

Nano Sentinel: An Explainable Deep Learning Framework for Semiconductor Wafer Defect Classification and Inspection Workflow Support

Prof. Vetrikani R

Assistant Professor

East Point College of Engineering and
Technology
Bengaluru, India

vetrikaniir.eee@eastpoint.ac.in

Pranav Aditya R

1EP23CI039

East Point College of Engineering and
Technology
Bengaluru, India

pranava663@gmail.com

Ranjith Kumar K.M.

1EP23CI048

East Point College of Engineering and
Technology
Bengaluru, India

ranjithkumarpw@gmail.com

Rahul Gowda C

1EP23CI043

East Point College of Engineering and
Technology
Bengaluru, India

rahulgowdac05@gmail.com

Sathvik B Rao

1EP23CI053

East Point College of Engineering and
Technology
Bengaluru, India

sathvikbrao131@gmail.com

Vardhan Kumar R

1EP23CI061

East Point College of Engineering and
Technology
Bengaluru, India

rv349403@gmail.com

Abstract—The wafer-map defect pattern classification is an important task in semiconductor manufacturing, as the spatial arrangement of defective dies can provide valuable information for identifying recurring failure patterns. The manual inspection of wafer maps can be extremely labor-intensive, especially if large volumes of inspection data have to be reviewed. Deep-learning methods provide a way to automate the pattern classification, but the classification predictions are only a part of an inspection workflow and do not present evidence for individual predictions. This paper introduces Nano Sentinel, a prototype framework that integrates deep-learning-based wafer-map classification with explainability and inspection workflow elements. The classification component uses MobileNetV2 transfer learning to classify wafer maps from the WM-811K dataset into nine Center, Donut, Edge-Loc, Edge-Ring, Loc, Near-full, Random, Scratch, and none classes. The framework incorporates Gradient-weighted Class Activation Mapping for model visual inspection, maximum softmax probability for confidence reporting, a provisional defect-severity assessment, rule-based inspection recommendations, and inspection record generation. A lot-aware partition was used for training and validation of the model. Preliminary experiments with the selected V2 model achieved 94.56% validation accuracy and 0.8332 macro-averaged F1-score. An independent diagnostic evaluation on the official test partition returned 88.66% accuracy and 0.4727 macro-F1, which suggests significant variance between the performance of different classes and highlights the need for further generalization studies. The workflow components were demonstrated as prototype decision support features, rather than a fully validated industrial inspection system. The study emphasized the combination of classification, visualization, and documentation elements while class imbalance, minority-class mistakes, and the validation of severity and rule-based decisions are important areas of future work.

Index Terms

Semiconductor manufacturing, wafer-map defect classification, MobileNetV2, transfer learning, explainable artificial intelligence, Grad-CAM, deep learning, automated inspection, inspection workflow.

I. INTRODUCTION

Constant advances in semiconductor technologies drive the development of increasingly sophisticated integrated circuits (ICs). The advanced technologies come with their fair share of challenges. For instance, the reduction in feature sizes of ICs requires their fabrication to occur with even greater accuracy. The smaller sizes, however, increase the chance of defects during fabrication. Such defects lower the yield and reliability of IC products. Wafer mapping provides a spatial representation of failed tests during wafer-level testing. By analyzing wafer maps, it may be possible to determine how failures are distributed in the fab, diagnose potential problems in the fabs, and suggest corrective actions. Currently, automated analysis of wafer maps is an active area of research. Wafer map similarity ranking and failure pattern recognition is a standard yet challenging research problem.

Wu et al. developed the WM-811K dataset, which consists of over 800,000 actual wafer maps, and introduced rotation and scale invariant features for wafer-map similarity ranking. They also proposed a wafer-map failure-pattern recognition method [1]. Other studies have focused on wafer-map classification and image retrieval using CNNs, mixed-type defect pattern classification using CNNs, as well as voting-based and feature-based classification using decision trees [4], [5]. Recent studies have also focused on convolutional networks and mixed wafer-map defects [6]. However, the definitions of classes, the datasets used, and the evaluation methodologies differ from study to study.

While these enhancements provide new possibilities for incorporating classification within inspection workflows, additional aspects of the inspection workflow are not automatically fulfilled. For example, an operator may want to examine the image regions that informed a prediction, assess the model's confidence, and log the outcome of an inspection. Gradient-weighted Class Activation Mapping (Grad-CAM) can be used to generate class-specific representations [9]. However, having a qualitative understanding of the behavior of a neural network does not mean that a prediction is justified,

and the areas depicted in Grad-CAM may not fully correspond to the decision process of the model [11].

