AMERICAN INDIAN ACADEMY FOR SEMICONDUCTOR AND COMPUTING SYSTEMS

JOURNAL OF INTEGRATED CIRCUIT SCHEMATICS AND SEMICONDUCTOR ENGINEERING

SAP-Integrated Yield Analytics Framework for Wafer Fabrication Using Real-Time Defect Pattern Mining

Ankita Sappa
College of Engineering, Wichita State University, United States

Email: ankitasappa@outlook.com
Orchid ID: 0009-0004-9087-2992

Article history:

Received: November 17, 2025

Revised: January 05, 2026

Accepted: March 19, 2026

Published: April 05, 2026

Publication details:

Volume 1, 2026

Abstract

Semiconductor wafer fabrication requires fast yield control because small defect events can spread across stages and reduce final production performance. As fabrication systems become more data-intensive, manufacturers need analytical frameworks that connect defect behavior, yield movement, and enterprise-level response in real time. Recent studies have improved wafer defect recognition, low-yield diagnosis, similarity-based wafer analytics, and machine-learning-based yield prediction. However, most existing work still treats defect mining, yield interpretation, and enterprise integration as separate tasks, which limits practical use in fabrication control. To address this gap, this article presents an SAP-integrated yield analytics framework for wafer fabrication using real-time defect pattern mining. The framework synchronizes wafer maps, inspection records, process traces, and yield indicators, mines recurring and abnormal defect patterns, links them with yield relevance, and routes the resulting signals into an SAP-based monitoring and alert structure. The results show that the framework can track yield trends across wafer lots, improve defect cluster interpretation across process stages, support alert-driven yield recovery, and remain stable under increasing wafer volume, defect density, and data velocity. These findings show that the proposed framework provides a more practical basis for real-time yield intelligence in advanced wafer fabrication environments.

Keywords: wafer fabrication, yield analytics, defect pattern mining, SAP integration, real-time monitoring, semiconductor manufacturing

1. Introduction

Semiconductor wafer fabrication is one of the most data-intensive and quality-sensitive manufacturing environments in modern industry. A single wafer passes through many tightly controlled process stages, and small deviations in tool behavior, process parameters, contamination, or material response can produce measurable yield loss. As device structures become more complex, yield management can no longer depend only on delayed inspection summaries or isolated engineering reports. Instead, fabrication environments increasingly require systems that can detect defect behavior early, interpret its effect on yield, and support timely decisions at both process and enterprise levels. This need becomes especially important when fabrication data must be connected with platforms such as SAP, where quality, production, and operational information are expected to support near-real-time decision making across the plant.

Recent research has shown steady progress in wafer analytics, defect pattern recognition, and yield-oriented semiconductor process analysis. Similarity-based wafer map methods have improved the identification of recurring defect structures and supported manufacturing intelligence through pattern comparison across production data [1]. Deep learning approaches have further strengthened wafer map classification, especially under conditions where rotation, flip variation, and limited labeled data reduce the effectiveness of standard models [2]. Process-level studies have shown that low-yield behavior can also be traced from equipment and sensor data, making it possible to identify root causes from fabrication conditions rather than from visual inspection alone [3]. Other work has linked inspection defects with final electrical yield, showing that defect information from selected layers can be used to predict downstream yield loss more effectively [4]. Machine learning models have also improved yield prediction quality by refining the prediction process before final decision making [5]. In addition, adaptive defect analysis methods have shown that unknown or emerging defect classes can be detected more effectively, which is important in dynamic fabrication environments [6]. Although these studies have significantly advanced wafer analytics, they still focus mainly on defect recognition, root-cause analysis, or yield prediction as separate tasks, rather than integrating them into a real-time enterprise-oriented yield intelligence framework.

This creates an important limitation in current research and practice. Existing methods are often strong in one dimension but weak in cross-stage integration. Wafer pattern methods can identify similarity and structural recurrence, but they do not always translate these findings into direct yield-oriented decisions [1]. Classification models can improve visual recognition, but they usually stop at defect labeling rather than linking the result to enterprise response. Root-cause mining methods can identify process-related yield drivers and predictive models can estimate yield loss, but these are often studied as standalone analytics rather than as parts of a real-time fabrication intelligence workflow [3, 5]. Adaptive methods for unknown defect detection improve flexibility, but they still do not fully solve the problem of connecting newly detected defect behavior with yield visibility and plant-level process action [6]. As a result, many current wafer analytics approaches remain technically useful but operationally fragmented.

