Title

Edge-AI Defect Classification for Semiconductor Images

Team Details

Team Idea ID: 994

Team Name: FabVision

SR. NO	ROLE	NAME	ACADEMIC YEAR
1	**Team Leader**	*Senthil Kumar Mahalingam*	*2024-2028*
2	Member 1	*Deepak Prabakaran*	*2024-2028*
3	Member 2	*Tharun G*	*2024-2028*
4	Member 3	*M KarthickRaja*	*2024-2028*

A team can have up to 4 members including the team leader. Add rows if necessary.

COLLEGE NAME

CHENNAI INSTITUTE OF TECHNOLOGY

TEAM LEADER CONTACT NUMBER

+91 9843111656

TEAM LEADER EMAIL ADDRESS

senthilkumarm.vlsi2024@citchennai.net

Problem Statement Addressed

Edge-AI Defect Classification for Semiconductor Images

DESCRIPTION / DETAILS

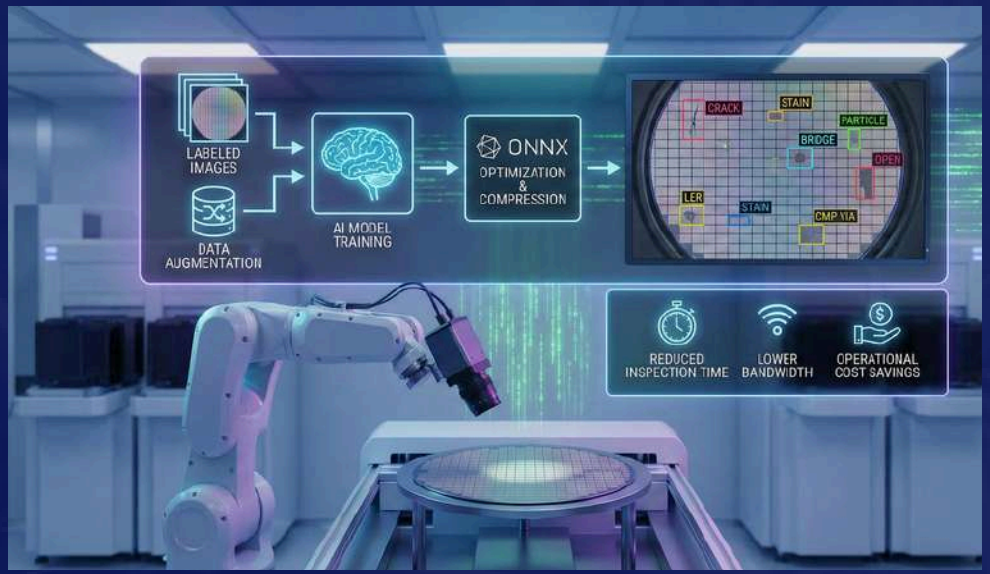
- *Semiconductor manufacturing needs accurate defect detection to avoid yield loss and device failure.*
- *Manual and cloud-based inspection methods are slow and add latency.*
- *This project uses Edge-AI to automatically classify defects from semiconductor images.*
- *A lightweight deep learning model runs directly on low-power edge devices.*
- *Enables real-time, fast, and reliable quality inspection.*

IESA
INDIA ELECTRONICS & SEMICONDUCTOR ASSOCIATION
India's ESDM Leadership and Knowledge Partner

