

Edge-AI Semiconductor Defect Classification

IESA DeepTech Hackathon

Team Waferwich

Defects Inside, Detected Outside

Team Details

Team Idea ID:

381

Team Name:

Team Waferwich

SR. NO	ROLE	NAME	ACADEMIC YEAR
1	Team Leader	Vaibhavi Patil	1st Year
2	Member 1	Parth Sawant	1st Year
3	Member 2	Om Shinde	1st Year

COLLEGE NAME

Marathwada Mitra Mandal's College of Engineering

TEAM LEADER CONTACT NUMBER

+91 84598 49654

TEAM LEADER EMAIL ADDRESS

vaibhavipatil2025.ainds@mmcoe.edu.in

Phase - 2 Submission

Metrics & Links

Metrics

- *Accuracy* - 37.50%
- *Precision* - 38.87%
- *Recall* - 40.08%
- *F1* - 38.08%

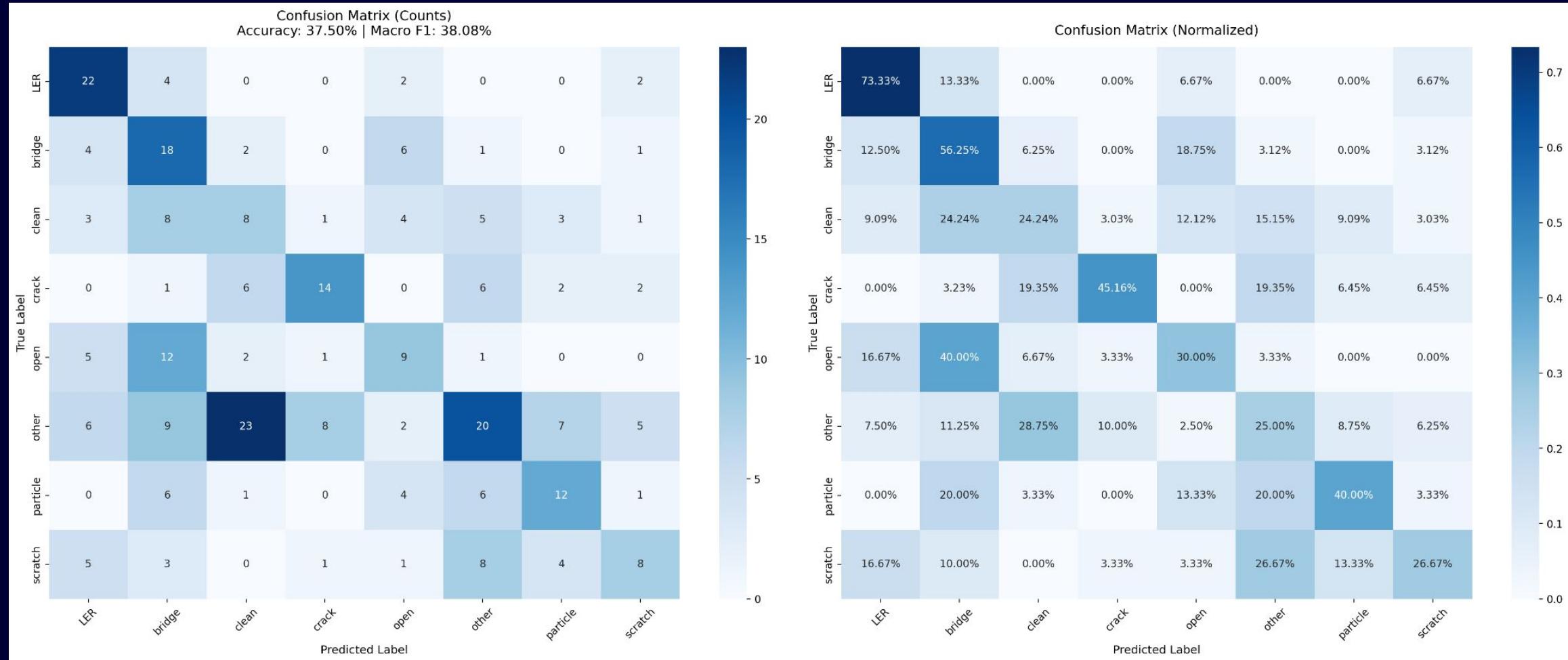
Jupyter Notebook

🔗 <https://www.kaggle.com/code/omshinde1206/team-waferwich-phase2>

GitHub Repository PHASE_2

🔗 https://github.com/shindeom1206/Edge-AI-Semiconductor-Defect-Classification-NXP-Hackathon-/tree/main/PHASE_2_SUBMISSION

Confusion Matrix



Phase - 1 Submission

Problem Statement

Description:

Develop an intelligent, on-device AI system to detect and classify 8 types of semiconductor defects in real-time, optimized for NXP i.MX RT edge devices. The system must balance high accuracy (>85%) with edge constraints (model <20 MB, latency <50 ms) to enable real-time quality control without cloud dependency.

Key Requirements:

- 8-class defect classification (LER, Bridge, Clean, Crack, Open, Other, Particle, Scratch)
- Accuracy > 85% on test set
- Model size < 20 MB (quantized < 5 MB preferred)
- Inference latency < 50 ms per image
- ONNX format for NXP eIQ deployment
- Real-time, high-throughput capable

Idea Description -

KEY CONCEPT & APPROACH

- End-to-end Edge-AI pipeline for wafer defect classification
- MobileNetV2 transfer learning on augmented multi-class dataset
