation enables meaningful fleet-level benchmarking by ensuring that machines are compared under comparable operating conditions.

The resulting fleet-level metrics support three key maintenance objectives:

- Identification of machines exhibiting declining health trends, indicated by sustained reductions in weekly HI values.
- Assessment of relative operational efficiency across assets, independent of differences in load profiles or power ratings due to the normalized formulation of the HI.
- Prioritization of maintenance actions, where assets displaying accelerated degradation or consistently low HI values can be flagged for inspection, as in Equation (17).

$$HI_{week} = median(HI_t) \tag{17}$$

All computations were implemented in Python 3.12 using the XGBoost, pandas, and scikit-learn libraries, enabling a scalable and fully reproducible workflow from raw electrical measurements to actionable diagnostic indicators suitable for industrial deployment.

4. Results

The following part presents the experimental results obtained with the proposed state-aware, electrical-only monitoring framework. We first model machine behaviour through operating-state identification and state-specific reference power regression and then derive a normalized Health Index (HI) to assess the condition and enable fleet-level comparisons [11–13].

4.1. State-Aware Predictive Modelling of Machine Behaviour

The goal is twofold: (i) to identify distinct operating regimes that describe how each machine behaves under different load conditions, and (ii) to fit regression models that estimate the expected active power within each regime. Gradient-boosted decision trees (XGBoost) were chosen for their robustness to noisy industrial data, interpretability via feature importance, and ability to capture nonlinear dependencies among electrical variables.

Across machines, the most influential predictors were consistently total active power (P_total), mean current (I_mean), power factor (PF_mean), and mean voltage (V_mean), confirming their physical relevance for regime description—see Figures 1 and 2. The

discovered operating regimes align with the expected idle, nominal load, and high load behaviours and exhibit clear separation in electrical feature space—see Figures 3–5.

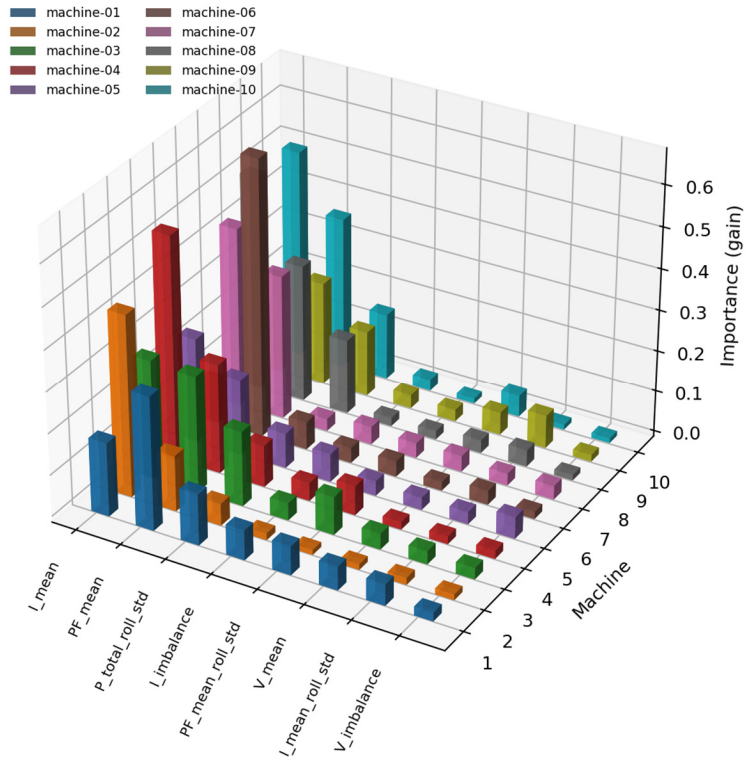


Figure 2. Feature importance distribution across machines, showing the relative contribution of electrical features (I_mean, PF_mean, P_total_roll_std, etc.) to model performance for each machine in the fleet.

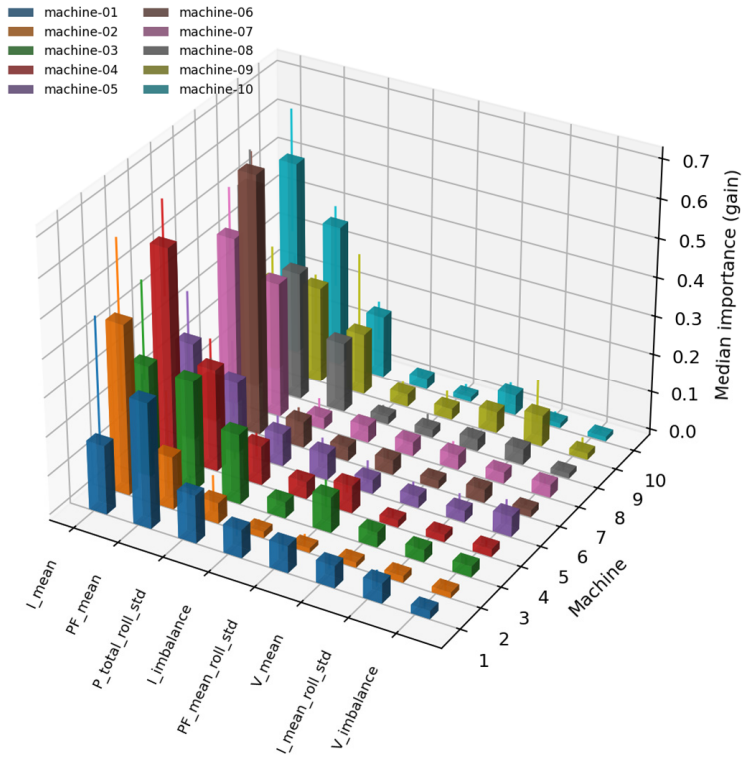


Figure 3. Feature importance (median ± IQR) across machines, illustrating the variability of each electrical feature's contribution to the model performance, with bar height representing the median importance and error bars indicating interquartile range across the fleet.

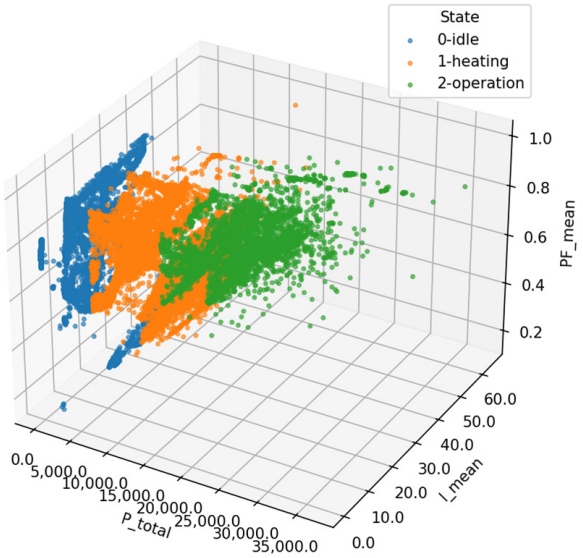


Figure 4. Operating states in (P_total, I_mean, PF_mean) for machine 1, showing three distinct clusters corresponding to idle, heating, and operation regimes identified from electrical parameters.