Idea Description

KEY CONCEPT & APPROACH

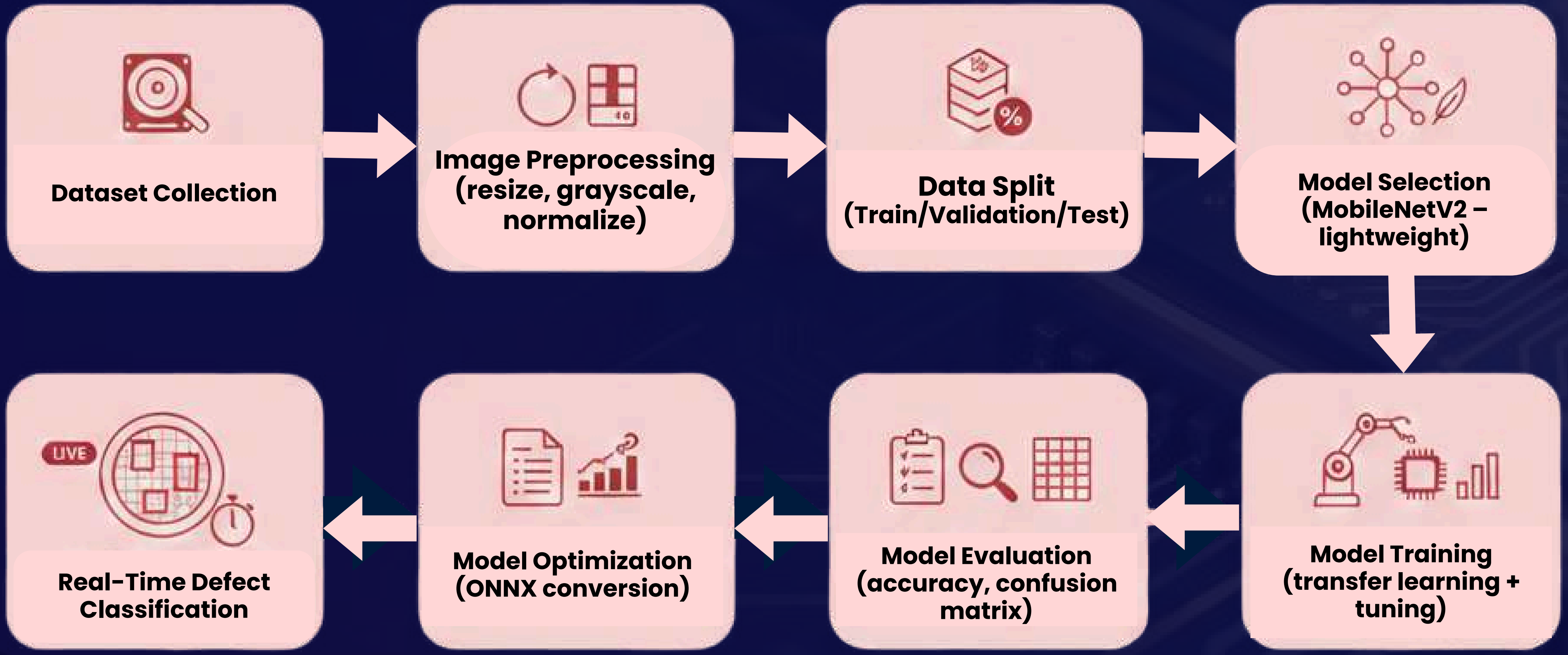
- *Edge-AI system for automated wafer defect classification*
- *Designed for real-time inspection environments*
- *Focus on low-latency & low-power execution*
- *Reduces manual review and cloud dependency*
- *Maintains high accuracy with compact models*



SOLUTION OVERVIEW

- *Train the model using labeled wafer and die images with data augmentation.*
- *Optimize the trained model using ONNX conversion and compression techniques.*
- *Classify multiple defect types such as cracks, stains, particles, bridge ,open ,LER, CMP vias.*
- *Reduce inspection time, bandwidth usage, and operational cost.*

Proposed Solution



SOLUTION DETAILS

- *A lightweight MobileNet-based deep learning model is trained on preprocessed semiconductor defect images using transfer learning, evaluated for accuracy, and optimized into an edge-compatible format (ONNX) to enable fast, real-time defect classification on low-resource devices.*

