electronics

MDPI

Article

A Serial-Number-Level Cumulative-Risk Framework for Yield Monitoring and Inspection Prioritization in Semiconductor Manufacturing

Seong Min Ryu [1], Young Shin Han [2,*], Jong Sik Lee [1] and Bo Seung Kwon [1]

[1] Department of Computer Engineering, Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Republic of Korea; smryu@inha.edu (S.M.R.); jslee@inha.ac.kr (J.S.L.); bskwon@inha.edu (B.S.K.)
[2] Frontier College, Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Republic of Korea
* Correspondence: hanys@inha.ac.kr

Abstract

In semiconductor fabrication, abnormal behavior may first appear in a small subset of serial numbers before it is reflected in lot-level yield metrics. We present a monitoring framework that detects measurement item outliers using Z-scores, aggregates them into a serial-level cumulative-risk score, provides exploratory views of lot- and site-level trends, and ranks high-risk serials for follow-up. The approach is evaluated on an industrial semiconductor manufacturing dataset comprising 14,142 unique serials (N_{serial} = 14,142). Because most TestResult labels are PASS, label-based yield shows little variation. In this setting, label-based yield alone is not informative enough for early monitoring, so we use outlier-based yield as the primary metric, defined as the proportion of serials with cumulative risk below the threshold ($R(s) < \tau$, where $\tau = 10$). A sensitivity study of the outlier threshold κ (σ-multiplier) shows that yield varies widely, from 61.66% at $\kappa = 3$ to above 99% at $\kappa \geq 7$. This result shows the trade-off between detection sensitivity and inspection workload. Case studies of top-ranked serials show two recurring patterns: cumulative risk is driven either by isolated extreme deviations or by the accumulation of moderate deviations across multiple items. These results support the use of the proposed score for inspection prioritization.

Keywords: semiconductor manufacturing; yield monitoring; early-warning monitoring; serial-level analysis; outlier detection; Z-score; cumulative-risk score; anomaly prioritization

Check for updates

Academic Editors: Yanpeng Shi, Yafei Ning and Jie Su

Received: 11 March 2026
Revised: 26 March 2026
Accepted: 27 March 2026
Published: 29 March 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) license.

1. Introduction

In semiconductor manufacturing, yield is a core indicator that directly affects production costs and is routinely used to assess process stability and product quality. As device scaling continues and process flows become more complex, production lines generate large volumes of serial-level test records, measurement item values, test-time data and contextual information such as lot, site, and tester IDs. In many cases, these data are archived for traceability or summary reporting, but are not fully used for early screening of localized abnormal behavior or for tracing where that behavior first emerges. In practice, yield loss is often recognized only after substantial scrap or rework has already occurred. This suggests the need for earlier monitoring in routine operations [1–4].

Conventional yield management typically relies on lot-level or wafer-level aggregated yields. While such aggregation is useful for tracking overall trends, it can mask localized abnormal behavior that repeatedly occurs in a small number of serial numbers [5]. In

production environments, anomalies associated with a specific tool, condition, or time window often concentrate in a limited subset of units. As these units accumulate, they may eventually appear as a measurable drop in aggregate yield. Early detection therefore requires quantifying abnormal behavior at a finer granularity and linking it to an operational yield indicator, rather than relying on lot averages alone.

Recent studies have applied machine-learning methods to yield prediction and anomaly detection in semiconductor manufacturing [1,6–8]. These approaches have shown potential, but in production environments they may still require substantial modeling effort, tuning, and maintenance. At the same time, recent work in explainable artificial intelligence has begun to improve the interpretability of such models in a manufacturing and semiconductor setting, for example, through SHAP-based analysis of process-quality drivers and yield-related factors [6,9]. However, the present study focuses on a simpler screening framework intended for direct tracing of abnormal serials in a PASS-dominated dataset [3,4]. The choice here was therefore not meant to dismiss learning-based methods, but to prioritize a transparent and easily traceable rule under the conditions of severe label imbalance. Two-step learning schemes have been proposed for manufacturing applications [10,11], and deep-learning-based virtual metrology and yield prediction have also been explored [12]. Domestic studies have examined wafer-yield prediction using virtual metrology variables [13] and adaptive online-update strategies for delayed measurements [14]. Recent optimization-based approaches have also been explored for manufacturing decision problems; however, the present study focuses on serial-level monitoring and inspection prioritization rather than process optimization itself.

We therefore propose a framework that does not depend on a complex predictive model, because in this setting the main need is not only detection but also straightforward tracing of abnormal serials. The method aggregates Z-score-based outlier magnitudes into a serial-level cumulative-risk score and links that score to lot/site aggregates.

The main contributions of this paper are threefold. First, it introduces a serial-number-level cumulative-risk framework for early-warning monitoring under severs PASS-dominant label imbalance. Second, it provides an operational way to review abnormal behavior across serial-, lot- and site-level views without relying on a complex predictive model. Third, it shows through κ-sensitivity analysis and case studies that the framework can support inspection prioritization in production data.

Section 2 defines the Z-score-based outlier detection method and the formulation of serial-level cumulative risk. Section 3 describes the monitoring framework and dashboard workflow. Section 4 presents case-study results on industrial manufacturing data, including sensitivity analyses and high-risk serial examples. Section 5 concludes the paper.

2. Definitions and Formulation

2.1. Statistical Normalization and Z-Score Outlier Detection

For a set of measurements, the mean and standard deviation are defined as

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{2}$$

Each measurement is normalized to a Z-score.

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{3}$$

Let κ denote the outlier threshold. For each measurement item, Z-scores are computed using the item-wise mean μ_j and standard deviation σ_j estimated from all non-missing records in the analysis window. These item-wise statistics are treated as fixed reference values within the selected analysis window. Accordingly, the present framework does not explicitly model within-window process drift or other forms of nonstationary behavior. Missing values are not imputed and are excluded from the calculation of μ_j, σ_j, and the corresponding Z-score. Item-wise normalization was necessary because the measurement items were recorded on different scales, so direct comparison was not meaningful without a common deviation rule. Such preprocessing considerations are also recognized as important in semiconductor manufacturing data analysis [15]. A measurement $x_{i,j}$ is flagged as an outlier when $|Z_{i,j}| \geq \kappa$.

$$|Z_i| \geq \kappa \tag{4}$$

Z-score–based outlier detection is widely used for statistical monitoring of manufacturing data and is commonly discussed as a baseline approach in semiconductor anomaly-detection studies [1,2,16]. It has also been reported that sensor noise and measurement uncertainty can affect monitoring performance [17]. Z-score was used here because the aim was to screen many measurement items with one consistent rule, not to tune a separate anomaly model for each case.