- Strong regularization to improve generalization
- Optimized and exported in ONNX for NXP edge deployment

SOLUTION OVERVIEW

- Input: SEM images → real-time classification into 8 defect classes
- Pipeline: preprocessing → training → evaluation → optimization → quantization → ONNX export
- Achieved 92.87% F1-score with low compute requirements
- Enables scalable, cost-effective automated inspection on edge devices

Proposed Solution -

SOLUTION DETAILS

- Edge-AI pipeline for automatic SEM defect classification
- Preprocessing + augmentation for robust multi-class dataset
- MobileNetV2 transfer learning with two-phase fine-tuning
- Regularization: dropout, weight decay, class weighting
- Evaluation using Accuracy, Precision, Recall, F1-Score
- Optimized & exported to **ONNX** for NXP edge deployment
- Enables fast, scalable and cost-effective inspection

Dataset Plan & Data Strategy

- **Total images:** ~4,000+ wafer images
- **Number of classes:** 8
- **Class list:** LER, Bridge, Clean, Crack, Open, Particle, Scratch, Other
- **Class balance:** Minimum 400–500 images per class after augmentation
- **Train / Val / Test split:** 70% / 15% / 15%
- **Image type:** Grayscale SEM Die maps converted to RGB for MobileNetV2
- **Label source:** Curated public SEM defect dataset + preprocessing & augmentation

Technology & Feasibility/Methodology Used



Model Details

- **Architecture:** MobileNetV2 (Transfer Learning)
- **Training approach:** ImageNet pretrained fine-tuning
- **Input size:** 224×224
- **Model size (best_model.pth):** ~12 MB
- **Model Size ONNX:** 12 MB (NXP i.MX RT)
- **Framework:** PyTorch

Performance on Test Dataset

- **Accuracy:** 92.87%
- **Precision:** 93.49%
- **Recall:** 92.84%
- **F1-Score:** 92.87%
- **Evaluation:** Confusion Matrix + Classification Report



