

Article

Data-Driven Calibration Scheduling for Measuring Instruments—A Smart Management Architecture

Marcel Behún ¹, Gabriel Galgócí ², Mária Kozlovská ² , Matúš Pohorenc ²  and Annamária Behúnová ^{3,*} 

¹ Faculty of Mining, Ecology, Process Control and Geotechnologies, Institute of Earth Resources, Technical University of Kosice, 042 00 Kosice, Slovakia; marcel.behun@tuke.sk

² Faculty of Civil Engineering, Institute of Construction Technology, Economics and Management, Technical University of Kosice, 042 00 Kosice, Slovakia; gabriel.galgoci@student.tuke.sk (G.G.); maria.kozlovska@tuke.sk (M.K.); matus.pohorenc@tuke.sk (M.P.)

³ Faculty of Mining, Ecology, Process Control and Geotechnologies, Institute of Logistics and Transport, Technical University of Kosice, 042 00 Kosice, Slovakia

* Correspondence: annamaria.behunova@tuke.sk

Abstract

Reliable calibration of measuring instruments is a foundational yet under-reported layer of the smart-manufacturing stack: without traceable calibrations, no downstream process can guarantee the metrological quality of the data it consumes. This research asks whether an administratively managed calibration record base is sufficient to support data-driven recalibration scheduling, and diagnoses the data-governance conditions such scheduling requires. A production multi-faculty university calibration database (3552 registered instruments, 371 calibration events, 2009–2025) is analyzed to characterize fleet inventory and compliance, quantify the historical recalibration late-rate with bootstrap and analytical 95% confidence intervals stratified by faculty and device type, benchmark supervised models against explicit baselines, and propose a transparent rule-based scheduler with an empirical faculty-level safety buffer. The historical late-rate is 47.8% (95% CI 34.1–61.9%), with exploratory inter-faculty heterogeneity, while supervised learning yields only a weak, uncertain signal and no useful deviation regression. Because 86.7% of the fleet lacks the baseline records needed to schedule it, the practical message for instrument managers is direct: completing the missing last-calibration and calibration-period fields is the highest-value first step, after which a transparent, data-governance-informed scheduler—not a black-box predictor—is the defensible deliverable, transferable to comparable research and manufacturing infrastructures.

Keywords: metrological data quality; measurement traceability; compliance analytics; small-sample inference; Industry 5.0



Academic Editors: Arkadiusz Gola
and César A. Collazos

Received: 12 June 2026

Revised: 23 July 2026

Accepted: 30 July 2026

Published: 2 August 2026

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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Smart manufacturing and the broader Industry 4.0/5.0 paradigm have placed data-driven decision support at the center of operations management [1–4]. Most attention has been directed at the visible “shop-floor” layer: condition monitoring, digital twins, predictive maintenance of machine tools, deep learning for fault diagnosis, and real-time analytics from IoT sensors [5–9]. A less glamorous but no less critical layer underpins all of these: the metrological layer that ensures every measurement entering the data pipeline—from a vernier caliper on a quality-assurance station to a spectrum analyzer

in a research laboratory—is traceable to a recognized standard and was acquired with a calibrated instrument [10–13].

In practice, the management of this metrological layer is dominated by spreadsheets, ad hoc reminders, and rigid time-based recalibration intervals dictated by manufacturer recommendations or local regulations. The scheduling rule “calibrate every T months” is universal but blunt: it ignores the actual usage profile, historical compliance behavior, and risk distribution across the fleet. A modern data-driven approach would replace, or at least augment, such heuristics with quantitative evidence drawn from the institution’s own calibration history [14–18].

The rapid proliferation of the machine-learning literature in predictive maintenance [19–21] has, however, also produced a documented tendency to overclaim: small sample sizes, inappropriate train/test splits, missing baselines, and a lack of statistical uncertainty quantification routinely yield reported accuracies that do not generalize [22,23]. Calibration data are particularly susceptible to this trap because the unit of analysis is a paired recalibration event, of which even mature institutions accumulate only a few hundred over a decade. Honest reporting in this regime requires explicit baselines, confidence intervals, and a willingness to declare a negative result when the data warrant it.

The overarching aim of this research is to determine to what extent a real, administratively managed calibration record base can support data-driven recalibration scheduling, and to identify the data-governance preconditions for it. The guiding question is therefore diagnostic rather than purely predictive: given the records an institution actually holds, which devices can be scheduled on evidence, and what must be corrected before supervised prediction becomes defensible? On this basis this research makes four contributions:

- This research describes the structure and content of the relational calibration-management database operated by a Slovak technical university across 11 faculties and 57 departments, and quantifies the inventory at the level of devices, events, suppliers and documents.
- An empirical, statistically grounded compliance analysis of recalibration timeliness is carried out at the institution level and stratified by faculty and device type, reporting bootstrap, Wilson and Clopper–Pearson 95% confidence intervals to handle the small-sample regime [24,25].
- A small set of supervised classifiers and regressors is benchmarked against explicit baselines under stratified k -fold cross-validation and permutation feature importance [26–29], and the outcome is reported honestly as exploratory, including the negative result of the regression task.
- A transparent, rule-based fleet-wide scheduler is proposed that combines the regulatory calibration period with an empirical safety buffer scaled to the faculty’s historical late-rate, and is applied to all 3552 devices, producing a six-class scheduling status that makes uncalibrated and incomplete sub-fleets explicit rather than hiding them behind a fabricated prediction.

It should be emphasized that the proposed architecture is a transparent, rule-based deployment layer rather than a validated operational prediction system: at the present data maturity its role is to make each device’s schedulability explicit and to expose the records that must be completed before predictive scheduling becomes defensible. Recent work on digital metrology situates calibration and instrument management within the Industry 4.0 digitalization agenda [30].

The paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the calibration database, the data-extraction pipeline, the exploratory analysis methodology, the modeling protocol with baselines, and the rule-based scheduler. Section 4

reports the descriptive, inferential and modeling results in turn. Section 5 discusses implications, limitations and deployment considerations. Section 6 concludes.

2. Literature Review

2.1. *Smart Manufacturing and Industry 4.0/5.0*

The notion of “smart manufacturing” emerged as the convergence of cyber–physical systems, the Industrial Internet of Things, advanced analytics, and additive and subtractive manufacturing [31–33]. Industry 5.0 has subsequently shifted the discourse toward human-centric, sustainable and resilient production [34,35]. Across both paradigms, the unifying assumption is that operational decisions are increasingly grounded in data. The literature on data quality and traceability in this context, however, remains comparatively thin [36–39].

Within the broader smart-manufacturing literature, instrument and tool management has historically received less attention than process monitoring or predictive maintenance of capital equipment [40–43]. The role of educational and research infrastructure in piloting Industry 4.0/5.0 enablers has likewise been highlighted [44], which provides the institutional context for the present case study. Recent contributions from the Slovak research community frame these paradigms in compatible terms: Knapčíková and Peraković [45] frame innovative technology approaches and challenges for Industry 4.0 in industrial engineering, while Peraković et al. [46] model the IoT-enabled ecosystem elements of Industry 4.0—directly relevant to the multi-faculty institutional context studied here.

