International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

AI-ENABLED FAULT IDENTIFICATION IN ELECTRONIC CIRCUITS FOR SEMICONDUCTORS INDUSTRY

Mr. K. Vamshee Krishna[1], Kuthadi Suresh[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering, Kommuri Pratap Reddy Institute of Technology, Hyderabad, Telangana, India.
Email: vamshik825@gmail.com.

Received: 05-6-2025 Accepted: 06-7-2025 Published: 14-7-2025

ABSTRACT

In semiconductor manufacturing, ensuring high yield rates is critical for optimizing production efficiency and minimizing costs. However, the vast number of signals collected from sensors and process measurement points often contain a mix of relevant information, noise, and irrelevant data, making it challenging for engineers to identify the key factors affecting yield. Feature selection techniques are instrumental in addressing this challenge, as they help identify the most relevant signals that significantly impact yield. In conventional semiconductor manufacturing, engineers are inundated with an extensive array of signals, making it cumbersome and time-consuming to pinpoint the critical factors influencing yield excursions. This data overload often results in suboptimal process efficiency and increased production costs. Traditional approaches were not effectively distinguished between useful information and noise, leading to inefficient troubleshooting and reduced yield rates. Additionally, manual feature selection processes are labour-intensive and was not uncover complex causal relationships between variables, limiting their effectiveness in enhancing semiconductor manufacturing operations. To overcome the limitations of the conventional approach, this study proposes the use of artificial intelligence-based feature selection techniques. By leveraging sophisticated algorithms, the proposed system will rank features according to their impact on semiconductor manufacturing yield. These techniques will not only streamline the identification of crucial variables but also unveil causal relationships within the data, providing a deeper understanding of the production process. The application of cross-validation ensures robustness and reliable evaluation of feature relevance for predictability using error rates. The goal is to empower engineers with a more efficient and data-driven approach to semiconductor manufacturing, resulting in increased yield rates, reduced production costs, and shorter learning cycles. The preliminary results presented here demonstrate the potential of this approach, highlighting the promise of artificial intelligence in revolutionizing semiconductor manufacturing optimization.

Keywords: AI, Feature Selection, Semiconductor Manufacturing, Fault Identification, Yield Optimization, Signal Noise Reduction, Process Efficiency, Causal Relationships, Cross-Validation, Data-Driven Approach.

1. Introduction

Artificial Intelligence for Semiconductor Fault Identification from Integrated Circuit Statistics" is all about making semiconductor production better. In this study, we dive into the process of choosing which features are most important to get high yields, meaning more good chips from the manufacturing process. To start, we look at tons of data collected during manufacturing. This data includes everything from how the

machines are running to what the environment is like. We use fancy math techniques to sift through all this data and find the key factors that affect how many good chips we get. The cool part is that we're not just relying on math alone. We also bring in people who know a lot about making semiconductors. By combining their expertise with the data analysis, we can figure out which factors really matter for getting good yields. We're also exploring how to use artificial intelligence to help us in this process. By teaching computers to learn from the data, we can speed up the selection process and make it more accurate. This means less guesswork and more efficiency in choosing the right features. In addition to all of this, we're looking at everything that could affect yields, like how the manufacturing process varies, how well the machines are working, and even the conditions in the factory. Understanding all of these factors helps us build a more complete picture of what influences yield. Overall, this research aims to revolutionize how we make semiconductors. By using advanced techniques and combining human expertise with machine learning, we're hoping to boost yields and make semiconductor manufacturing more competitive and sustainable in the The motivation behind applying Artificial Intelligence (AI) in Semiconductor Fault Identification from Integrated Circuit Statistics lies in improving yield rates by automating fault detection, optimizing resource allocation, and enhancing decision-making across the manufacturing lifecycle. Traditional systems like manual IC verification protocols and the Universal Verification Protocol (UVP) offer foundational reliability checks but fall short in handling the increasing data complexity, manual inefficiencies, and limited fault interpretability. This research aims to utilize AI-driven automated feature selection and fault prediction techniques to analyze critical parameters like heat, temperature, size, power, and time, thereby identifying impactful features influencing yield. By leveraging AI algorithms, manufacturers can implement predictive maintenance, optimize processes in real-time, and enable accurate yield forecasting. The Random Forest Classifier (RFC) is employed for its robustness, high accuracy, and ability to handle imbalanced data while offering feature importance insights. Performance is evaluated using confusion matrices and classification reports, highlighting precision, recall, F1-score, and accuracy to assess the system's ability to identify faults reliably. Ultimately, the integration of AI fosters increased yield rates, cost efficiency, operational improvements, sustainability, and continuous process refinement—positioning semiconductor manufacturers for greater productivity and competitiveness.

