ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 3 Issue 2 June 2023 / Page No: 120-134
Paper Id: JETA-V3I6P107 / Doi: 10.56472/25832646/JETA-V3I6P107

Original Article

Smart Semiconductor Wafer Inspection Systems: Integrating AI for Increased Efficiency

Jyothi Swaroop Arlagadda Narasimharaju
Hardware Engineer, Intel, United States of America (USA).

***Abstract**: The semiconductor industry has received the pressure of the need to develop techniques for higher efficiency and accuracy of wafer inspection processes. It has been a problem to inspect the complexity of the semiconductor wafers with traditional inspection systems and therefore sophisticated solution is required. This paper looks at the evaluation of Artificial Intelligent (AI) in semiconductor wafer inspection systems to improve the outcome. Applying the ML and Computer Vision approaches in AI allows automation of defect identification, sorting, and enhanced yield levels. From the points of methodology, the study offers a thorough analysis of the current research and development in the field of AI practices within wafer inspection and the effects that improvements have had on the manufacturing process. Some conclusions from experiment research and development show that the semiconductor organization's distance in the speed of inspection time and the ratio of defect detection is notably enhanced, thus supporting the concept of AI convergence in the semiconductor organization.*

***Keywords**: Semiconductor, Wafer Inspection, Artificial Intelligence (AI), Machine Learning (ML), Computer Vision, Defect Detection, Yield Improvement, Deep Learning.*

I. INTRODUCTION

Semiconductor plays a significant role in today's technologically advanced society. It is applied in areas such as consumer electronics, telecommunication, computing and automobile industries, aerospace industries, etc. These materials are in products ranging from Handy gadgets such as Mobile phones, laptops, and computers to complicated items like medical equipment, satellites and more. Ongoing improvement in the industry results from the steady demand for smaller, swifter, and more efficient electronics. The basic operation that occurs in the manufacturing of semiconductors is the inspection of the semiconductor wafers. Wafers are thin wafers of semi-conductive material like silicon on which microelectrical devices are created. These wafers must, therefore, be of high quality without any flaws because it is common knowledge that faults are likely to lead to the malfunction or production of wires and other electronic accessories of poor quality in other production lines. Besides, the checking of wafers can be regarded as a strategic operation as it assists in ensuring that all products are good enough for customers to use and not instances where companies produce many poor-quality products and spend a lot of money in convincing customers to engage in purchase.

The past and current forms of inspection methods have largely involved optical microscopes and the basic methods of inspection. The usage of these methods, to some extent, can be possible though these methods have some drawbacks. Traditional inspection of the optical surface involves using one's sight to scan through the surface of the wafer with the intention of identifying such flaws as scratches and particles. However, these approaches can contribute to the improvement of the efficiency of the inspection process and reduce the influences of human aspects; nevertheless, these approaches do not reveal enough power to identify and learn significant complicated defects and further recognize the changes in the inspection environment. Hence, the conventional method of inspection on structures ultimately fails in the display of the quality that is sought in the integrated circuits producing industries.

Figure 1: Wafer Inspection Systems

A. Need for Advanced Inspection Systems

a) Increasing Complexity and Miniaturization

So far as semiconductor devices are becoming smaller and have a more complicated structure, the traditional inspection technique is confronted with great difficulties. These challenges require multi-disciplinary technology solutions which assure high precision and reliability to the intended defect in Figure 2.

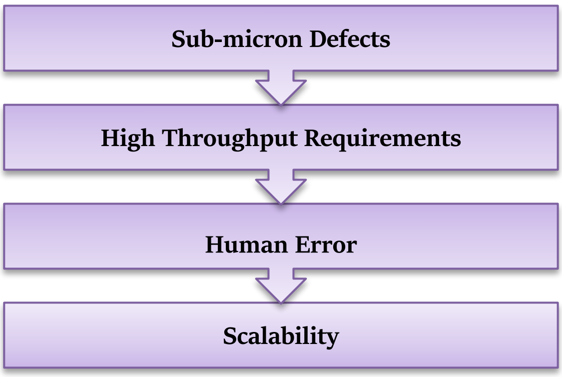

Figure 2: Need for Advanced Inspection Systems

b) Sub-micron Defects:

The elements of articles of smaller nodes and the complexity of semiconductor devices' disadvantage result in sub-micron defects and thus cannot be easily inspected.

c) High Throughput Requirements:

The drive to produce goods more quickly raises the demand for high-speed inspection equipment. Traditional methods are sometimes identified as bottlenecks because they take time to complete their processes.

d) Human Error:

This system exposes some vagaries of human-made techniques of inspection and commonplace forms of automation that are, liable to faults and may fail to identify obvious imperfections.

e) Scalability:

When the manufacturing of the wafers increases, then the traditional techniques prove to be inadequate in accommodating the volume and diversification of the wafers.