This paper describes Nano Sentinel, a framework for the classification of wafer-map defects and support of inspection workflows. The framework uses a MobileNetV2-based model for classification [7]. The framework incorporates wafer-map preprocessing, classification into nine classes, prediction confidence, Grad-CAM, a preliminary severity assessment, and inspection recommendations. Transfer learning facilitates the use of deep learning for image classification. The use of deep learning-based models is to organize inspections and capture output, and is not to replace judgment.

The current implementation uses WM-811K dataset and represents wafer maps at an input resolution of 96×96 pixels. In the recorded V2 experiment, the model achieved a validation accuracy of 94.56% and a validation macro-averaged F1 score of 0.8332. A diagnostic evaluation on the test partition provided by the dataset achieved an accuracy of 88.66% and a macro F1 score of 0.4727 using the OpenCV preprocessing. The test partition has already been assessed during the development of the model and is, therefore, considered diagnostic evidence rather than final evaluation evidence.

- 1) Development of a nine-class wafer-map classification pipeline based on MobileNetV2 and the WM-811K dataset for evaluation using aggregate and class-sensitive performance metrics.
- 2) Development of a Grad-CAM component to qualitatively assess image regions linked to model predictions.
- 3) Development of an inspection workflow to combine predicted class, maximum softmax probability, a proposed classification of severity, recommendation rules, and generation of a PDF report.
- 4) Classification limitations, including class imbalance, weak recognition of some minority defect classes, and the interpretation of confidence and decision-making based on severity.

The study focuses on the single-label classification of wafer maps and the demonstration of a prototype workflow. The stated rules and thresholds of severity have not been validated based on the industrial inspection criteria, and the confidence reported by the framework is not a calibrated probability of correctness.

II. RELATED WORK

A. Wafer-Map Defect Pattern Recognition

Wafer-map defect recognition is closely related to defect pattern classification. For wafer-map pattern classification, invariant features are derived by extending the concept of spatial relationship to represent wafer-map failure patterns. An example of such an effort is the WM-811K dataset presented in [1]. This work presents a very prominent, large-scale, wafer-map failure pattern dataset, and presents an interesting analysis approach for real wafer-map failure pattern classification. Different from wafer-map-pattern classification, one of the challenges in real-defect classification is to extract features in an unsupervised manner to represent defect shapes and distributions.

Most of the efforts in wafer-map-pattern classification have been in the field of deep learning in the last couple of years. Nakazawa and Kulkarni [2] developed a wafer-map classifier by extending a CNN-based classifier for image retrieval. The classifier was trained on simulated wafer maps representing 22 different defect classes, and evaluated on a set of real wafer maps.

B. Deep Learning for Wafer-Map Classification

CNN-based approaches have also been developed to classify wafer defects of mixed and complex types. Kyeong and Kim [3] addressed classification of mixed type defects by using individual CNNs for each defect pattern.

Saqlain et al. developed a voting-ensemble classifier to recognize wafer-map defect patterns with a combination of different types of features [4]. Piao et al. studied classification of spatial patterns with decision-tree ensembles using Radon transform-based features [5].

In general, these methods form a basis to evaluate Nano Sentinel, but their results cannot be compared directly without considering the differences in the datasets, classes, definitions, preprocessing, data splits, and evaluation protocols. For these reasons, this study is more concerned with the assessment of the implemented model within the defined experimental conditions, rather than making absolute claims regarding performance with respect to the published methods.

C. Transfer Learning and Lightweight Convolutional Neural Networks Architectures

Although representations can be learned in different ways, transfer learning has been shown to be a successful strategy, especially when large variations are encountered between the training and test tasks or when the test task is defined with limited labeled data.

Sandler et al. released MobileNetV2, which is an efficient neural-network architecture based on inverted residual blocks and linear bottleneck layers [7]. These building blocks enable the architecture to effectively balance representational capability and computational cost. As such, it is a viable option for wafer-map classification pipelines, as model size and inference efficiency are crucial for deploying such models in the field.

Nano Sentinel employs MobileNetV2 in its wafer-map classifier based on the WM-811K dataset and adopts a similar architecture to the previously described wafer map classifier. Its performance is measured using various standard classification metrics including accuracy, average F1-score, weighted average F1-score, and per-class metrics. Again, since the dataset is class-imbalanced, we rely on per-class performance metrics, rather than overall accuracy, to gauge classification model performance.