2.2. *Calibration, Metrology and Compliance*

The international metrology framework distinguishes between calibration, which establishes the relation between values indicated by a measuring instrument and the corresponding values realized by a standard, and verification, which declares that an instrument satisfies stated requirements. The conceptual boundary between calibration and verification has been re-examined in recent metrology practice [47]. ILAC G24 and the NIST handbook recommend that calibration intervals be re-evaluated based on accumulated evidence rather than held constant indefinitely. The EURACHEM/CITAC guidance similarly emphasizes the role of measurement-uncertainty information in compliance assessment [48]. In practice, however, periodic schedules are rarely revised in response to compliance data [49], partly because the underlying historical records are seldom analyzed systematically. Empirical work on the metrological layer itself remains comparatively rare. Trebuňa and Hagara [50] report systematic experimental vibration and modal analysis, illustrating the measurement effort required to characterize mechanical systems in laboratory settings, and Knapčíková et al. [51] address sustainable material circularity of components within a modern manufacturing enterprise—a theme complementary to the data-governance findings reported in Section 4.

2.3. *Data-Driven and Predictive Maintenance*

The machine-learning maintenance literature is by now extensive [52–55]. Random forests, gradient boosting, and increasingly explainable variants [56] dominate published benchmarks. Recent bibliometric reviews confirm the continued expansion of machine-learning methods in this domain [57]. Methodological reviews, however, have raised persistent concerns: cross-validation failure with small samples, leakage from improperly designed splits, and a missing culture of explicit baselines. These warnings are particularly germane to calibration data, where the analysis unit (a paired recalibration event) accumulates slowly and the dominant outcome (PASS) creates extreme class imbalance. The applied condition-monitoring literature has further shown that in industrial maintenance settings simple statistical descriptors and rule-based thresholds often remain the

dominant operational tool [58]. Analogously, asset and equipment information systems for construction and industrial enterprises increasingly rely on structured BIM-anchored database records as the substrate for any subsequent analytics or scheduling layer [59,60]. These observations motivate the deliberately conservative, baseline-anchored modeling protocol adopted in Section 3.5.

2.4. Position of the Present Work

Two gaps motivate the present research. First, the smart-manufacturing literature is dominated by sensor-rich, high-frequency datasets. The equally important case of administratively structured, low-frequency calibration records receives little methodological attention. Second, existing predictive-maintenance benchmarks rarely engage with the small-sample reality faced by individual institutions, and almost never report negative results. The present paper addresses both gaps with a transparent, evidence-driven case study and a deliberately conservative reporting style. Taken together, the reviewed literature establishes the methods—condition-based maintenance and supervised learning—and the accompanying small-sample pitfalls, but leaves largely unexamined the prior question of whether the administrative record base is complete enough to support any of them. It is this precondition that the present work foregrounds. The Slovak smart-manufacturing literature reviewed in Sections 2.1–2.3 supplies the immediate institutional context for the architecture proposed here.

3. Experimental Research

3.1. The Calibration Database

The calibration database is a production SQL Server schema operated by a Slovak technical university to manage its fleet of measuring and laboratory instruments. Its schema comprises 15 domain-specific tables (plus an auxiliary job-scheduling layer) organized around four central entities: devices (the inventory), device calibrations (the event log), calibration period codebook (the regulatory periodicity catalog), and companies (manufacturers and suppliers). Auxiliary tables capture device categories, components, parts, documents, departmental responsibilities, and units of measurement. Because the instruments are de-identified for data-protection reasons, individual device photographs are not reproduced. Each device is characterized only by its categorical type, measurement type and owning faculty. The device-type and measurement-type labels are anonymized categorical enumerations. Their physical mapping is withheld under the same data-protection policy, so device types are referred to by their coded values throughout.

The dataset analyzed here was extracted via a series of 21 parameter-free SQL queries executed against a read-only mirror of the production database. All queries return tabular, de-identified outputs. No personal data or restricted institutional information are included. Table 1 summarizes the headline counts of the extract.

Table 1. Headline counts of the calibration-database extract used in this research (snapshot, 2025).

Entity	Count
Active devices (in service)	3544
Decommissioned devices	8
Total registered devices	3552
Calibration events	371
Devices with ≥ 1 calibration	325
Distinct faculties	11
Distinct departments	57
Distinct manufacturers/suppliers	265

Table 1. *Cont.*

Entity	Count
Calibration documents on file	207
Parts in stock	45
Earliest device registration	2009
Most recent calibration event	2025

3.2. Data-Extraction Pipeline

The extraction pipeline produces twenty-one tab-separated files. Four are central to the present analysis: the device inventory (3552 rows $\times$ 43 columns), the calibration event log (371 rows $\times$ 27 columns), a derived paired-event table for compliance analysis (see Section 3.3), and a summary-statistics table that consolidates the headline counts in Table 1. The remaining files contribute reference data on calibration periods, device categories, companies, departmental responsibilities, documents, calibration result distributions, year-by-year aggregates, and date ranges.

All downstream processing was performed in Python 3.14 with pandas 3.0.2 [61], numpy 2.4.4 [62], scikit-learn 1.8.0 [63], matplotlib 3.10.8 and standard scientific-Python libraries. The complete analysis pipeline is fully reproducible against the twenty-one extract files and regenerates every number, table and figure reported here. The extraction queries and the analysis notebook are available from the corresponding author on request.

3.3. Compliance Analysis Unit and Sample Size

The fundamental unit of compliance analysis is a paired recalibration event: a calibration record together with its predecessor for the same device, allowing computation of the actual inter-calibration interval, the expected interval implied by the assigned CalibrationPeriodMonths, and the deviation between the two. A device contributes a paired event only if it has been calibrated at least twice. In the snapshot analyzed here, this yields $n = 46$ paired events across 46 unique devices. Of the 325 devices with at least one calibration, 46 have two or more records, and these 46 devices account for exactly 92 calibration events (two each). No device carries a third record, so retaining the latest pair per device discards no information. The reduction from 371 events to 46 paired events therefore reflects the structural scarcity of repeated calibrations rather than an analytical exclusion. This small sample size is the dominant constraint on the inferential analysis and is treated as such throughout. Because these 46 repeatedly calibrated devices are precisely those that carry sufficient records, they need not be representative of the full fleet of 3552 devices. The observed late-rate and faculty-level heterogeneity are therefore conditional on record availability and may be affected by record-completeness bias.

For each paired event, the following quantities are computed:

1. DaysBetweenCalibrations = CalibrationDate – PrevCalibrationDate;
2. ExpectedIntervalDays = CalibrationPeriodMonths $\times$ 30.44;
3. DeviationDays = DaysBetweenCalibrations – ExpectedIntervalDays;
4. CalibratedOnTime $\in \{0, 1\}$, where 1 indicates the new calibration occurred before the validity expiry of the previous one;
5. DaysOverdueOrEarly = days between actual recalibration and the validity-expiry date.

3.4. Empirical Late-Rate Estimation

To characterize compliance, the population late-rate is estimated, $p = 1 - \mathbb{E}[\text{CalibratedOnTime}]$, together with two 95% confidence intervals appropriate to small-sample binomial inference: the Wilson score interval and a non-parametric bootstrap with

$B = 5000$ resamples. Stratified estimates are reported by faculty and by device type for groups with at least three observations. Because the per-stratum counts are small, each proportion is additionally reported with a Clopper–Pearson exact 95% interval, between-group contrasts are assessed with Fisher’s exact test, and all stratified estimates are treated as exploratory.