2.2. Serial-Number–Level Cumulative Risk

Let s denote the set of serial numbers. For a given serial s, let O_s be the set of Z-scores that are flagged as outliers:

$$O_s = \{Z_{s,j} \mid |Z_{s,j}| \geq \kappa\} \tag{5}$$

where s denotes a serial, j denotes a measurement item, and O_s is the set of outlier Z-score observed for serial s.

To quantify how abnormal behavior accumulates at the serial level, we define the cumulative risk $R(s)$ as

$$R(s) = \sum_{Z \in O_s} |Z| \tag{6}$$

Equation (6) reflects both the number and magnitude of outlier events. For each serial, $R(s)$ is computed by summing outlier magnitudes across measurement items. This definition was intended to retain both patterns seen in practice: repeated moderate deviations and isolated extreme excursions within the same serial.

A linear sum was used as the baseline aggregation rule because it preserves both frequency and magnitude information while keeping the contribution of each outlier item directly traceable. The present study does not claim that this rule is superior to all alternative weighting schemes; rather, it is adopted here as the most transparent starting point for serial-level screening.

For dashboard-level prioritization, the cumulative-risk can also be converted into a normalized ranking score on a 0–1 scale. Let R_{min} and R_{max} denote the minimum and maximum cumulative risk values among the serials under review. The normalized score is defined as

$$\text{Score}(s) = \frac{R(s) - R_{min}}{R_{max} - R_{min}} \tag{7}$$

where values closer to 1 indicate higher inspection priority. When $R_{max} = R_{min}$, the score is set to 0 for all serials.

2.3. Risk-Based Yield Monitoring Metric

Given the serial-level cumulative risk $R(s)$, a threshold τ is used to classify serials as Normal or Risk. A serial is classified as Risk when $R(s) \geq \tau$.

$$s \in \begin{Bmatrix} S_{\{normal\}}, R(s) < \tau \\ S_{\{risk\}}, R(s) \geq \tau \end{Bmatrix} \quad (8)$$

where S_{noraml} and S_{risk} denote sets of normal and risk serials, respectively, and τ is the risk threshold.

Let N_{serial} be the total number of serials and N_{normal} the number classified as normal. Yield (%) is defined as follows.

$$\text{Yield}(\%) = \frac{N_{normal}}{N} \times 100 \quad (9)$$

When N_{serial} is the total number of serials and N_{normal} is the number of serials classified as normal.

Equation (9) represents the fraction of normal serials and relates serial-level deviation to yield degradation [18].

In this work, yield is reported in two ways. Label-based yield is defined from the final pass/fail label as Yield $= 1 - N_{fail}/N_{serial}$, where N_{fail} is the number of unique serials labeled FAIL. Outlier-based yield is defined as the proportion of serials classified as normal under the risk threshold τ. The two metrics play different roles. FAIL labels mark final outcomes, whereas outlier-based yield is used to screen cumulative abnormal deviation before a final failure becomes visible.

3. Proposed Framework

3.1. Framework Overview

Figure 1 shows the workflow from Z-score-based outlier detection to serial-level risk aggregation and yield monitoring.

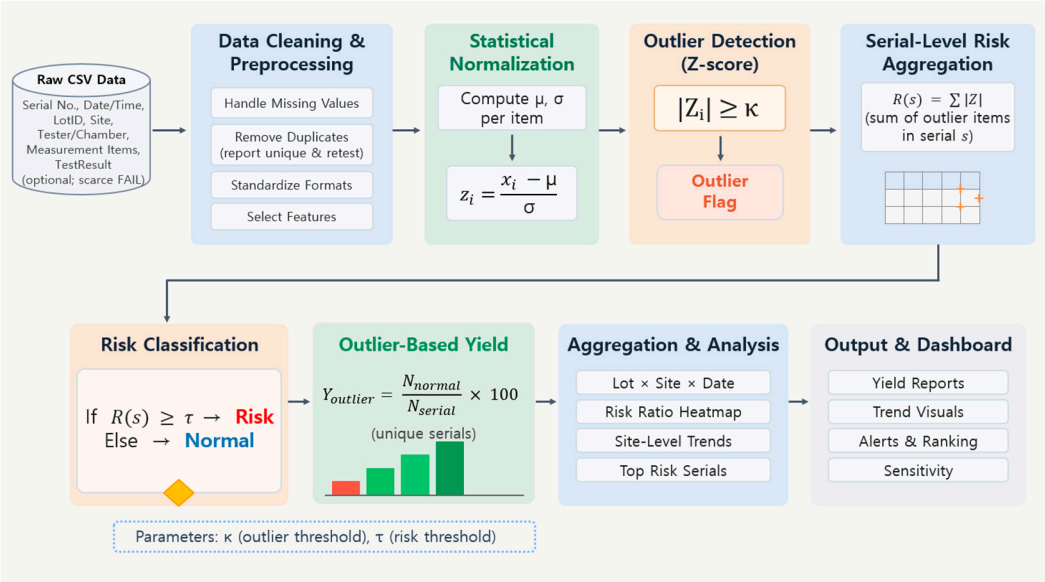


Figure 1. Overall workflow of the proposed serial-level yield-monitoring framework.

The raw dataset includes serial number, date/time, LotID, site, tester/chamber, measurement items, and outlier flags. Records are aggregated by LotID × site × date to compare lot/site trends over time. Table 1 summarizes the dataset.

Table 1. Summary of the dataset used in this study.

Item	Description
Raw data fields	Serial number, date/time, LotID, site, tester/chamber, measurement items, and outlier flags
Aggregation level	LotID × site × date
Number of unique serials	n = 14,142
Number of site × date groups	119
Yield views reported	Label-based yield; outlier-based yield
Primary monitoring indicator	Outlier-based yield in Equation (9)
Label characteristic	The final label (TestResult) is overwhelmingly PASS; FAIL cases are scarce, so label-based yield shows little variation

The main monitoring indicator is the serial-level cumulative risk $R(s)$, computed by summing outlier magnitudes across measurement items for each serial. Both label-based and outlier-based yield are reported, with the latter used as the primary metric. The basic unit of analysis is a measurement record defined by serial number, date/time, lot, site, and measurement items. Depending on the analysis, both a non-deduplicated view and a unique serial view are reported.

As shown in Table 2, the framework consists of five steps.

Table 2. Proposed monitoring framework.