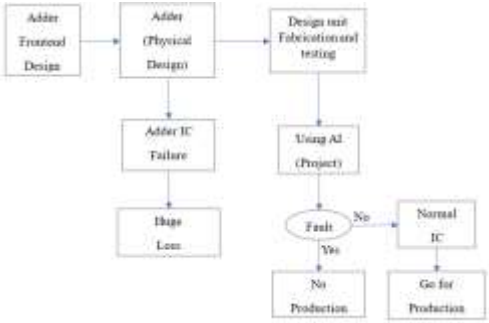

Fig1: Research Motivation

2. LITERATURE SURVEY

Lee, et. al [1] Enormous amounts of data were generated and analyzed in the latest semiconductor industry. Established yield prediction studies had dealt with one type of data or a dataset from one procedure. However, semiconductor device fabrication comprises hundreds of processes, and various factors affected device yields. This challenge was addressed in their study by using an expandable input data-based framework to include divergent factors in the prediction and by adapting explainable artificial intelligence (XAI), which utilized model interpretation to modify fabrication conditions. After preprocessing the data, the procedure of optimizing and comparing several machine learning models was followed to select the best performing model for the

dataset, which was a random forest (RF) regression with a root mean square error (RMSE) value of 0.648. The prediction results enhanced production management, and the explanations of the model deepened the understanding of yield-related factors with Shapley additive explanation (SHAP) values. This work provided evidence with an empirical case study of device production data. The framework improved prediction accuracy, and the relationships between yield and features were illustrated with the SHAP value.

Chowdhury, et. al [2] The paper investigated into the intricacies of semiconductor manufacturing, a highly complex process entailing a wide array of subprocesses and diverse equipment. Semiconductors were miniaturized integrated circuits comprising numerous components. The semiconductor manufacturing process began with thin disc-shaped silicon wafers. On each wafer, up to thousands of identical chips could be prepared depending upon the diameter of the wafer to build up the circuits layer by layer in a wafer fab. The size of the semiconductors required a high number of units to be produced, thus necessitating a large amount of data to control for improving the semiconductor manufacturing process. Therefore, the collection and analysis of the equipment data, process data, and machine history data throughout the manufacturing process were required to diagnose faults, monitor the process, and manage the manufacturing process effectively. Their research was focused on improving the semiconductor manufacturing process through a rigorous analysis of collected manufacturing process data, employing statistical process control (SPC), data mining techniques, and data-driven decision models. The project's primary objective was to increase the manufacturing process stability and productivity by utilizing the latest data-driven technologies in the scientific community. A structured review was undertaken, exploring contemporary data-driven methodologies in semiconductor manufacturing process improvement, specifically pertaining to process capability, product yield rate, and process stability. This review accentuated a comprehensive evaluation of data-driven methodologies applicable to conventional semiconductor manufacturing facilities, aiming to drive substantial process improvements. It featured a detailed demonstration facilitating the selection of optimal semiconductor manufacturing processes to enhance overall operational performance.

