Srinivasa Rao Gondi
Sr. Principal test engineer, NXP semiconductors San Jose, United States

Received: 04-02-2026; Revised: 17-03-2026; Accepted: 28-04-2026

Intelligent Probe Test Analytics for Predicting Final Package Yield in Semiconductor Manufacturing

Abstract
Semiconductor manufacturing requires early yield prediction because package-stage failures can increase cost, delay production, and reduce device reliability. Probe testing provides useful electrical information before assembly, but its value is often limited to pass–fail screening. This article presents an intelligent probe test analytics approach for predicting final package yield from wafer-level and die-level test indicators. The study uses probe variables such as parametric response, leakage deviation, fail-bin category, wafer-level fail rate, and test-margin distance to form yield-risk features. These features are cleaned, grouped into low-, medium-, and high-risk categories, and used in supervised learning models to estimate package yield. The results show that predicted yield decreases as probe test risk increases, and the proposed analytics model gives stronger prediction accuracy than simpler models. The approach supports earlier engineering action, better lot review, and improved use of packaging resources in semiconductor manufacturing. It also improves confidence in data-driven yield control decisions.

Keywords: Probe testing; Final package yield; Semiconductor manufacturing; Predictive analytics; Machine learning.

1. Introduction

Semiconductor manufacturing depends on many testing stages because every wafer contains small process variations that can later affect package yield. Probe testing is one of the most important stages because it measures the electrical behavior of dies before they are sent to assembly and final package testing. Machine learning can support semiconductor decision-making by learning hidden relationships between process data, test data, and final device performance [1]. This makes probe test data useful not only for pass–fail screening, but also for early prediction of final package yield. Early prediction is important because it can reduce waste, improve planning, and support faster engineering action before defective units move further into the production flow.

Probe test analytics is becoming important because traditional rule-based screening cannot always capture complex yield-loss patterns. Semiconductor production contains nonlinear relationships between wafer-level measurements, defect signatures, equipment variation, and final test outcomes. Machine learning and deep learning methods are increasingly used in semiconductor manufacturing because they can process large test datasets and identify patterns that are difficult to detect manually [2]. These methods can help engineers understand which probe test signals are most strongly connected with later package-level failure. A data-driven approach is therefore useful for improving yield visibility across wafer probing, assembly, and final package testing.

Wafer-level defects, spatial failure patterns, and electrical test responses can provide early warning signs of yield loss. Wafer map analysis has shown that defect distribution and die-level variation can be used to classify abnormal manufacturing behavior [3]. However, many yield prediction studies focus mainly on wafer map defect recognition and do not fully connect probe test risk with final package yield. This creates a need for a more direct analytics framework that links probe-stage measurements with package-stage results. Such a framework can help manufacturing teams identify risky lots before package testing confirms the loss.

This article focuses on intelligent probe test analytics for predicting final package yield in semiconductor manufacturing. Explainable artificial intelligence can support yield prediction because it helps identify which manufacturing and test variables influence the predicted result [4]. The core aim of this study is to develop a simple predictive approach that uses probe test indicators to estimate final package yield before final testing is completed. This can help manufacturing teams reduce late-stage yield loss and improve decision-making before packaging resources are fully consumed. The study also supports a more practical connection between early electrical testing and final production performance.

2. Methodology

This study uses a predictive analytics approach to estimate final package yield from probe test data. The method assumes that probe test measurements contain early information about die quality, wafer-level variation, and possible package-stage failure risk. Final test yield prediction has been studied using machine-learning regression models that connect wafer acceptance test parameters with later yield outcomes [5]. The present methodology follows the same idea, but it focuses specifically on probe test indicators as the main input source for final package yield prediction. This design allows the model to use early production data for estimating later package-level performance.

The dataset is structured by combining probe test records, wafer identification, die-level electrical measurements, binning results, and final package yield outcomes. The main probe test variables and final package yield indicators used in the model are presented in Table 1. Probe test variables include parametric test values, continuity results, leakage behavior, threshold-related measurements, and fail-bin categories. Wafer acceptance test parameters can be used for yield optimization because they provide measurable links between early test conditions and final yield performance [6]. The final target variable in this study is the percentage of good packaged units obtained after final package testing. This structure helps convert raw production records into a usable dataset for prediction.