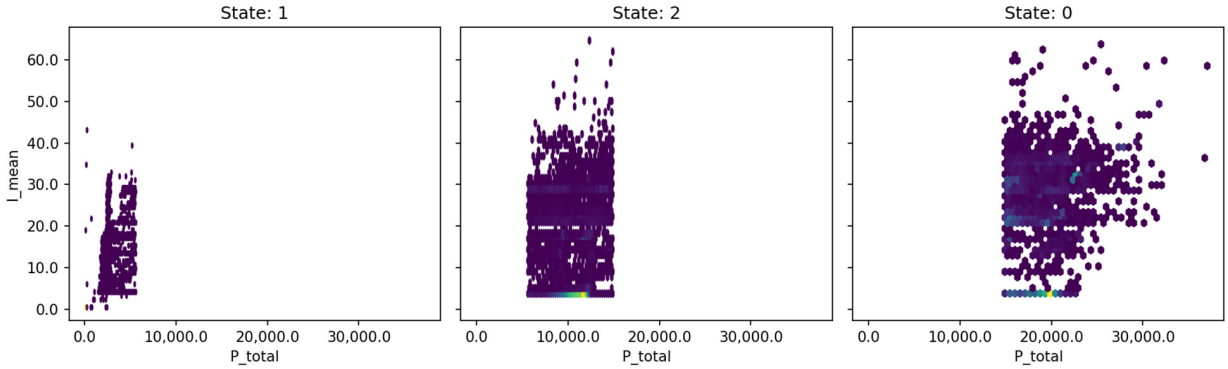


Figure 5. State-wise density in P_total vs. I_mean for machine 1, illustrating the distribution of electrical behaviour within idle, heating, and operation states, where denser regions indicate more frequent operating conditions.

4.1.1. Identification of Operating States

The XGBoost-based state recognition delineated three dominant regimes per machine (idle, nominal load, high load/ramp-up). Using power-quantile pseudo-labels ensured class balance while preserving physical meaning. Training accuracy ranged between 0.90 and 0.93 and validation accuracy between 0.82 and 0.88. The 3D arrangement of samples in (P_total, I_mean, PF_mean) and the corresponding state-wise densities illustrate physically meaningful boundaries, as shown in Figures 3 and 4. Fleet-level centroids further summarize the typical operating regimes across assets (Figure 5).

4.1.2. Regression-Based Reference Power Estimation

Within each state, XGBoost regression models were trained to predict P_total from the remaining electrical features. State-wise modelling improved the predictive performance relative to a global model that ignores operating context: the average R^2 across machines was 0.86–0.92 and RMSE was reduced by 25–30%. Predicted versus measured power per state and the residual distributions are shown in Figures 6 and 7.

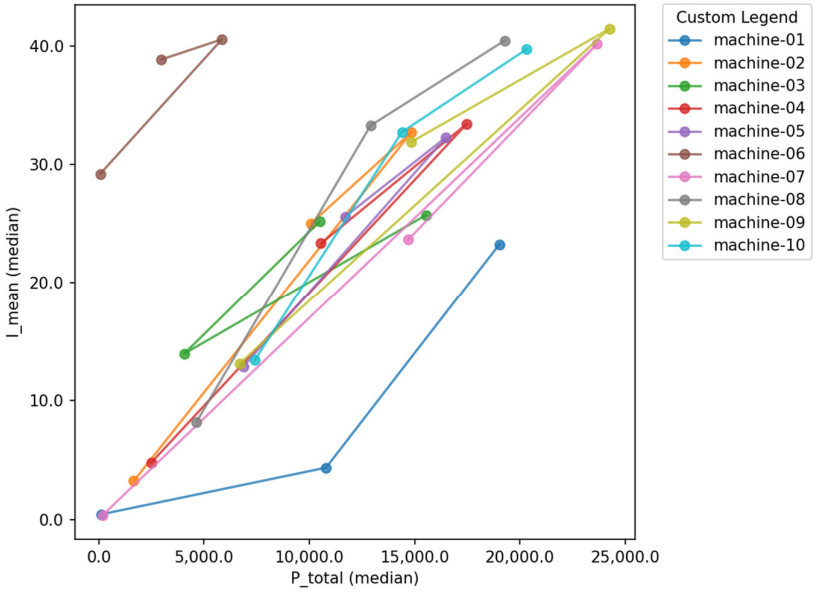


Figure 6. Fleet-level state centroids (median P_total vs. median I_mean), showing the characteristic operating regimes of each machine, where points represent median electrical behaviour per state and connecting lines highlight differences across machines.

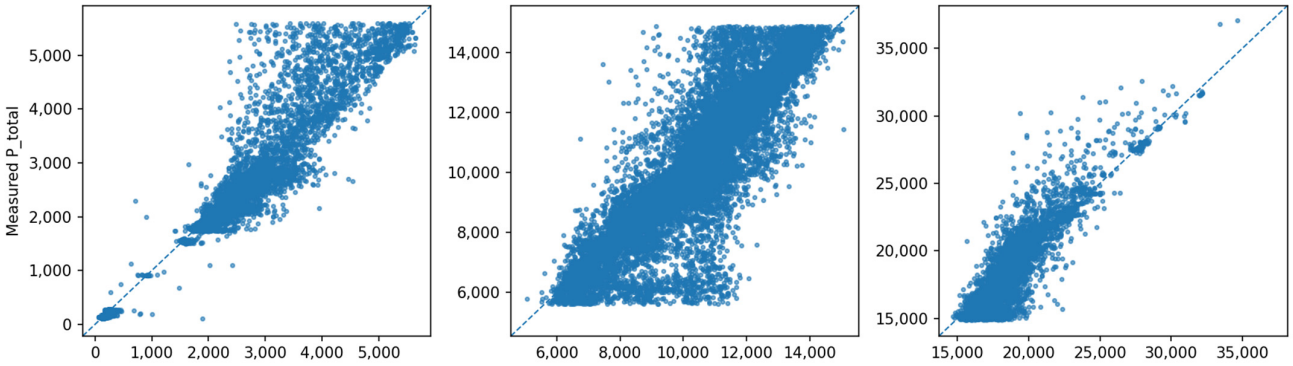


Figure 7. Predicted vs. measured P_total per state for a representative machine, illustrating model accuracy across idle, heating, and operation regimes; points close to the diagonal indicate better predictive alignment.