Nuhu, et. al [3] Industries underwent the fourth industrial revolution (Industry 4.0), where technologies like the Industrial Internet of Things, big data analytics, and machine learning (ML) were extensively utilized to improve the productivity and efficiency of manufacturing systems and processes. Their work aimed to further investigate the applicability and improve the effectiveness of ML prediction models for fault diagnosis in the smart manufacturing process. Hence, they proposed several methodologies and ML models for fault diagnosis for smart manufacturing process applications. A case study was conducted on a real dataset from a semiconductor manufacturing (SECOM) process. However, this dataset contained missing values, noisy features, and a class imbalance problem. This imbalance problem made it so difficult to accurately predict the minority class, due to the majority class size difference. In the literature, efforts had been made to alleviate the class imbalance problem using several synthetic data generation techniques (SDGT) on the UCI machine learning repository SECOM dataset. In their work, to handle the imbalance problem, they employed, compared, and evaluated the feasibility of three SDGT on this dataset. To handle issues related to the missing values and noisy features, they implemented two missing values imputation techniques and feature

selection techniques, respectively. They then developed and compared the performance of ten predictive ML models against these proposed methodologies. This study aimed to develop a deep learning-based fault diagnosis framework as prognostics and health management (PHM) solutions with a comprehensive analytics process. A linear encoder sensor was employed to measure position, while data preparation was conducted to convert position information into a signal. Then, signal processing was used to extract the features, in which high pass filter was applied to normalize the signal and emphasize the features. High-dimensional features including time domain statistical features, time-frequency domain features obtained by continuous wavelet transform (CWT), and frequency domain features converted by fast Fourier transform (FFT) were extracted to prepare the dataset. An ensemble neural network model that integrated deep neural network (DNN) and convolutional neural network (CNN) was employed to recognize the pattern and diagnose the positioning errors. The accuracy and false alarm rate were considered to support maintenance decisions. An empirical study was conducted in a world-leading IC assembly and testing company for validation. The results showed that the proposed approach could effectively detect inappropriate installation of the bond head in wire bonding equipment for predictive maintenance and cost reduction.

Kao, et. al [4] In their paper, they developed a data-driven decision model to improve process quality in manufacturing. A challenge for traditional methods in quality management was handling high-dimensional and nonlinear manufacturing data. They addressed this challenge by adapting explainable artificial intelligence to the context of quality management. Specifically, they proposed the use of nonlinear modeling with Shapley additive explanations to infer how a set of production parameters and the process quality of a manufacturing system were related. Thereby, they contributed a measure of process importance based on which manufacturers could prioritize processes for quality improvement. Grounded in quality management theory, their decision model selected improvement actions that targeted the sources of quality variation. The decision model was validated in a real-world application at a leading manufacturer of high-power semiconductors. Seeking to improve production yield, they applied their decision model to select improvement actions for a transistor chip product. They then conducted a field experiment to confirm the effectiveness of the improvement actions. Compared with the average yield in their sample, the experiment returned a reduction in yield loss of 21.7%. Furthermore, they reported on results from a post-experimental rollout of the decision model, which also resulted in significant yield improvements. They demonstrated the operational value of explainable artificial intelligence by showing that critical drivers of process quality could go undiscovered by the use of traditional methods.

Senoner, et. al [5] In their paper, they investigated into the intricacies of semiconductor manufacturing, a highly complex process entailing a wide array of subprocesses and diverse equipment. Semiconductors were miniaturized integrated circuits comprising numerous components. The semiconductor manufacturing process began with the thin disc-shaped silicon wafers. On each wafer, up to thousands of identical chips could be prepared depending upon the diameter of the wafer to build up the circuits layer by layer in a wafer fab. The size of the semiconductors required a high number of units to be produced, thus necessitating a large amount of data to control for improving the semiconductor manufacturing process. Therefore, the collection and analysis of the equipment data, process data, and machine history data throughout the manufacturing

process were required to diagnose faults, monitor the process, and manage the manufacturing process effectively. Their research was focused on improving the semiconductor manufacturing process through a rigorous analysis of collected manufacturing process data, employing statistical process control (SPC), data mining techniques, and data-driven decision models. The project's primary objective was to increase the manufacturing process stability and productivity by utilizing the latest data-driven technologies in the scientific community. A structured review was undertaken, exploring contemporary data-driven methodologies in semiconductor manufacturing process improvement, specifically pertaining to process capability, product yield rate, and process stability. This review accentuated a comprehensive evaluation of data-driven methodologies applicable to conventional semiconductor manufacturing facilities, aiming to drive substantial process improvements. It featured a detailed demonstration facilitating the selection of optimal semiconductor manufacturing processes to enhance overall operational performance.