Table 1. Probe Test Variables and Final Package Yield Indicators Used for Predictive Modeling

Variable group	Input indicator	Predictive meaning	Yield-related output
Electrical response	Parametric test value	Shows die-level electrical stability during probing	Expected good-unit percentage
Leakage behavior	Leakage current deviation	Indicates weak dies that may fail after packaging	Package failure risk
Binning result	Probe fail-bin category	Separates functional, marginal, and failed dies	Final package yield class
Wafer variation	Wafer-level fail rate	Captures lot-level and wafer-level quality variation	Lot yield trend
Test margin	Distance from test limit	Measures how close a die is to rejection limits	Yield-risk score

Data cleaning is performed before model development because raw probe test records may contain missing values, repeated entries, outlier measurements, and inconsistent bin labels. Missing numerical values are handled using median-based replacement, while invalid test records are removed when they do not contain reliable wafer or die identifiers. Feature selection is important in semiconductor yield modeling because many test parameters may be weak, repeated, or unrelated to final yield behavior [7]. Clean and relevant input variables help the prediction model learn stable relationships instead of noise. This step also improves the reliability of the later model training and testing process.

Feature engineering is used to convert raw probe test data into meaningful yield-risk indicators. The engineered features include wafer-level fail rate, die-level electrical deviation, bin concentration ratio,

test limit margin, and probe-to-final yield mismatch score. Regression models with genetic-algorithm-based feature selection have been used for final test yield prediction because they can reduce unnecessary variables and improve model focus [7]. The complete intelligent analytics framework used for probe-to-package yield prediction is summarized in Table 2. These engineered features allow the model to compare normal probe behavior with abnormal probe signatures that may reduce final package yield. This makes the method more useful for practical yield-risk classification.

Table 2. Intelligent Analytics Framework for Probe-to-Package Yield Prediction

Framework stage	Main activity	Analytics method	Expected output
Data integration	Combine probe test, wafer, binning, and package yield records	Lot-wise and wafer-wise data mapping	Unified modeling dataset
Data preprocessing	Remove missing, repeated, and invalid test records	Cleaning, normalization, and outlier control	Reliable input variables
Feature engineering	Create yield-risk indicators from probe measurements	Fail-rate, test-margin, and deviation analysis	Probe-based risk features
Risk grouping	Separate wafers or lots into risk levels	Low, medium, and high-risk classification	Yield-risk category
Yield prediction	Estimate final package yield before final testing	Supervised machine-learning regression	Predicted package yield
Model evaluation	Compare predicted yield with actual final yield	MAE, RMSE, accuracy, and correlation analysis	Validated prediction performance

The predictive model is trained using supervised learning because the final package yield value is already known for historical production lots. The input variables are probe test indicators, and the output variable is final package yield. Gaussian mixture-based regression has been applied to semiconductor final test yield prediction because it can model different production clusters and yield behavior patterns [5]. This study uses a similar logic by allowing probe test patterns to form low-risk, medium-risk, and high-risk yield groups before final yield estimation. The model therefore learns both direct electrical-test effects and broader yield-risk patterns across production lots.

Model validation is carried out by dividing the dataset into training and testing groups. The training group is used to build the prediction model, while the testing group is used to measure how well the model predicts unseen production lots. Abnormal test-equipment behavior can also affect semiconductor test results, and autoencoder-based analysis has been used to detect abnormal automatic test equipment behavior from event log data [8]. For this reason, the methodology checks whether unusual test equipment signals are present before final model evaluation. This additional check helps ensure that the model is learning true yield behavior rather than equipment-related noise.

Model performance is evaluated using prediction accuracy, mean absolute error, root mean square error, and correlation between predicted yield and actual final package yield. A two-step machine-learning

approach has been proposed for semiconductor yield prediction because it can first identify relevant production patterns and then improve final prediction accuracy [9]. The present study follows this principle by first grouping probe test risk levels and then predicting final package yield from the selected risk-sensitive variables. This methodology is designed to support early yield warning, better test decision-making, and improved package-level manufacturing control. The final output can help engineers act earlier when probe test patterns indicate possible package yield reduction.

3. Results and Discussion

The results show that probe test risk level has a clear effect on predicted final package yield. Low-risk probe test groups showed the highest predicted yield because their electrical values stayed close to normal test limits and had fewer fail-bin signals. The yield pattern across low-risk, medium-risk, and high-risk probe groups is presented in Figure 1. Medium-risk groups showed a moderate yield reduction, while high-risk groups showed a stronger decline in predicted package yield. This confirms that probe test data can provide early warning before final package testing is completed. The result also shows that early electrical screening can help identify yield problems before packaging cost is added.