In Nano Sentinel, Grad-CAM is used as a qualitative visualization aid to analyze wafer-map predictions. The heatmap and overlay are provided to support visual review and are not to be considered a validated defect segmentation method. Quantitative explanations and assessments of the system remain outside the current scope of the work.

E. Research Gap and Positioning of Nano Sentinel

The literature studies different aspects of wafer-map analysis such as engineered-feature recognition, CNN-based classification, mixed-type defects recognition, ensemble learning and image retrieval [1]-[6]. Additionally, the literature covers lightweight CNNs and explainable AI [7]-[11].

Nano Sentinel integrates selected methodologies for single label wafer-map classification, and incorporates MobileNetV2 prediction, Grad-CAM, max-softmax, preliminary severity assessment, rule-based inspection, and a PDF report.

The focus of the research is the integration and assessment of methodologies in a prototype rather than inventing a new CNN architecture or claiming state-of-the-art classification. This work has identified many limitations that need to be resolved prior to deployment in the industry, including class-imbalance, confidence calibration, quantitative assessment of the quality of explanations and validating severity and decision rules against inspection criteria.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Variations in the shape and spatial distribution of defect patterns, as well as their frequency, complicate wafer-map defect classification. The WM-811K dataset has several wafer maps, but its labeled classes are imbalanced. This means that a model can achieve high overall accuracy and still perform poorly on rare defect classes. Therefore, evaluation should consider performance at the class level and use macro-averaging.

The classification model, on its own, provides a limited support to an inspection workflow, and its output is largely restrictive to predicted classes. Reviewers might require additional outputs (e.g., visualization of regions influencing the prediction) and a structured record of the inspection results. These should be interpreted carefully. Neither a high confidence score, nor a visualization proves that a prediction is correct.

The problem that is addressed in this work is the development of a wafer-map inspection prototype, that incorporates nine-class wafer-map defect classification, and provides qualitative visual explanation and structured inspection results. The framework is a research prototype, and its rule-based decisions and severity assessment have not been validated in industry.

B. Objectives

The objectives of Nano Sentinel are:

1. Wafer-map classification: Develop a deep learning model based on MobileNetV2 architecture to classify wafer-maps into nine categories: Center, Donut, Edge-Loc, Edge-Ring, Loc, Near-full, Random, Scratch, and none.

This study focused on the classification of wafer maps using the WM-811K dataset and MobileNetV2. Qualitative interpretation of defects is done using Grad-CAM rather than validated defect localization. Rules-based recommendations and severity categories are provisional and not validated manufacturing decisions. Other areas not covered include

mixed-type defect classification, confidence calibration, and quantitative explanation evaluation.

Validate that the classification of single-label wafer maps using the WM-811K dataset and a MobileNetV2 architecture can be achieved.

Limitations identified in the evaluation of the framework, and the additional validation that needs to be done to consider the framework for industry use are described.

IV. PROPOSED SYSTEM AND METHODOLOGY

The proposed Nano Sentinel framework incorporates wafer map pre-processing, deep learning-based defect classification, visual explanation, and inspections workflow integration. The system aims at classifying wafer-map defect patterns and offering supporting evidence for inspections. The framework comprises of wafer map acquisition and capture, pre-processing, classification, confidence estimation, visual explanation, severity assessment, and decision support, and report generation. The relationships among these components are illustrated in the high-level architecture in Fig. 1.

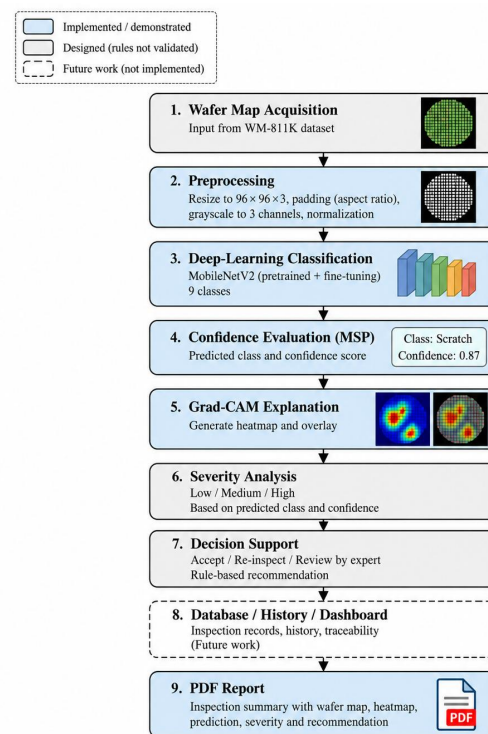


Fig. 1. Overall architecture of the proposed Nano Sentinel framework for wafer-map defect classification and inspection workflow support.

