Team Details

Team Name: **NEXUS**

SR. NO	ROLE	NAME	ACADEMIC YEAR
1	**Team Leader**	Annie Darling Kanmani A	3rd Year
2	Member 1	Mridulaa Kishore	3rd Year

COLLEGE NAME

SRM Madurai College for Engineering and Technology

TEAM LEADER CONTACT NUMBER

+91 7845242804

TEAM LEADER EMAIL ADDRESS

anniedarling60@gmail.com

GitHub Repository

https://github.com/kanmaniannie180-hub/edge-ai-wafer-defect-classification/tree/main

Prototype / Simulation Video

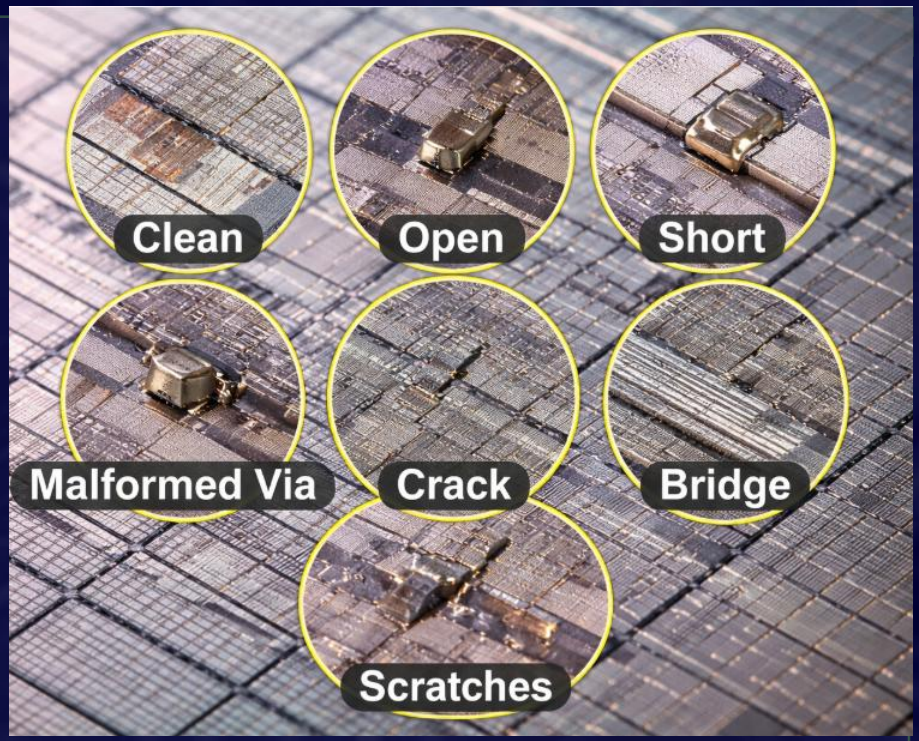
https://drive.google.com/file/d/1pWAH54snOZnXsOuz2sV-4Bh3qW7buV9A/view?usp=sharing

Problem Statement Addressed

Selected the problem statement your idea addresses

DESCRIPTION / DETAILS

- Semiconductor wafer and die defects reduce manufacturing yield and reliability
- Manual inspection is time-consuming and prone to human error
- Cloud-based AI solutions introduce latency and power overhead
- Real-time defect detection requires low-power, compact AI models
- Edge AI enables fast and efficient on-device inspection