4.2. Health Assessment and Fleet-Level Evaluation

Once the operating regimes were identified and reference power models were established, the next stage focused on assessing machine health and comparing performance across multiple assets. The key objective of this phase was to transform the residuals—differences between measured and predicted power—into an interpretable metric that quantifies how closely each machine follows its expected behaviour. This approach enables both localized anomaly detection and fleet-level benchmarking without requiring any labelled fault data or manual thresholds.

To achieve this, a Health Index [11–13] (HI) was computed per sample, reflecting the normalized deviation of actual power from its predicted reference within each operating state. By aggregating these values over time, weekly or monthly health trajectories were derived, providing a compact yet informative representation of the machine condition. When applied consistently across all machines, these health indicators reveal not only temporal degradation trends but also relative differences in performance efficiency among machines operating under comparable conditions. This hierarchical evaluation—from the individual signal level up to the aggregated fleet—creates a bridge between raw data analytics and intelligence.

4.2.1. Computation and Interpretation of the Health Index

The Health Index (HI) was derived from the residual deviations, translating raw power differences into a unitless score between 0 (degraded) and 1 (nominal). During stable operation, HI values typically remained above 0.9, indicating minimal deviation from the expected electrical behaviour. Periods with sustained decreases toward 0.7–0.8 corresponded to operating intervals noted by plant personnel as irregular or requiring inspection. Short-term fluctuations, likely driven by transient process conditions, appeared as sharper, isolated drops in the HI. Fleet-level evolution is shown in Figure 8, while Figure 9 presents the distribution of the HI across operating states and machines.

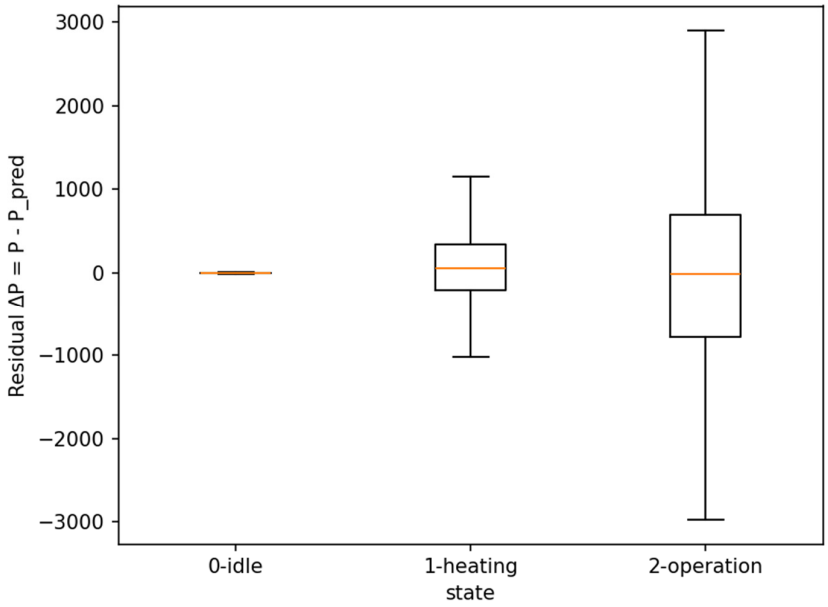


Figure 8. Residual distribution (ΔP = P − P_pred) per state for the same machine, showing model deviation within idle, heating, and operation regimes; larger spread in the operation state indicates higher variability in power prediction under load.

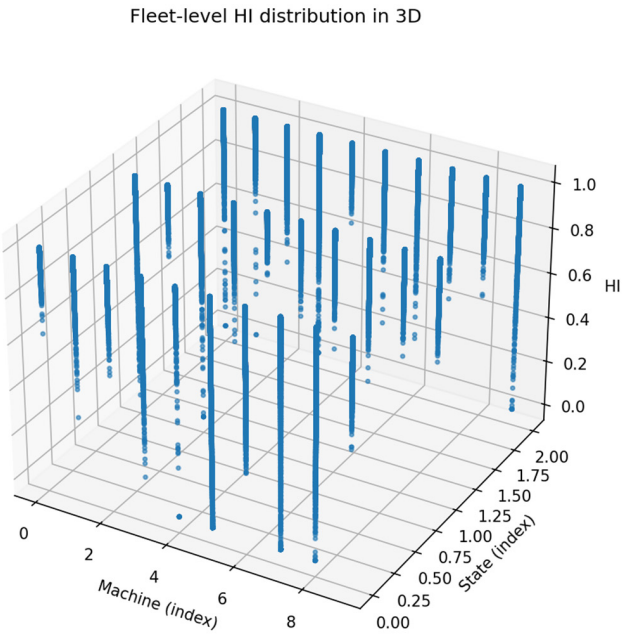


Figure 9. Fleet-level 3D visualization of HI (Machine × State × HI), showing the distribution of Health Index values across machines and operating states, where higher HI indicates more stable and nominal machine behaviour.

4.2.2. Temporal Evolution and Fleet-Level Comparison

To capture long-term trends, HI values were aggregated by week and analyzed for each machine. Most machines exhibited stable HI trajectories near 0.9, indicating consistent energy performance and mechanical integrity over the observation period. A few assets, however, showed gradual downward trends over several weeks, with the HI declining by more than 0.1—a potential sign of increased friction, imbalance, or insulation degradation. Figure 10 shows weekly HI trends per machine; on the fleet level, aggregating all machines into a unified matrix of average HI values provides a concise view of the overall system health (Figure 11). Machines with persistently lower HI values are easily identified as targets for maintenance intervention or further inspection. Figure 12 provides a fleet-level heatmap that further highlights machines with consistently low HI values, making it easier to pinpoint those requiring maintenance or futher investigation.