3.5. Modeling Protocol

A small set of supervised models is benchmarked against explicit baselines, strictly under five-fold stratified cross-validation. No held-out test set is reported, because at $n = 46$ a single test split is uninformative. Two tasks are considered:

1. Classification of CalibratedOnTime (binary). Models: a stratified baseline Dummy-Classifier, a most-frequent baseline, a standardized logistic regression with class weighting, a random forest, and a gradient boosting machine.
2. Regression of DeviationDays. Models: mean and median dummy baselines, a standardized Ridge regression, a random forest regressor, and a gradient boosting regressor.

Features were drawn from the calibration period (months), device type, measurement type, faculty identifier, calibration sequence number, calibration price, and binary administrative flags (categorized, imported, portable). A pre-modeling feature audit discarded constants (MeasureType, CalSeq) and flagged near-constants (>95% on the dominant value) for cautious interpretation. Permutation importance was used in place of raw coefficients to compare feature contributions across model families. Reported metrics are mean $\pm$ standard deviation across folds.

3.6. Rule-Based Fleet Scheduler

For deployment, a transparent rule is constructed that is applicable to every device in the fleet, including those without a complete history:

$$\text{RecommendedDate}_d = \text{LastCal}_d + \text{Period}_d - B(F_d),$$

where $B(F_d) = \min(\text{MAX_BUFFER}, \hat{p}_{F_d} \cdot \text{MAX_BUFFER})$ is a faculty-level safety buffer linearly scaled by the empirical late-rate $\hat{p}_{F_d}$ of the faculty to which device d belongs, capped at $\text{MAX_BUFFER} = 30$ days. The 30-day cap is a managerial risk-tolerance parameter rather than a fitted value; it bounds how far a high-late-rate faculty is recalled ahead of the nominal date, keeping the earliest recall within one month of the regulatory due date so that recall cost is not materially increased. A sensitivity check over 15-, 30- and 60-day caps leaves the fleet-status decomposition unchanged, because the actionable devices are already well past their nominal date. The buffer therefore tunes recall timing, not the coverage diagnosis. Devices with a missing LastCal or missing CalibrationPeriodMonths are not assigned a recommended date. Instead, they are flagged into one of three “incomplete” status classes (NEEDS BASELINE CALIBRATION, NO PERIOD ASSIGNED, NO PERIOD & NO HISTORY). The remaining devices receive one of three “actionable” status classes (OVERDUE, DUE SOON (<90 d), SCHEDULED) based on DaysUntilNominal. The full breakdown is reported in Section 4.5.

4. Results

4.1. Inventory Overview

Figure 1 presents the inventory composition across faculties, device types and measurement types, together with the top fifteen departments by device count. Distribution across the eleven faculties is highly skewed. The largest faculty holds approximately one third of the fleet, and the top five departments together account for close to half of the fleet

(44.8%). Device-type 1 dominates the fleet (approximately 70% of records, 2484 of 3552), with type 0 and type 3 forming smaller sub-populations.

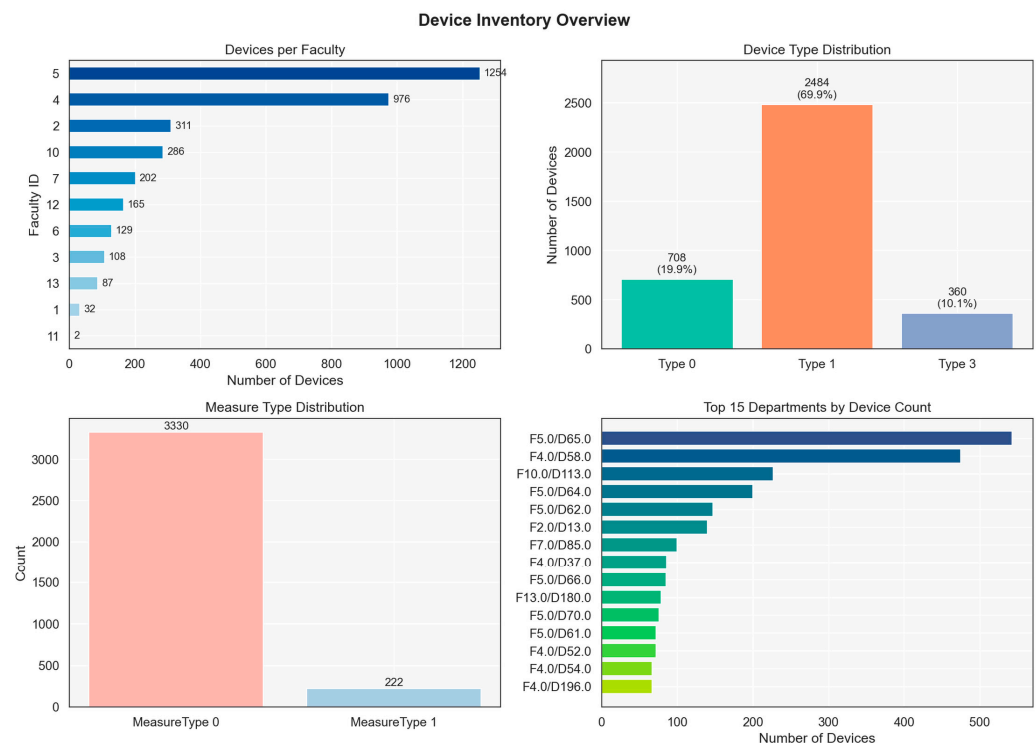


Figure 1. Inventory composition across faculties, device types, measurement types and the top fifteen departments by device count.

The inventory composition exposes a structural concentration that conditions every subsequent finding: a single faculty owns roughly one third of the fleet and the top five departments together own close to half of the registered devices (44.8%). Operationally, this means that any compliance improvement targeted at the heaviest holders propagates to most of the institution, while light-tail faculties contribute relatively little absolute risk even if their per-device behavior is poor. The dominance of device-type 1 (approximately 70% of records) further implies that the calibration regime can be optimized for a small number of recurring instrument categories rather than a fragmented long tail. The most consequential observation, however, is data-governance: of 3552 registered devices only 325 (9.1%) carry any calibration record at all and only 80 (2.2%) have a populated Last-CalibrationDate. This is the single largest finding of the research and the principal reason why the proposed scheduler must explicitly handle missing-history sub-fleets rather than fabricate predictions.

4.2. Calibration Interval Analysis

Figure 2 reports the four core compliance distributions on the paired-event sub-sample ($n = 46$): the actual inter-calibration interval, the deviation from the expected interval (as a boxplot to expose the single outlier of +1345 days), the on-time vs. late split with Wilson 95% CI, and the days overdue/early relative to validity expiry. Display ranges for the histograms are clipped at the 95th percentile to prevent the single outlier from collapsing the visible distribution. The outlier is annotated explicitly.

The four interval-distribution panels jointly characterize the empirical compliance regime on the paired-event sub-sample ($n = 46$). The actual inter-calibration interval is centered close to the regulatory period but heavily right-skewed. The deviation boxplot exposes a single +1345-day outlier that dominates the upper tail and that is annotated

explicitly rather than clipped silently. The on-time vs. late split—24 vs. 22 events—yields the headline 47.8% late-rate (Wilson 95% CI 34.1–61.9%, bootstrap 95% CI 32.6–63.0%). Half of all events deviate by no more than four days from the planned interval, so the failure mode is not a uniform laxity but a concentrated minority of severely late recalibrations. This shape is precisely the regime in which a faculty-scaled safety buffer is more useful than a single global early-recall offset.