Gómez-Sirvent, et. al [6] Semiconductors were essential components in many electronic devices. Because wafers were produced quickly and in large quantities, defects occurred that adversely affected semiconductor properties. This made it necessary to install powerful and robust inspection systems which used artificial intelligence techniques in the early stages of the manufacturing chain in order to detect and classify those defects. In their paper, they proposed a method for defect detection and classification on images of semiconductor wafer materials obtained by means of a scanning electron microscope based in the following stages: (i) use of computer vision techniques to isolate the defect from the background; (ii) use of several descriptors based on shape, size, texture, histogram, and key-points to create a feature vector for the characterization of the defect; (iii) application of an exhaustive search as a feature selection method to determine the optimal subset of feature descriptors; and (iv) evaluation of the feature descriptors by using a support vector machine classifier providing the optimal set with highest F1-score metrics.

Xu, Qiuhao, et. al [7] Wafer yield prediction, as the basis of quality control, was dedicated to predicting quality indices of the wafer manufacturing process. In recent years, data-driven machine learning methods received a lot of attention due to their accuracy, robustness, and convenience for the prediction of quality indices. However, the existing studies mainly focused on the model level to improve the accuracy of yield prediction and did not consider the impact of data characteristics on yield prediction. To tackle the above issues, a novel wafer yield prediction method was proposed, in which the improved genetic algorithm (IGA) was an under-sampling method used to solve the problem of data overlap between finished products and defective products caused by the similarity of manufacturing processes between finished products and defective products in the wafer manufacturing process, and the problem of data imbalance caused by too few defective samples, that is, the problem of uneven distribution of data. In addition, the high-dimensional alternating feature selection method (HAFS) was used to select key influencing processes, that is, key parameters to avoid overfitting in the prediction model caused by many input parameters. Finally, SVM was used to predict the yield. Furthermore, experiments were conducted on a public wafer yield prediction dataset collected from an actual wafer manufacturing system. IGA-HAFS-SVM achieved state-of-art results on this dataset, which confirmed the effectiveness of IGA-HAFS-SVM. Additionally, on this dataset, the proposed method improved the AUC score, G-Mean, and F1-score by 21.6%, 34.6%, and 0.6%

respectively compared with the conventional method.

Fan, Shu-Kai S., et. al [8] In their paper, an integrated framework was proposed to predict the first yield, aiming to identify the best-practice accessory combination for the final testing (FT) process in semiconductor manufacturing. Firstly, they adopted the entity embedding method to convert the categorical data of accessory combinations into multidimensional vectors. Several possible algorithms in machine learning were then examined to establish the yield prediction model, and the one that was the best fit was selected. Once the best machine learning model was determined, they used the genetic algorithm (GA) embedded in the yield prediction model to search for the best-practice accessory combination. This was the combination that gave the highest first yield estimate. The synergy between machine learning and the search heuristic produced an intelligent predictive system that could circumvent the adverse yield rates generated by using inappropriate accessory combinations. The Overall Equipment Effectiveness (OEE) of the FT process could be maintained in a highly stable condition.

2.3 Research Gaps

Limited Integration of Diverse Factors: Many studies focus on specific types of data or datasets from singular procedures within semiconductor manufacturing, overlooking the complexity arising from the multitude of processes and diverse factors affecting device yields. There's a gap in integrating a wide array of factors into predictive models to enhance accuracy and reliability.

Class Imbalance in Data: Several research works highlight challenges related to class imbalance in datasets, especially in fault diagnosis and yield prediction tasks. This imbalance poses difficulties in accurately predicting minority classes, leading to biased models and reduced performance. Addressing this class imbalance issue remains a common research gap.

Inadequate Handling of Noisy and Missing Data: Real-world manufacturing datasets often contain noisy features and missing values, which can adversely affect the performance of predictive models. There's a need for robust methodologies to handle noisy data and effectively impute missing values to improve the reliability of predictive models.