B. Role of AI in Wafer Inspection

a) The Relation between AI and the Field of Wafer Inspection

Regarding such matters, Machine Learning (ML) and Computer Vision provide some of the most effective working solutions because the operatives' input is not necessary, timelines are reduced significantly, and outcomes are quite precise [2].

b) Automated Defect Detection:

With the help of the newly implemented AI programs, a company can process vast amounts of data and, at the same time, obtain nearly on-line results of the inspection and the imperfections which an inspector might overlook.

c) Reduction of Human Error:

With regards to the evaluation of the defects, those elements that relate to the fluctuation and//or irregularity of the human eyes are absent when AI is applied in the detection of similar. An AI system will always provide one solution when invoked, and at the time the procedure is being performed, the specific program under discussion will generate the same reaction.

d) Increased Speed and Efficiency:

Quicker than when the authentication is carried out separately, and the throughput rate is elevated when the enumerated systems under the category of AI major dominants are the ones to be used for validating a food product, a document, or any other object.

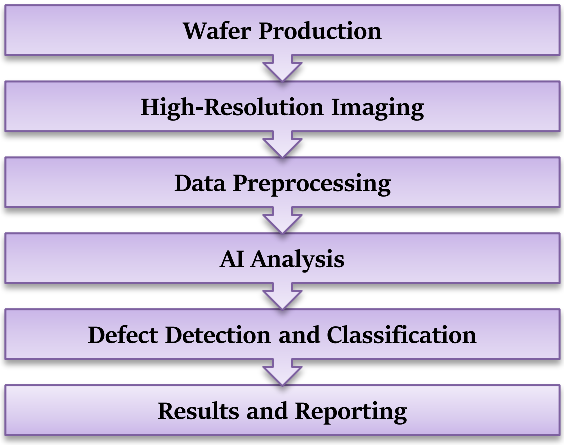

Figure 3: AI-Driven Wafer Inspection Process

i) Wafer Production

- Purpose: This stage is about the production of semiconductor wafers, which is done through further processing.
- Wafer Generation: Materials are also obtained and undergo processes to provide the product of silicon wafers.
- Parameter Setting: Some of the critical settings of the entire process of wafers' manufacturing, including temperature, pressure, and chemical makeup, are predetermined and controlled.
- Output: Wafer inventory in top quality that has not been touched or has not passed any inspection center yet.

ii) High-Resolution Imaging

- Purpose: Obtain high-resolution of the wafer surface using high-end imaging techniques.
- Optical Microscopy: Uses light to produce images of high resolution.
- Scanning Electron Microscopy (SEM): Offers photomicrographs with better resolution compared to optical microscopy.
- Atomic Force Microscopy (AFM): It utilized a mechanical probe to contact the surface, hence determining atomic detail.
- Output: Optical micrographs of the wafer surface to a high level of detail.

iii) Data Preprocessing

- Purpose: Preprocess captured image data by denoising it and ensuring that the variables range through zero.
- Noise Reduction: To eliminate irrelevant noise from the images, apply filters.
- Normalization: Normalize the image data for a standard scale to perform the subsequent steps of analysis on the data.
- Output: Images that have been preprocessed for the AI.

iv) AI Analysis

- Purpose: After obtaining images which are preprocessed, use AI techniques on the image to identify the defects.
- Machine Learning (ML): Use the methods of machine learning designed for evaluation of the images and search for the patterns and less regularities in them.
- Deep Learning (DL): Employ better DL models, for example, CNNs to improve the analysis of the images.
- Output: Findings of the analysis of the item that identifies flaws and variations from the acceptable level.