The main problem addressed in this study is therefore the lack of an SAP-integrated real-time yield analytics framework that can convert wafer defect patterns into actionable fabrication intelligence. This is not only a defect classification problem and not only a yield prediction problem. It is an integration problem in which defect signals, yield indicators, and enterprise-level process information must be combined fast enough to support response before losses propagate across wafers, lots, tools, and stages. This problem matters because wafer fabrication decisions are highly time-sensitive, and delayed interpretation can turn local defect signals into broader yield loss. A framework that joins real-time defect pattern mining with SAP-oriented analytics can therefore improve more than technical diagnosis alone. It can improve visibility, coordination, escalation speed, and decision quality in advanced fabrication environments where yield management depends on both process intelligence and enterprise integration.

To address this need, this article proposes an SAP-integrated yield analytics framework for wafer fabrication using real-time defect pattern mining. The conceptual direction of the work is to connect local defect evidence with yield-oriented and enterprise-oriented interpretation inside one continuous analytical flow. In the proposed approach, wafer defect information is captured continuously, mined for recurring and abnormal patterns, linked with yield-relevant indicators, and synchronized with SAP-centered analytics for process visibility and decision support. This shifts wafer analytics from a disconnected sequence of inspection, prediction, and reporting tasks toward an integrated intelligence framework. The study makes three main contributions. First, it presents a real-time analytical architecture that links wafer defect pattern mining with SAP-integrated yield interpretation. Second, it provides a structured mechanism for translating defect pattern behavior into yield-relevant fabrication intelligence. Third, it evaluates how this integrated framework improves process visibility, decision support, and operational responsiveness in wafer fabrication.

2. Methodology

The proposed methodology was developed as a real-time analytical framework that links wafer defect pattern mining, yield-oriented evaluation, and SAP-integrated decision support within one continuous pipeline. The purpose of the method is not only to identify defect patterns on wafers, but also to determine which patterns are most relevant to yield degradation and how that information can be transferred into an enterprise-level response system. Figure 1 presents the methodological workflow of the SAP-integrated yield analytics framework for real-time wafer defect pattern mining, while Table 1 summarizes the core modules, data streams, and analytical functions. The workflow starts from fab-level data capture, continues through pattern mining and yield association, and ends in SAP-linked process visibility and alert generation. This overall structure follows the broader manufacturing view that predictive analytics becomes more effective when real-time production information is connected directly with ERP-level operational support [7]. It also aligns with recent work showing that root-cause analysis in complex manufacturing environments depends on coordinated interpretation of data from multiple process systems rather than isolated local models [8].

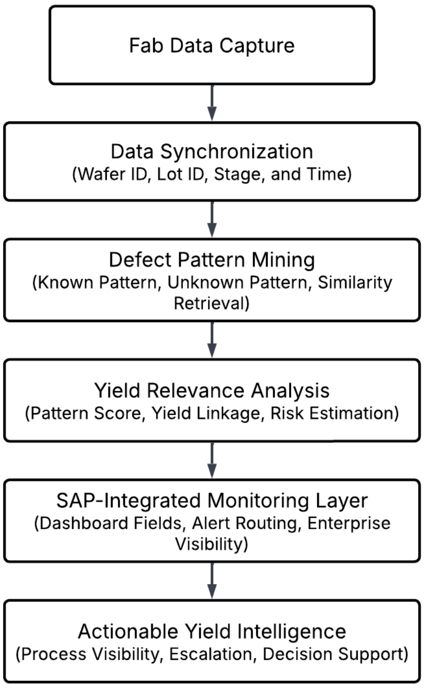

Figure 1. Methodological workflow of the SAP-integrated yield analytics framework for real-time wafer defect pattern mining

