IESA DeepTech Hackathon

Team Details

Team Name: *SilicoLattice*

SR. NO	ROLE	NAME	ACADEMIC YEAR
1	**Team Leader**	*Jishu Mapdar*	*3rd*
2	Member 1	*Debabrata Das*	*3rd*
3	Member 2	*Malay Roy*	*3rd*

COLLEGE NAME

National Institute of Technology Durgapur

TEAM LEADER CONTACT NUMBER

+91 7363964326

TEAM LEADER EMAIL ADDRESS

jishumapdar.nitdgp@gmail.com

Problem Statement Addressed

Edge-AI Defect Classification for Semiconductor Images and our idea addresses

With the Scaling of technology node, billions of wafers are produced each year, generating large volumes of inspection images from SEM and other tools, making manual analysis slow, expensive, and error-prone.

Wafer Defects vary in nature like "Opens, Bridges, CMP scratches, cracks, LER, Incomplete Etch, *etc" making pass-fail inspection insufficient for practical manufactures.*

In industrial practice, defect categorization is important because different defect types represent different manufacture/process issue and require different handling strategies.

Without proper classification, usable or partially defect wafers may be separately categorized and used for different productions . Example: Wafers are categorized by "Performance Binning", where chips with minor defects (like LER) are sold as lower-tier SKUs (e.g., i5 vs. i9) instead of being discarded, while higher-risk dies are diverted to less-demanding consumer markets to maximize total yield.

As the inspection system faces latency, bandwidth and scalability issues, a properly introduced Defect Classification system can help at this inspection stage and help in continuous real-time monitoring.

Idea – SEM Image classification using Multibranch Model

KEY CONCEPT & APPROACH

Scalable edge inspection: SEM images are processed locally using shared backbone, enabling system to scale large wafer images.
Shared Feature Extraction: MobileNetV3 backbone runs once per image, ensuring scalable reuse across defect branches.
Early Filtering and Energy Efficiency: A Classifier(1) is used to separate Other images, Normal and Defective images early, using a particular **Threshold Confidence Level**, below this level, every sample is detected as "**Other category**" preventing unnecessary downstream computation.
Hierarchical Defect Grouping: From separated Defective samples, computation is done using our Model to check defect types Based on groups, improving accuracy while keeping model bounded.
Energy-efficient: Since only the required classifier branch is activated per image, minimizing MAC operation and power consumption.

SOLUTION OVERVIEW

Scalability: SEM images are normalized to *1* × 224 × 224 grayscale, enabling high-volume inspection without cloud bandwidth limits.
Latency and Cost Reduced: As lightweight MobileNetV3 performs feature extraction on edge hardware, expensive and slow manual process is unnecessary.
Unnecessary Processing Avoided: Classifier(1) filters Normal and Other images early, reduces computation load and enables energy efficient monitoring.
Defect Diversity: Classifier(2) groups similar defects, supporting precise defect categorization, matching real manufacture handling strategies.
Modular Branch Design: Each defect Group is handled by independent binary classifier, allowing new defect Classes to be added without disrupting existing branches.

Proposed Solution – Implement the solution in NXP FRDM i.MX 93

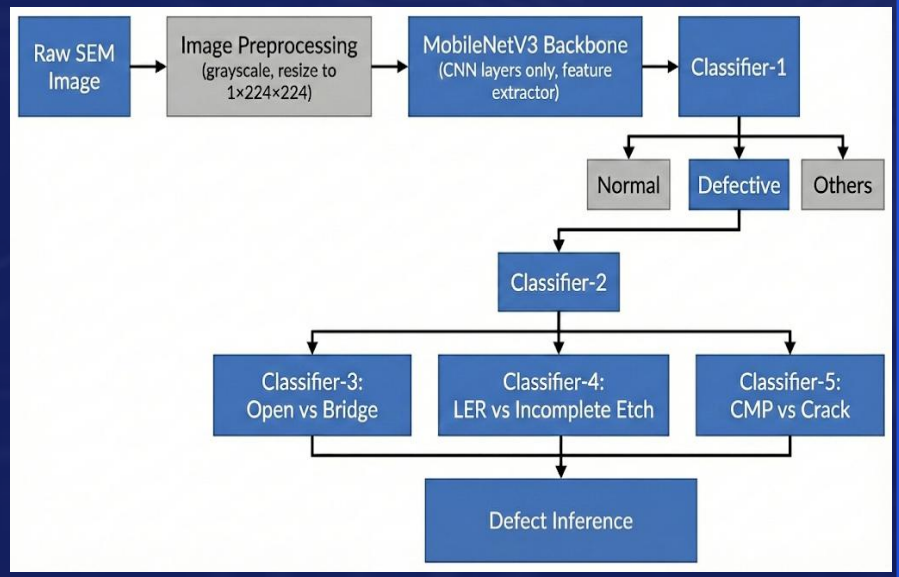
Technology Stack and Feasibility

Input and Data Handling : SEM Wafer Images normalized to 1 × 224 × 224 for uniform Processing

Board Implementation : Trained Model exported to ONNX and then to TensorFlow Lite format using the NXP eIQ ML toolkit, model is integrated into Yocto-based Linux Environment on NXP i.MX 93 board. TensorFlow Lite run time is used to execute the model using board's Processor and NEON optimized acceleration support, making model operations faster. Then the model is transferred to board via peripherals and can be used for the defect inference.