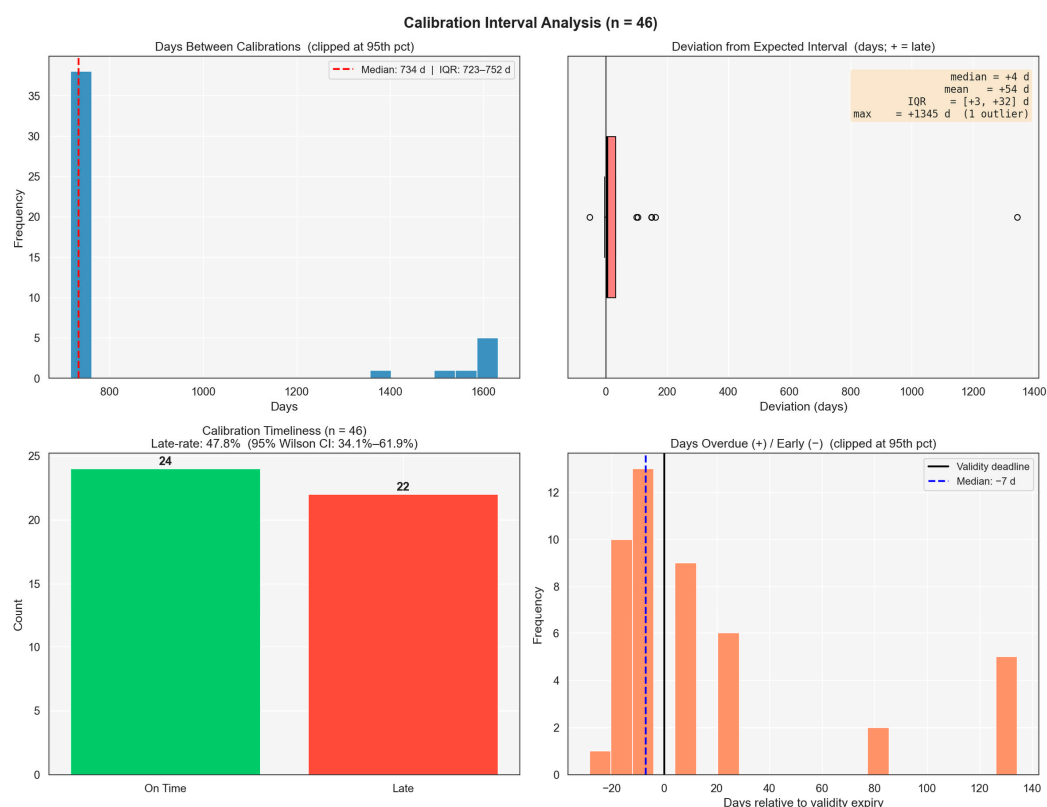


Figure 2. Calibration interval analysis ($n = 46$ paired events): the days between calibrations, deviation from expected interval, on-time vs. late split with Wilson 95% CI, and days overdue/early. The histograms are clipped at the 95th percentile. The single +1345 day outlier is annotated. The circles in the deviation panel (top right) are boxplot outliers, that is, individual paired events lying more than $1.5 \times \text{IQR}$ beyond the quartiles.

The headline statistics are summarized in Table 2. The empirical late-rate is 47.8% with a Wilson 95% CI of 34.1–61.9% (Clopper–Pearson 32.9–63.1%) and a bootstrap 95% CI of 32.6–63.0%. Half of all events deviate by no more than four days from the planned interval, but the upper tail is heavy: a single device exceeded its expected interval by more than three years.

Table 2. The empirical compliance statistics on the paired-event sub-sample ($n = 46$).

Statistic	Value
Paired recalibration events (n)	46
On-time recalibrations	24 (52.2%)
Late recalibrations	22 (47.8%)
Wilson 95% CI for late-rate	34.1–61.9%
Bootstrap 95% CI for late-rate ($B = 5000$)	32.6–63.0%
Median deviation	+4 days
Inter-quartile range of deviation	[+3, +32] days
Maximum deviation (single outlier)	+1345 days

4.3. Empirical Late-Rate Stratified by Faculty and Device Type

Figure 3 reports the bootstrap-based late-rate estimates stratified by faculty and device type, restricted to groups with at least three observations. The stratification reveals quantitatively meaningful and, at these sample sizes, exploratory heterogeneity:

1. Faculty 13: late-rate 87.5% with bootstrap 95% CI [62.5%, 100%] ($n = 8$);
2. Faculty 12: late-rate 39.5% with bootstrap 95% CI [23.7%, 55.3%] ($n = 38$).

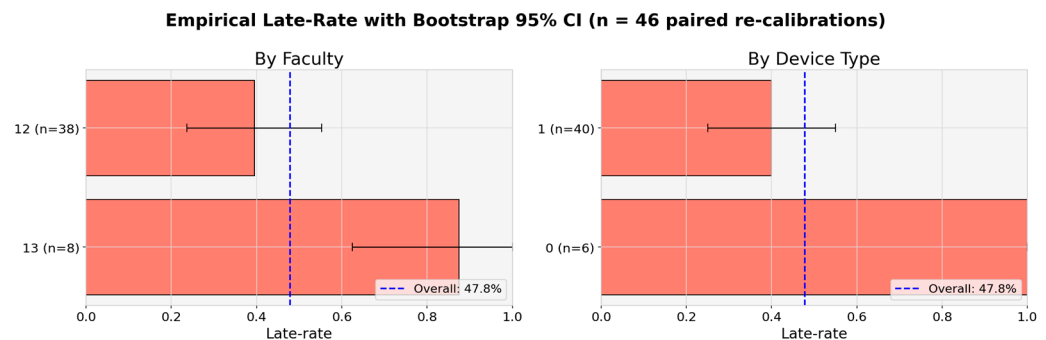


Figure 3. Empirical late-rate with bootstrap 95% confidence intervals, stratified by faculty (**left**) and device type (**right**). Strata with $n < 3$ are omitted. The dashed blue line marks the overall late-rate.

The two faculty confidence intervals do not overlap, and the contrast is significant by Fisher's exact test (odds ratio 10.7, $p = 0.020$), consistent with compliance behavior that is not uniform across organizational units. Given the small strata ($n = 8$ and $n = 38$), the result is nonetheless reported as exploratory. By device type, type 0 (six observations) recorded 100% late events, while type 1 (forty observations) returned 40.0% (95% CI 25.0–55.0%). These two strata provide a preliminary empirical basis for the faculty-scaled safety buffer in the rule-based scheduler (Section 4.4).

Stratification by faculty produces non-overlapping bootstrap confidence intervals for the two largest strata: Faculty 13 records an 87.5% late-rate (95% CI 62.5–100%, $n = 8$) while Faculty 12 records 39.5% (95% CI 23.7–55.3%, $n = 38$). The non-overlapping intervals, corroborated by Fisher's exact test ($p = 0.020$), point to heterogeneity consistent with non-uniform compliance across the institution—reported as exploratory at these sample sizes. Pooling the fleet would mask this heterogeneity and motivates the faculty-level safety buffer of Section 3.6. The device-type panel reinforces the point—type 0 (six observations) recorded 100% late events while type 1 (forty observations) returned 40.0% (95% CI 25.0–55.0%)—and the type 0 versus type 1 contrast is significant by Fisher's exact test ($p = 0.008$), though it is interpreted with caution given the small per-stratum sample sizes and the absence of correction for multiple comparisons. These preliminary empirical risk bands inform the way the rule-based scheduler scales its recall-day offset.