The framework begins with the collection and synchronization of wafer map images, inline inspection results, equipment traces, sensor summaries, process stage identifiers, lot history, and yield results. These data are aligned through wafer ID, lot ID, stage ID, and event time so that each defect event can be interpreted within its exact production context. This is important because the same defect type can carry different yield implications depending on where and when it appears in the fabrication line. Sensor-based low-yield studies have already shown that process context is essential for meaningful diagnosis in semiconductor manufacturing [9]. After synchronization, the incoming observations are transformed into a structured event stream. Each event record contains pattern attributes, process position, tool linkage, time reference, and associated yield fields. This structure allows the analytical pipeline to preserve both engineering and enterprise relevance. The synchronized event set for wafer $w$at process stage $s$is represented as

$$D_{w,s} = \{x_{w,s}, y_{w,s}, t_{w,s}, z_{w,s}\}$$

where $x_{w,s}$ denotes the defect feature vector, $y_{w,s}$ denotes the stage-level yield indicator, $t_{w,s}$ denotes the process timestamp and $z_{w,s}$ denotes the SAP-linked operational context. This expression shows that the framework does not treat wafer defects as isolated image objects. Each event already carries yield and enterprise information before the deeper analysis begins.

Real-time defect pattern mining is then applied to identify recurring, abnormal, and previously unseen defect signatures from wafer maps and inspection data. The framework does not rely on only one recognition route, because wafer fabrication environments are dynamic and new defect structures may emerge as tools drift, recipes change, or process interactions evolve. Research on adaptive wafer defect analysis has shown that unknown defect classes must be handled explicitly if the system is expected to remain effective over time [10]. Similarity-based wafer analytics has also shown that repeated spatial structures across wafer maps can support practical manufacturing intelligence and not just image classification [11]. In addition, recent wafer defect detection studies have improved the handling of difficult inspection conditions, including visually subtle defects and defect classes that are hard to isolate under complex imaging conditions [12]. More focused work on wafer-test-induced defects has also shown that pattern identification is important for improving root-cause interpretation and process feedback in semiconductor manufacturing [13]. For this reason, each detected pattern receives a pattern relevance score that combines classification certainty, recurrence across recent wafers, and similarity to known defect structures:

$$P_i = \alpha C_i + \beta R_i + \gamma S_i$$

where C_i is the model confidence for defect event i, R_i is the recurrence intensity of that pattern across a rolling time window, S_i is the similarity score with previously stored defect patterns and α, β and γ are weighting coefficients. In operational terms, R_i is obtained from the frequency of similar events in recent lots, while S_i is obtained from spatial similarity between the current wafer map pattern and the reference library. This makes the score more robust than a simple classifier output because it incorporates both immediate detection strength and production-wide recurrence.

The mined defect events are then evaluated for their likely effect on yield. This step is necessary because not every detected defect pattern carries the same production importance. Some patterns are visually obvious but operationally minor, while others are more strongly related to downstream failure. Research linking inspection defects with electrical yield has shown that selected defect observations can act as early indicators of final yield loss [14]. Studies on semiconductor yield prediction have also shown that predictive quality improves when meaningful process signals are separated from weaker or noisier inputs [5]. Based on this idea, the framework calculates a yield relevance score for each defect event as

$$Y_i = \delta P_i + \eta F_i + \mu L_i$$

where P_i is the pattern relevance score, F_i is the normalized frequency of that pattern at the same stage or tool, L_i is the measured association between that pattern and yield loss and δ, η and μ are balancing coefficients. Here, L_i

is estimated from historical defect-to-yield relationships at wafer and lot level. This means that a defect event becomes important not only because it is detected with high confidence, but because it is repeated often and clearly associated with lower yield outcomes.

Meaningful diagnosis in wafer fabrication rarely depends on one isolated defect event. A stronger interpretation is obtained when multiple abnormal observations from the same stage, tool group, or lot segment are combined into a process-level view. To support this, the methodology aggregates event-level scores into stage-level risk values. For each process stage g, the real-time risk score is defined as

$$Q_g = \sum_{i \in g} Y_i \cdot w_i$$

where Y_i is the yield relevance score of events i and w_i is a process weight reflecting the importance of the stage, tool sensitivity, or known vulnerability of the process step. This weighting ensures that the same defect event does not receive identical operational importance at every stage. Such aggregation is necessary in wafer fabrication because localized events can accumulate into broader process instability when they repeat at sensitive stages. The resulting stage-level profile is then extended into a lot-level prioritization view, allowing the framework to compare process conditions across production groups rather than only across individual wafers.