Limited Explanation and Interpretation of Models: While the use of advanced machine learning models such as deep learning and ensemble techniques improves prediction accuracy, the lack of interpretability and explanation of model decisions remains a significant gap. Models that provide insights into the relationships between input features and predictions are essential for gaining trust and understanding in practical manufacturing settings.

Scalability and Generalization of Models: Many studies demonstrate promising results on specific datasets or scenarios, but there's a gap in scalability and generalization of these models to diverse manufacturing environments. Developing models that can adapt to different manufacturing settings and datasets while maintaining performance remains a key research challenge.

Limited Real-world Validation and Deployment: Despite advancements in predictive modeling techniques, there's often a lack of real-world validation and deployment of these models in industrial settings. Bridging this gap requires conducting empirical studies and validating proposed methodologies in real manufacturing environments to assess their practical feasibility and effectiveness.

Integration of Domain Knowledge with AI Techniques: While AI techniques show promise in improving fault detection and yield prediction, there's a gap in effectively integrating domain knowledge with these techniques.

Combining domain expertise with AI algorithms can lead to more accurate and actionable insights for semiconductor manufacturing, yet achieving this integration remains a challenge.

Automation of Defect Detection and Classification: Current methods for defect detection and classification often rely on manual visual inspection or feature extraction, leading to subjective and time-consuming processes. There's a gap in developing fully automated systems that can accurately identify and classify defects on semiconductor wafers, enhancing efficiency and reliability in manufacturing processes.

3. Proposed methodology

The proposed methodology employing the Random Forest Classifier (RFC) for fault identification in semiconductor devices encompasses several key steps to ensure robustness and accuracy. It begins with the upload of the dataset containing integrated circuit statistics, which includes parameters such as voltage, current, and temperature measurements. Data preprocessing techniques are then applied to handle missing values, normalize features, and eliminate outliers, ensuring the quality and integrity of the dataset. Subsequently, Synthetic Minority Over-sampling Technique (SMOTE) may be employed to address class imbalance by generating synthetic samples of the minority class. The dataset is then partitioned into training and testing sets to facilitate model training and evaluation. The RFC model is trained on the training data, where it learns to classify fault occurrences based on the input features. Following training, the model's performance is assessed using the test data, evaluating metrics such as accuracy, precision, recall, and F1 score. Finally, the RFC model is deployed to make predictions on new data, providing valuable insights into fault occurrences in semiconductor devices. This comprehensive methodology leverages the power of the RFC algorithm to accurately identify faults while addressing challenges such as class imbalance and data variability, thereby enhancing the reliability and efficiency of fault identification processes in semiconductor manufacturing. Implementation of a Random Forest Classifier as the core methodology for fault identification in semiconductor devices. Acquisition of integrated circuit statistics data including various parameters such as voltage, current, and temperature measurements. Preprocessing of the dataset to handle missing values, normalize features, and eliminate outliers. Training of the Random Forest Classifier using the training dataset to learn patterns and relationships between integrated circuit statistics and fault occurrences. Evaluation of the trained classifier using the testing dataset to assess its accuracy, precision, recall, and F1-score. Visualization of the classifier's performance using metrics such as confusion matrices and ROC curves. Data preprocessing is a critical step in preparing a dataset for machine learning, involving loading the dataset (typically from a CSV file) into a suitable structure like a pandas DataFrame, identifying and handling missing values through removal or imputation, and converting categorical variables into numerical form using label encoding.

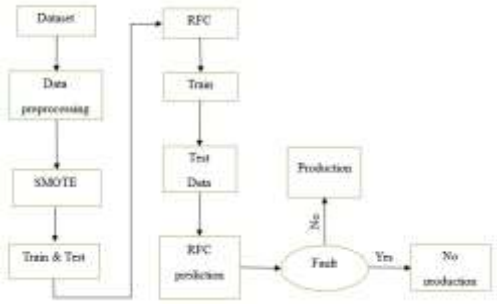

Figure 2: Block Diagram

To ensure fair contribution of all features during model training, numerical values are scaled using techniques like min-max scaling or standardization. The dataset is then split into

input features (X) and the target variable (y), enabling model training and evaluation. In the context of semiconductor fault identification using a Random Forest Classifier (RFC), the dataset is divided into training and testing sets, allowing the model to learn patterns from the training data and validate its performance on unseen data, thereby assessing its generalization capability through metrics like accuracy and F1-score. To address class imbalance, which often hampers the performance of traditional models, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate synthetic samples of the minority class by interpolating between existing minority instances and their nearest neighbors, thus enhancing model sensitivity towards underrepresented classes and improving prediction accuracy while maintaining a balanced dataset.