v) Defect Detection and Classification

- Purpose: Carry out the classification of the defects which exist in the wafer based on the results which have been analyzed by the AI system.
- Defect Detection: Run AI algorithms to detect positions with defects on the exterior of wafers.
- Classification: The defects to be checked should be split into separate categories, for instance; Abrasion, Foreign matter and distortions of pattern.
- Output: Defect maps that provide further amplified information of the object to be subjected to repair and which are segmented into action strategies.

vi) Results and Reporting

- Purpose: This should be done while neatly writing the analysis and the results of the defect detection process into papers.
- Result Compilation: The completion of the data collected from the understanding of the part of a product using the AI and from the results of the defect detection phase.
- Report Generation: Develop detailed documentation of the outcomes of the research on the following issues: kinds of defects, areas in which they can be found, and potential consequences.
- Reporting: Provide the reports of such results to personnel such as engineers and/or the quality control section or department.
- Output: Thus, when assessing a machine or any other related item, checklists and other writings act as archives of the observations in case of decision-making or making effective changes.

C. Pattern Recognition and Anomaly Detection:

Specifically, the experts appreciated the identity of additional abilities of AI systems per se to traditional ones: it has provided the tools to make the comparison fast and without an issue the quite apparent differences in the overall appearance of the wafers. RNN, as well as CNNs, could learn the mentioned feature in a way that can help to detect, such as segmenting the normal as well as defective areas.

D. Continuous Improvement:

As for the special discussion-worthy circumstances, it should be noted that some types of artificial intelligence arrangements can be tamed and supplied with new datasets; thus, they can advance in terms of accuracy and the ability to bark a new form of defects and kinds of manufacturing. Advantage of Smart Semiconductor Wafer Inspection Systems: Integrating AI for Increased Efficiency:

a) Enhanced Detection Accuracy

- In fact, techniques such as ML and DL are even able to target minute problems that I may not be able to identify with the naked eye.
- Impact: Lesser false positives and false negatives that help in separating the exceptional wafers from the average and passing them through manufacturing.

b) Increased Inspection Speed

- When comparing the application of artificial intelligence to the manner through which people attempt to inspect images.
- Impact: This, in turn, reduces the time taken in inspection, and consequently, throughput is enhanced, meaning that the time taken in the production is reduced.

c) Reduction in Human Error

- The application of AI in analyzing and inspecting the process eliminates the effect of management, the variability, and the inconsistency of the human inspector while on the field.
- Impact: Increase the possibility of disclosing the defects in a wider scope and thus affect the general improvement of quality.

d) Scalability

- That is, the use of AI systems enables the organization to state that even if the number of wafers increases, the time for the inspection or required workers is the same.
- Impact: This makes it possible that as the market demands grow, the size of the manufacturing operations can easily expand.

e) Cost Efficiency

- Overriding the challenging factors of time-associated concerns and randomness, the implementation of AI in inspecting systems lowers the costs related to rework, scrap, and warranty claims efficiently.
- Impact: Builds cost of production; hence, the likelihood of making higher profits is high.

f) Continuous Improvement

- what is important is that AI systems are not static and there is always a possibility to feed new data into the AI on occasion, thus boosting its performance from time to time.
- Impact: SEMI-GENERALLY Real-time tracking of advancements in the semiconductor along with the change of improvement/variety of the defect type that the inspection process should cover.

g) Comprehensive Data Analysis

- It holds information that might transit to analyze substantially bigger data sets to enable patterns and trends of surface defects to be identified.
- Impact: It assists in maintenance plans, in minoring prediction, and in the enhancement of usage of the resources that in general, enhances the efficiency of manufacturing processes.

II. LITERATURE SURVEY

A. Traditional Wafer Inspection Methods

In traditional inspection methods, the use of microscopy and especially human interaction in the process of inspection is common. All these methods, although useful in specific applications, are prone to human error and are incapable of sensing sub-micron-level defects. To overcome these limitations, there are gadgets referred to as Automated Optical Inspection (AOI) systems, though the increasing complexity of the semiconductor wafers also disadvantages these systems.

B. Optical Microscopy

Optical microscopy is one of the first methods of semiconductor wafer inspection that started with the development of the process. It consists of visually examining a wafer for defects with the help of high-power lenses. Namely, it is helpful in identifying larger defects, though not suitable for sub-micron defects as the resolution is not as high.