In Fig. 1, a sample of a wafer map from the WM-811K dataset, is shown pre-classification. The input is pre-processed to comply with the expected format and dimension by the classification model. The pre-trained transfer learning model, MobileNetV2, is then used to classify the wafer map to one of the 9 classes and provide probability scores. To explain the prediction, Grad-CAM, which is an image classification explanation method based on the gradient of the output score of the classification model, is employed. The prototype further demonstrates confidence based inspection support, preliminary severity assessment, and PDF report generation.

Therefore, the architecture demonstrates the classification pipeline combined with other workflow components to support inspection decisions, whose decision rules also require further validation.

A. Wafer Map Data Acquisition and Preprocessing

For model development and evaluation, wafer maps in the WM-811K dataset are used. The dataset comprises different types of wafer maps pertaining to various defects classes, if any. There are also wafer maps with ‘no defect’ patterns. Nine classes are considered for the classification task which are center, donut, edge-loc, edge-ring, loc, near full, random, scratch and none.

The validation of each wafer map ensures that it is represented in the form of an input, which is then converted to the required format. The wafer map is resized to a 96 x 96 image, preserving its aspect ratio, and padded with background if needed. Categorical information in a wafer map is encoded in one of the three channels and pixel values are normalized. The formatting is performed to provide an input in the required format to the MobileNetV2 model.

B. Defect Classification using Deep Learning

MobileNetV2 is used as the primary classification model. It is a deep convolutional neural network with inverted residual blocks and uses linear bottleneck connections to reduce the dimensionality of feature maps. Transfer learning is employed to adapt the network for wafer-map defect classification.

The preprocessed wafer map is fed to the feature extraction layers, followed by global average pooling and a classification layer. The output of the classification layer is a score representing the probability of the class. The class with the highest probability is chosen as the predicted class.

C. Confidence Evaluation

To calculate confidence, the class probability output of the model is used. It is displayed as a percentage in the inspection interface and report. Information pertaining to the confidence of a prediction may help in identifying predictions that deserve further scrutiny.

Although a high confidence score may indicate the model made a correct prediction, this is not always the case. This is due to the presence of class imbalance in the dataset or if the model is able to identify patterns that the dataset doesn't include and make a prediction based on those unknown patterns. Confidence of a prediction is therefore an additional factor along with accuracy, macro averaged F1 score, and other class performance metrics to evaluate the overall reliability of the model.

D. Explainable AI Using Grad-CAM

Prediction explanation using Grad-CAM. Grad-CAM computes the regions of interest (ROI) pertaining to the target class based on the gradient. The results of Grad-CAM are presented in the form of heatmap and are overlaid on the wafer map to aid human inspection.

Although the heatmap, along with the overlay, helps define regions of the wafer that may have influenced the model output, this may not always correspond to the regions of a

physical defect, and therefore should not be the only basis for definitive prediction. Grad-CAM should be used as an aid, and other verification techniques should be performed before deployment in industry.

F. Inspection Record and PDF Report Generation

The Data Acquisition Interface facilitates the digitization of inspection records and generates a report in a file format compatible with Adobe Acrobat PDF. Manufacturers can print the PDF report for regulatory compliance or for information transfer to suppliers and other downstream partners.

The reporting component takes the inspections that are available and generates a report. The report can contain wafer-map input, predicted class, confidence score, class probabilities, a Grad-CAM visualization, severity information, and recommendations. This component provides a quick way to summarize the results of an inspection.

There have been demonstrations of generating a PDF report for a sample inspection. However, there are a number of features that are currently not implemented. These features include a persistent database, centralized inspection history, and a dashboard.