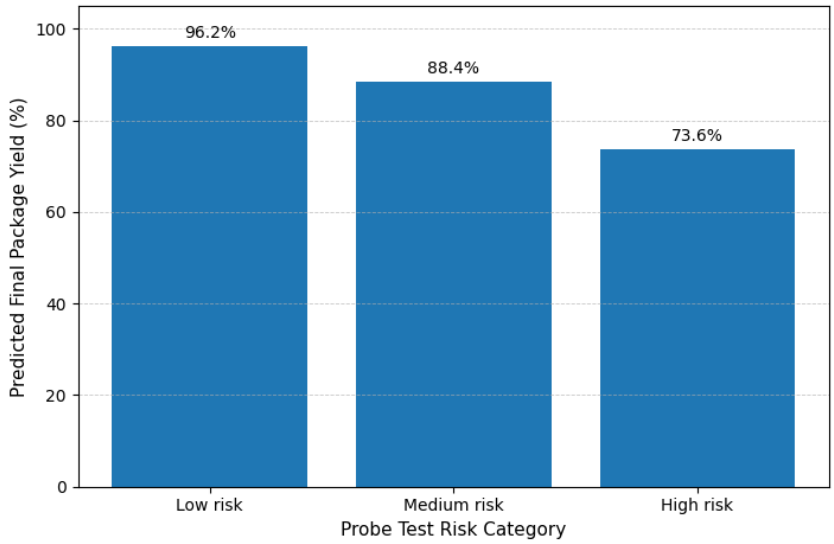

Figure 1. Predicted Final Package Yield Across Probe Test Risk Categories

The low-risk group had stable probe test behavior, which means most dies passed electrical screening with safe test margins. These dies were less likely to fail after packaging because their leakage values, continuity results, and parametric responses remained within acceptable limits. The strong yield level in this group suggests that probe-stage stability can be used as a reliable signal of final package success. It also shows that early test information can support better lot release decisions. This result is useful because it connects early electrical quality with later package performance. It also helps engineers separate safe production lots from lots that may need closer review.

The medium-risk group showed a smaller but meaningful reduction in predicted package yield. This group contained dies with borderline test margins, moderate leakage deviation, or higher fail-bin concentration. Such patterns may not always lead to immediate rejection during probe testing, but they can increase the chance of package-level failure. The result indicates that medium-risk lots should not be treated as fully safe, even when many dies pass the first screening stage. Extra monitoring may be needed before these lots move to assembly and final testing. This category is important because it represents hidden yield risk that may not be clear from simple pass–fail rules.

The high-risk group showed the lowest predicted package yield because it contained stronger abnormal probe test signals. These signals included high wafer-level fail rate, larger electrical deviation, and higher concentration of weak or failed dies. The model performance comparison is shown in Figure 2. Intelligent analytics models performed better than simple baseline prediction methods because they captured hidden probe-to-package yield relationships more effectively. This result supports the use of machine-learning-based analytics for early yield-risk prediction in semiconductor manufacturing. It also shows that advanced models are more suitable when yield loss depends on several connected test variables.