4.4. Exploratory Predictive Models with Explicit Baselines

Table 3 reports the five-fold cross-validation ROC-AUC for the classification of Calibrate-dOnTime. The standardized logistic regression delivers the best mean CV-AUC (0.655 ± 0.18), modestly above the stratified dummy baseline (0.40 ± 0.18). Tree-based ensembles do not improve on the linear model at this sample size, and the scale of the standard deviations (driven by ten observations per fold) cautions against any strong claim of validated predictive performance. Under the same five-fold protocol the best classifier attains a balanced accuracy of 0.61 ± 0.08 and a PR-AUC of 0.62 against a positive-class prevalence of 0.52 (PR-AUC for the operationally relevant late class is correspondingly weaker), and its out-of-fold confusion matrix recovers only six of the 22 late events (16 missed). Supervised classification is therefore reported as exploratory and not yet deployable at this sample size rather than as a core predictive contribution. These additional metrics are collected in Table 4.

Table 3. Classification of CalibratedOnTime under stratified five-fold CV. Mean $\pm$ standard deviation across folds.

Model	ROC-AUC (Mean $\pm$ Std)	Accuracy (Mean $\pm$ Std)
Baseline (stratified)	0.40 $\pm$ 0.18	0.41 $\pm$ 0.18
Baseline (most-frequent)	0.50 $\pm$ 0.00	0.52 $\pm$ 0.04
Logistic Regression (scaled)	0.655 $\pm$ 0.18	0.61 $\pm$ 0.13
Random Forest	0.61 $\pm$ 0.16	0.59 $\pm$ 0.14
Gradient Boosting	0.58 $\pm$ 0.15	0.57 $\pm$ 0.12

Table 4. Additional exploratory classification metrics for the best classifier (standardized logistic regression) under stratified five-fold cross-validation. The last two rows summarize the out-of-fold confusion matrix.

Metric	Value
Balanced accuracy (5-fold CV)	0.61 $\pm$ 0.08
PR-AUC (5-fold CV)	0.62 (positive-class prevalence 0.52)
On-time events correctly recovered (out-of-fold)	23 of 24
Late events correctly recovered (out-of-fold)	6 of 22 (16 missed)

Table 5 reports the regression results for DeviationDays. No machine-learning model beats the median dummy baseline. This is reported as a negative result. At $n = 46$, with a single outlier of +1345 days, the variance of the target overwhelms any signal the available features can extract.

Table 5. Regression of DeviationDays under stratified five-fold CV (target binarized by sign for stratification). Mean $\pm$ standard deviation across folds.

Model	MAE (Days, Mean $\pm$ Std)	R^2 (Mean $\pm$ Std)
Baseline (mean)	99.6 $\pm$ 86	−0.05 $\pm$ 0.10
Baseline (median)	60.5 $\pm$ 47	−0.06 $\pm$ 0.10
Ridge (scaled)	110.7 $\pm$ 92	−0.18 $\pm$ 0.20
Random Forest	73.4 $\pm$ 50	−0.10 $\pm$ 0.15
Gradient Boosting	78.2 $\pm$ 53	−0.12 $\pm$ 0.16

4.5. Rule-Based Fleet-Wide Schedule

Applying the rule of Section 3.6 to the full fleet of 3552 devices produces the six-class status decomposition shown in Figure 4. The dominant class is NO PERIOD & NO HISTORY: 3078 devices (86.7% of the fleet) lack both a LastCalibrationDate and a CalibrationPeriodMonths and therefore cannot be scheduled at all. A further 394 devices (11.1%) have a period but no history (NEEDS BASELINE CALIBRATION), and 71 devices (2.0%) have a recorded last-calibration date but no period assignment in the codebook (NO PERIOD ASSIGNED). Only nine devices (0.3%) admit a recommended date under the current rule, and all of them are already past their nominal recalibration date, classifying them as OVERDUE. None of the actionable devices fall into the DUE SOON or SCHEDULED classes, indicating that the records currently usable by the scheduler are precisely those whose periodic review has been allowed to lapse the longest.

The fleet-wide decomposition shows that the dominant operational issue today is not scheduling sophistication but baseline data coverage: 86.7% of devices fall into NO PERIOD & NO HISTORY and a further 11.1% into NEEDS BASELINE CALIBRATION. Only nine devices (0.3%) currently admit a recommended date under the rule, and all of them are already past their nominal recalibration date. The right-hand panel makes the design choice of the scheduler concrete: faculties with the highest historical late-rate receive the largest empirical buffer (capped at 30 days), so that high-risk organizational units are

recalled earlier than low-risk ones for the same regulatory period. As noted in Section 3.6, varying this cap between 15 and 60 days does not change the six-class decomposition (Table 6): the buffer is intended to govern recall timing for schedulable devices, whereas the fleet-level result is driven by baseline-record coverage.

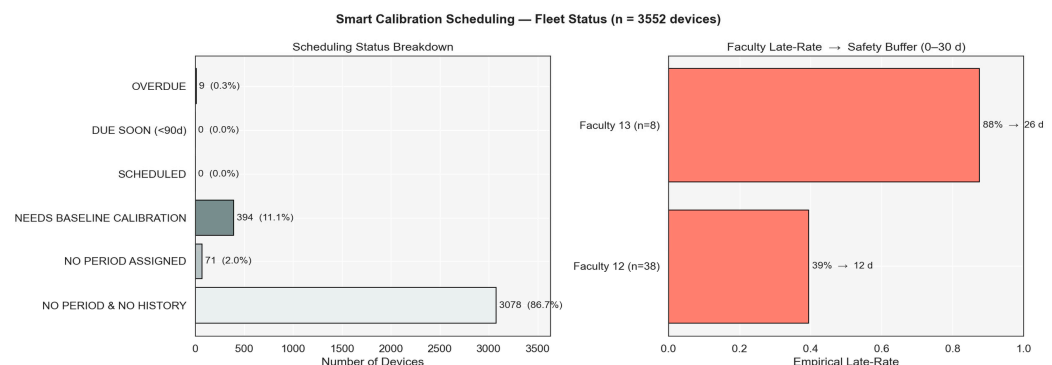


Figure 4. Smart calibration scheduling—fleet-wide status (left) and faculty empirical late-rate mapped to the linear safety buffer (right).

Table 6. The sensitivity of the fleet-status decomposition to the safety-buffer cap. The six-class counts are identical for 15-, 30- and 60-day caps: within this snapshot no cap can move a device across a class boundary, because every schedulable device is already overdue. The buffer would affect recall timing only once devices approach their nominal date; on the present data it therefore changes no scheduling decision.

