

## ABSTRACT

As semiconductor manufacturing advances toward sub-7nm process nodes, the complexity of wafer fabrication has surpassed the capabilities of traditional automated inspection systems. While modern Deep Learning models achieve high classification accuracy, they often operate as “black boxes,” lacking the semantic interpretability required for industrial root-cause analysis.

This research investigates wafer defect detection through comparative evaluation of convolutional neural networks (CNNs), vision transformers (ViTs), and hybrid CNN-ViT architectures using the WM-811K dataset. Preliminary experimental results demonstrate that hybrid architectures provide a superior balance between local feature extraction and global spatial reasoning compared to standalone models.

Building on these findings, this dissertation proposes HyViT-TM, a multi-scale fusion framework that integrates a ResNet-50 convolutional backbone with a Data-efficient Image Transformer (DeiT) to capture both localized defect textures and wafer-scale spatial dependencies. To address the interpretability gap, the framework incorporates a Text Mapping (TM) module based on the Aya Vision 8B multimodal large language model, which converts visual defect evidence into engineering-oriented diagnostic narratives.

A key contribution of this research is the emulation of expert diagnostic reasoning through the generation of simulated technical narratives. By aligning Grad-CAM++ saliency heatmaps with process-specific linguistic embeddings, the TM module translates visual signals into structured engineering hypotheses. These narratives provide a computational proof-of-concept for AI-assisted root-cause analysis by contextualizing spatial defect patterns with plausible semiconductor process mechanisms, such as thin-film non-uniformity or process imbalance.

Experimental evaluation demonstrates that the HyViT-TM framework achieves 97.18% classification accuracy and a macro F1-score of 0.86, outperforming standalone CNN and transformer baselines. The generated textual narratives further provide interpretable explanations aligned with semiconductor process knowledge, enabling the transformation of visual inspection outputs into actionable diagnostic insights. This work establishes a scalable pathway for integrating interpretable artificial intelligence into semiconductor manufacturing environments, bridging the gap between automated defect detection and expert-level engineering analysis.