Computational Knowledge Discovery from Semiconductor Process Metrology and Electrical Test Data

James Wilson

Australia

Abstract

This study develops a computational knowledge discovery framework for extracting meaningful process insights from semiconductor metrology measurements and electrical test datasets. The proposed approach integrates critical dimension measurements, thin-film thickness, overlay accuracy, wafer defect distributions, and electrical parameters to investigate relationships between fabrication variability and device performance. Advanced data preprocessing techniques are employed to address missing observations, measurement noise, and high-dimensional dependencies. Principal Component Analysis and clustering algorithms are incorporated to identify hidden process patterns, while association rule mining reveals relationships among manufacturing parameters and electrical test outcomes. Supervised learning models are utilized to predict device performance variations and identify influential process characteristics. Statistical validation and feature importance analysis are designed to assess model reliability and interpretability across different manufacturing conditions. The framework supports early detection of process abnormalities, improved yield analysis, and systematic knowledge extraction, contributing to intelligent process monitoring and data-driven decision-making in semiconductor manufacturing.

Keywords: Semiconductor Process Metrology; Electrical Test Data; Knowledge Discovery; Computational Analytics; Manufacturing Yield Optimization.

1. Introduction

Enterprise applications depend on reliable diagnostics to detect runtime errors, failed workflows, API issues, slow transactions, and integration faults [1]. Logs are the main evidence source for understanding what happened before, during, and after an incident. Effective application diagnostics require structured logging, traceability, contextual metadata, error classification, and operational visibility [2]. However, fixed logging strategies often create two problems. Low-detail logs may miss important causes, while high-detail logs may generate excessive storage cost and alert noise [3].

The main challenge is that diagnostic needs change during runtime. A normal transaction may require only basic logs, but a failing workflow, repeated exception, or security-sensitive event may need deeper traces [4]. Adaptive logging improves this process by changing log depth according to system condition, transaction risk, and error behavior [5]. AI-assisted monitoring can identify when additional diagnostic detail is needed and when logging can return to normal levels. This article proposes an adaptive logging framework for enterprise application diagnostics.

2. Methodology

The proposed framework captures log metadata from application servers, API gateways, databases, background jobs, authentication modules, and integration services [6]. It records event type, severity, timestamp, user role, transaction ID, request path, error frequency, response time, affected component, and workflow state. These features are converted into diagnostic context profiles. Adaptive logging requires context because the same event may be low-risk in one workflow but critical in another [7].

The logging controller classifies runtime states as normal, warning, error-prone, performance-risk, security-sensitive, workflow-critical, or incident-active [8]. Based on this classification, it adjusts log level, trace depth, parameter capture, correlation ID tracking, and diagnostic enrichment. Lightweight scoring and policy rules are used so logging decisions remain fast and explainable. Sensitive fields are masked before storage [9], and transaction-critical events are linked with workflow and database identifiers for easier diagnosis.

The framework is evaluated using simulated enterprise applications involving APIs, database workflows, authentication events, scheduled jobs, and service integrations. Static logging is compared with adaptive logging under repeated exceptions, slow requests, failed jobs, API timeout, suspicious login, and workflow interruption scenarios [10]. Evaluation metrics include diagnosis time, log volume reduction, root-cause evidence quality, missed-event rate, storage overhead, alert usefulness, and administrator acceptance rate. Feedback from resolved incidents, ignored logs, and administrator notes is used to refine logging policies over time.

3. Results and Discussion

The results show that adaptive logging improves diagnostics by capturing richer evidence only when runtime risk increases. In the baseline condition, static logs either lack enough detail or produce large volumes of low-value records. With the proposed framework, detailed traces are activated for high-risk events and reduced during normal execution. The strongest improvement appears in intermittent failures, where deeper logs are needed only around the failure window.

The discussion shows that adaptive logging must balance diagnostic depth with privacy and storage cost. Excessive logging can expose sensitive data or overload monitoring systems, while weak logging may hide failure causes. The main limitation is that log-level decisions depend on accurate runtime classification. Regular policy review and masking controls are necessary for safe long-term use.

4. Conclusion

This article presented an adaptive logging framework for enterprise application diagnostics. The framework adjusts log detail using runtime context, risk scoring, workflow state, and error behavior. It improves troubleshooting by increasing diagnostic evidence during abnormal conditions while reducing unnecessary log noise during normal operation. The study shows that adaptive logging can strengthen enterprise diagnostics when combined with structured metadata, privacy controls, and continuous incident feedback.

References

1. Jiang, D., Lin, W., & Raghavan, N. (2020). A novel framework for semiconductor manufacturing final test yield classification using machine learning techniques. *Ieee Access*, *8*, 197885-197895.
2. Jiang, D., Lin, W., & Raghavan, N. (2021). A Gaussian mixture model clustering ensemble regressor for semiconductor manufacturing final test yield prediction. *Ieee Access*, *9*, 22253-22263.
3. Tin, T. C., Tan, S. C., & Lee, C. K. (2022). Virtual metrology in semiconductor fabrication foundry using deep learning neural networks. *IEEE Access*, *10*, 81960-81973.
4. Gondi, S. R. (2024). Digital Twin Frameworks for Real-Time Process Monitoring and Yield Prediction in Semiconductor Fabs. Emerging Digital Systems, 11, 37-43.
5. Kang, S., Kim, D., & Cho, S. (2016). Efficient feature selection-based on random forward search for virtual metrology modeling. *IEEE Transactions on Semiconductor Manufacturing*, *29*(4), 391-398.
6. 吕佑龙, 许鸿伟, 郑城, 张洁, & 郑鹏. (2020). # br# Wafer Acceptance Test Key Parameter Identification Based on Hybrid Feature Selection Method. *China Mechanical Engineering*, *31*(16), 1978-1984.
7. 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.

8. Sheelam, G. K., & Nandan, B. P. (2021). Machine Learning Integration in Semiconductor Research and Manufacturing Pipelines. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), DOI*, *10*.
9. Espadinha-Cruz, P., Godina, R., & Rodrigues, E. M. (2021). A review of data mining applications in semiconductor manufacturing. *Processes*, *9*(2), 305.
10. rao Gondi, S. (2019). Predictive Modeling of Wafer Acceptance Test Parameters for Semiconductor Yield Enhancement. Emerging Digital Systems, 6, 33-38.