Scheduling Status	15-Day Cap	30-Day Cap	60-Day Cap
OVERDUE	9	9	9
DUE SOON (<90 d)	0	0	0
SCHEDULED	0	0	0
NEEDS BASELINE CALIBRATION	394	394	394
NO PERIOD ASSIGNED	71	71	71
NO PERIOD & NO HISTORY	3078	3078	3078

5. Discussion

5.1. What the Data Support

The strongest and most transferable result of this work is a data-governance finding: at the present data maturity most instruments cannot be scheduled algorithmically because the baseline records required to do so are absent. The interpretation below separates that governance finding from the narrower scheduling-method contribution, and reads the supervised-learning results strictly as exploratory.

Three findings are supported by the evidence at a level appropriate for publication. First, the institution-level recalibration late-rate is 47.8% with a Wilson 95% CI of [34.1%, 61.9%], indicating that roughly half of all paired recalibration events occur after the previous validity has lapsed. Evidence-based determination of calibration intervals [64] and uncertainty-based scheduling of recalibration dates [65] are increasingly advocated in recent metrology research as alternatives to fixed calendar-based practice. Second, this aggregate figure conceals inter-faculty heterogeneity: at the smallest sample sizes in the present data, two faculty confidence intervals do not overlap and the contrast is significant by Fisher's exact test ($p = 0.020$), pointing to compliance that is not uniform across the institution, reported as exploratory at these sample sizes. Third, the fleet's data-governance state—9.1% calibration coverage and 2.2% populated LastCalibrationDate—is the dominant constraint on any data-driven scheduling architecture. This is consistent with empirical findings that data quality governs the value of data-driven maintenance decisions across the asset life cycle [66]. Without addressing this gap, even an

ideal model would be confined to a small fraction of the fleet. The binding constraint is data completeness, not model sophistication.

These findings are consistent with patterns reported in the broader Slovak smart-manufacturing literature: faculty- and unit-level heterogeneity in calibration and condition-monitoring practice has been documented in adjacent academic and industrial settings [50,51,58]. The present results extend that picture by quantifying the heterogeneity with non-overlapping bootstrap confidence intervals on a single, end-to-end institutional dataset.

5.2. What the Data Do Not Support

Two claims that a less cautious analysis would have made cannot be supported. The cross-validated ROC-AUC for the best classifier is 0.655 ± 0.18 , a modest but uncertain improvement over the stratified baseline (0.40 ± 0.18). At $n = 46$ this is consistent with both “weak signal” and “no signal” and must therefore be reported as exploratory. More starkly, no machine-learning regressor improves on the median baseline for DeviationDays. In a body of research where small-sample studies routinely report inflated metrics, the explicit reporting of this negative result is itself a contribution.

5.3. Why a Rule-Based Scheduler Is the Right Deployment

Given the constraints established above, the practical engineering deliverable of this work is the rule-based scheduler of Section 3.6, not a black-box ML model. The rule has four properties that make it appropriate for the present setting: it is transparent (auditable in a single line of arithmetic), bounded (safety buffer capped at thirty days), empirically grounded (the buffer is scaled by the institution’s own historical late-rate, not by an external prior), and honest about missing data (devices without sufficient history are explicitly flagged rather than implicitly imputed). These properties mirror the design principles increasingly emphasized in the explainable-AI literature and align with the human-centric framing of Industry 5.0. A comparable risk-based approach to modeling metrological support for an instrument fleet has recently been proposed [67].

5.4. Implications for Smart-Manufacturing Architecture

The finding that 86.7% of the fleet cannot currently be scheduled algorithmically because of missing baseline data has direct architectural implications. A smart-manufacturing instrument-management platform should:

- Treat the LastCalibrationDate and CalibrationPeriodMonths fields as first-class data-quality indicators and surface their completion rate in operational dashboards, not as optional metadata.
- Provide a six-class scheduling status (visualized in Figure 4) rather than a binary “due/not due” indicator, so that incomplete records are visible and actionable rather than hidden.
- Retain historical compliance information per faculty/device type to support the empirical safety buffer described here and revise the buffer periodically as the historical sample grows.
- Defer the introduction of supervised models until the historical sample is sufficiently large for a defensible split (a working threshold of $n \geq 200$ paired events per stratum, applied where feasible, is a pragmatic target broadly consistent with small-sample guidance).

These architectural recommendations are consistent with the trajectory of recent Slovak contributions on BIM-enabled facility and equipment management [59,60] and on the sustainable use of measurement and calibration records in smart-manufacturing systems [51], all of which identify the data-governance layer—completeness, traceability and historical retention of asset records—as the binding constraint on data-driven decision-making, in-

dependent of the specific machine-learning method eventually applied. Recent evidence corroborates this: data-governance and data-quality rules materially improve analytics in smart-factory settings [68], and automated machine-readable calibration workflows such as digital calibration certificates are emerging as a practical substrate for such governance [69].

5.5. Limitations

Several limitations of this research must be stated explicitly. (i) The data originate from a single Slovak technical university. Cross-institutional generalization is not warranted without comparable extracts from other sites. (ii) The compliance sub-sample is small ($n = 46$ paired events). This is addressed through bootstrap, Wilson and Clopper–Pearson confidence intervals and through the exclusive use of cross-validation in Section 4.4, but no inferential technique can wholly compensate for the underlying scarcity. (iii) The CalibrationResult field is severely imbalanced (370 PASS/1 FAIL), so failure prediction is structurally infeasible at present and is not attempted. (iv) The rule-based scheduler treats all faculties whose historical sample is below the $n \geq 3$ stratification threshold with the institution-wide buffer. A finer-grained buffer will become defensible as the dataset grows. (v) No real-world A/B test of the proposed scheduler against the current calendar-based practice has yet been performed. Such an evaluation is the natural follow-up. (vi) The study objects are de-identified, so their physical device categories and measurement contexts are not disclosed. The analysis therefore reports statistical association rather than established cause, and the inter-faculty differences are not given a causal interpretation. (vii) Because usage intensity and environmental exposure are not recorded, the reasons behind the differentiated compliance behavior across faculties and device types cannot be resolved and are left to future, context-aware data collection. Concretely, future research should (a) replicate the analysis on comparable extracts from other institutions to test external validity, (b) instrument the calibration workflow to capture usage intensity and environmental exposure as predictive features, (c) re-estimate the supervised models once several hundred paired events have accrued (a working target of roughly 200 per stratum where feasible), and (d) run a controlled comparison of the rule-based scheduler against current calendar-based practice on operational outcomes such as overdue rate and recall cost. In summary, the three headline limitations of this research are: a single-institution dataset; the absence of external validation on independent sites; and a deliberately heuristic, rule-based scheduler whose adoption of learned models is intentionally deferred until the historical sample is large enough to support them under cross-validation.

6. Conclusions

This research presented an empirical, statistically grounded analysis of a real multi-faculty calibration database operated by a Slovak technical university and proposed a transparent, data-governance-informed rule-based scheduler. Taken as a whole, the analysis shows that the binding obstacle to data-driven scheduling at this institution is record completeness rather than algorithm choice. Stepwise, the evidence establishes three points: first, the historical recalibration late-rate is 47.8% (Wilson 95% CI 34.1–61.9%); second, compliance varies across faculties (Fisher’s exact $p = 0.020$) but the stratified estimates remain exploratory at the available sample sizes; and third, 86.7% of the fleet cannot be scheduled algorithmically today because baseline records are missing, while supervised modeling yields only a weak, uncertain classification signal and no useful deviation regression. The practical implication for instrument managers is that a transparent rule-based scheduler, coupled with deliberate completion of last-calibration and calibration-period fields, is the defensible interim deliverable, and that these conditions transfer to comparable research and manufacturing infrastructures. Directions for extending the work—external replication

across institutions, usage- and condition-aware features, and a controlled evaluation against calendar-based practice that includes qualitative assessment beyond purely quantitative outcomes—are set out among the limitations in Section 5.5.