Step	Description	Output
1	Load process measurement data (CSV)	Raw dataset
2	Compute Z-scores and detect outliers by $\lvert Z_i \rvert \geq \kappa$	Outlier flags
3	Aggregate serial-level cumulative risk $R(s)$	Risk scores
4	Classify serials into Normal/Risk by applying threshold τ to cumulative risk $R(s)$, according to Equation (8)	Normal/Risk sets
5	Monitor yield and highlight high-risk serials	Yield and top risk list

3.2. Outlier Detection and Distribution-Based Monitoring

For each measurement item, the Z-score outlier results are visualized together with the distribution histogram, the mean μ, the standard deviation σ, and the threshold band $\pm\kappa\sigma$, so that deviations from the normal range can be identified at a glance (Figure 2). In Figure 2, the blue bars indicate histogram bins associated with the highest outlier region, highlighting where extreme observations are concentrated. Because κ controls sensitivity, both the number of flagged serials and outlier-based yield vary substantially with κ [16,19].

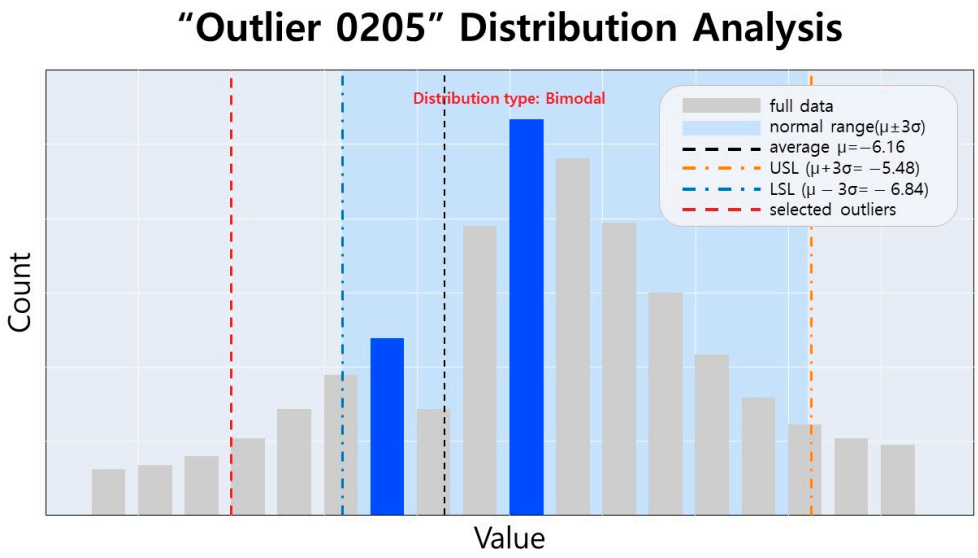


Figure 2. Distribution-based visualization of Z-score outlier threshold ($\mu \pm \kappa\sigma$) for a representative measurement item.

3.3. Tracking High-Risk Serials and Dashboard Linkage

As shown in Figure 3, serials are ranked in descending order of cumulative risk R(s) after Z-score-based outlier detection and threshold-based screening. When a high-risk serial is selected, the linked table and plots show the corresponding outlier intervals and context. This allows direct review of the serials contributing most strongly to abnormal behavior [2,3].

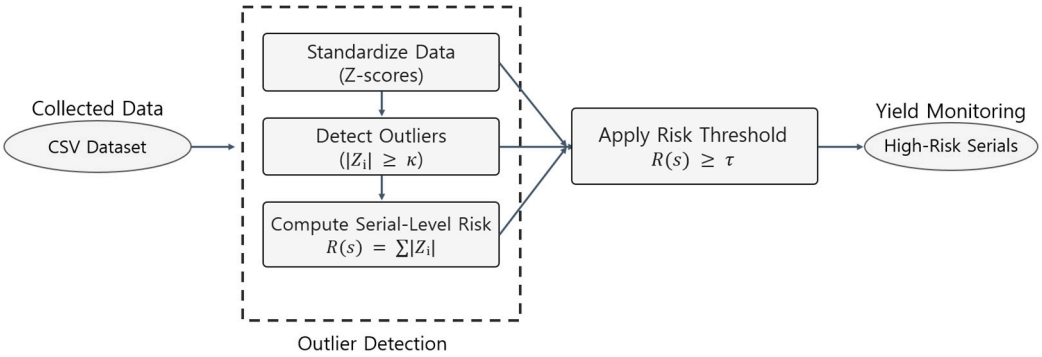


Figure 3. Simplified flow of outlier detection, cumulative-risk computation, and risk-serial identification.

Figure 4 illustrates how ranked risk serials are reviewed in a linked dashboard environment. In this study, risk serials are treated as early-warning candidates rather than confirmed FAIL cases. The risk threshold τ is set according to monitoring goals and dataset characteristics. In addition to cumulative risk R(s), Table 3 reports a normalized ranking score defined in Section 2.2. This score is a relative prioritization index on a 0–1 scale and does not represent the probability of final failure.

Table 3. Example of top-ranked risk serials in CSV1 ($\kappa = 3$, $\tau = 10$).

Serial Number	Outliers	Mean(|Z|)	Sum(|Z|)=R(s)	Class	Max|Z|	Score(0–1)
S1	1	40.651	40.651	Risk	40.65	1.000
S2	15	5.001	75.014	Risk	23.15	1.000
S3	7	7.912	55.383	Risk	24.12	1.000

Table 3. *Cont.*

Serial Number	Outliers	Mean(|Z|)	Sum(|Z|)=R(s)	Class	Max|Z|	Score(0–1)
S4	3	10.647	31.941	Risk	24.59	1.000
S5	1	26.071	26.071	Risk	26.07	1.000
S6	2	12.002	24.004	Risk	20.86	0.998
S7	1	21.307	21.307	Risk	21.31	0.998
S8	2	12.243	24.486	Risk	19.91	0.998
S9	2	11.685	23.370	Risk	19.73	0.998
S10	3	9.625	28.876	Risk	18.85	0.997