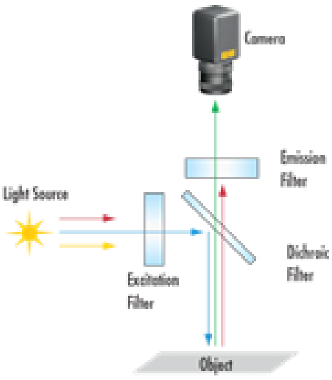

Figure 4: Optical Microscopy Setup [4]

Microscopy System has some relationships with the Light Source, Stage, Objective Lens, Camera, Image Processing, and User Interface classes; thus, these components are a part of the Microscopy System.

C. Human Inspection

A person analyzes wafers with a naked eye, microscope or employing automatic means. Peculiar to this method is human inconsistency and the possibility of errors given the increasing complexity of the wafers. Also, human inspection is highly subjective, and the same part can be inspected differently at different times or by different personnel, causing variation when it comes to defects.

D. Automated Optical Inspection (AOI)

AOI systems are a sort of automatic inspection system that uses high-resolution cameras as well as image processing algorithms. These systems can inspect the wafers faster than human inspectors, but there are some issues with relevance to discovering diminutive or complex defects. The integration of this system however, needs constant adjustment and the need for frequent calibration.

E. AOI System Workflow

1. Initialize System: A sub-step of the system initialization which consists of synchronization.
2. Load Inspection Parameters: Place the loading of the parameters necessary for the inspection.
3. Parameters Valid: Determine if the parameters are valid.
   - Yes: Continue to go and begin the inspection as a process.
   - No: Exception handling and inform the operator.
4. Start Inspection Process: Start the assessment exercise.
5. Capture Image: Capture an image for view purposes.
6. Process Image: Display image. File processing.
7. Analyze Features: There are also features in the processed image that need to be analyzed as well.
8. Defects Found: Better see if some of them are defective.
   - Yes: Register and mark the defects.

- No: When you are certain that the item is passing, label it.

9. Generate Report: A report needs to be produced based on the analysis.
10. Shutdown System: After inspection, power off the system.

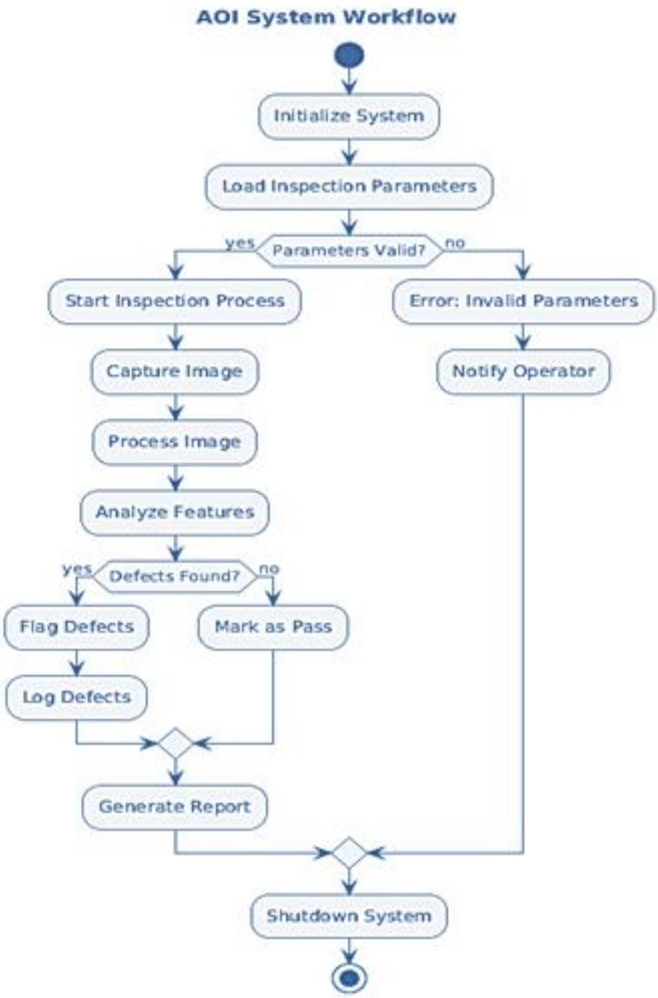

Figure 5: AOI System Workflow

Table 1: Comparison of Traditional Inspection Methods

Method	Strengths	Limitations
Optical Microscopy	High magnification	Limited to larger defects
Human Inspection	Flexible and adaptable	Prone to human error and inconsistency
AOI	Automated and faster than human inspection	Challenges with sub-micron and complex defects