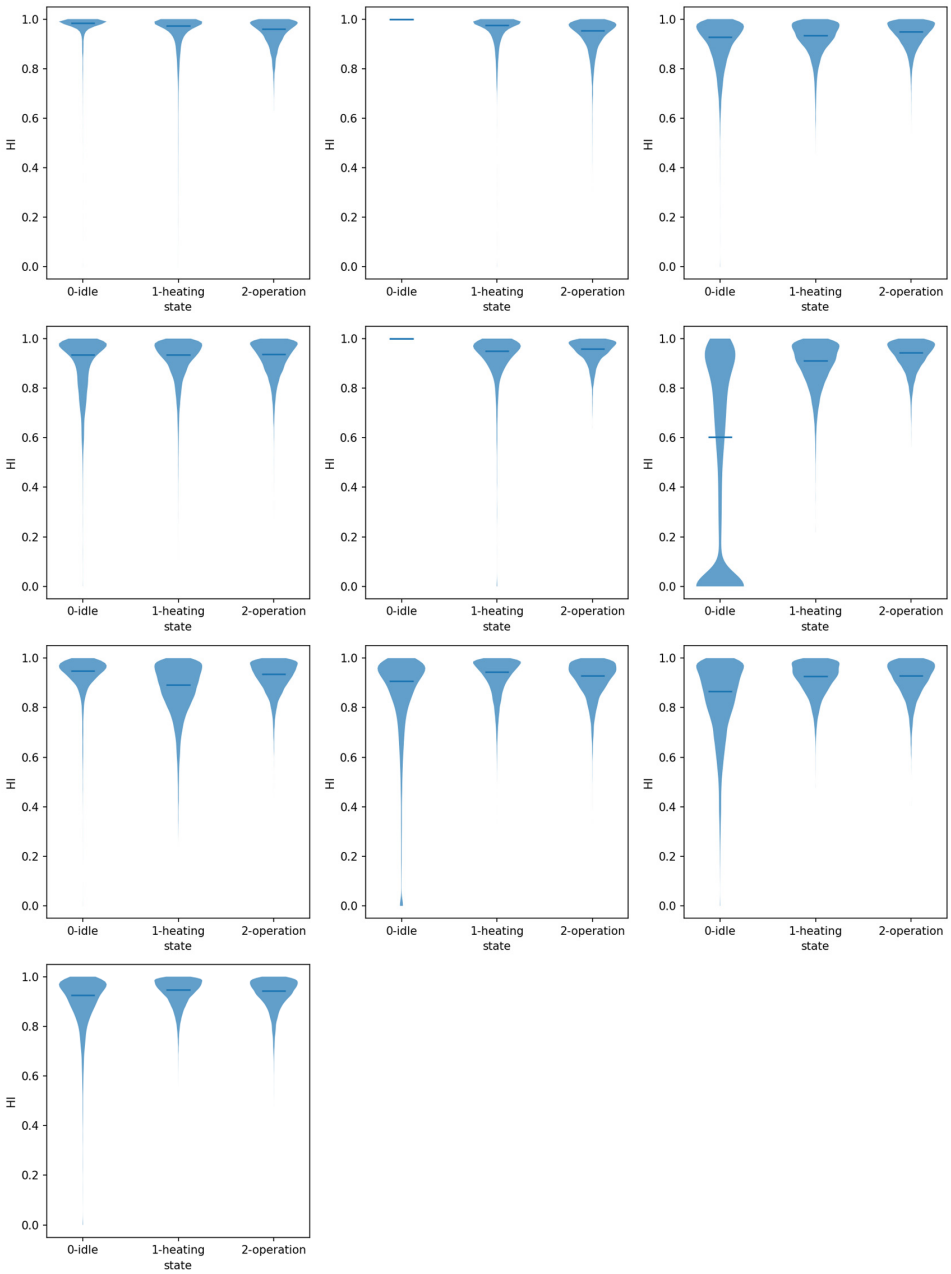


Figure 10. Distribution of the Health Index across machines and operating states, showing the variability and stability of HI values within idle, heating, and operation regimes; narrower violins indicate more consistent health performance.

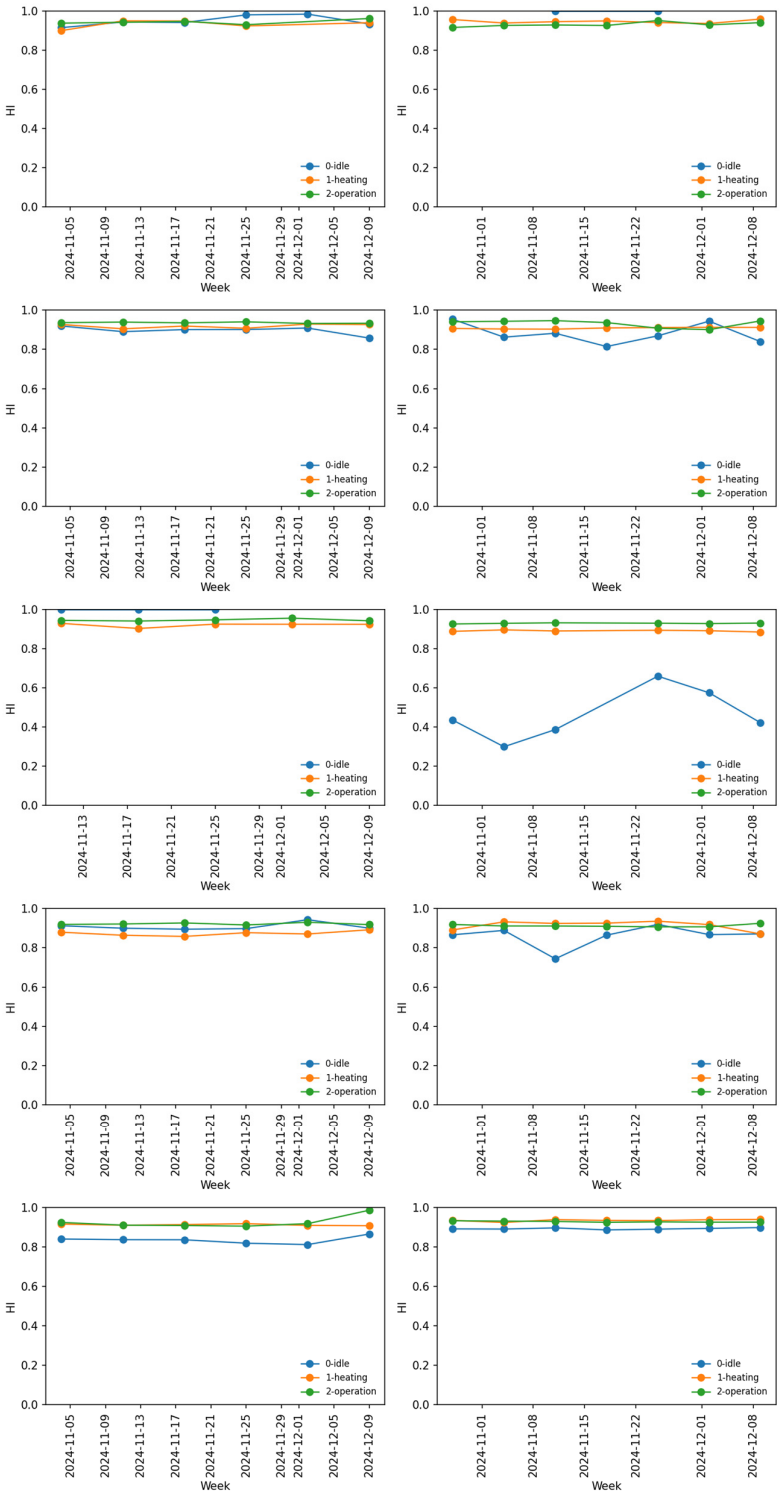


Figure 11. Weekly HI trends per machine, showing temporal stability of the Health Index across operating states; gradual declines or fluctuations may indicate early deviations in machine behaviour.