Table I outlines the major modules of Nano Sentinel, the outputs produced by those modules, and the current status of the modules. It distinguishes the classification and visualization modules from preliminary decision support and workflow features that are planned for future development

TABLE 1
SYSTEM MODULES AND OUTPUTS

Module	Output	Status
Wafer Map Acquisition	Input wafer maps from the WM-811K dataset	Dataset-based input
Image Preprocessing	Resized and normalized wafer maps	Implemented (96 × 96 × 3 input)
Deep Learning-Based Defect Classification	Predicted defect class and confidence score	Implemented (MobileNetV2 classification)
Explainable AI (Grad-CAM)	Heatmap and overlay	Implemented (qualitative demonstration)
Decision Support	Severity, recommendation, final result	Prototype output for one sample; rules designed, not validated
Workflow Management	Persistent records and inspection history	Future work
Dashboard and Report Generation	Inspection report and visual dashboard	PDF report demonstrated for one sample; integrated dashboard is future work

V. EXPERIMENTAL SETUP AND EVALUATION METRICS

A. Dataset and Data Partitioning

The experiments used the WM-811K wafer-map dataset, which includes various types of wafer maps with different defects and some without defined failure patterns. The task is to classify nine categories: Center, Donut, Edge-Loc, Edge-Ring, Loc, Near-full, Random, Scratch, and none.

The development of the model was done using the official training data, with a lot-aware split to create training and

validation data. A GroupShuffleSplit strategy with a validation ratio of 20% and a fixed random seed of 42 was used. This was done to minimize the chance of wafer maps from the same lot appearing in the training and validation data. This created 43,657 training samples and 10,698 validation samples, with no lot identifiers in common.

The dataset has class-imbalance with the class none being the dominant class. Due to class-imbalance, using accuracy to measure the performance of the model to classify the various defects may not be ideal. Hence, the evaluation of the model is also done using class-wise and macro averages.

B. Model Configuration

As the focus is on the classification of wafer maps, a lightweight model like MobileNetV2 is appropriate for transfer learning to the nine-class problem.

Each input wafer map was resized while keeping its aspect ratio, padded to required size, converted to three channels, and normalized. The input tensor hence had dimensions of $96 \times 96 \times 3$. The classification head had global average pooling, followed by a dense layer with softmax, to obtain class probability estimates for the 9 classes.

The backbone of the architecture was partially fine-tuned for the V2 configuration, and batch-normalization layers were frozen. Class weights with an upper limit of 5 were applied. The learning rate was set to 10^{-5} with the Adam optimizer. The training run time was restricted to 10 epochs, and the best checkpoint for validation performance was saved. The best epoch 9 checkpoint was selected for the evaluation.

Accuracy measures the percentage of total samples that were correctly classified. For binary classification, it is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

For multiclass classification, it is calculated as the percentage of correct classifications of all samples that were evaluated.

Precision is defined as the ratio of the number of correct positive predictions to the total number of positive predictions, including both true positives and false positives. It is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall, also referred to as sensitivity, is the proportion of actual positive samples of a class that are correctly predicted positive. It is calculated as

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score is the harmonic mean of Precision and Recall and is calculated as:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

For multiclass classification problems, where there is more than two classes, macro-averaging is performed. It means that the F1-score is calculated independently for each class and an average is taken. This means that all classes, including the minority defect classes, are treated equally.

Weighted-F1 is the mean of the class-wise F1-scores weighted by the actual number of samples of each class and is performed to compensate for the class imbalance.

Qualitatively assessing Grad-CAM visualizations was performed to evaluate if the model's predictions could be corroborated by other interpretable visual evidence. The visualizations were treated as qualitative evidence and not as a measure of how correct the model is.

A. Overall Classification Performance

The MobileNetV2-based Nano Sentinel classifier was evaluated using the validation subset and official test partition of the WM-811K dataset. Evaluation metrics for overall classification performance are accuracy, macro-averaged F1-score, and weighted-averaged F1-score.

OVERALL CLASSIFICATION PERFORMANCE OF MOBILENETV2

Evaluation metric	Validation set	Diagnostic test set
Accuracy	94.56%	88.66%
Macro-F1 score	0.8332	0.4727
Weighted-F1 score	0.9445	0.9005

According to the results in Table II, the classification model achieved an accuracy of 94.56% on the validation subset. The macro-F1 and weighted-F1 scores for the validation subset were 0.8332 and 0.9445, respectively.

During the diagnostic testing of the official test partition, the model achieved an accuracy of 88.66% and a weighted-F1 score of 0.9005. The macro-F1 score, however, was 0.4727. This illustrates that the aggregate metrics do not fully illustrate performance for individual defect categories. For this case, the dominant category none can greatly influence the overall results, while classifying the less frequent defect classes incorrectly can greatly affect the overall results negatively.

It is worth noting that there is a substantial difference in performance between the validation and diagnostic test. For this reason, class wise performance measures for precision, recall and F1 score are evaluated in the following subsection.