Example of wafer surface defects during inspection

Idea Description - Edge-AI Based Semiconductor Wafer Defect Classification

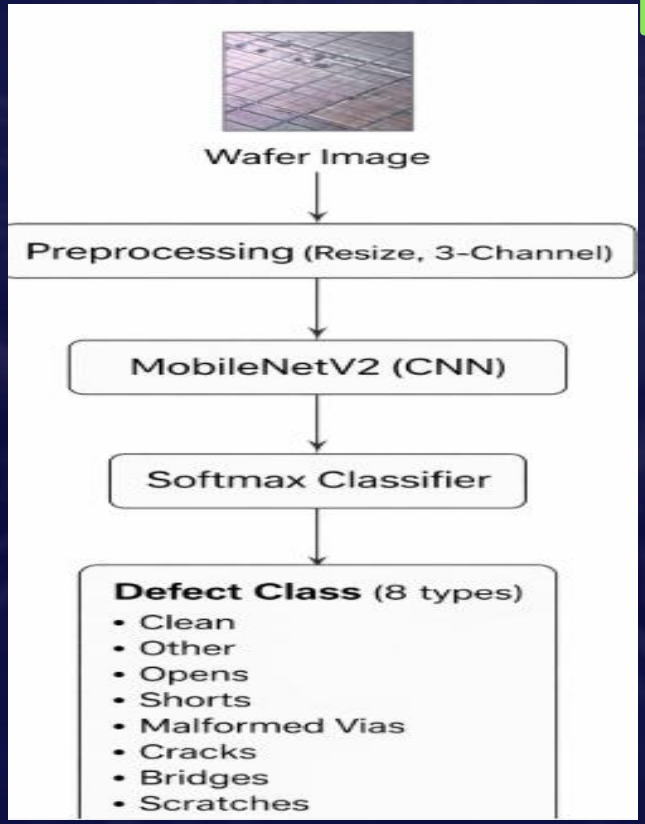
Provide a brief summary of your idea, including the key concept and approach. Provide a brief overview of your solution and how it addresses the problem statement.

KEY CONCEPT & APPROACH

- Develop an AI-based image classification system to automatically identify semiconductor wafer defects
- Use a lightweight convolutional neural network optimized for edge deployment
- Apply transfer learning to achieve reliable performance with limited labeled data
- Focus on model portability and compact size to enable real-time, low-power inference

SOLUTION OVERVIEW

- Automates wafer defect detection, reducing reliance on manual and error-prone inspection processes
- Uses a compact deep learning model to meet low-power and real-time edge constraints
- Enables on-device inference, eliminating latency and dependency on cloud connectivity , internet connectivity.
- Provides scalable and consistent defect classification across multiple defect types