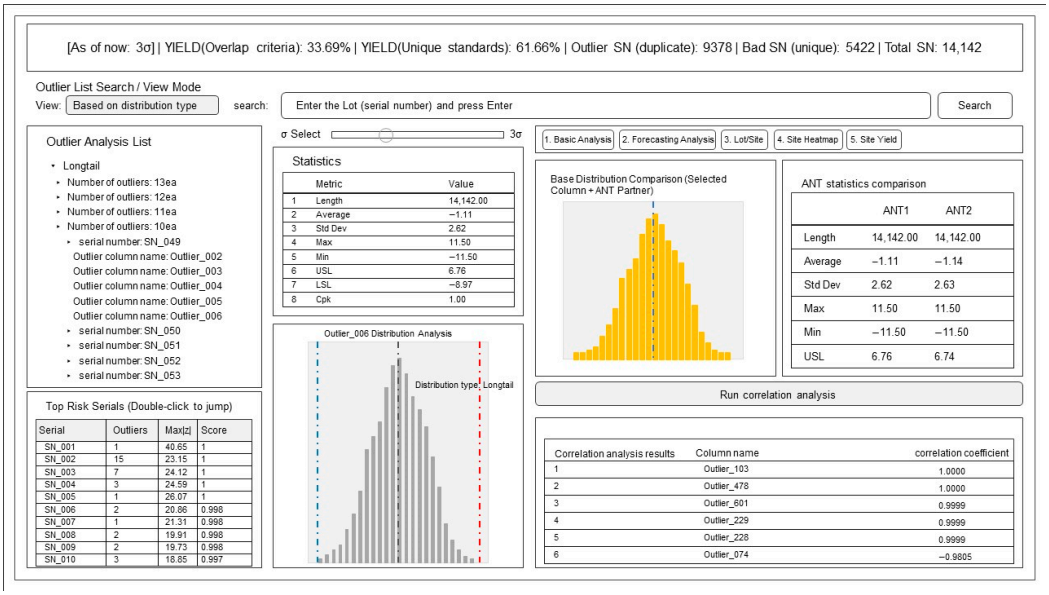


Figure 4. Overall dashboard structure for serial-level yield monitoring.

3.4. Linking Serial-Level Signals to Lot/Site Monitoring

Aggregating serial-level measurements at the lot and site levels allows temporal trends and localized shifts to be examined. Figure 5 plots daily mean testTime(s) by lot or site together with the global mean and the normal band. Intervals outside the band are treated as candidates for further investigation. Among the available variables, testTime was used as a monitoring signal because shifts in test duration can indicate changes in operating conditions at the lot/site level.

The lot/site view is used to identify time windows that deviate from the baseline, after which the ranked risk serials within the same LotID × site × date windows are reviewed. In this way, serial-level risk is examined in the context of aggregated lot/site behavior rather than as an isolated signal. The global mean and standard deviation are computed from all available aggregated daily values in the analysis window.

A site-level outlier-based yield heatmap further highlights spatio-temporal patterns, including date–site combinations where yield decreases. In Figure 6, low-$Y_{outlier}$ regions appear locally, indicating periods in which cumulative outlier burden is concentrated. These regions provide the starting points for follow-up review of the associated serials. Risk serials are used for prioritization rather than final-failure labeling.

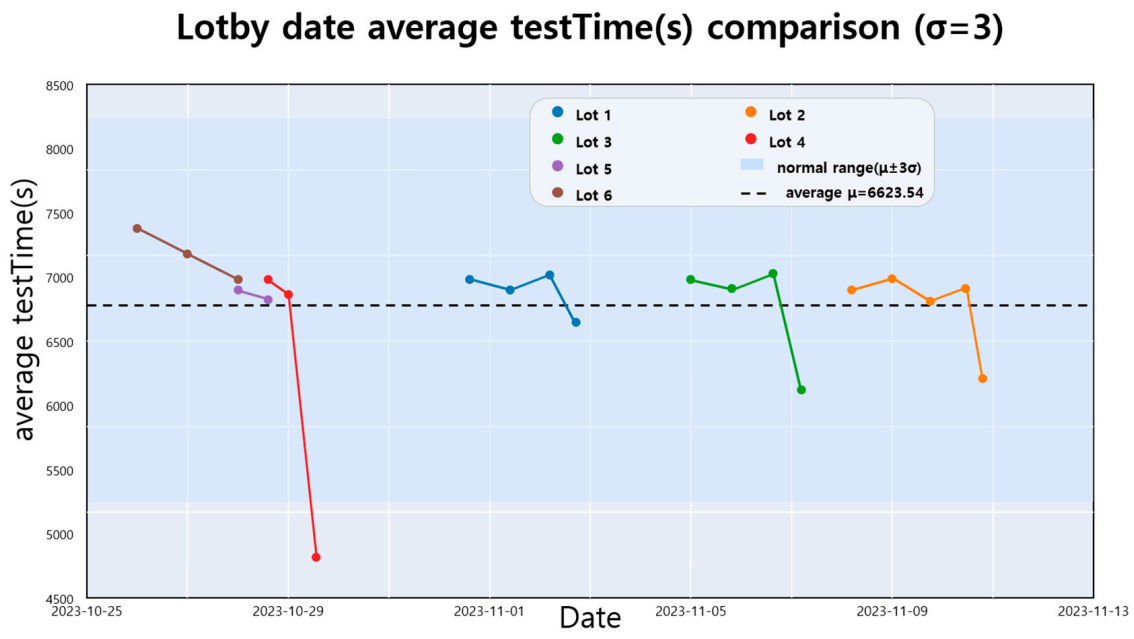


Figure 5. Daily aggregated testTime(*s*) by lot (or site) with global μ and $\mu \pm 3\sigma$ band.