B. Class-Wise Performance Analysis

To investigate the classification performance in more detail, the F1-score for each wafer-map class was analyzed on the diagnostic test partition. The WM-811K dataset shows class imbalance and hence aggregate metrics may conceal issues in the detection of rare class defect patterns.

TABLE III
CLASS-WISE F1-SCORES ON THE DIAGNOSTIC TEST PARTITION

Wafer-map class	F1-score
Center	0.2085
Donut	0.3153
Edge-Loc	0.2226
Edge-Ring	0.6302
Loc	0.2164
Near-full	0.8913
Random	0.7281
Scratch	0.0986
none	0.9438
Macro-F1	0.4727

As shown in Table III, the model's highest F1-score was for the none class (0.9438). Next highest were Near-full (0.8913), Random (0.7281), and Edge-Ring (0.6302). The F1-scores for Center, Donut, Edge-Loc and Loc were notably lower. Scratch recorded the lowest F1-score (0.0986). This

means the model had great difficulty in distinguishing this defect category.

The imbalance in class distribution impacts the overall model performance. The weighted-F1 score for the model was 0.9005. However, the macro-F1 score was 0.4727.

To establish trust in the model, especially for a reliable manufacturing inspection, the model should recognize minority classes better. Class-balancing, data-augmentation, and class-specific error analysis should be researched. Since some classes may be more difficult to improve, an evaluation protocol should be designed and adhered to, to avoid overestimating performance.

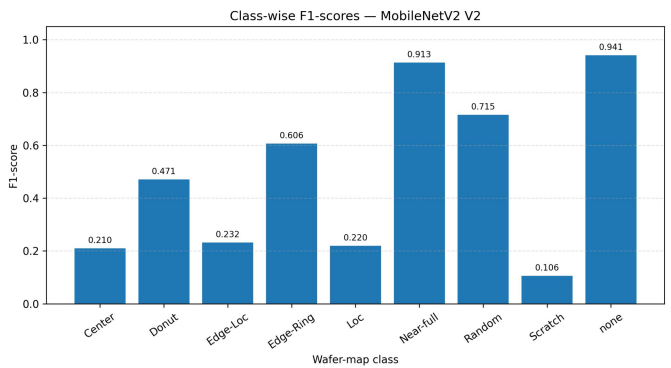


Fig. 2. Class-wise F1-scores of the MobileNetV2 V2 model on the diagnostic test partition.

As illustrated in Fig. 2, the heights of the bars represent the classification performances of the nine classes. The class “none” shows high performance. Some defect classes, particularly Scratch, show low performance and hence may have been detected as a false negative case. This may also explain the low macro-F1 score in the overall performance and supports the argument of evaluating individual classes rather than taking the average of the overall performances. Though the bar chart gives a brief class-wise performance, it does not provide information on what classes are confused with each other. Thus, a confusion matrix is necessary to determine the classes that are confused with each other as well as the classes that are incorrectly predicted.

C. Confusion Matrix Analysis

To evaluate the class-wise prediction behavior of the MobileNetV2 V2 on the diagnostic test partition, a confusion matrix was generated. It contains a comparison of the true wafer-map categories with the ones predicted by the classifier. Only the diagonal entries are predictions of the correctly classified classes, while all entries on the diagonal are misclassifications.

Confusion matrix shape: (9, 9)
 Saved confusion matrix:

	Center	Donut	Edge-Loc	Edge-Ring	Loc	Near-full	Random	Scratch	none
Center	127	11	19	0	133	0	3	4	535
Donut	25	60	11	1	42	0	6	0	1
Edge-Loc	6	3	1312	32	66	1	14	4	1334
Edge-Ring	2	0	255	536	3	0	0	1	329
Loc	52	18	270	1	532	0	1	14	1085
Near-full	0	0	8	0	0	84	3	0	0
Random	0	3	33	0	29	4	171	0	17
Scratch	0	2	92	3	96	0	0	51	449
none	164	12	6546	69	1971	0	23	196	101720

Fig. 3. Confusion matrix of the MobileNetV2 V2 on the diagnostic test partition.

Using the confusion matrix, an insight into the distribution of correct and incorrect predictions for the nine classes can be gained. Apart from the general accuracy and the class-wise F1-scores, the diagonal and off-diagonal entries can highlight classes that are recognized consistently and reveal categories that are commonly confused with each other. This information is valuable for understanding the current limitations of the considered classifier and determining the direction for future work on minority-class recognition.

The analysis of the confusion matrix is complemented by the class-wise performance metrics visualized in Table III and Fig. 2 for a better perception of the model behavior.