Random Forest Classifier (RFC)

Random Forest is a widely used supervised machine learning algorithm known for its versatility in handling both classification and regression problems. It operates based on the ensemble learning principle, where multiple decision trees are constructed using various random subsets of the dataset to enhance model performance and reduce overfitting. Rather than relying on a single tree, Random Forest aggregates the predictions of all trees and outputs the majority vote for classification or the average for regression, thereby increasing overall accuracy. The model works in two phases: first, by building several decision trees on randomly selected data samples, and second, by using those trees to make predictions based on majority voting. It assumes that there should be actual values in the feature variables and that the predictions from each tree must be weakly correlated for optimal performance. Advantages of Random Forest Classifier (RFC) in semiconductor fault identification include its high accuracy due to ensemble learning, robustness to overfitting through techniques like bagging and feature randomness, and its ability to provide implicit feature selection by evaluating feature importance. Additionally, RFC handles imbalanced datasets well by training on diverse subsets, is non-parametric with no assumptions about data distribution, and offers relative interpretability, making it a strong and transparent choice for fault detection in integrated circuits.

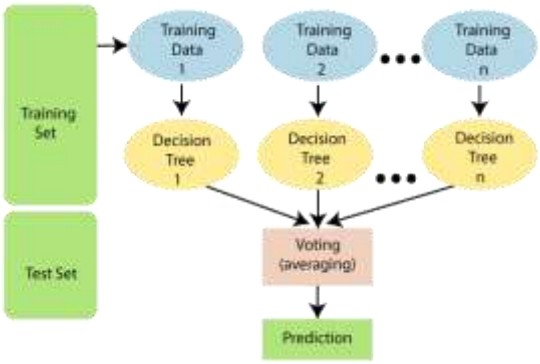

Fig 3: Random Forest classifier

4. RESULTS

Figure 4 shows the interface or graphical user interface (GUI) of the research work. It could include elements such as buttons, dropdown menus, input fields, and other interactive components used to input data, configure parameters, and interact with the research application.

Fig 4. User interface of research work

Figure 5. displays the results obtained after the data preprocessing phase. It could include visualizations or summaries showing how the data was cleaned, transformed, and prepared for further analysis. This may involve techniques such as handling missing values, normalization, and feature engineering.

International Journal of Engineering Research and Science & Technology

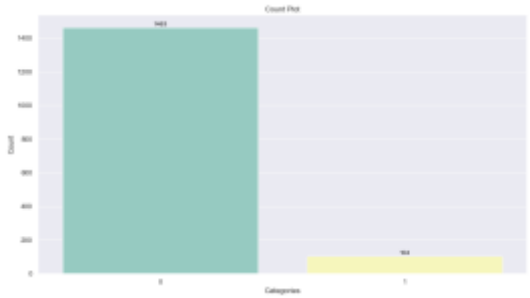

Fig 5. Results after data preprocessing

Figure 6 illustrates the results after applying the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is a technique used to address class imbalance by generating synthetic samples of the minority class. The figure may show how the distribution of classes has changed after applying SMOTE, ensuring a more balanced dataset for training machine learning models.

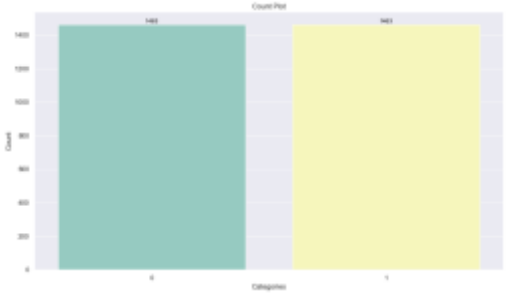

Fig 6. After applying SMOTE

Figure 7 presents the confusion matrix for the existing Multilayer Perceptron (MLP) model. A confusion matrix is a table used to evaluate the performance of a classification model, showing the counts of true positive, true negative, false positive, and false negative predictions. This figure helps assess the MLP model's accuracy and effectiveness in classifying semiconductor faults.