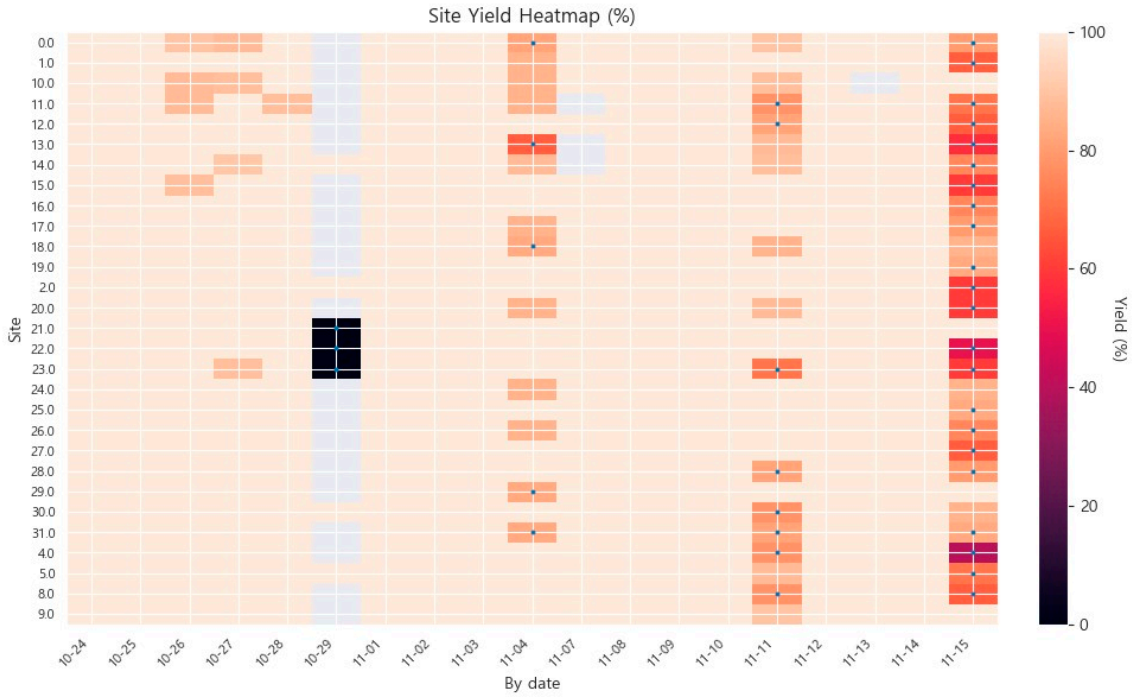


Figure 6. Site-level outlier-based yield heatmap ($\kappa = 3$, $\tau = 10$) for identifying spatio-temporal degradation patterns.

4. Results and Discussion

4.1. Dataset and Experimental Setup

The proposed framework was evaluated using real manufacturing data from a semiconductor process stored in CSV format. The primary dataset used for evaluation, referred to as CSV1, contains the serial-level measurements and contextual fields listed in Table 1. Additional datasets are denoted as CSV2, CSV3, and so on. The dataset includes serial numbers, test time (testTime), process/test measurements, and contextual fields such as lot and site. In this setting, testTime is monitored because abnormal increases or local shifts in test duration can reflect changes in process or test conditions, even when final FAIL labels

remain scarce. Prior work has noted that virtual metrology can be strengthened in CMP by combining model-based and data-driven approaches [20]. The outlier threshold κ and the risk threshold τ were selected for monitoring, and sensitivity to κ was examined. Because anomaly detection can raise false alarms when process anomalies and sensor failures are not distinguished, approaches to separate these effects have also been reported [4,18].

The dataset contains $N_{serial} = 14,142$ unique serials. Because the TestResult labels are overwhelmingly PASS, label-based yield shows limited variation. We therefore report both label-based yield and outlier-based yield, where the latter is defined as the fraction of serials classified as normal under the cumulative-risk rule.

We compare the risk-serial ratio, the mean cumulative risk aggregated by lot/site, and yield under different settings of κ and τ. Lower κ increases outlier sensitivity and raises the risk-serial ratio, whereas higher κ concentrates alarms on a smaller subset of serials. Changing τ shifts the decision boundary for risk classification and therefore changes the yield metric even under the same outlier flags. Quantitative results on κ-sensitivity and case studies of top risk serials are presented in Sections 4.2 and 4.3. The framewok is evaluated in terms of three aspects: sensitivity of outlier-based yield to κ, concentration of risk serials across site x date windows, and case-based inspection relevance of the highest-ranked serials.

4.2. Sensitivity Analysis of κ and τ

This section examines the sensitivity of the proposed framework to two key operational parameters: the outlier threshold κ and the risk threshold τ. We report (i) the distribution of the risk-serial ratio (RiskRatio) across site $\times$ date groups under the default setting, (ii) the sensitivity of outlier-based yield to κ, and (iii) the sensitivity of risk-serial classification to τ. Together, these results show how alarm volume changes as early-warning sensitivity increases.

4.2.1. Sensitivity to κ (σ-multiplier): Outlier-Based Yield and Risk-Serial Ratio

We evaluate the sensitivity of outlier-based yield by varying the outlier threshold κ. A serial is classified as Risk if its cumulative risk satisfies $R(s) \geq \tau$ under the outlier threshold κ, and outlier-based yield is computed as $Y_{outlier} = N_{normal}/N_{serial}$. The risk threshold is fixed at $\tau = 10$. Here, "outlier-based yield" refers to the proportion of serials with $R(s) < \tau$. RiskRatio denotes the fraction of risk serials within each site $\times$ date group. Across 119 site $\times$ date groups, the median RiskRatio is 0.10, with an average of 0.1329, as summarized in Table 4. In some windows, RiskRatio reaches 1.0, indicating strong concentration of risk serials in specific date–site combinations.

Table 4. Descriptive statistics of risk-serial ratio (RiskRatio) for $\kappa = 3$, $\tau = 10$ (site $\times$ date groups).

κ = 3	τ = 10
Count (=number of Site $\times$ Date groups)	119.000000
Mean	0.132921
Std	0.148754
Min	0.000000
25%	0.037241
50%	0.100000
75%	0.165476
max	1.000000

4.2.2. Key Findings

Table 5 summarizes the outlier-based yield for unique serials ($N_{serial} = 14,142$) across different values of κ. At $\kappa = 3$, the yield is 61.66%. As κ increases, the yield rises to 86.67% at $\kappa = 4$, 95.28% at $\kappa = 5$, and 98.26% at $\kappa = 6$, and exceeds 99% for $\kappa \geq 7$. In contrast, $\kappa = 2$ yields only 13.63%, indicating that the detection setting is too sensitive for practical use.

Table 5. Sensitivity of unique-serial outlier-based yield to $\kappa(\sigma)$ ($\tau = 10$, $N_{serial} = 14,142$).

$\kappa(\sigma)$	$Yield_{outlier}$ (%)	N_{risk} (Unique Serial)	N_{normal}	N_{total} (=14,142)
2	13.63	12,214	1928	14,142
3	61.66	5422	8720	14,142
4	86.67	1885	12,257	14,142
5	95.28	667	13,475	14,142
6	98.26	246	13,896	14,142
7	99.24	108	14,034	14,142
8	99.41	83	14,059	14,142
9	99.53	66	14,076	14,142

4.2.3. Interpretation and Discussion

The results indicate a clear trade-off. κ was selected by examining how many serials were actually flagged at each setting, rather than by imposing a single statistically optimal cutoff. Smaller κ produces more risk serials, whereas larger κ concentrates alarms on fewer serials. In this dataset, κ between 3 and 5 appears to be the most workable range, because it preserves sensitivity while avoiding an excessive number of flagged serials [1,3,4].