F. Utilizing AI in Manufacturing

Application of BIG DATA and Artificial intelligence systems for the manufacturing industry AI in manufacturing has shape-shifted almost all industries, such as automotive, electronics, and pharmaceutical industries. These IoT integrated systems may use AI in analyzing large data for processes' enhancement, predicted maintenance or quality enhancement of a product. For AI in semiconductor manufacturing, the application areas include predictive maintenance, process optimization, and defect detection, which show enormous possibilities for improving the wafer inspection systems.

a) Predictive Maintenance:

Data gathered by sensors has been used by AI to predict when equipment is likely to break down used in predictive maintenance. This, in turn, leads to increased efficiency in the use of production facilities in the different manufacturing processes and, thus, minimum time on maintenance.

b) Process Optimization:

Manufacturing processes are improved with the help of AI algorithms analyzing production data and finding out the weak points. This results in high efficiency, lower costs and better finished goods or services in the organization.

c) Quality Control:

AI improves QC, as it scrutinizes products and defects that the human eye might overlook, thus making products of better quality and uniformity.

G. AI in Semiconductor Wafer Inspection

As for present research to use AI methods in the field of wafer inspection, several methods, such as supervised learning, unsupervised learning and deep learning have been applied. Supervised learning techniques need training data containing attributes of defects to train models for detecting and categorizing them. Even clustering methods, which belong to unsupervised learning, can find anomalies without having any labeled data. The most vital area of deep learning is Convolutional Neural Networks (CNNs) that has proved quite effective and efficient in image-based defect detection than the standard techniques [1].

a) Supervised Learning:

Supervised learning is applied when models are trained on labeled datasets in which the input images contain the tags of the defects. Some of the popular classification methods include Support Vector Machines (SVM) as well as Decision Trees. These models can be quite precise when it comes to classifying defects. However, they need labeled training data to be precise.

b) Unsupervised Learning:

Clustering and anomaly detection do not rely on training data as they come under the category of unsupervised learning. The above methods are used to find features in the images of wafers and, thus, can be used to detect new forms of defects.

c) Deep Learning:

Convolutional neural network is a branch of deep learning and has been considered outstanding in image-based defect detection. Thanks to that, CNNs can extract the hierarchical features of the wafer images and the defects, even if they are as small as a pinhead. What makes this possible is transfer learning where the existing models are trained on datasets specific to the use.

H. Key Challenges and Solutions

However, there are some issues when applying AI in a wafer inspection system. Albeit the promising improvement in wafer inspection systems using AI, the following are the challenges that are found in the application: Such limitations are the requirements of the significant labeled dataset, various types of defects, and computation capability for real-time inspection. Researchers are facing several challenges to tackle these challenges, researchers are improving the algorithms, transferring learning and exploring the capability of edge computing.

There is a need for large, labeled datasets. All AI models rely on labeled data with the understanding that the more labeled data there is, the better the AI models. Such datasets can be relatively expensive as well as time-consuming to obtain and annotate. Some of the possibilities are the following: data augmentation that allows increasing the size of a set artificially, using the synthetic data received by simulation.

I. Complexity of Defect Types

The defects that can appear on semiconductor wafers can be of different types, which complicate the model's ability to generalize. Newer models are being created with the ability to address a variety of defect scenarios as well as including the knowledge of experts in the learning procedures.

Table 2: Types of Semiconductor Defects

Defect Type	Description
Scratches	Linear marks on the wafer surface
Particles	Foreign particles embedded in the wafer
Pattern Defects	Irregularities in the wafer pattern
Cracks	Fractures in the wafer material

a) Computational Resource:

Real-time key point detection is a very computationally intensive process since it must analyze high-resolution images in real-time. Some of the solutions are algorithm optimization, utilization of purposes built-in hardware such as GPU and edge computing which involves computation of data near the sources.

b) Edge Computing:

Edge computing refers more to data processing nearer to the data rather than at a main data center. This cuts the time of processing and enables administrators or managers to make decisions as they happen, which is essential in inspection systems for wafers.