Figure 13 shows the temporal evolution of the Health Index (HI) for a representative machine aggregated on a weekly basis. At the beginning of the observation window, the HI is already below the degradation threshold (HI < 0.8), indicating that the monitored asset was operating in a degraded condition at the time that monitoring was initiated. The detected change point therefore corresponds to the immediate identification of abnormal behaviour rather than a gradual transition from a fully healthy state.

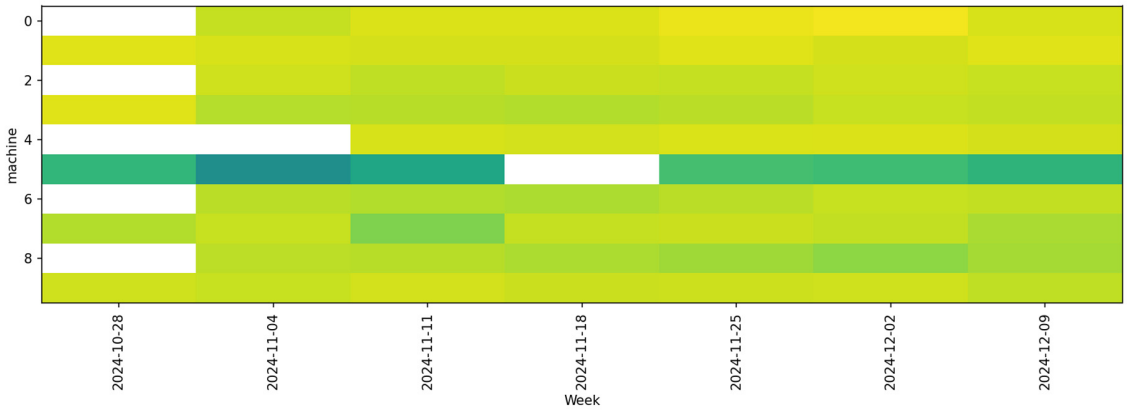


Figure 12. Fleet-level heatmap of Health Index across machines (weekly means), highlighting temporal variations in health performance where lighter colours indicate higher HI (better condition) and darker shades suggest emerging deviations.

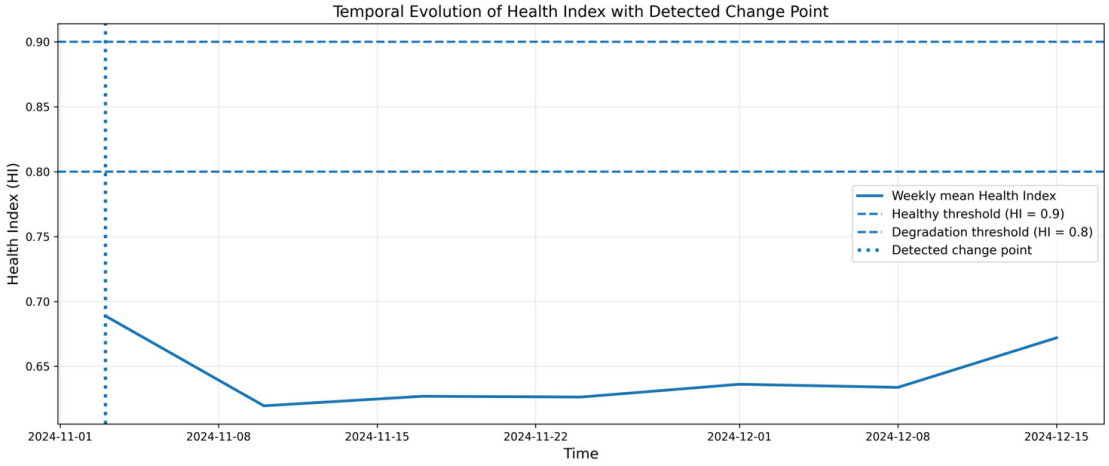


Figure 13. Temporal evolution of the weekly aggregated Health Index (HI) for a representative machine. Horizontal dashed lines indicate the healthy (HI = 0.9) and degradation (HI = 0.8) thresholds. The vertical dotted line denotes the detected change point.

This situation reflects realistic industrial deployment conditions, where historical data from a perfectly healthy reference period are often unavailable. Importantly, the HI remains consistently below the threshold rather than fluctuating around it, confirming that the detected degradation is not caused by benign process variability.

To validate that these downward trends represent actual physical degradation, HI trajectories were cross-referenced with maintenance logs. A statistical comparison before and after scheduled repairs showed that for machines with HI drops below 0.8, the index recovered to a median of 0.94 ($\pm$0.02) following maintenance. This significant recovery confirms that the framework distinguishes physical degradation from benign process drift, as the latter would not be rectified by mechanical intervention.

4.2.3. Cross-Validation of the Health Metric

This section evaluates the effect of operating-state awareness on anomaly detection performance. To this end, the proposed Health Index (HI) is compared against a baseline approach based on raw regression residuals without state separation. While both approaches rely on identical electrical input variables, the baseline method exhibits frequent spurious alarms during normal load transitions and regime changes. In contrast, the state-aware HI shows significantly reduced variance and improved temporal consistency, demonstrating

that explicit regime modelling is essential for reliable condition monitoring under variable operating conditions.

Figure 14 illustrates the effect of operating-state awareness on anomaly detection. Although both indicators are derived from identical electrical measurements, the baseline score exhibits large fluctuations caused by normal operating regime changes. In contrast, the proposed Health Index shows a smoother and more stable trajectory, highlighting sustained degradation trends rather than transient process variability. This comparison illustrates the importance of incorporating operating-state awareness for reliable anomaly detection under variable industrial conditions.