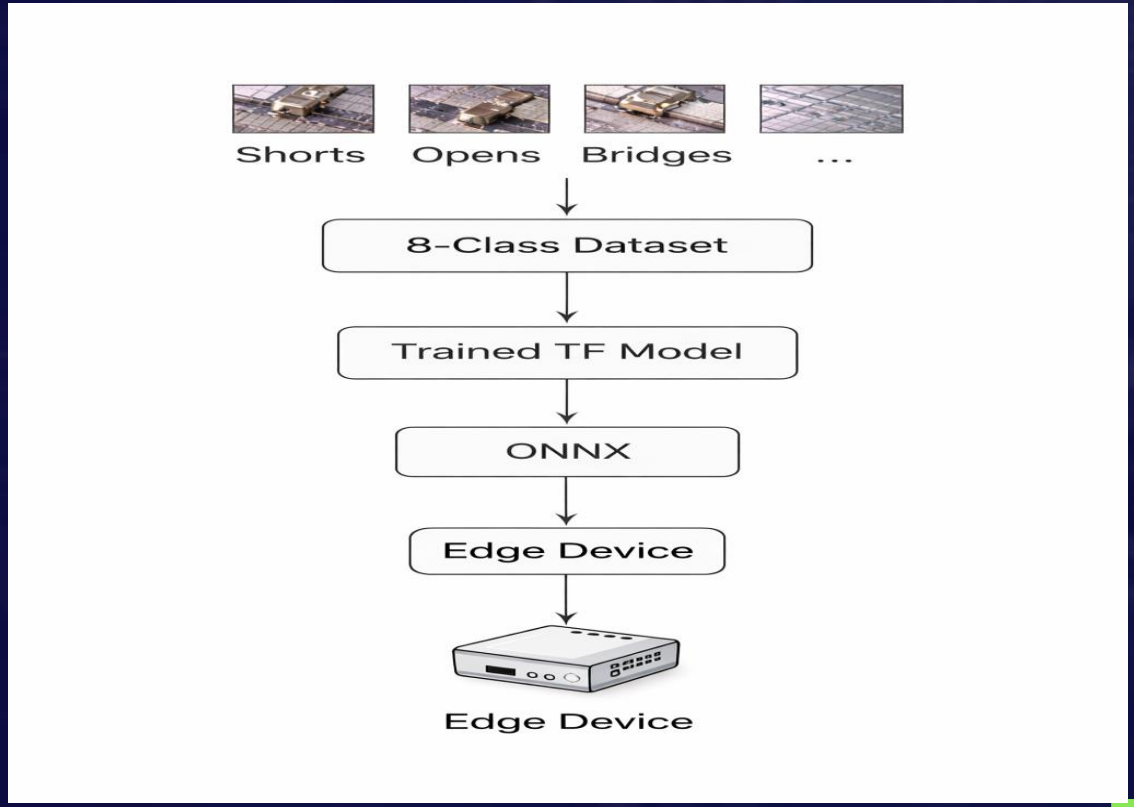


Proposed Solution - Edge-AI Based Semiconductor Wafer Defect Classification

Describe your idea in detail. Include the methodology, technologies involved, and how it addresses the chosen problem statement.

SOLUTION DETAILS

- The proposed solution uses a convolutional neural network to classify semiconductor wafer defects from inspection images
- A lightweight MobileNetV2 architecture is employed to meet edge deployment constraints
- Transfer learning with ImageNet pretraining enables stable training on limited defect data
- Images are resized to 224×224 and converted to 3-channel input for model compatibility
- The trained model is exported to ONNX format to ensure portability across edge platforms
- The solution is designed for real-time, low-power inference on embedded devices.



Innovation and Uniqueness

Highlight what makes your idea unique or innovative compared to existing solutions.

KEY INNOVATION

- The innovation is a **novel application of lightweight edge AI** for real-time semiconductor wafer defect classification
- The solution enables **fully offline, on-device inference without internet connectivity**, eliminating cloud dependency
- An **edge-first, portable workflow using ONNX** allows deployment on low-power embedded platforms

COMPETITIVE ADVANTAGE

- Enables **real-time, offline defect detection**, removing cloud latency and recurring costs
- Uses a **lightweight, compact model** that lowers power and hardware requirements for edge deployment
- Provides **automated, consistent classification**, improving efficiency over manual or rule-based inspection

Impact and Benefits

Explain how your solution will make an impact, such as improving performance, reducing costs, increasing efficiency, or solving other challenges.

- Enables **real-time, offline wafer defect detection**, improving inspection speed and manufacturing yield
- Reduces **inspection cost and power consumption** by eliminating cloud dependency and heavy hardware
- Provides **consistent and reliable defect classification**, minimizing human error in quality control

- **2× faster inspection**
- **Reduced inspection latency**
- **Lower operational cost**
- **Improved inspection consistency**
- **Compact model size (< XX MB)**

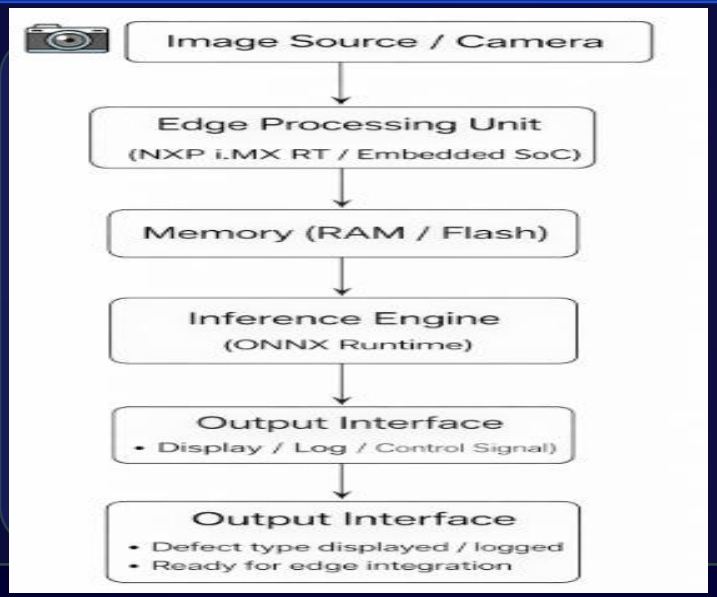
Technology & Feasibility/Methodology Used