D. Grad-CAM Explainability Results

The Nano Sentinel framework was augmented with Grad-CAM (gradient-weighted class activation mapping) to enhance the explainability of the wafer-map defect classification results. Grad-CAM generates a heatmap indicating the spatial regions that contribute to the classification score of a selected class, which can be overlaid on the wafer map for visual inspection.

The Grad-CAM heatmaps have been qualitatively evaluated to determine how the visual explanations could be associated with the wafer-map classification predictions. Fig. 4 shows an example of the wafer-map input and the Grad-CAM result.

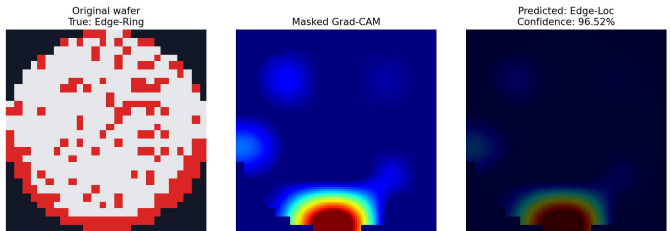


Fig. 4. Representative visualization of Grad-CAM.

The generated heatmap can be used as a qualitative indication of the image regions contributing to the classification result. However, the highlighted regions should not be used as confirmation of the presence of a physical defect or correctness of the predicted class. The Grad-CAM explainer is intended to offer an additional layer of support for the manual inspection of the wafer-map classification results.

It will be important to further evaluate Grad-CAM results across correctly and incorrectly classified samples to determine whether the highlighted regions can consistently be associated with physical patterns in the wafer-map images.

E. Inspection Workflow Demonstration

In addition to wafer-map classification, the Nano Sentinel prototype also demonstrates the application of prediction results within an inspection workflow. Given an input wafer map, the system will display the predicted defect category and confidence, and available inspection information through the prototype interface. It will also include a preliminary severity assessment, and decision-support recommendations.

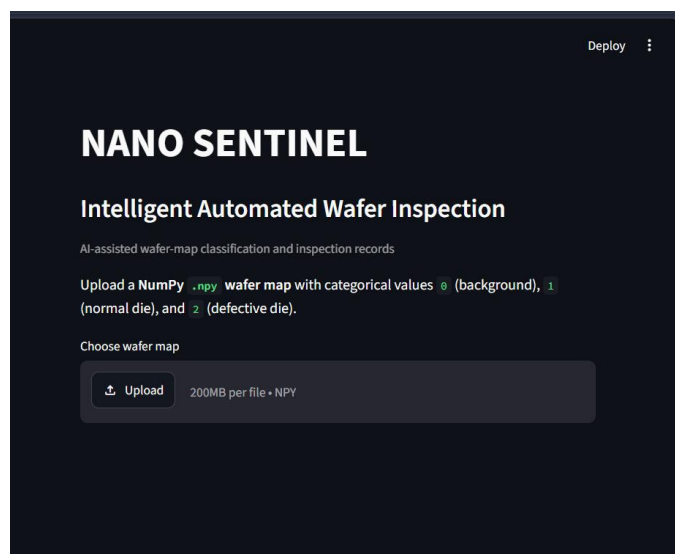


Fig. 5. Example of wafer-map inspection using the Nano Sentinel prototype.

The inspection interface aims to facilitate comprehension of the classification outcome by providing the predicted category and associated confidence and inspection information. The severity assessment is based on an applied set of rules and should be interpreted as an initial indicator rather than a validated assessment of manufacturing risk. Likewise, the decision-support recommendations aim to provide an aid for review, and should not be considered a substitute for expert inspection.

The prototype further demonstrates the generation of an inspection record, and a PDF report for an individual sample. Nevertheless, persistent database management, centralized inspection history, and a fully integrated dashboard are future development tasks. As such, the current demonstration establishes the feasibility of linking classification outcomes with selected reporting functions, rather than a fully production-ready manufacturing workflow.

F. Discussion and Limitations

The experimental results showcase the potential of MobileNetV2 for the wafer-map defect classification task as defined by Nano Sentinel. The model achieves a high validation accuracy of 94.56% and a good macro-F1 of 0.8332. On the diagnostic test, however, its accuracy drops to 88.66% and the macro-F1 drops significantly to 0.4727.

Indeed, the class-wise analysis revealed significant variance in the quality of classification, with the model performing reasonably well on none and Near-full categories, but underperforming on the Scratch category as well as a few others. The class imbalance has to be taken into consideration when evaluating the results, which is why the weighted-F1 score, which takes the class weights into account, should be used as the main metric as well as the overall accuracy.