Innovation and Uniqueness

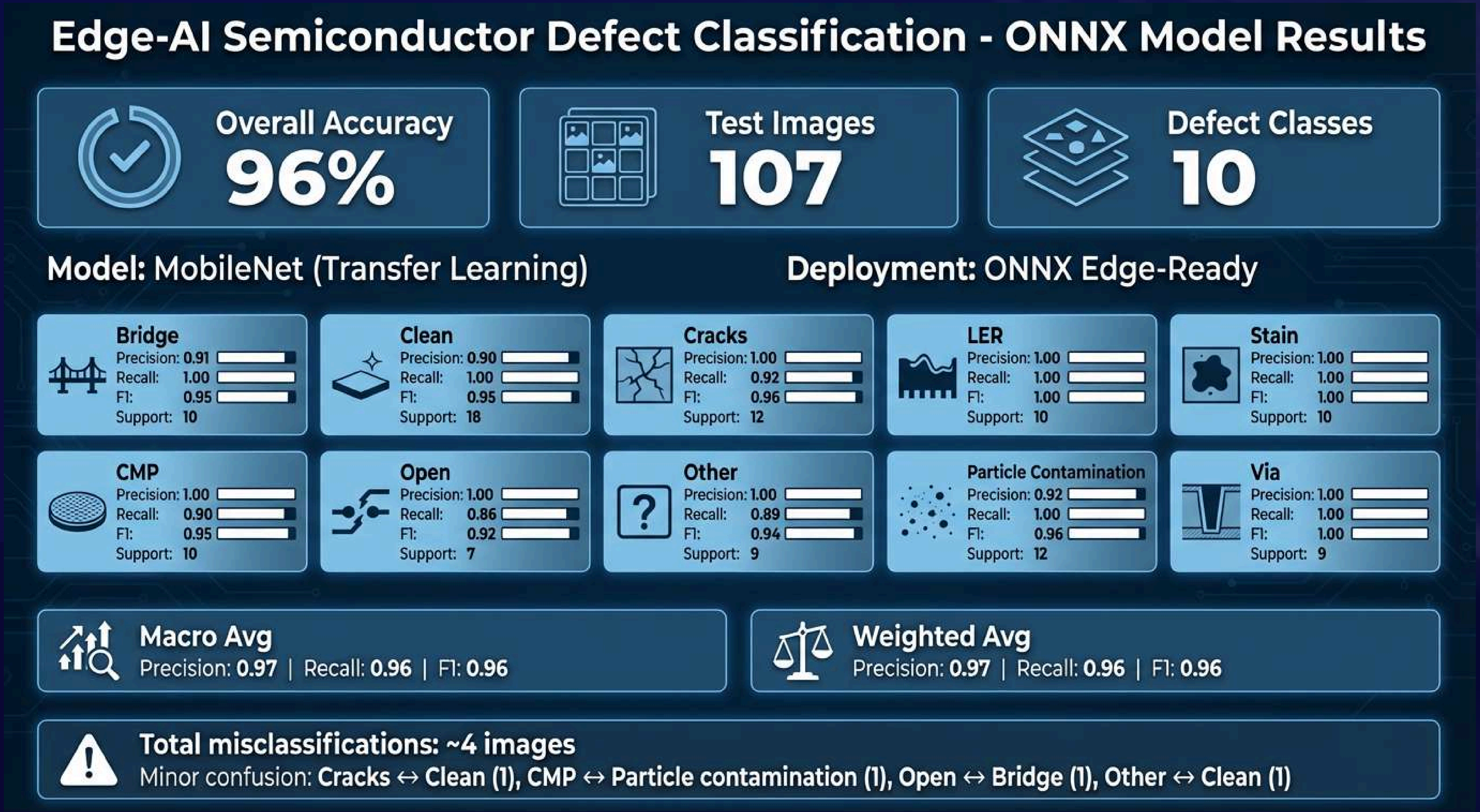
KEY INNOVATION

- *Edge-AI based semiconductor defect classification using lightweight MobileNet models*
- *Transfer learning + model optimization for low-size, high-accuracy performance*
- *Edge-ready model conversion ONNX built into the pipeline from the start*
- *Designed specifically for wafer/defect image categories, not generic image tasks*