Feasibility : TensorFlow Lite, ONNX, NXP eIQ SDK are existing tools with proper workflow and hardware support available for model conversion and deployment on i.MX 93 board . The system itself is feasible as it uses edge-optimized CNN with H-Swish and RELU functions to minimize no. of MAC (very resource and energy intensive) Operations.

Model Architecture



Hardware Components

Development Tools

Dataset Plan & Class Design

Total images planned/current: 2400

Image type: Grayscale

Train/Val/Test split: 80/10/10

Labelling method/source: public dataset and some generated **images.**

No. of classes: 7 + others(if Confidence Low)

Class list:
Opens, Bridges, CMP scratches, cracks, LER, Incomplete Etch

Class balance plan: The dataset establishes symmetry between 1,200 normal and 1,200 defect images to ensure unbiased binary classification, while 150 Other images teach the model to ignore non-critical SEM artifacts. By partitioning defects into three morphological groups and then six specific classes of 200 images each, the plan prevents over-fitting to any single failure mode. This hierarchical structure maintains a high sensitivity for granular defects while ensuring the model learns a robust definition of structural integrity.

Baseline Model & Results(Phase 1)

Technology & Feasibility/Methodology Used

Model details

Architecture: MobileNetV3-Small

Training approach: transfer learning

Input size: *1* × 224 × 224

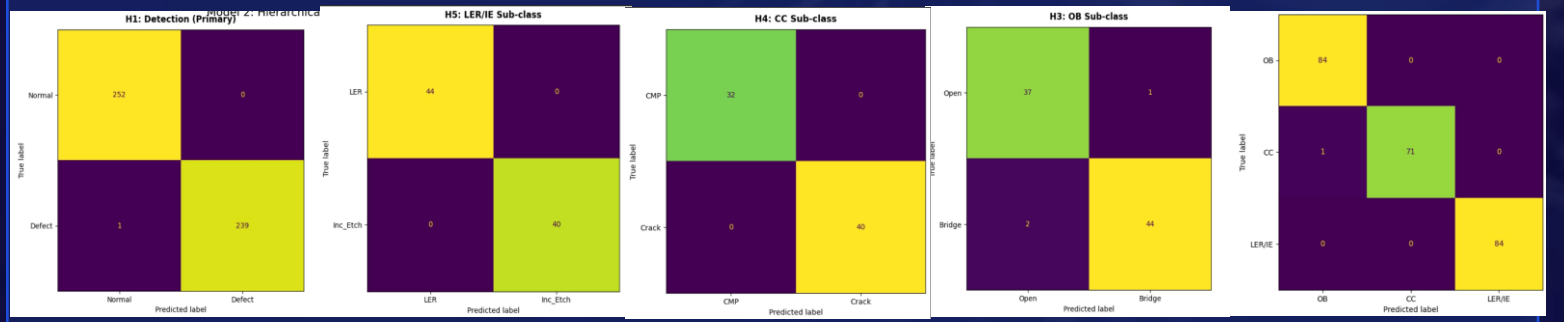
Model size: 1.83 MB

Framework: PyTorch

Metrices

Accuracy : Classifier1(H1): 94.33, Classifier2(H2)-98.88, Classifier3(H3)-97.65, Classifier4(H4)-100, Classifier5(H5)-100.

Precision/Recall: Classifier1: 0.89/1.00, Classifier2-1.00/0.895, Classifier3-0.96/1.00, Classifier4-1.00/0.97, Classifier5-1.00/1.00,



Confusion Matrix

Artifacts & Links for Research and References

GitHub Repository

GitHub Repository

ONNX Model Link

ONNX.com

Dataset Link:

Dataset_train.file

Research Background & Methodology

Our study utilizes raw SEM images from fabrication labs to identify pattern-dependent defects, such as dishing and erosion caused by non-uniform CMP. To ensure high-tier classification, marginal wafers are categorized via performance binning—similar to commercial processor grading—rather than being discarded. We expanded this dataset through targeted augmentation and employed transfer learning to improved detection accuracy.

References & Citations

{Ref 1: Deep-learning-based-defect-classification-and-detection-in-sem-images}

{Ref 2: https://researchdata.ntu.edu.sg/dataset.xhtm}