Author Contributions: Conceptualization, all authors; methodology, A.B.; software and formal analysis, M.P.; data curation, M.B. and M.P.; writing—original draft preparation, M.K. and G.G.; writing—review and editing, A.B.; visualization, G.G.; supervision, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This paper presents partial research results of the project supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under contract no. KEGA 075TUKÉ-4/2024 “Customization of Higher Education through the Implementation of Industry 4.0 Tools—Visualization of Mining Processes for Practical Education of the Study Program Earth Resources Management” and KEGA 054TUKÉ-4/2024 “Circular Construction Academy, Low Carbon and Green Solutions: Educational Platform on Sustainable Construction Transition”. This paper presents partial research results of the project supported by the Slovak Research and Development Agency under contract no. APVV-22-0576 “Research of digital technologies and building information modeling tools for designing and evaluating the sustainability parameters of building structures in the context of decarbonization and circular construction” and the EU Recovery and Resilience Plan for Slovakia (Project No. 09I04-03-V02-00051).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this research are available on request from the corresponding author.

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

References

1. Kang, H.S.; Lee, J.Y.; Choi, S.; Kim, H.; Park, J.H.; Son, J.Y.; Kim, B.H.; Noh, S.D. Smart Manufacturing: Past Research, Present Findings, and Future Directions. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2016**, *3*, 111–128. [\[CrossRef\]](#)
2. Lasi, H.; Fettke, P.; Kemper, H.-G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [\[CrossRef\]](#)
3. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, Conception and Perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [\[CrossRef\]](#)
4. Leng, J.; Sha, W.; Wang, B.; Zheng, P.; Zhuang, C.; Liu, Q.; Wuest, T.; Mourtzis, D.; Wang, L. Industry 5.0: Prospect and Retrospect. *J. Manuf. Syst.* **2022**, *65*, 279–295. [\[CrossRef\]](#)
5. Tao, F.; Cheng, J.; Qi, Q.; Zhang, M.; Zhang, H.; Sui, F. Digital Twin-Driven Product Design, Manufacturing and Service with Big Data. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 3563–3576. [\[CrossRef\]](#)
6. Lu, Y.; Liu, C.; Wang, K.I.-K.; Huang, H.; Xu, X. Digital Twin-Driven Smart Manufacturing: Connotation, Reference Model, Applications and Research Issues. *Robot. Comput.-Integr. Manuf.* **2020**, *61*, 101837. [\[CrossRef\]](#)
7. Wang, J.; Ma, Y.; Zhang, L.; Gao, R.X.; Wu, D. Deep Learning for Smart Manufacturing: Methods and Applications. *J. Manuf. Syst.* **2018**, *48*, 144–156. [\[CrossRef\]](#)
8. Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital Twin in Manufacturing: A Categorical Literature Review and Classification. In *IFAC-PapersOnLine*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 51, pp. 1016–1022. [\[CrossRef\]](#)
9. Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of Digital Twin about Concepts, Technologies, and Industrial Applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [\[CrossRef\]](#)
10. JCGM 200:2012; International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM). Joint Committee for Guides in Metrology: Sèvres, France, 2012.
11. ISO/IEC 17025:2017; General Requirements for the Competence of Testing and Calibration Laboratories. International Organization for Standardization: Geneva, Switzerland, 2017.
12. ILAC-G24:2022; Guidelines for the Determination of Calibration Intervals of Measuring Instruments. International Laboratory Accreditation Cooperation: Silverwater, Australia, 2022.
13. Taylor, J.K.; Opperman, H.V. *Handbook for the Quality Assurance of Metrological Measurements*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.

14. Jardine, A.K.S.; Lin, D.; Banjevic, D. A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [\[CrossRef\]](#)
15. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [\[CrossRef\]](#)
16. Li, Z.; Wang, Y.; Wang, K.S. Intelligent Predictive Maintenance for Fault Diagnosis and Prognosis in Machine Centers: Industry 4.0 Scenario. *Adv. Manuf.* **2017**, *5*, 377–387. [\[CrossRef\]](#)
17. Zonta, T.; da Costa, C.A.; da Rosa Righi, R.; de Lima, M.J.; da Trindade, E.S.; Li, G.P. Predictive Maintenance in the Industry 4.0: A Systematic Literature Review. *Comput. Ind. Eng.* **2020**, *150*, 106889. [\[CrossRef\]](#)
18. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.d.P.; Basto, J.P.; Alcalá, S.G.S. A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [\[CrossRef\]](#)
19. Çınar, Z.M.; Abdussalam Nuhu, A.; Zeeshan, Q.; Korhan, O.; Asmael, M.; Safaei, B. Machine Learning in Predictive Maintenance Towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability* **2020**, *12*, 8211. [\[CrossRef\]](#)
20. Achouch, M.; Dimitrova, M.; Ziane, K.; Sattarpanah Karganroudi, S.; Dhoub, R.; Ibrahim, H.; Adda, M. On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. *Appl. Sci.* **2022**, *12*, 8081. [\[CrossRef\]](#)
21. Nunes, P.; Santos, J.; Rocha, E. Challenges in Predictive Maintenance—A Review. *CIRP J. Manuf. Sci. Technol.* **2023**, *40*, 53–67. [\[CrossRef\]](#)
22. Varoquaux, G. Cross-Validation Failure: Small Sample Sizes Lead to Large Error Bars. *NeuroImage* **2018**, *180*, 68–77. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine Learning Algorithm Validation with a Limited Sample Size. *PLoS ONE* **2019**, *14*, e0224365. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [\[CrossRef\]](#)
25. Wilson, E.B. Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **1927**, *22*, 209–212. [\[CrossRef\]](#)
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
27. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
28. Altman, N.; Krzywinski, M. Points of Significance: Ensemble Methods: Bagging and Random Forests. *Nat. Methods* **2017**, *14*, 933–934. [\[CrossRef\]](#)
29. Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. *Commun. Comput. Inf. Sci.* **2020**, *1323*, 417–431. [\[CrossRef\]](#)
30. Garcia, I.; Ferreira, J.; Oliveira, M.; Macedo, E.; Navarro, M.; Peixoto, J. Innovative Approaches in Digital Metrology: Advancing Calibration and Management through Industry 4.0 Technologies. *Acta IMEKO* **2026**, *15*, 1–5. [\[CrossRef\]](#)
31. Lee, J.; Bagheri, B.; Kao, H.-A. A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems. *Manuf. Lett.* **2015**, *3*, 18–23. [\[CrossRef\]](#)
32. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [\[CrossRef\]](#)
33. Kusiak, A. Smart Manufacturing. *Int. J. Prod. Res.* **2018**, *56*, 508–517. [\[CrossRef\]](#)
34. Nahavandi, S. Industry 5.0—A Human-Centric Solution. *Sustainability* **2019**, *11*, 4371. [\[CrossRef\]](#)
35. Maddikunta, P.K.R.; Pham, Q.-V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A Survey on Enabling Technologies and Potential Applications. *J. Ind. Inf. Integr.* **2022**, *26*, 100257. [\[CrossRef\]](#)
36. Moyne, J.; Iskandar, J. Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing. *Processes* **2017**, *5*, 39. [\[CrossRef\]](#)
37. Lu, Y. Industry 4.0: A Survey on Technologies, Applications and Open Research Issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10. [\[CrossRef\]](#)
38. Ghahramani, M.; Qiao, Y.; Zhou, M.; O'Hagan, A.; Sweeney, J. AI-Based Modeling and Data-Driven Evaluation for Smart Manufacturing Processes. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 1026–1037. [\[CrossRef\]](#)
39. Piccarozzi, M.; Aquilani, B.; Gatti, C. Industry 4.0 in Management Studies: A Systematic Literature Review. *Sustainability* **2018**, *10*, 3821. [\[CrossRef\]](#)
40. Antosz, K.; Kulisz, M.; Husár, J. Evaluation and Comparison of Selected Machine Learning Methods for Improving Maintenance Processes. In *IFAC-PapersOnLine*; Elsevier: Amsterdam, The Netherlands, 2024; Volume 58, pp. 85–90. [\[CrossRef\]](#)
41. Tao, F.; Qi, Q.; Liu, A.; Kusiak, A. Data-Driven Smart Manufacturing. *J. Manuf. Syst.* **2018**, *48*, 157–169. [\[CrossRef\]](#)
42. Mourtzis, D. Simulation in the Design and Operation of Manufacturing Systems: State of the Art and New Trends. *Int. J. Prod. Res.* **2020**, *58*, 1927–1949. [\[CrossRef\]](#)
43. Sezer, E.; Romero, D.; Guede, F.; Macchi, M.; Emmanouilidis, C. An Industry 4.0—Enabled Low-Cost Predictive Maintenance Approach for SMEs. In *Proceedings of the IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8. [\[CrossRef\]](#)
44. Husár, J.; Knapčíková, L.; Balog, M. Implementation of Material Flow Simulation as a Learning Tool. In *Lecture Notes in Mechanical Engineering*; Springer: Cham, Switzerland, 2019; pp. 33–41. [\[CrossRef\]](#)