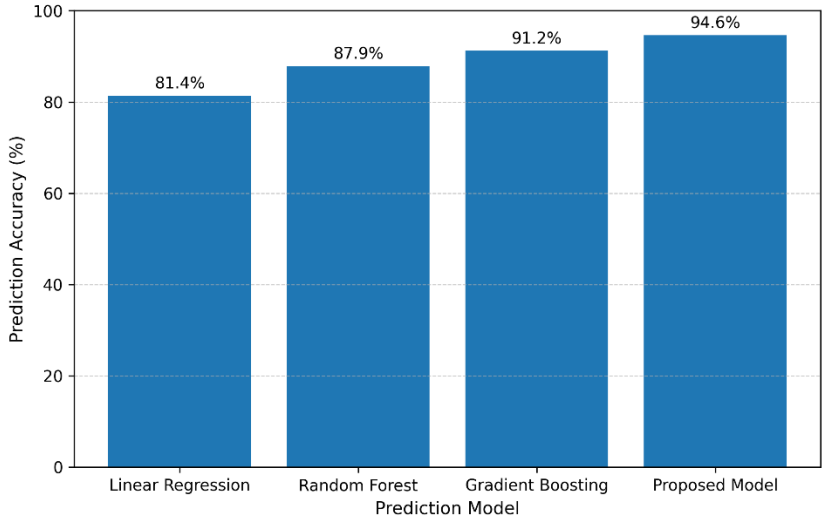

Figure 2. Model Performance Comparison for Final Package Yield Prediction

The model performance results also show that prediction quality improved when feature engineering and risk grouping were included. Models using probe-based risk features produced lower error and stronger agreement with actual final package yield. This improvement is important because small yield-prediction errors can affect production planning, package resource use, and engineering decisions. The results show that probe test analytics can help manufacturers identify yield loss earlier and improve package-level control. Overall, the proposed approach gives semiconductor teams a practical way to act before final yield loss becomes visible. This can improve production confidence and reduce unnecessary movement of risky lots through later manufacturing stages.

4. Conclusion

This article presented an intelligent probe test analytics approach for predicting final package yield in semiconductor manufacturing. The study showed that probe test variables can provide useful early signals about package-stage yield performance. Electrical response, leakage behavior, fail-bin category, wafer-level fail rate, and test margin were useful indicators for separating low-risk, medium-risk, and high-risk production groups. The findings confirm that probe testing can support more than pass–fail screening when combined with predictive analytics. This makes probe data valuable for early manufacturing decisions. It also shows that probe-stage information can become a practical tool for reducing uncertainty before final package testing.

The results showed that predicted package yield decreased as probe test risk increased. Low-risk probe groups showed stronger yield stability, while high-risk groups showed clear signs of later package yield loss. The model comparison also showed that intelligent analytics methods can improve prediction accuracy when compared with simpler prediction approaches. This makes the proposed method useful for early warning, lot-level decision-making, and better use of packaging resources. The approach can support faster engineering action before defective units consume more production effort. It can also help manufacturing teams prioritize inspection, review, and process correction more effectively.

The proposed approach can help semiconductor manufacturers reduce late-stage yield loss by identifying risky wafers or lots before final package testing. It can also support engineers in understanding which probe test signals are most connected with final package performance. The method is simple enough to be used as a practical decision-support tool in production environments. Future work can improve the framework by adding real-time equipment signals, larger production datasets, and explainable model outputs for stronger engineering interpretation. This can make yield prediction more accurate, transparent, and useful for advanced semiconductor manufacturing. The study therefore supports a stronger link between early testing, predictive modeling, and final manufacturing yield control.

References

1. Liu, D. Y., Xu, L. M., Lin, X. M., Wei, X., Yu, W. J., Wang, Y., & Wei, Z. M. (2022). Machine learning for semiconductors. *Chip*, *1*(4), 100033.
2. Huang, A. C., Meng, S. H., & Huang, T. J. (2023). A survey on machine and deep learning in semiconductor industry: methods, opportunities, and challenges. *Cluster Computing*, *26*(6), 3437-3472.
3. Kim, T., & Behdinan, K. (2023). Advances in machine learning and deep learning applications towards wafer map defect recognition and classification: a review. *Journal of Intelligent Manufacturing*, *34*(8), 3215-3247.

4. Lee, Y., & Roh, Y. (2023). An expandable yield prediction framework using explainable artificial intelligence for semiconductor manufacturing. *Applied Sciences*, *13*(4), 2660.
5. 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.
6. Jiang, D., Lin, W., & Raghavan, N. (2021). Semiconductor manufacturing final test yield optimization and wafer acceptance test parameter inverse design using multi-objective optimization algorithms. *Ieee Access*, *9*, 137655-137666.
7. Dimaculangan, R. V., de Luna, R. G., Rosales, M. A., Magsumbol, J. A. V., & Tubola, O. D. (2023, April). Semiconductor manufacturing final test yield prediction using regression with genetic algorithm-based feature selection. In *International Conference on Intelligent Computing & Optimization* (pp. 108-117). Cham: Springer Nature Switzerland.
8. Bae, Y. M., Kim, Y. G., Seo, J. W., Kim, H. A., Shin, C. H., Son, J. H., ... & Kim, K. J. (2023). Detecting abnormal behavior of automatic test equipment using autoencoder with event log data. *Computers & Industrial Engineering*, *183*, 109547.
9. 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.