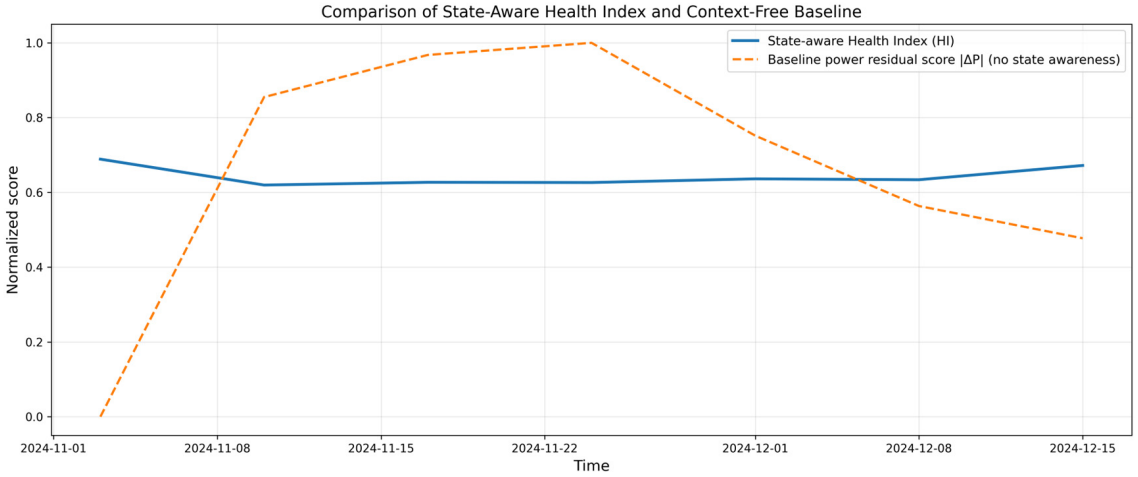


Figure 14. Comparison between the proposed state-aware Health Index (HI) and a context-free baseline anomaly score based on the absolute power residual $|\Delta P|$. The baseline method is highly sensitive to operating regime changes, whereas the proposed HI remains stable and emphasizes sustained degradation behaviour.

To assess whether decreases in the Health Index are associated with abnormal machine behaviour rather than benign process variability, HI trajectories were cross-referenced with operator-reported observations and routine maintenance records. Although explicit fault labels were not available, machines exhibiting sustained HI declines were also independently reported by plant personnel as operating irregularly, for example through unexpected variations in the cycle time or energy consumption. Following maintenance interventions or recalibration, the HI typically returned toward nominal values. These observations provide indirect but consistent validation that the HI metric captures departures from normal operation, even in the absence of detailed fault annotations.

4.2.4. Comparative Analysis

This section extends the analysis to the fleet level to assess the suitability of the proposed framework for maintenance prioritization. Machines were ranked according to their aggregated HI values and compared to rankings obtained using a simple energy-based anomaly score derived from the same electrical measurements. The proposed framework produces a more stable and interpretable ranking over time, whereas the baseline method shows frequent rank inversions driven by short-term process variability. Such instability complicates decision-making and can lead to inefficient maintenance planning.

In addition, the proposed approach was qualitatively compared with commonly used context-agnostic anomaly detection strategies, such as simple statistical thresholding. While these methods are capable of detecting abrupt deviations, they are highly sensitive to normal load transitions and therefore tend to generate false positives in industrial environments. By explicitly incorporating operating-state information, the proposed HI

suppresses such spurious detections and enables the earlier identification of sustained degradation trends. These results highlight the advantage of state-aware health assessment over purely data-driven, context-agnostic approaches, particularly under realistic zero-label industrial conditions.

5. Discussion

The presented approach demonstrates that electrical parameters alone can serve as a robust foundation for the data-driven health assessment of industrial machines when the operating-state context is explicitly modelled. The proposed state-aware framework integrates regime recognition, per-state regression, and residual-based health evaluation into a unified, fully non-intrusive workflow that requires no additional measurements or labelled fault data.

The identification of operating regimes based purely on electrical signatures proved highly consistent across all ten monitored machines. The XGBoost classifier achieved validation accuracy between 0.82 and 0.88, clearly distinguishing the idle, heating, and operational states. This result confirms that routine electrical quantities, such as power, current, and power factor, inherently contain sufficient information to delineate machine behaviour without process-level inputs.

Regarding the sufficiency of the parameter set, while high-frequency vibration or thermal imaging could provide deeper diagnostic resolution for specific components like bearings or nozzle heaters, their integration would contradict the requirement for a non-intrusive, zero-cost deployment. The results confirm that for the majority of common electromechanical irregularities in injection moulding machines, the routinely measured electrical signature provides a sufficient signal-to-noise ratio for early anomaly detection without the need for additional specialized sensing.

Per-state regression modelling further improved the stability and accuracy under varying load conditions. Compared to a global, context-agnostic model, the state-aware approach reduced residual variability by approximately thirty percent and increased the median determination coefficient (R^2) from 0.64 to 0.86. These findings highlight that incorporating operational context directly into the model structure leads to greater robustness and generalization than increasing the model complexity or data volume alone.

While previous research has explored state recognition and fleet analytics in isolation, this work integrates these components into a cohesive, system-level framework designed for strict industrial constraints. Unlike traditional AIoT or deep learning approaches that rely on specialized sensing and large labelled fault datasets, our framework achieves robust predictive maintenance using only standard electrical parameters. By prioritizing this synergistic integration over mere algorithmic complexity, the solution remains cost-effective and non-intrusive, making it directly applicable to legacy machine fleets where zero-label constraints are a practical reality. The normalized Health Index (HI), derived from residual deviations, provided a reproducible and interpretable measure of machine behaviour over time. Stable assets maintained HI values above 0.9, while those exhibiting mechanical or electrical deviations showed gradual decreases toward 0.7–0.8. This indicates that the framework can reveal emerging irregularities early and with reduced false detections compared to purely statistical or unsupervised approaches. The smooth evolution of the HI also minimizes sensitivity to normal load transitions, which often trigger false alarms in context-free detection schemes.

While previous research has explored state recognition, regression modelling, and fleet-level analysis separately, this work represents the first integration of these components into a cohesive, system-level framework for predictive maintenance. The novelty of this approach lies not in the individual methods themselves, but in how they are combined to

create a system that operates under strict industrial constraints—particularly the lack of labelled fault data and the reliance on standard electrical parameters, such as active power, voltage, and current. This system-level integration enables scalable and cost-effective machine health monitoring without the need for additional sensors or labelled fault data. Compared with traditional Electrical Signature Analysis and existing machine learning models, the proposed framework combines physical interpretability with scalability and computational efficiency. Frequency- and wavelet-based ESA methods, though effective for stationary signals, remain sensitive to parameter tuning and are less robust under dynamic operation. Deep learning approaches offer high predictive power but rely on large, labelled datasets and limited explainability. The gradient-boosted modelling strategy used here achieves comparable accuracy on standard metering data while maintaining transparency, interpretability, and full compliance with ISO 17359 and ISO 13379 principles for data-driven diagnostics.