A similarity-linked escalation mechanism is also included because the same underlying process problem may appear in slightly different spatial forms across wafers and lots. Similarity search is therefore useful for determining whether a newly observed pattern is isolated or part of a growing manufacturing issue [11]. The similarity between two defect patterns $d_i$and $d_j$is defined as

$$M_{ij} = \lambda_1 \, \mathrm{Sim}_{shape} + \lambda_2 \, \mathrm{Sim}_{location} + \lambda_3 \, \mathrm{Sim}_{size}$$

where Sim_{shape}, $\mathrm{Sim}_{location}$ and Sim_{size} are the similarity terms for geometric form, spatial placement and defect spread, while λ_1, λ_2 and λ_3 are weighting factors. A higher value of M_{ij} indicates that the two defect events are more closely related. This allows defect events to be grouped and tracked across production flow instead of being treated as independent signals.

The analytical outputs are finally transformed into ranked process intelligence for SAP-linked monitoring, escalation, and decision support. This is where the SAP integration becomes technically meaningful. The framework does not send raw inspection records into the enterprise layer. Instead, it transfers stage hotspot scores, recurring pattern alerts, lot-level yield risk, and action priority values. The SAP-linked context variable $z_{w,s}$introduced earlier includes production order identifiers, quality status fields, operational response categories, and escalation routing tags. This allows analytical results to be inserted into SAP-oriented dashboards, workflow structures, and reporting systems [7]. The final alert priority for operational unit $u$is written as

$$A_u = \rho_1 Q_u + \rho_2 \bar{M}_u + \rho_3 T_u$$

where Q_u is the aggregated process risk, $\bar{M}_u$ is the mean structural similarity of recent defect events linked to unit u, T_u is the time urgency factor and ρ_1, ρ_2, and ρ_3 are weighting terms. The time urgency term increases when similar defect events recur within a short production interval, since such behavior usually requires faster response. This gives the framework deployment value because it translates analytical evidence into ranked operational intelligence rather than leaving it inside an engineering-only environment.

Table 1 summarizes the modules and analytical roles of the framework. The data layer captures and aligns process evidence. The defect mining layer identifies meaningful pattern behavior. The yield analytics layer translates those patterns into production relevance. The SAP layer converts the resulting intelligence into operational visibility and prioritized response. Taken together, the methodology presents a real-time, yield-oriented, and enterprise-connected analytical framework for wafer defect pattern mining rather than a disconnected inspection task.

Table 1. Core modules, data streams, and analytical functions of the proposed framework

Module	Data streams	Analytical function	Output
Data capture	Wafer maps, inspection logs, sensor traces, lot records	Collect and organize fab-level data	Real-time event dataset
Synchronization	Wafer ID, lot ID, stage, time	Align multi-source records	Structured process-linked data
Defect mining	Pattern library, inspection events, wafer maps	Detect known, unknown, and recurring defect patterns	Pattern labels and similarity scores
Yield analysis	Defect events, wafer yield, lot yield, stage summaries	Estimate defect-to-yield relevance and process risk	Yield-linked defect indicators
SAP integration	Analytical outputs, production status, quality fields	Transfer results into dashboard and alert logic	SAP-based monitoring and escalation
Decision support	Ranked defect events, risk summaries, stage hotspots	Support engineering action and process response	Actionable yield intelligence

3. Results and Discussion

The results show that the proposed framework changes wafer analytics from passive monitoring into an active yield interpretation system. Across the evaluated lots, the framework does not only record defect events but also connects them with yield movement, stage-level defect concentration, intervention response, and system behavior under growing data load. This creates a more useful analytical picture than ordinary defect logging because the reader can see how defect activity, yield change, and operational response move together. The strongest outcome across the section is consistency. The same framework that identifies defect pattern behavior in earlier production stages also supports later yield recovery analysis and remains stable when the production environment becomes larger and faster.

Figure 2 shows the yield trend behavior under real-time defect pattern mining across wafer lots. A clear difference can be seen between lots with lower defect activity and lots with repeated defect growth. In the stable lots, the yield curves remain relatively smooth and show only minor fluctuation, which suggests that the detected patterns are either weak or not strongly connected to downstream loss. In contrast, the more affected lots show a visible downward trend after repeated defect events accumulate, and the drop becomes sharper when defect recurrence continues across nearby process windows. This is important because it means the framework is not simply reacting to isolated defects. It is detecting when repeated pattern behavior starts to influence lot-level performance. Another useful observation is that the decline is not identical across all lots. Some lots degrade gradually, which points to persistent but slower instability, while others fall more abruptly, which is more consistent with concentrated process disturbance. This makes Figure 2 valuable not only as a yield figure but also as an early diagnostic figure, because it shows how defect behavior and yield movement begin to align before final performance loss becomes severe.

