Ensemble Learning Framework for Predicting Wafer Acceptance Test Parameters from Manufacturing Data

Thomas Becker

Germany

Abstract

This study proposes an ensemble learning framework for predicting Wafer Acceptance Test (WAT) parameters from semiconductor manufacturing data to improve process monitoring, electrical performance estimation, and production quality. The proposed methodology integrates fabrication process variables, equipment sensor measurements, metrology records, and historical electrical test results to establish relationships between manufacturing conditions and wafer-level electrical characteristics. Ensemble regression techniques, including Random Forest, Gradient Boosting, and Stacking Regression, are employed to predict critical WAT parameters such as threshold voltage, leakage current, sheet resistance, and contact resistance. Feature engineering and statistical preprocessing techniques are incorporated to address nonlinear dependencies, measurement noise, and variations across manufacturing batches. Model performance is assessed using coefficient of determination, root mean square error, mean absolute error, and cross-validation procedures. Feature importance analysis supports identification of influential manufacturing variables. The proposed framework facilitates early electrical performance prediction, process optimization, and data-driven quality assurance in semiconductor fabrication.

Keywords: Ensemble Learning; Wafer Acceptance Testing; Semiconductor Manufacturing; Predictive Modeling; Electrical Parameter Estimation.

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, 9]. 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 [10], 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 [11]. 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. rao Gondi, S. (2023). Data-Driven Process Analytics for Early Defect Detection in High-Volume Semiconductor Fabrication. Cross-Disciplinary Knowledge Systems, 10, 37-42.
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. 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.
4. Gondi, S. R. (2026). Predictive Yield Modeling Using Machine Learning and Process Metrology Data in Semiconductor Fabrication. Cross-Disciplinary Knowledge Systems, 13, 1-8.
5. Lim, M. Y. S., Sharma, A., Chin, C. S., Yip, T. C. M., & Ong, J. Y. S. (2022, June). Prediction of wafer map categories using wafer acceptance test parameters in semiconductor manufacturing. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 136-144). Cham: Springer International Publishing.
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. Fan, A., Huang, Y., Xu, F., & Bom, S. (2023). Soft-sensing regression model: From sensor to wafer metrology forecasting. *Sensors*, *23*(20), 8363.

8. Xu, H., Zhang, J., Lv, Y., & Zheng, P. (2020). Hybrid feature selection for wafer acceptance test parameters in semiconductor manufacturing. *Ieee Access*, *8*, 17320-17330.
9. Xu, H. W., Qin, W., & Sun, Y. N. (2022). An improved XGBoost prediction model for multi-batch wafer yield in semiconductor manufacturing. *IFAC-PapersOnLine*, *55*(10), 2162-2166.
10. rao Gondi, S. (2019). Predictive Modeling of Wafer Acceptance Test Parameters for Semiconductor Yield Enhancement. Emerging Digital Systems, 6, 33-38.
11. Wei, M. C., Shen, C. W., & Hsieh, H. P. (2025, March). Spatial Modeling with Automated Machine Learning and Gaussian Process Regression Techniques for Imputing Wafer Acceptance Test Data. In *2025 Design, Automation & Test in Europe Conference (DATE)* (pp. 1-7). IEEE.