Describe the technologies, methodologies, or tools you plan to use to implement your idea.

IMPLEMENTATION STRATEGY

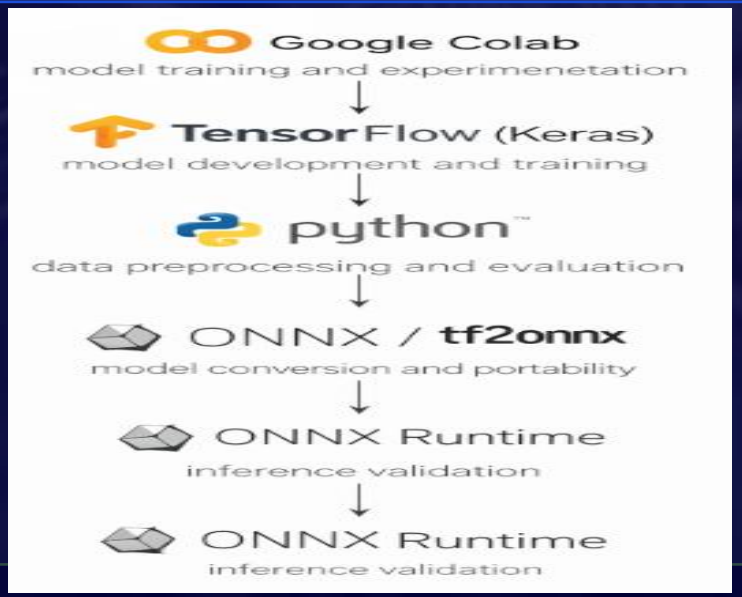
Technical Stack & Feasibility
Software: TensorFlow (Keras), Python, ONNX, ONNX Runtime
Architecture: MobileNetV2-based CNN with transfer learning and 224×224 inputs
Deployment: ONNX export for offline, edge-ready inference
Feasibility: Lightweight model ensures low power, small memory footprint, and easy edge deployment

Software Architecture



Hardware Components

Development Tools



Research and References

Research Background & Methodology

Briefly describe the research foundation or scientific principles supporting your idea.

Transfer Learning Baseline: Used a pretrained MobileNetV2 (ImageNet) with a frozen backbone to establish a stable and reproducible Phase-1 baseline suitable for edge deployment.
Dataset Strategy: Eight defect classes organized using folder-based labeling with fixed train/validation/test splits to ensure consistent and unbiased evaluation.
Training Discipline: Trained for a maximum of 25 epochs using Adam (LR = 1e-4) with early stopping; data augmentation applied only to the training set.
Evaluation Approach: Performance assessed using overall accuracy, per-class precision and recall, and a confusion matrix to analyze defect-wise behavior.
Key Insight: Strong performance on well-defined classes (e.g., clean, cracks), while lower recall on rare or ambiguous defects highlights data limitations rather than model failure, guiding Phase-2 improvements.

References & Citations

List key papers, articles, or data sources.

•**Wafer Map Failure Pattern Recognition Using Deep Learning**
Source: IEEE
URL: https://ieeexplore.ieee.org/document/8113459
•**Deep Learning for Semiconductor Defect Classification: A Review**
Source: Springer
URL: https://link.springer.com/article/10.1007/s10489-020-01744-0
•**WM-811K Wafer Map Dataset**
Source: Kaggle
URL: https://www.kaggle.com/datasets/qingyi/wm811k-wafer-map