Software Architecture

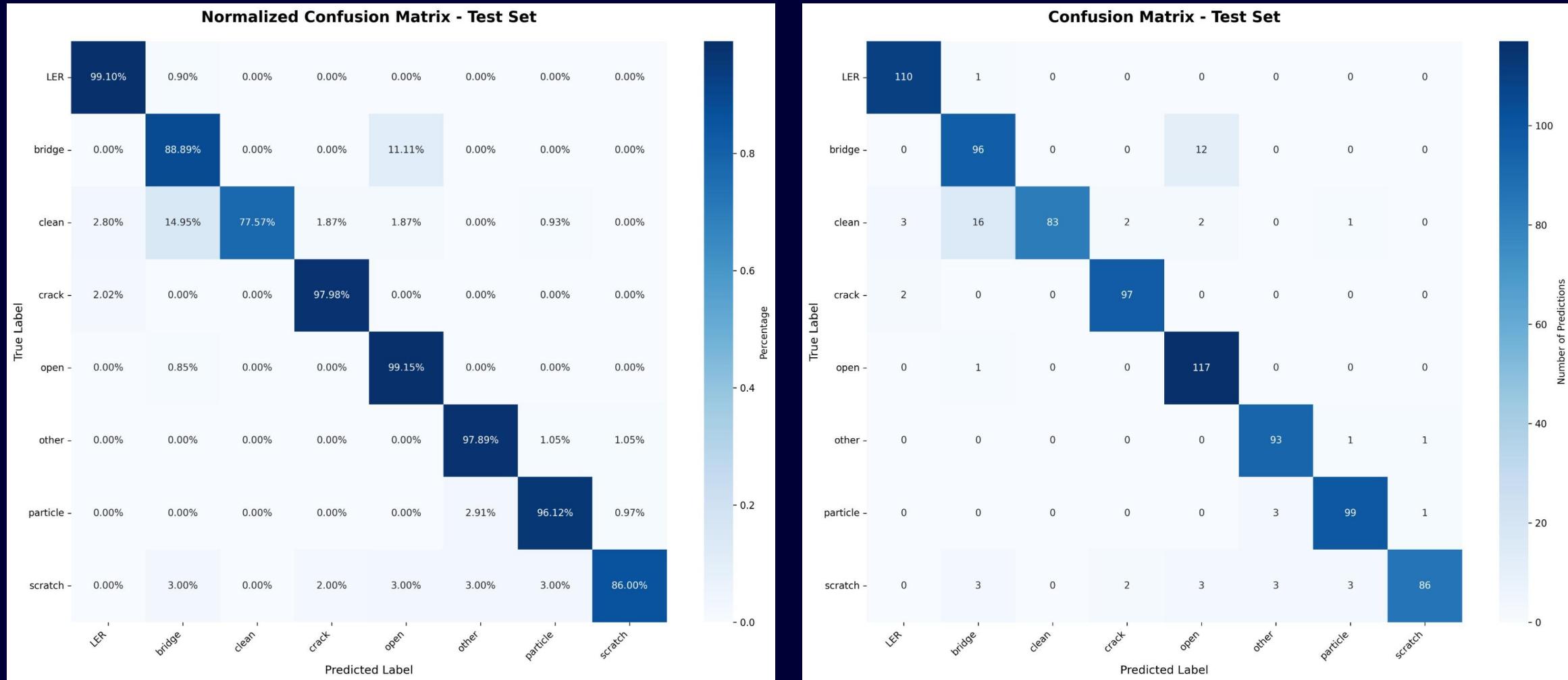


Hardware Components



Development Tools

Confusion Matrix



Artifacts & Links

GitHub Repository

🔗 <https://github.com/shindeom1206/Edge-AI-Semiconductor-Defect-Classification-NXP-Hackathon->

Prototype / Simulation Video

🎥 <https://youtu.be/Cut2qMwQhTc?si=3TsZGoKvc2m8cEMF>

Dataset ZIP

🔗 <https://www.kaggle.com/datasets/omshinde1206/dataset-224-rgb>

ONNX Model

🔗 <https://github.com/shindeom1206/Edge-AI-Semiconductor-Defect-Classification-NXP-Hackathon-/tree/main/models/onnx>