4.2.4. Why Outlier-Based Yield Is Emphasized

Because PASS labels dominate the dataset, label-based yield provides limited information for early-warning monitoring. We therefore use the outlier-based metric as the primary indicator and adopt $\kappa = 3$ and $\tau = 10$ as the default setting.

Table 6 compares the cumulative-risk-based view with simpler monitoring views derived from the same data.

Table 6. Practical comparison of monitoring views in the studied data.

Monitoring View	Primary Signal	Reflects Severity	Supports Serial Ranking	Suitable for Early-Warning Screening Under Scarce FAIL Labels
Label-based yield	Final pass/fail label (TestResult)	No	No	Limited
Outlier-based yield	Fraction of serials classified as normal under the cumulative-risk rule	Indirectly	No	Yes
Cumulative-risk-based serial monitoring	Serial-level cumulative risk $R(s)$ aggregated from outlier magnitudes	Yes	Yes	Yes

4.2.5. Sensitivity to τ Under Fixed κ (Risk-Classification Robustness)

Table 5 shows the sensitivity to κ with τ fixed at 10, whereas Table 7 shows the sensitivity to τ with κ fixed at 3. Neither κ nor τ were chosen through formal statistical optimization in the present study. Instead, both were treated as operational screening parameters and examined through sensitivity analysis. Practically, κ determines which deviations are counted as outliers and therefore affects $R(s)$, whereas τ sets the decision boundary for classifying a serial as Risk.

Table 7. Sensitivity of risk-serial classification to τ under $\kappa = 3$.

τ	Number of Risk Serials	Number of Normal Serials	Risk-Serial Ratio
5	2133	12,009	0.150827
7	1497	12,645	0.105855
10	822	13,320	0.058125
12	566	13,576	0.040023
15	395	13,747	0.027931

As shown in Table 7, increasing τ reduces the number of risk serials and decreases the risk-serial ratio, reflecting a stricter alarm policy. This τ-sensitivity analysis complements Table 5 by separating sensitivity driven by outlier detection (κ) from sensitivity driven by threshold-based screening (τ), and thus helps determine alarm volume under the default κ setting.

4.3. Case Analysis and Prioritization of Top Risk Serials

After examining parameter sensitivity in Section 4.2, we analyze the serial-level cumulative risk $R(s)$ as an operational prioritization index. This section presents case analyses of top-ranked risk serials to show how $R(s)$ captures both the frequency and severity of outlier deviations and how these profiles support inspection prioritization and root-cause reasoning.

4.3.1. Prioritizing Risk Serials Using Cumulative Risk

We next examine the highest-ranked risk serials. Table 3 summarizes, for $\tau = 10$, the number of outliers, the mean outlier magnitude, the cumulative risk, and the Normal/Risk classification. The ranking reflects both the frequency and severity of abnormal deviations [2].

4.3.2. Spike-Type Anomalies (Single, High-Severity Deviations)

As shown in Table 3, some serials have only one or two outliers but still rank highly because the deviation itself is large; Rank 1, for example, has a single outlier with $Z = 40.651$. This shows that $R(s)$ can be dominated by a single extreme deviation.

4.3.3. Accumulation-Type Anomalies (Multi-Item, Moderate Deviations)

Conversely, serials with many outliers can accumulate large risk even when the item-wise Z-scores are not extreme. In Table 3, S2 has 15 outliers with $\text{mean}(|Z|) = 5.001$, yet its cumulative risk reaches 75.014. This type of profile differs from spike-type anomalies and is more consistent with a distributed shift across multiple items than with a single isolated excursion.

4.3.4. How to Interpret and Use 'Risk Serials' Operationally

In this paper, a risk serial is treated as an early-warning flag rather than a confirmed failure case. High-ranked serials are therefore reviewed first together with associated lot, site, and tester context. In practice, the ranking may also be used to organize an engineering inspection queue for follow-up review.

4.3.5. Scope Under Limited Labels

Because FAIL labels are rare in this dataset, the evaluation focuses less on supervised prediction accuracy and more on whether the framework identifies serials worth reviewing.

4.4. Discussion, Limitations, and Practical Implications

Because final FAIL labels are scarce in CSV1, label-based yield provides limited variability, which constrains supervised validation against final outcomes. Under this class imbalance, standard predictive validation metrics such as precision or false positive rate are not sufficiently informative for the present dataset. A related limitation is that the Z-score reference statistics are estimated over the analysis window, so the framework assumes that the selected window is sufficiently stable for pooled standardization. In addition, because the framework relies on fixed statistical thresholds, its operating characteristics may change across datasets with different distributions or process regimes. The discussion therefore focuses on monitoring utility, limitations, and future validation needs.

4.4.1. Limited Variability of Label-Based Yield

In CSV1, FAIL cases are scarce, and label-based yield therefore shows limited variation. Accordingly, rather than focusing on final-failure prediction, we evaluate the framework in terms of early-warning screening, the sensitivity of outlier-based yield, and inspection prioritization [3,4]. The present dataset also does not provide a consistent surrogate label that can be used as a reliable proxy for non-pass outcomes. For this reason, the framework is assessed here as an early-screening and prioritization tool rather than as a supervised predictor of final failure. For the same reason, the present study does not attempt to report precision, false positive rate, or related supervised validation metrics as primary evidence of effectiveness.

4.4.2. Operational Meaning of κ and the Trade-Off

Section 4.2 shows that the σ-multiplier κ strongly affects both the number of risk serials and outlier-based yield.

In deployment, κ is the parameter that most directly controls how many serials are pulled into review. Reporting κ-sensitivity alongside the default setting provides a clearer basis for selecting κ under site-specific monitoring policies. This dependence on fixed thresholds should therefore be regarded as a limitation when the framework is transferred to new datasets or operating regimes.

4.4.3. Measurement Anomalies vs. Process Anomalies

Risk serials may reflect both process shifts and measurement issues, including sensor or tester conditions. Risk flags should therefore be interpreted together with contextual factors such as equipment condition, lot/site, and retest codes. Where available, methods that distinguish process anomalies from sensor failures can be integrated to reduce false alarms [16,18].

Another limitation is that the current aggregation assigns equal weight to all measurement items. In practice, some items may be more closely related to yield loss than others, and item-specific weighting is left for future work.

In some cases, high-risk serials may reflect measurement instability or temporary tested as review candidates rather than confirmed defect cases, because the present framework is intended for early screening. This is one reason why the ranking should be interpreted together with lot, site, tester, and retest context.