45. Knapčíková, L.; Peraković, D. Innovative Approaches in Technology Challenges in the Context of Industry 4.0. *Wirel. Netw.* **2022**, *28*, 427–429. [\[CrossRef\]](#)
46. Peraković, D.; Periša, M.; Cvitić, I.; Zorić, P. Identification of the Relevant Parameters for Modeling the Ecosystem Elements in Industry 4.0. In *EAI/Springer Innovations in Communication and Computing*; Springer: Cham, Switzerland, 2020; pp. 111–123. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Levin, S. Calibration and Verification of Measuring Instruments: Conceptual Transformation. *Meas. Tech.* **2022**, *64*, 871–882. [\[CrossRef\]](#)
48. Williams, A.; Magnusson, B. *Eurachem/CITAC Guide: Use of Uncertainty Information in Compliance Assessment*, 2nd ed.; Eurachem: Bucharest, Romania, 2021. Available online: <https://www.eurachem.org> (accessed on 12 June 2026).
49. Mobley, R.K. An Introduction to Predictive Maintenance. In *Plant Engineering Series*; Butterworth-Heinemann: Oxford, UK, 2002. [\[CrossRef\]](#)
50. Trebuňa, F.; Hagara, M. Experimental Modal Analysis Performed by High-Speed Digital Image Correlation System. *Measurement* **2014**, *50*, 78–85. [\[CrossRef\]](#)
51. Knapčíková, L.; Bobko, D.; Karakai, M. Material Circularity of Selected Components in the Context of Modern Enterprise. In *Advances in Manufacturing V; Lecture Notes in Mechanical Engineering*; Springer: Cham, Switzerland, 2026; pp. 204–215. [\[CrossRef\]](#)
52. Ran, Y.; Zhou, X.; Lin, P.; Wen, Y.; Deng, R. A Survey of Predictive Maintenance: Systems, Purposes and Approaches. *arXiv* **2019**. [\[CrossRef\]](#)
53. Korba, P.; Šváb, P.; Vereš, M.; Lukáč, J. Optimizing Aviation Maintenance through Algorithmic Approach of Real-Life Data. *Appl. Sci.* **2023**, *13*, 3824. [\[CrossRef\]](#)
54. Ouadah, A.; Zemmouchi-Ghomari, L.; Salhi, N. Selecting an Appropriate Supervised Machine Learning Algorithm for Predictive Maintenance. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 4277–4301. [\[CrossRef\]](#)
55. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [\[CrossRef\]](#)
56. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Vithi, N.; Chibaya, C. Advancements in Predictive Maintenance: A Bibliometric Review of Diagnostic Models Using Machine Learning Techniques. *Analytics* **2024**, *3*, 493–507. [\[CrossRef\]](#)
58. Sloboda, A.; Piľa, J.; Sloboda, O.; Hovanec, M.; Rácek, B.; Korba, P.; Hajduová, Z. Tribotechnical Diagnostics of Aircraft Emergency Hydrogenerator NP-27 T. *Appl. Sci.* **2026**, *16*, 4395. [\[CrossRef\]](#)
59. Mandičák, T.; Mésároš, P.; Spišáková, M.; Kanáliková, A. Exploitation of Document Management Systems in Construction Industry. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2022; Volume 1252, p. 012077. [\[CrossRef\]](#)
60. Kozlovská, M.; Petkanič, S.; Vranay, F.; Vranay, D. Enhancing Energy Efficiency and Building Performance through BEMS–BIM Integration. *Energies* **2023**, *16*, 6327. [\[CrossRef\]](#)
61. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*; SciPy: Austin, TX, USA, 2010; pp. 56–61. [\[CrossRef\]](#)
62. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [\[CrossRef\]](#)
64. Kopke, V.; Mourao, S.; Brito, M. Determination of the Calibration Interval of Measuring Instruments: Which Method Should I Use? *IEEE Instrum. Meas. Mag.* **2024**, *27*, 59–62. [\[CrossRef\]](#)
65. Schaudé, J.; Hausotte, T. Uncertainty-Based Determination of Recalibration Dates. *Int. J. Metrol. Qual. Eng.* **2024**, *15*, 2. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Hinrichs, M.; Prifti, L.; Schneegass, S. Data-Driven Decision-Making in Maintenance Management and Coordination throughout the Asset Life Cycle: An Empirical Study. *J. Qual. Maint. Eng.* **2023**, *30*, 202–220. [\[CrossRef\]](#)
67. Khramenkov, V.; Khayrullin, R. Risk-Based Approach in the Problems of Modeling Metrological Support for a Measuring Instrument Fleet. *Meas. Tech.* **2024**, *66*, 819–826. [\[CrossRef\]](#)
68. Tufano, A. Data Governance in Smart Factories: Consistency Rules for Improved Data Quality in Logistics & Operations. *Manuf. Lett.* **2023**, *37*, 57–60. [\[CrossRef\]](#)
69. Bernien, M.; Straka, M.; Bock, T.; Niepraschk, R. Digital Calibration Certificates and Automated Calibration Workflow in the Vacuum Metrology Group of PTB. *Meas. Sens.* **2025**, *38*, 101490. [\[CrossRef\]](#)

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.