Fleet-level aggregation of the Health Index extended the analysis from single-machine evaluation to collective performance monitoring. By comparing weekly HI trajectories across machines, the framework enabled quantitative benchmarking and identification of assets with emerging performance deviations. Its lightweight implementation—requiring only a few minutes for training and less than 100 milliseconds for inference—ensures compatibility with existing SCADA and energy monitoring infrastructures and supports deployment on edge devices without specialized hardware.

Regarding the comparison with alternative modelling approaches listed in Table 2, a direct quantitative benchmark is not applicable, as the methods differ fundamentally in their data requirements and modelling assumptions. Unsupervised techniques such as K-Means and GMM do not provide stable regime boundaries under class imbalance, making them unsuitable for supervised per-state regression. Deep recurrent architectures (LSTM/RNN) require high-frequency waveform data and large labelled datasets, which are not available in the present industrial setting. Random Forests provide lower resolution and weaker separation between operating regimes, which led to higher residual variance in preliminary tests. For these reasons, XGBoost offered the best trade-off between robustness, interpretability and suitability for the proposed state-aware residual framework.

A notable limitation of the current framework is the potential for a self-referential loop, as electrical parameters (specifically active power) are used both for operating-state recognition and as targets for Health Index regression. In theory, a degradation that manifests as a significant power shift could be misclassified as a legitimate state transition, potentially masking the anomaly. However, this risk is mitigated by the multi-dimensional nature of the input features—including current, voltage, and power factor—which provides physical redundancy. Furthermore, the fleet-level aggregation acts as a critical safeguard; since the health baseline is derived from the median behaviour of multiple identical machines, an individual asset's degradation cannot "redefine" the global state boundaries, ensuring that significant deviations remain detectable.

Overall, the results confirm that the developed state-aware, electrical-only modelling framework provides a reliable, interpretable, and scalable foundation for continuous condition assessment in industrial environments. By linking data-driven modelling with physically meaningful indicators, the approach bridges the gap between raw electrical signals and actionable diagnostic information. Future work will focus on expanding the framework toward online adaptation, the inclusion of frequency-domain indicators, and transferability across heterogeneous machine types, further strengthening its role as a core enabler of intelligent and sustainable industrial operation.

6. Conclusions

This study presented a zero-label, state-aware framework for the health assessment of industrial assets using only Electrical Signature Analysis. By integrating automated regime identification with state-specific XGBoost regression models, the proposed Health Index (HI) effectively isolates mechanical and hydraulic degradation from normal operational load variations.

Key Findings:

- The method successfully detected emerging degradation trends 3–5 days before maintenance intervention without requiring any historical fault labels.
- The state-aware approach significantly reduced false positive rates compared to context-agnostic anomaly detection methods (e.g., Isolation Forest).
- The use of standard electrical parameters makes the framework highly scalable and deployable across various industrial fleets without additional sensor investments.

Challenges and Constraints

While robust, the framework's performance depends on the initial definition of the "healthy baseline." If the baseline period contains undetected minor defects, the reference dispersion may be inflated, reducing the sensitivity of the HI. Furthermore, extremely high-frequency transients (sub-millisecond) are not captured by the 5 Hz sampling rate, which may limit the detection of certain rapid electrical switchgear faults.

Future Work

Future research will focus on the automatic adaptation of the baseline (transfer learning) to account for seasonal environmental changes (e.g., ambient temperature effects on hydraulic oil viscosity). Additionally, we aim to extend the methodology to multi-asset correlation, where the health of one machine is cross-referenced with its neighbours in the fleet to further isolate external grid-level disturbances from local mechanical wear.

Author Contributions: Conceptualization, D.H. and P.T.; methodology, D.H. and P.T.; software, D.H.; validation, D.H. and P.T.; formal analysis, D.H.; investigation, D.H.; resources, D.H.; data curation, D.H.; writing—original draft preparation, D.H.; writing—review and editing, P.T.; visualization, D.H.; supervision, P.T.; project administration, P.T.; funding acquisition, P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences, grant number VEGA 1/0770/25: "Research on the transformation of current SME companies to the level of sustainable and resilient smart companies using advanced exponential technologies in the context of Industry 4.0 and 5.0".

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed during the current study are not publicly available due to industrial confidentiality agreements but are available from the corresponding author on reasonable request. A small, anonymized excerpt of the processed dataset is included in Appendix A for reference.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

HI	Health Index
XGBoost	Extreme Gradient Boosting
ESA	Electrical Signature Analysis

IIOT	Industrial Internet of Things
MAD	Median Absolute Deviation
RMS	Root Mean Square
PF	Power Factor
DSP	Digital Signal Processing
PHM	Prognostics and Health Management
IF	Isolation Forest
SCADA	Supervisory Control and Data Acquisition

Appendix A

Table A1. Sample of the dataset showing time, machine performance (P_total, I_mean, PF_mean), operational state, and corresponding Health Index (HI) values for a specific machine.

_time	Machine_public	P_total	I_mean	PF_mean	State_name	HI
7 November 2024 22:45	1	24,704	36.83	0.98	operation	0.96
7 November 2024 22:45	1	20,285	36.8	0.98	operation	0.88
7 November 2024 22:46	1	20,105	26.78	0.96	operation	0.91
7 November 2024 22:47	1	18,914	34.9	0.97	operation	0.86
7 November 2024 22:47	1	24,681	34.82	0.98	operation	0.95
7 November 2024 22:47	1	14,329	34.88	0.94	heating	0.88
7 November 2024 22:48	1	18,804	28.61	0.96	operation	0.99
7 November 2024 22:49	1	18,694	26.68	0.96	operation	0.97
7 November 2024 22:49	1	23,224	36.18	0.97	operation	0.99
7 November 2024 22:50	1	24,653	34.80	0.98	operation	0.94
7 November 2024 22:50	1	24,706	34.83	0.98	operation	0.97
7 November 2024 22:52	1	18,987	26.72	0.98	operation	0.96
7 November 2024 22:52	1	26,210	36.88	0.98	operation	0.95
7 November 2024 22:53	1	18,742	26.77	0.96	operation	0.99
7 November 2024 22:53	1	18,737	26.77	0.96	operation	0.98
7 November 2024 22:54	1	18,691	26.73	0.96	operation	0.96
7 November 2024 22:54	1	18,744	26.69	0.96	operation	0.99
7 November 2024 22:55	1	18,912	34.73	0.98	operation	0.82
7 November 2024 22:56	1	18,777	26.68	0.96	operation	0.99
7 November 2024 22:56	1	18,706	26.66	0.96	operation	0.97