The computation is relatively light, since once item-wise normalization is done, the remaining steps are thresholding and serial-level aggregation. This makes the framework easy to extend to larger datasets, as long as the normalization statistics are refreshed for the analysis window being used.

4.4.4. Future Work: Lead-Time Validation and Monitoring Extension

Future work will focus on lead-time validation and the evaluation of early-warning performance using longer-term datasets with sufficient FAIL events. This would allow quantitative assessment of how far in advance risk serials appear before yield loss, together with evaluation using metrics such as Precision/Recall and AUPRC. Practical operating guidelines will also be refined by jointly considering κ, τ, and retest context to balance alarm volume and early-detection capability.

Beyond the scope of this paper, the framework may be extended by incorporating virtual metrology, uncertainty-aware process modeling, and other process-specific monitoring approaches studied in semiconductor manufacturing [7,20–28]. A further direction is to reduce reliance on fixed statistical thresholds such as κ and τ. In particular, threshold-free and label-free pipelines based on unsupervised clustering or similarity-based grouping may offer a more adaptive way to identify abnormal process states across changing operation conditions. Future work may also consider wafer/bin-map defect pattern analysis to capture spatial defect signatures and incorporate rule-based or fuzzy-inference components as post hoc explanation layers to improve interpretability [29–40].

5. Conclusions

We proposed a serial-level monitoring framework for early-warning yield monitoring that aggregates outlier information into a cumulative-risk score $R(s)$. By combining Z-score-based outlier detection with serial-level risk aggregation, the framework prioritizes high-risk serials and links them to lot/site trends.

Because label-based yield showed limited variation in this dataset, we treated outlier-based yield as the more informative monitoring view and examined its sensitivity to the outlier threshold κ. Outlier-based yield changed substantially with κ, from 61.66% at $\kappa = 3$ to above 99% at $\kappa \geq 7$, highlighting the trade-off between early-warning sensitivity and false-alarm burden. Case analyses showed that $R(s)$ can be dominated either by spike-type anomalies or by accumulation-type anomalies. This supports the score as a basis for inspection prioritization.

A key limitation of this study is the scarcity of final FAIL labels, which limits validation against final-failure outcomes. Even so, the proposed framework remains useful for early-warning monitoring, serial prioritization, and parameter selection through κ-sensitivity analysis. Future work will extend the validation to datasets with richer failure events and refine operational guidelines that incorporate τ selection and retest context.

Author Contributions: Conceptualization, Y.S.H.; methodology, J.S.L.; writing—original draft preparation, S.M.R.; writing—review and editing, S.M.R. and B.S.K.; supervision, J.S.L.; funding acquisition, Y.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP)—Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2026-RS-2023-00259678).

Data Availability Statement: This work was supported by data provided by STATS ChipPAC Korea (SCK). We are grateful to SCK for their valuable contribution to this research.

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

References

1. Lee, Y.; Roh, Y. An Expandable Yield Prediction Framework Using Explainable Artificial Intelligence for Semiconductor Manufacturing. *Appl. Sci.* **2023**, *13*, 2660. [CrossRef]
2. Susto, G.A.; Terzi, M.; Beghi, A. Anomaly Detection Approaches for Semiconductor Manufacturing. *Procedia Manuf.* **2017**, *11*, 2018–2024. [CrossRef]
3. Maitra, V.; Su, Y.; Shi, J. Virtual Metrology in Semiconductor Manufacturing: Current Status and Future Prospects. *Expert Syst. Appl.* **2024**, *249*, 123559. [CrossRef]
4. Wang, B.; Mao, Z. Outlier Detection Based on Gaussian Process with Application to Industrial Processes. *Appl. Soft Comput.* **2019**, *76*, 505–516. [CrossRef]
5. Chen, C.-Y.; Chang, S.-C.; Liao, D.-Y. Equipment Anomaly Detection for Semiconductor Manufacturing by Exploiting Unsupervised Learning from Sensory Data. *Sensors* **2020**, *20*, 5650. [CrossRef]
6. Fan, S.K.S.; Hsu, C.-Y.; Tsai, D.-M.; He, F.; Cheng, C.-C. Data-Driven Approach for Fault Detection and Diagnostic in Semiconductor Manufacturing. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 1925–1936. [CrossRef]
7. Lee, C.-Y.; Chang, K.; Ho, C. Autoencoder-Based Detector for Distinguishing Process Anomaly and Sensor Failure. *Int. J. Prod. Res.* **2024**, *62*, 7130–7145. [CrossRef]
8. Busch, R.; Czerner, P.; Wahl, M.; Choubey, B. Yield Prediction in Semiconductor Manufacturing Using Two-Step Machine Learning. *Int. J. Prod. Res.* **2025**, 1–18. [CrossRef]
9. Nam, W.S.; Kim, S.B. A Prediction of Wafer Yield Using Product Fabrication Virtual Metrology Process Parameters in Semiconductor Manufacturing. *J. Korean Inst. Ind. Eng.* **2015**, *41*, 572–578. [CrossRef]
10. Lee, J.H.; Ha, S.H. Recognizing yield patterns through hybrid applications of machine learning techniques. *Inf. Sci.* **2009**, *179*, 844–850. [CrossRef]
11. Salem, M.; Taheri, S.; Yuan, J.-S. An Experimental Evaluation of Fault Diagnosis from Imbalanced and Incomplete Data for Smart Semiconductor Manufacturing. *Big Data Cogn. Comput.* **2018**, *2*, 30. [CrossRef]
12. Lee, H.; Kim, Y.; Kim, C.O. A deep learning model for robust wafer fault monitoring with sensor measurement noise. *IEEE Trans. Semicond. Manuf.* **2017**, *30*, 23–31. [CrossRef]
13. Fan, S.-K.S.; Hsu, C.-Y.; Jen, C.-H.; Chen, K.-L.; Juan, L.-T. Defective wafer detection using a denoising autoencoder for semiconductor manufacturing processes. *Adv. Eng. Inform.* **2020**, *46*, 101166. [CrossRef]
14. Cai, H.; Feng, J.; Yang, Q.; Li, W.; Li, X.; Lee, J. A virtual metrology method with prediction uncertainty based on Gaussian process for chemical mechanical planarization. *Comput. Ind.* **2020**, *119*, 103228. [CrossRef]
15. Cai, H.; Feng, J.; Zhu, F.; Yang, Q.; Li, X.; Lee, J. Adaptive virtual metrology method based on Just-in-time reference and particle filter for semiconductor manufacturing. *Measurement* **2021**, *168*, 108338. [CrossRef]
16. Jebri, M.A.; El Adel, E.; Graton, G.; Ouladsine, M.; Pinaton, J. The Impact of the Virtual Metrology on a Run-to-Run control for a Chemical Mechanical Planarization process. *IFAC-Pap.* **2017**, *50*, 6154–6159. [CrossRef]
17. Jeong, M.K.; Choi, J.; Son, Y.; Kang, J. Deep Learning Based Virtual Metrology and Yield Prediction in Semiconductor Manufacturing Processes. In Proceedings of the PHM Society Asia-Pacific Conference 2017, Jeju, Republic of Korea, 12 July 2017. [CrossRef]
18. Di, Y.; Jia, X.; Lee, J. Enhanced Virtual Metrology on Chemical Mechanical Planarization Process using an Integrated Model and Data-Driven Approach. *Int. J. Progn. Health Manag.* **2020**, *8*. [CrossRef]
19. Sarpietro, R.E.; Pino, C.; Coffa, S.; Messina, A.; Palazzo, S.; Battiato, S.; Spampinato, C.; Rundo, F. Explainable Deep Learning System for Advanced Silicon and Silicon Carbide Electrical Wafer Defect Map Assessment. *IEEE Access* **2022**, *10*, 99102–99128. [CrossRef]
20. Zheng, H.; Sherazi, S.W.A.; Son, S.H.; Lee, J.Y. A Deep Convolutional Neural Network-Based Multi-Class Image Classification for Automatic Wafer Map Failure Recognition in Semiconductor Manufacturing. *Appl. Sci.* **2021**, *11*, 9769. [CrossRef]
21. Tsai, T.-H.; Wang, C.-Y. Wafer map defect classification using deep learning framework with data augmentation on imbalance datasets. *EURASIP J. Image Video Process.* **2025**, *2025*, 6. [CrossRef]
22. Shin, E.; Yoo, C.D. Efficient Convolutional Neural Networks for Semiconductor Wafer Bin Map Classification. *Sensors* **2023**, *23*, 1926. [CrossRef]
23. Lee, J.H.; Moon, I.C.; Oh, R. Similarity Search on Wafer Bin Map Through Nonparametric and Hierarchical Clustering. *IEEE Trans. Semicond. Manuf.* **2021**, *34*, 464–474. [CrossRef]