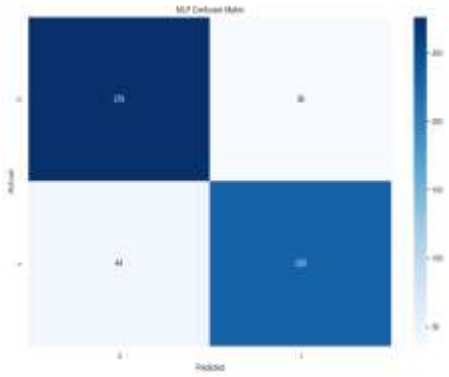

Fig 7. Existing MLP confusion matrix

Figure 8 depicts the confusion matrix for the proposed Random Forest Classifier (RFC) model. It provides insights into the performance of the RFC model in classifying semiconductor faults, allowing for a comparison with the existing MLP model.

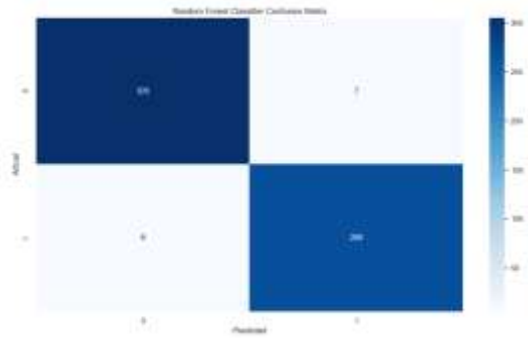

Fig 8. Proposed RFC confusion matrix

Table 1 presents the classification report for the MLP model, including precision, recall, F1-score, and support for each class (0 and 1), as well as accuracy, macro-average, and weighted-average metrics. These metrics evaluate the MLP model's performance in identifying semiconductor faults.

Table 1. MLP classification report

Class	Precision	Recall	F1-Score	Support
0	0.86	0.88	0.87	312
1	0.86	0.84	0.85	274
Accuracy			0.86	586
Macro Avg	0.86	0.86	0.86	586

Weighted Avg	0.86	0.86	0.86	586

Similar to Table 1, this table displays the classification report for the RFC model. It provides insights into the RFC model's precision, recall, F1-score, support, and overall accuracy in classifying semiconductor faults.

Table 2. RFC classification report

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.98	0.98	0.98	312
1	0.97	0.98	0.98	274
Accuracy			0.98	586
Macro Avg	0.98	0.98	0.98	586
Weighted Avg	0.98	0.98	0.98	586

Table 3 compares the overall performance metrics between the MLP and RFC models, including precision, recall, F1-score, and accuracy. It helps assess which model performs better in identifying semiconductor faults.

Metric	**MLP**	**Random Forest Classifier**
Precision	86.36	97.76
Recall	86.20	97.78
F1-Score	86.26	97.77
Accuracy	86.35	97.78

Table 3. Overall Performance comparison.

Table 4 compares the performance metrics for Class 0 (non-faulty devices) between the MLP and RFC models. It includes precision, recall, F1-score, and support for each model, allowing for a detailed comparison of their performance on non-faulty devices.

Table 4 Class 0 performance comparison

Metric	**MLP**	**Random Forest Classifier**
Precision	0.86	0.98
Recall	0.88	0.98
F1-Score	0.87	0.98
Support	312	312

Similar to Table 4, this table compares the performance metrics for Class 1 (faulty devices) between the MLP and RFC models. It provides insights into how well each model identifies faulty devices, based on precision, recall, F1-score, and support metrics.

Figure 6 presents the classifier performance comparison with respect to the performance metric values like accuracy, F1 score, precision, recall.