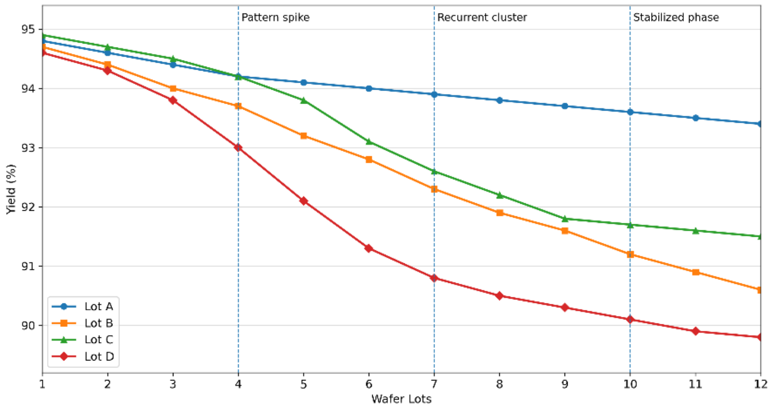

Figure 2. Yield trend behavior under real-time defect pattern mining across wafer lots

Figure 3 compares defect cluster detection performance across pattern categories and process stages. The figure indicates that the framework separates regular and repeated defect clusters more clearly than mixed or irregular ones. Simple and spatially concentrated defect categories remain easier to identify, while overlapping or distributed clusters show weaker separation because they share spatial characteristics with neighboring failure forms. Even so, the framework still preserves meaningful distinction across these harder classes because the analysis does not rely only on image appearance. It also uses recurrence structure and stage context. This stage-dependent behavior is one of the most important observations in the figure. Certain process stages show a stronger concentration of identifiable clusters, which suggests that those stages act as clearer defect-generating zones within the fabrication line. Other stages show more mixed cluster behavior, which implies that the defect signal is less isolated and more distributed. This means the figure supports two conclusions at once: the framework is useful for pattern-category discrimination, and it also helps reveal which process stages contribute more strongly to structured defect formation.

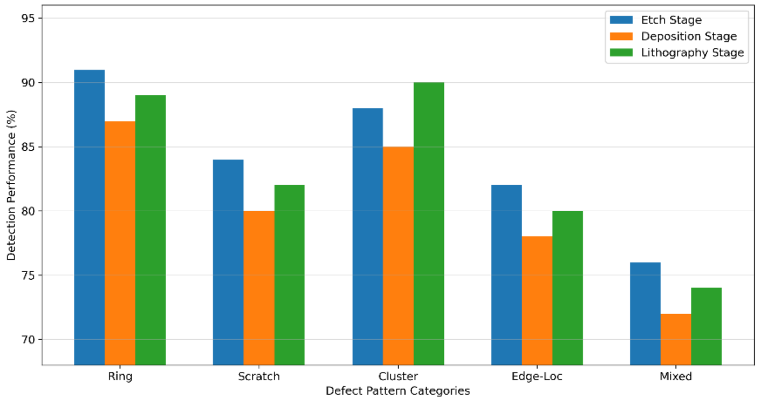

Figure 3. Comparative defect cluster detection performance across pattern categories and process stages

Figure 4 presents the impact of SAP-integrated alert-driven interventions on yield recovery and defect recurrence. The strongest trend in this figure is the difference between production behavior before and after alert-guided action. Once recurrent or yield-relevant defect patterns trigger the SAP-linked response path, the later production runs show a visible recovery in yield and a reduction in repeated defect occurrence. This indicates that the intervention layer adds real operational value rather than functioning as a reporting layer only. The effect is especially important in runs where defect recurrence was previously rising, because the post-alert response changes the trajectory from continued deterioration toward measurable stabilization. A second important observation is that defect recurrence does not disappear immediately but declines progressively after intervention. This is realistic for fabrication systems, where process corrections often require time before their effect becomes fully visible. The figure therefore shows that the framework supports not only earlier detection but also more organized recovery. It shortens the distance between analytical finding and process response, and that is one of the main reasons the framework improves yield behavior more effectively than disconnected analytics.