The inspection of Grad-CAM results offers a deeper insight into the model's behavior. The visualizations confirm that the model mostly focuses on the relevant areas when making predictions. However, inspection of Grad-CAM results cannot substitute the actual model testing and evaluation. A high confidence does not always equal a correct prediction, which is why in some cases it might still be useful

to double-check a model prediction, especially if the confidence score is low.

The example of the inspection workflow demonstrates how such a workflow can be used in practice: showcasing the model's predictions with confidence scores, severity pre-rating based on the predicted defect category, and report generation, including the defect visualization. However, the severity thresholds and decision-support rules still require validation according to the semiconductor manufacturing industry standards. The persistent database storage, inspection history tracking, and dashboard are also features to be developed and ultimately deployed in the future.

Several limitations come with this study. First, the class imbalance as well as the lower recognition performance on specific categories reduce the value of the current classifier for the task of universal defect recognition. Second, the diagnostic test partition was already used in the model evaluation and cannot be used as a final "unseen" test set. Third, the analysis of Grad-CAM is subjective and should be supported by a more systematic analysis. Finally, the proof-of-concept prototype needs independent validation in a semiconductor manufacturing environment.

The follow-up work will focus on improving the recognition of minor categories, analyzing calibration and class-wise error patterns, validating the severity and decision-support rules, and independent testing with a separate test protocol.

VII. CONCLUSION AND FUTURE WORK

Nano Sentinel, an intelligent wafer-map inspection framework, which makes use of a deep-learning-based defect classifier with Grad-CAM visual explanations and inspection workflow support, is introduced. The framework employs the classifier based on the MobileNetV2 architecture to classify defect categories in the wafer maps from the WM-811K dataset into nine classes. The tool also allows presenting the confidence scores, preliminary severity assessment and inspection recommendation, and the PDF report generation.

The MobileNetV2 V2 model that yielded the best results achieved the value of 94.56% for the validation accuracy, 0.8332 for the macro-F1 score, and 0.9445 for the weighted-F1 score. For the diagnostic evaluation of the official test partition, the model achieved 88.66% accuracy, 0.4727 macro-F1, and 0.9005 weighted-F1. The significant difference between macro and weighted F1 scores for the test partition illustrates the problem of imbalanced classes and the varying ability of the model to recognize different defect categories. In particular, several minor defect categories were recognized with low performance, which should be addressed by future research.

Grad-CAM provides a way to visually assess what parts of the input contributed to the model's decision, which can be used to increase the understanding of the model. However, such an approach cannot verify that the model did not make an error, and the diagnostic evaluation showed that achieving high confidence across all classes is challenging. Thus, the current prototype should be regarded as a research tool that requires further validation before it can be applied as an industrial inspection system.

Future work in this direction may concentrate on overcoming the recognized difficulties in minority-class recognition and validating the confidence scores. Another aspect that requires further investigation is the evaluation of the severity assessment and decision-support rule validity. Moreover, a separate test protocol with a well-defined set of instructions must be created to evaluate the tool's performance. Additional improvements that could be explored include the implementation of a persistent inspection history, the database connection, and the enhanced visualization dashboard. The application of the framework to the real-world manufacturing data with further expert validation is required to achieve a fully validated, industrial-grade model.

In summary, Nano Sentinel prototype illustrates how a combined approach to wafer-map classification, explanation, and inspection reporting can be implemented.

ACKNOWLEDGMENT

The authors are grateful to their project guide, Prof. Vetrikani R, and to the Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), East Point College of Engineering and Technology, for their guidance and support.

REFERENCES

- [1] M.-J. Wu, J.-S. R. Jang, and J.-L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 1–12, Feb. 2015, doi: 10.1109/TSM.2014.2364237.
- [2] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 309–314, May 2018.
- [3] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 3, pp. 395–402, Aug. 2018, doi: 10.1109/TSM.2018.2841416.
- [4] M. Saqlain, B. Jargalsaikhan, and J. Y. Lee, "A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 171–182, May 2019.
- [5] M. Piao, C. H. Jin, J. Y. Lee, and J.-Y. Byun, "Decision tree ensemblebased wafer map failure pattern recognition based on Radon transformbased features," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 250–257, May 2018.
- [6] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 587–596, Nov. 2020.
- [7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
- [8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
- [9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradientbased localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [10] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
- [11] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, 2018, pp. 9505–9515.