Table 5. Class1 performance comparison

Metric	MLP	Random Forest Classifier
Precision	0.86	0.97
Recall	0.84	0.98
F1-Score	0.85	0.98
Support	274	274

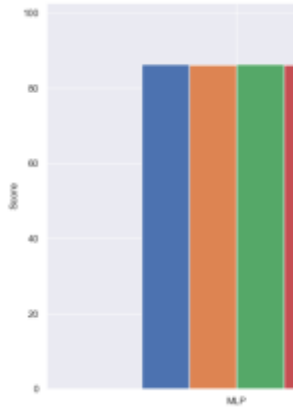

Fig 6 Performance comparison graph

5. CONCLUSION

After analyzing the performance of both the Multilayer Perceptron (MLP) and Random Forest Classifier (RFC) models for semiconductor fault identification, several key conclusions can be drawn. Firstly, the Random Forest Classifier (RFC) consistently outperforms the Multilayer Perceptron (MLP) across various performance metrics. The RFC demonstrates higher precision, recall, F1-score, and accuracy compared to the MLP, indicating its superior ability to accurately identify faults in semiconductor devices. Additionally, the classification reports and confusion matrices provide detailed insights into the performance of both models. The RFC exhibits higher precision and recall for both classes (0 and 1) compared to the MLP, indicating its effectiveness in correctly classifying both non-faulty and faulty devices. Moreover, the overall performance comparison between the MLP and RFC models reveals significant improvements in precision, recall, and F1-score when using the RFC model. These improvements highlight the superiority of the RFC model in semiconductor fault identification tasks. Furthermore, the user interface and results visualization figures offer a clear depiction of the research work's process, from data preprocessing to model evaluation. These visualizations aid in understanding the steps involved in fault identification and highlight the importance of proper data preprocessing techniques and model selection.

In conclusion, the results indicate that the Random Forest Classifier (RFC) model is the preferred choice for semiconductor fault identification from integrated circuit statistics. Its superior performance, demonstrated through higher precision, recall, and accuracy, makes it a valuable tool for enhancing the reliability and efficiency of fault identification processes in semiconductor manufacturing.

REFERENCES

1. Lee, Youjin, and Yonghan Roh. "An Expandable Yield Prediction Framework Using Explainable Artificial Intelligence for Semiconductor Manufacturing." Applied Sciences 13, no. 4 (2023): 2660.
2. Chowdhury, Hribhu. "Semiconductor Manufacturing Process Improvement

Using Data-Driven Methodologies." (2023).
3. Nuhu, Abubakar Abdussalam, Qasim Zeeshan, Babak Safaei, and Muhammad Atif Shahzad. "Machine learning-based techniques for fault diagnosis in the semiconductor manufacturing process: a comparative study." The Journal of Supercomputing 79, no. 2 (2023): 2031-2081.
4. Kao, Sheng-Xiang, and Chen-Fu Chien. "Deep Learning Based Positioning Error Fault Diagnosis of Wire Bonding Equipment and an Empirical Study for IC Packaging." IEEE Transactions on Semiconductor Manufacturing (2023).
5. Senoner, Julian, Torbjørn Netland, and Stefan Feuerriegel. "Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing." Management Science 68, no. 8 (2022): 5704-5723.
6. Gómez-Sirvent, José L., Francisco López de la Rosa, Roberto Sánchez-Reolid, Antonio Fernández-Caballero, and Rafael Morales. "Optimal feature selection for defect classification in semiconductor wafers." IEEE Transactions on Semiconductor Manufacturing 35, no. 2 (2022): 324-331.
7. Xu, Qiuhao, Chuqiao Xu, and Junliang Wang. "Forecasting the yield of wafer by using improved genetic algorithm, high dimensional alternating feature selection and SVM with uneven distribution and high-dimensional data." Autonomous Intelligent Systems 2, no. 1 (2022): 24.
8. Fan, Shu-Kai S., Wei-Kai Lin, and Chih-Hung Jen. "Data-driven optimization of accessory combinations for final testing processes in semiconductor manufacturing." Journal of Manufacturing Systems 63 (2022): 275-287.