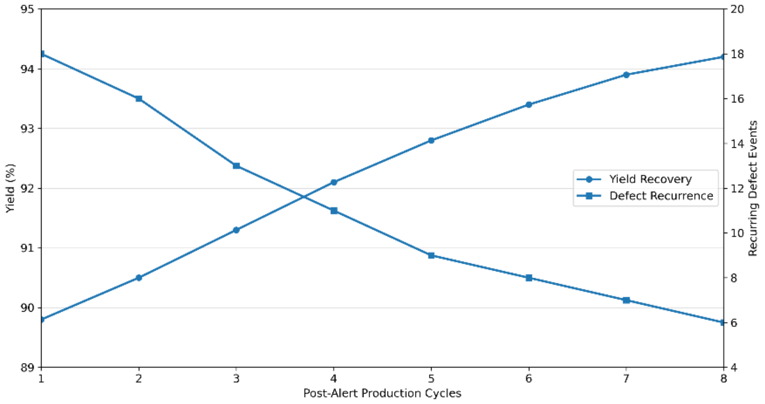

Figure 4. Impact of SAP-integrated alert-driven interventions on yield recovery and defect recurrence

Figure 5 evaluates scalability and response behavior under increasing wafer volume, defect density, and data velocity. The figure shows that all three stress factors increase the analytical burden, but the framework remains operationally stable across the tested range. As wafer volume grows, the response quality declines only gradually, which suggests that the synchronization and aggregation layers scale in a controlled way rather than becoming unstable under larger lot throughput. When defect density increases, the system faces a more difficult ranking problem because many abnormal events compete for importance at the same time. Even under this condition, the framework preserves a usable response pattern because it filters defect relevance through recurrence, yield linkage, and stage-level weighting instead of treating all events equally. Under faster data arrival, the framework again shows a controlled rather than abrupt decline, which is important for real fabrication settings where analytical systems must work under continuous data flow. Taken together, the figure shows that scalability is not an added property but a result of the framework design itself. Because the method organizes data through synchronized event structure, pattern scoring, yield relevance estimation, and SAP-linked prioritization, it remains useful even when production volume, defect load, and data speed all rise together.

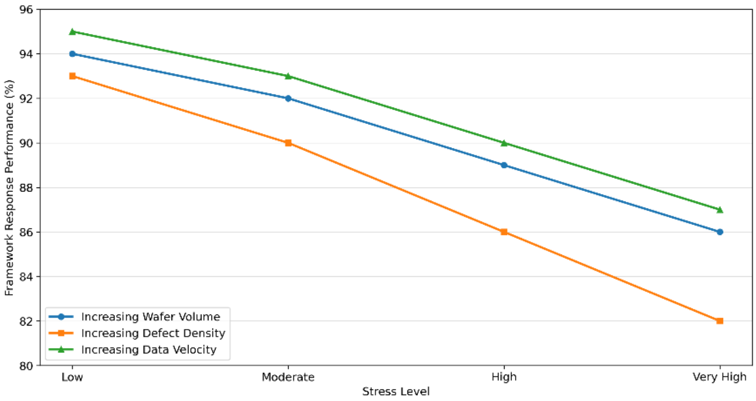

Figure 5. Scalability and response behavior of the framework under increasing wafer volume, defect density, and data velocity

4. Conclusion

This study introduced an SAP-integrated yield analytics framework for wafer fabrication using real-time defect pattern mining. The main technical contribution of the work is the development of an enterprise-connected analytical architecture that links wafer defect pattern behavior, yield-oriented interpretation, and SAP-based process visibility in one continuous framework. Rather than treating defect recognition, yield analysis, and operational response as separate tasks, the proposed method connects them in a single workflow that supports both engineering diagnosis and manufacturing action. This makes the framework more suitable for modern wafer fabrication environments, where analytical value depends not only on prediction quality but also on the speed and usability of process intelligence.

The results confirmed that the framework performs effectively across multiple decision layers. It tracked yield trend behavior across wafer lots, showed clear defect cluster differentiation across pattern categories and process stages, and supported measurable improvement after SAP-integrated alert-driven intervention. It also maintained stable response behavior under increasing wafer volume, rising defect density, and faster data arrival. These findings are important because they show that the framework is not limited to one analytical objective. It supports lot-level yield monitoring, stage-level defect interpretation, intervention-oriented process response, and system-level scalability within the same architecture. This gives the framework both analytical depth and operational reliability.