III. METHODOLOGY

A. Data Collection and Preprocessing

a) High-Resolution Imaging

The first process, therefore, in building an automated inspection system based on artificial intelligence is to gather images of semiconductor wafers in high resolution. Ahead of time, use high-resolution cameras to capture images that reveal different sorts of defects such as; scratches and particles or pattern mismatch. The quality is rather an important issue as these images are used in AI models therefore, the quality determines the models' performance in Figure 6.

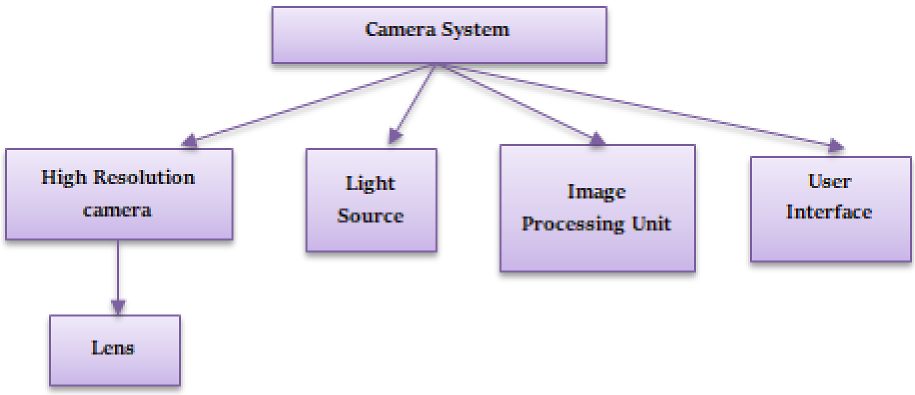

Figure 6: High-Resolution Camera Setup

High-resolution cameras and light sources are associated with the camera system, meaning they are components of a camera system, while the image processing unit and user interface also have a similar relation to the proposed camera system. In the class diagram, a solid line with a diamond shape on the class end shows that a high-resolution camera has a composition relationship with a lens as the camera has a lens.

i) Data Annotation:

The gathered images are manually annotated by experts, where the defects found in the food images are described and classified. This labeled data used in this work is the ground truth for the training of supervised learning models. This process requires drawing rectangles around the defects so there is no mistake in labeling, which forms the basis of the created dataset.

ii) Data Augmentation:

To improve the stability and transfer ability of AI models, data augmentation strategies are used. These techniques are used to generate more data from the given data set by creating variants of the same data, such as rotated, scaled, flipped, and noisy data. Augmentation assists the models in not deviating from transformations and enhances their performance when it comes to unseen data.

B. AI Model Development

i) Supervised Learning Models

In this kind of learning, the models are trained using the annotated datasets. Some of the familiar models that are used are called Support Vector Machines (SVMs) and Decision Trees: the latter categorizes defects by their characteristics. To ensure the effectiveness of these models, they undergo metrics such as accuracy, precision, recall, and F1-score in Figure 7.

1. Define Problem: Describe in detail the problem that needs to be addressed with the help of machine learning.
2. Collect Data: Accumulate all the information that will allow one to solve the problem under consideration. This can refer to any kind of data and method of gathering data.
3. Preprocess Data: Cleans the data and format it to a suitable form that will be easy for analysis for the real estate firm. In this step, some of the common processes include how to handle missing data, how to normalize data or standardize it and feature engineering.
4. Split Data into Training and Test Sets: Divide the dataset into two parts. An example is that one of the sets must be used in training the model while the other set is only used in testing the trained model. This helps in ascertaining the accuracy that the model offers to the new set of data.