COMPETITIVE ADVANTAGE

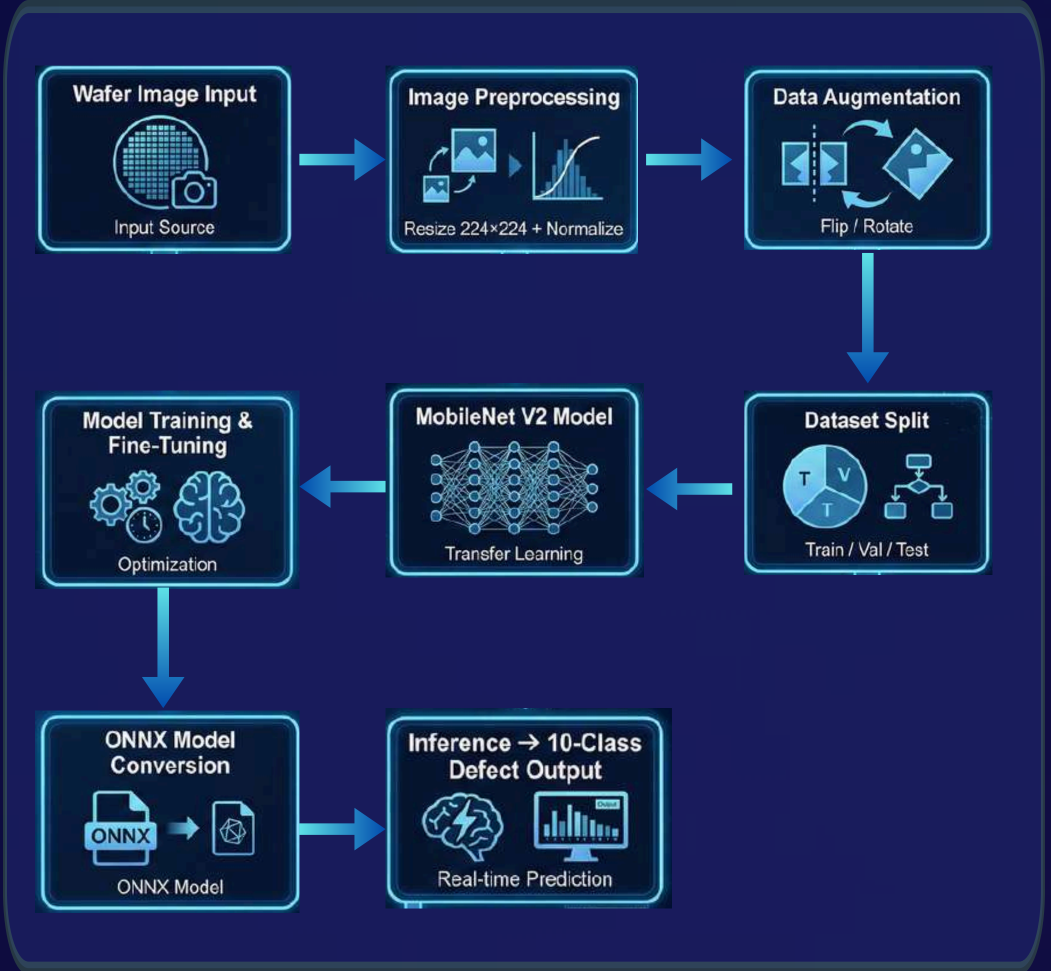
- *Faster inference with low compute compared to heavy CNN/cloud models*
- *Lower deployment cost — runs on low-resource edge hardware*
- *No cloud dependency → reduced latency and better data privacy*
- *Scalable and easy to integrate into real-time inspection workflows*
- *High accuracy with small model size for practical industrial use*

Impact and Benefits



IESA INDIA ELECTRONICS & SEMICONDUCTOR ASSOCIATION
India's ESDM Leadership and Knowledge Partner

Technology & Feasibility/Methodology Used



- Python
- TensorFlow / Keras
- ONNX Runtime
- OpenCV
- NumPy / Scikit-learn
- matplotlib
- GitHub
- ubuntu