24. Kim, D.; Kang, P. Dynamic Clustering for Wafer Map Patterns Using Self-Supervised Learning on Convolutional Autoencoders. *IEEE Trans. Semicond. Manuf.* **2021**, *34*, 444–454. [CrossRef]
25. Jin, C.; Na, H.; Piao, M.; Pok, G.; Ryu, K. A Novel DBSCAN-Based Defect Pattern Detection and Classification Framework for Wafer Bin Map. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 286–292. [CrossRef]
26. Kim, T.S.; Lee, J.W.; Lee, W.K.; Sohn, S.Y. Novel method for detection of mixed-type defect patterns in wafer maps based on a single shot detector algorithm. *J. Intell. Manuf.* **2022**, *33*, 1715–1724. [CrossRef]
27. Yoon, S.; Kang, S.H. Semi-automatic wafer map pattern classification with convolutional neural networks. *Comput. Ind. Eng.* **2022**, *166*, 107977. [CrossRef]
28. Yu, J.; Zheng, X.; Liu, J. Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map. *Comput. Ind.* **2019**, *109*, 121–133. [CrossRef]
29. Park, H.-J.; Koo, Y.-S.; Yang, H.-Y.; Han, Y.-S.; Nam, C.-S. Study on Data Preprocessing for Machine Learning Based on Semiconductor Manufacturing Processes. *Sensors* **2024**, *24*, 5461. [CrossRef]
30. Park, S.Y.; Kim, T.S. Fuzzy Inference System for Interpretable Classification of Wafer Map Defect Patterns. *Electronics* **2026**, *15*, 130. [CrossRef]
31. Han, X.; Miller, M.; Moyne, J.; Vogl, G.; Penkova, A.; Jia, X. A Comparative Study of Semiconductor Virtual Metrology Methods and Novel Algorithmic Framework for Dynamic Sampling. *IEEE Trans. Semicond. Manuf.* **2025**, *38*, 232–239. [CrossRef]
32. Sim, H.-E.; Lee, M.-U.; Hong, S. Virtual Metrology of Multiple Dielectric Layer Thickness for 3D-NAND Deposition Process. *IEEE Trans. Semicond. Manuf.* **2025**, *38*, 240–250. [CrossRef]
33. Joo, J.; Yang, K.W.; Choi, Y.J.; Min, B.; Kim, C.O. Virtual Metrology Modeling for Wafer Edges via Graph Attention Networks. *IEEE Trans. Semicond. Manuf.* **2023**, *36*, 359–366. [CrossRef]
34. Breidung, M.; Rothe, T.; Lauff, A.; Thieme, P.; Langer, J.; Günther, M.; Kuhn, H. Process data-driven machine learning for non-uniformity prediction and virtual metrology in chemical mechanical planarization. *J. Intell. Manuf.* **2025**, 1–18. [CrossRef]
35. Deivendran, B.; Masampally, V.S.; Nadimpalli, N.R.V.; Runkana, V. Virtual metrology for chemical mechanical planarization of semiconductor wafers. *J. Intell. Manuf.* **2024**, *36*, 1923–1942. [CrossRef]
36. Yeh, K.-T.; Tsai, H.-Y. Mechanically modeling chemical mechanical planarization from wafer to molecular scales: A review and discussion of future direction. *J. Mech.* **2024**, *40*, 769–773. [CrossRef]
37. Xu, Q.; Yu, N.; Essaf, F. Improved Wafer Map Inspection Using Attention Mechanism and Cosine Normalization. *Machines* **2022**, *10*, 146. [CrossRef]
38. Choi, J.E.; Seol, D.H.; Kim, C.Y.; Hong, S.J. Generative Adversarial Network-Based Fault Detection in Semiconductor Equipment with Class-Imbalanced Data. *Sensors* **2023**, *23*, 1889. [CrossRef]
39. Jeong, I.; Lee, S.-Y.; Park, K.; Kim, I.; Huh, H.; Lee, S. Wafer map failure pattern classification using geometric transformation-invariant convolutional neural network. *Sci. Rep.* **2023**, *13*, 8127. [CrossRef]
40. Senoner, J.; Netland, T.; Feuerriegel, S. Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing. *Manag. Sci.* **2022**, *68*, 5704–5723. [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.