References

1. Issa, R. Review of Fault Diagnosis Methods for Induction Machines. *Energies* **2024**, *17*, 2728. [CrossRef]
2. Hamani, K.; Kheldoun, A.; Moulahoum, S.; Bessous, N.; Benslama, M.; Cherif, A. Advancements in Induction Motor Fault Diagnosis and Condition Monitoring: A Comprehensive Review. *Sensors* **2025**, *25*, 5942. [CrossRef] [PubMed]
3. Kumar, R.R.; Andriollo, M.; Cirrincione, G.; Cirrincione, M.; Tortella, A. A Comprehensive Review of Conventional and Intelligence-Based Approaches for the Fault Diagnosis and Condition Monitoring of Induction Motors. *Energies* **2022**, *15*, 8938. [CrossRef]
4. Glowacz, A.; Sulowicz, M.; Zielonka, J.; Li, Z.; Glowacz, W.; Kumar, A. Acoustic fault diagnosis of three-phase induction motors using smartphone and deep learning. *Expert Syst. Appl.* **2025**, *262*, 125633. [CrossRef]
5. Benbouzid, M.E.H. A review of induction motors signature analysis as a medium for faults detection. *IEEE Trans. Ind. Electron.* **2000**, *47*, 984–993. [CrossRef]
6. Bhosinak, S.; Audomsi, S.; Angkawisittpan, N.; Photong, C.; Sa-Ngiamvibool, W. Optimized Machine Learning for Induction Motor Fault Diagnosis Using Vibration and Frequency-Domain Features. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 28584–28590. [CrossRef]
7. Halder, S.; Bhat, S.; Zychma, D.; Sowa, P. Broken Rotor Bar Fault Diagnosis Techniques Based on Motor Current Signature Analysis—A Review. *Energies* **2022**, *15*, 8569. [CrossRef]
8. Antonino-Daviu, J.A.; Riera-Guasp, M.; Roger-Folch, J.; Molina, M.P. Validation of a New Method for the Diagnosis of Rotor Bar Failures via Wavelet Transform in Industrial Induction Machines. *IEEE Trans. Ind. Appl.* **2006**, *42*, 990–996. [CrossRef]

9. Douglas, H.; Pillay, P.; Ziarani, A. Detection of Broken Rotor Bars in Induction Motors Using Wavelet Analysis. In Proceedings of the IEEE International Electric Machines and Drives Conference, 2003. IEMDC'03, Madison, WI, USA, 1–4 June 2001.
10. Cocca, P.; Rossi, A.; Albertini, M. Anomaly detection using electrical signature analysis and machine learning: Application to a CNC mill. *IFAC-PapersOnLine* **2024**, *58*, 502–507. [CrossRef]
11. *ISO 17359:2018*; Condition Monitoring and Diagnostics of Machines—General Guidelines. ISO: Geneva, Switzerland, 2018.
12. *ISO 13379-1:2025*; Condition Monitoring and Diagnostics of Machine Systems—Data Interpretation and Diagnostics Techniques—Part 1. ISO: Geneva, Switzerland, 2025.
13. *ISO 13379-2:2015*; Condition Monitoring and Diagnostics—Data-Driven Applications. ISO: Geneva, Switzerland, 2015.
14. Pohakar, P.; Gandhi, R.; Hans, S.; Sharma, G.; Bokoro, P.N. Analysis of Multiple Faults in Induction Motor Using Machine Learning Techniques. *E-Prime-Adv. Electr. Eng. Electron. Energy* **2025**, *12*, 101007. [CrossRef]
15. Hassan, O.E.; Amer, M.; Abdelsalam, A.K.; Williams, B.W. Induction motor broken rotor bar fault detection techniques based on fault signature analysis—A review. *IET Electr. Power Appl.* **2018**, *12*, 895–907. [CrossRef]
16. Douglas, H.; Pillay, P.; Ziarani, A.K. Broken Rotor Bar Detection in Induction Machines with Transient Operating Speeds. *IEEE Trans. Energy Convers.* **2005**, *20*, 135–141. [CrossRef]
17. Li, Y.; Lin, J.; Niu, G.; Wu, M.; Wei, X. A Hilbert–Huang Transform-Based Adaptive Fault Detection and Classification Method for Microgrids. *Energies* **2021**, *14*, 5040. [CrossRef]
18. Chisedzi, L.P.; Muteba, M. Detection of Broken Rotor Bars in Cage Induction Motors Using Machine Learning Methods. *Sensors* **2023**, *23*, 9079. [CrossRef] [PubMed]
19. Chen, J.; Wang, Y.; He, Y. A Method for Broken Rotor Bars Diagnosis Based on Sum-of-Squares of Current Signals. *Appl. Sci.* **2020**, *10*, 5980. [CrossRef]
20. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
21. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
22. Artigao, E.; Honrubia-Escribano, A.; Gómez-Lázaro, E. Current signature analysis to monitor DFIG wind turbine generators. *Renew. Energy* **2018**, *116*, 5–14. [CrossRef]
23. Chai, B.X.; Gunaratne, M.; Ravandi, M.; Wang, J.; Dharmawickrema, T.; Di Pietro, A.; Jin, J.; Georgakopoulos, D. Smart Industrial Internet of Things Framework for Composites Manufacturing. *Sensors* **2024**, *24*, 4852. [CrossRef] [PubMed]
24. Artigao, E.; Honrubia-Escribano, A.; Gomez-Lazaro, E. Condition monitoring of a wind turbine DFIG through current signature analysis. *J. Phys. Conf. Ser.* **2017**, *926*, 012008. [CrossRef]
25. NIST/SEMATECH. Median Absolute Deviation (MAD). 2016. Available online: https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/mad.htm (accessed on 30 October 2025).
26. NIST/SEMATECH. H15 Robust Location/Scale Estimator. 2011. Available online: https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/h15.htm (accessed on 30 October 2025).
27. NIST/SEMATECH. Detection of Outliers—Modified Z-Score Rule. 2012. Available online: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm (accessed on 30 October 2025).
28. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley: Boston, MA, USA, 1977.
29. Iglewicz, B.; Hoaglin, D.C. *How to Detect and Handle Outliers*; ASQC Press: Milwaukee, WI, USA, 1993.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.