Software Architecture

Hardware Components

Development Tools

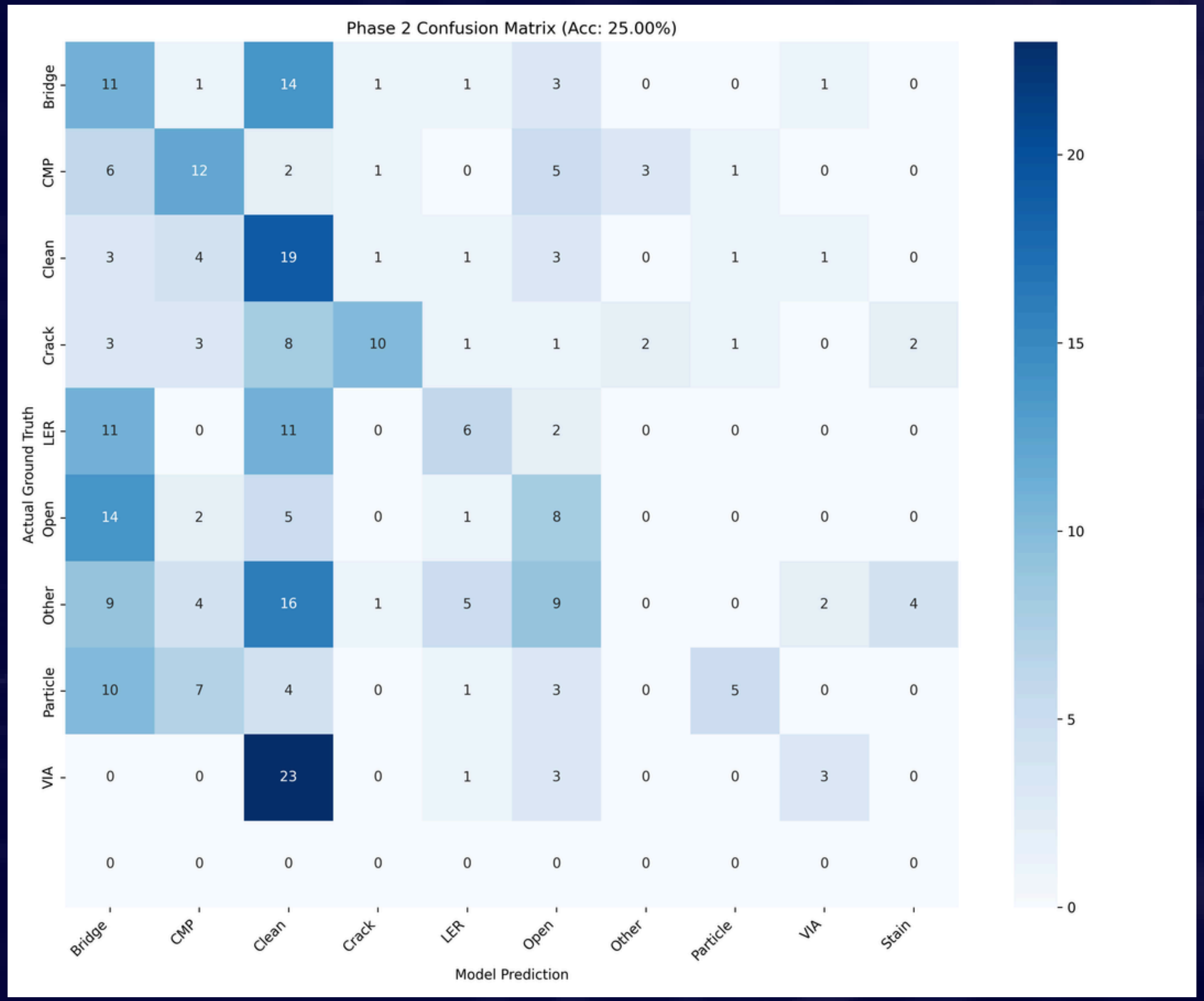
Phase2(Hackathon Dataset Evaluation):

Metrics:

Arguments: ()
📋 Detailed Classification Report:

	precision	recall	f1-score	support
Bridge	0.16	0.34	0.22	32
CMP	0.36	0.40	0.38	30
Clean	0.19	0.58	0.28	33
Crack	0.71	0.32	0.44	31
LER	0.35	0.20	0.26	30
Open	0.22	0.27	0.24	30
Other	0.00	0.00	0.00	50
Particle	0.62	0.17	0.26	30
VIA	0.43	0.10	0.16	30
micro avg	0.26	0.25	0.25	296
macro avg	0.34	0.26	0.25	296
weighted avg	0.31	0.25	0.23	296

Confusion Matrix:



Note:
Some images labeled as "Other" in the test set were part of defect categories in training.
This creates a label inconsistency affecting evaluation metrics.

GitHub & Dataset Folder

GitHub Repository

https://github.com/Senthil7271/edge-ai-wafer-defect-classifier

Dataset Folder

https://drive.google.com/drive/folders/1nrMQLJqcVQ7cgJZH3JaNslrel6j5rm_e?usp=sharing

Research and References

Research Background & Methodology

- *Semiconductor wafer defects such as cracks, LER, particle contamination, stains, vias, and CMP show distinct visual patterns that can be automatically learned using CNN-based models.*
- *The system uses MobileNetV2 with transfer learning to achieve high accuracy with low computational requirements.*
- *The trained model is optimized and compressed (ONNX) to enable real-time edge deployment with low latency.*
- *This approach enables faster automated inspection, improving early defect detection and overall manufacturing yield.*

References & Citations

- Semiconductor Wafer Defect Detection using Deep Learning – https://primerascientific.com/pdf/psen/PSEN-04-097.pdf
- Inspection and Classification of Semiconductor Wafer Surface Defects Using CNN Deep Learning Networks - https://www.mdpi.com/2076-3417/10/15/5340
- "Wafer Map Failure Pattern Classification Using Geometric Transformation-Invariant CNN - https://primerascientific.com/pdf/psen/PSEN-04-097.pdf