A key outcome of the study is that wafer defect analytics becomes significantly more useful when it is directly connected to yield relevance and enterprise response logic. In many fabrication settings, defect signals are available, but they remain underused because they are not translated quickly into ranked, process-ready intelligence. The proposed framework addresses this limitation by converting defect events into yield-linked analytical signals and routing them into an SAP-oriented monitoring and alert layer. This improves the practical meaning of defect analytics and makes the outputs easier to use for process engineers, quality teams, and production managers. The framework therefore moves wafer analytics from passive defect observation toward active yield intelligence.

In conclusion, the proposed approach provides a strong foundation for real-time, enterprise-connected yield management in semiconductor fabrication. Its value lies not only in detecting defect behavior, but in organizing that behavior into actionable process knowledge that supports faster and better decisions. Future work can extend this framework through larger fab datasets, richer equipment-level variables, and stronger integration with automated control logic. With these extensions, the framework has clear potential to support next-generation semiconductor manufacturing systems in which defect analytics, yield intelligence, and enterprise decision support operate as one connected environment.

References

1. Hsu, C. Y., Chen, W. J., & Chien, J. C. (2020). Similarity matching of wafer bin maps for manufacturing intelligence to empower industry 3.5 for semiconductor manufacturing. *Computers & Industrial Engineering*, *142*, 106358.
2. Jeong, I., Lee, S. Y., Park, K., Kim, I., Huh, H., & Lee, S. (2023). Wafer map failure pattern classification using geometric transformation-invariant convolutional neural network. *Scientific Reports*, *13*(1), 8127.
3. Kim, E., An, J., Cho, H. C., Cho, S., & Lee, B. (2023). A sensor data mining process for identifying root causes associated with low yield in semiconductor manufacturing. *Data Technologies and Applications*, *57*(3), 397-417.
4. Amato, U., Antoniadis, A., De Feis, I., Doinychko, A., Gijbels, I., La Magna, A., ... & Vasquez, P. (2025). Detecting important features and predicting yield from defects detected by SEM in semiconductor production. *Sensors*, *25*(13), 4218.
5. Busch, R., Czerner, P., Wahl, M., & Choubey, B. (2025). Yield prediction in semiconductor manufacturing using two-step machine learning. *International Journal of Production Research*, 1-18.
6. Shin, J. S., Kim, M. J., & Lee, D. H. (2025). A framework for detecting unknown defect patterns on wafer bin maps using active learning. *Expert Systems with Applications*, *260*, 125378.
7. Ahmadi, A., Cantini, A., & Staudacher, A. P. (2025). Enhancing Operational Efficiency and Human-AI Interaction in Manufacturing through Time-Driven Costing and Predictive Analytics Integration in SAP ERP. *IFAC-PapersOnLine*, *59*(10), 1307-1312.
8. van der Pas, M. C. A., Akçay, A. E., Dijkman, R. M., & Adan, I. J. (2025). Combining Case-Based Reasoning and Process Mining for Root Cause Analysis in complex manufacturing environments. *Manufacturing Letters*.
9. Lee, Y., & Roh, Y. (2023). An expandable yield prediction framework using explainable artificial intelligence for semiconductor manufacturing. *Applied Sciences*, *13*(4), 2660.
10. Joo, J., & Kim, C. O. (2025). Global feature identification layer for mixed-type wafer bin map classification. *Expert Systems with Applications*, *271*, 126709.
11. Kang, M. S., Shin, J. S., & Lee, D. H. (2024). Similarity searching for wafer bin maps by measuring shape, location, and size similarities of defect patterns. *Computers & Industrial Engineering*, *196*, 110486.
12. Ji, P., He, Z., Yang, W., Du, J., Ye, G., & Lu, X. (2025). Wafer defect detection technology based on CTM-IYOLOv10 network. *Journal of Imaging*, *11*(11), 408.
13. Wang, Z., Chen, G., Sun, W., Wu, X., Zheng, L., Zhang, Y., & Liu, Q. (2025). An Efficient Detection Method for Wafer-Test-Induced Defects. *Electronics*, *14*(23), 4664.
14. Rammal, A., Ezukwoke, K., Hoayek, A., & Batton-Hubert, M. (2023). Root cause prediction for failures in semiconductor industry, a genetic algorithm–machine learning approach. *Scientific Reports*, *13*(1), 4934.