5. Select Algorithm: Choose the right type of machine learning according to the type of problem to solve, whether it is classification or regression or according to the type of data set.
6. Train Model on Training Data: Use the data gained to train the selected type of machine learning algorithm. This means altering the values that it employs for the arrival of the predictions in a way that may ensure that there is an improved accuracy of the prediction that is done.
7. Evaluate Model on Test Data: Finally, when training of the model is done use the model to predict on the test set to determine the efficiency of the model. In classification-based problems, the common measures which are used most often are; Accuracy, Precision, Recall, F1 score and others.
8. Model Performance Acceptable: It is necessary to decide whether the existing criteria for assessing the constructed model's efficiency are satisfying all.
   - Yes: Transform to the implementation of the model.
   - No: After that, proceed to hyperparameters tuning.
9. Tune Hyperparameters: Fine-tune the hyperparameters of the model - For this next block of code, create a new copy of the binary for further work. This can be done by methods like the grid search or actually random search type of methods.
10. Retrain Model on Training Data: Training process for the provided model Reshape training data Re-apply transformations to the training dataset Retrain model on the generated training data
11. Reevaluate Model on Test Data: This means that four values should be reevaluated on test data, and the last four values of the model should be calculated from scratch using the same test data.
12. Deploy Model: If the model adequately fits the data, then it should be moved to a production area in which it is used to make forecasts on new data.
13. Monitor Model Performance: Also, one should monitor the performance of the model whenever this is in use with a view to ascertaining its performance when used for an extended time. This includes determining the conditions under which the model is mis-specified, re-estimation of the model and management of new data.

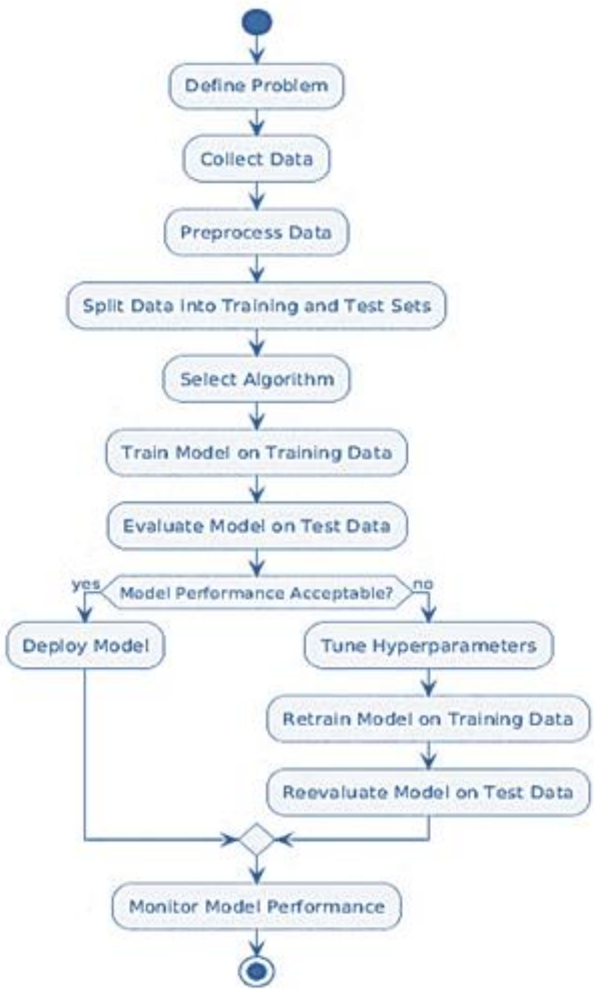

Figure 7: Supervised Learning Workflow

C. Deep Learning Models

CNNs are used in most image-based defect identification professions because the deep learning models are linked to them. Thus, architectures like ResNet and Inception are used when training the model with the wafer images to further its capacity to distinguish the hierarchical characteristics that define the different types of defects. Convolutional layers, for the help of the transfer learning procedures on the large image databases are used in order not to depend on the labeled images.

D. System Integration

Hardware and Software Components: The developed AI models are incorporated into an automatic surface inspection system of the wafer produced. The acquisition system involves the required units to capture the images of the semiconductor die; the pre-processing system, which involves the required unit to pre-process the images of the semiconductor die; and the real-time analysis system which requires the unit to analyze the defect and classify the same. This implies that integration of the operation offers improved functionality as well as reliability in the aspect of the inspection.

a) Inspection Process:

The inspection process involves several steps. In turn, the inspection presupposes several procedures, including the procedure of detailed examination of the object.

b) Image Capture:

Newcomers are made with high-reducing cameras.

c) Image Preprocessing:

In images, post-processing is carried out on the specific feature, enhancing the feature of the image and, during the process, eradicating the noise.

d) Defect Detection and Classification:

The AI model then is to make a determination of what part is defective out of the images that have already been preprocessed, and then the defects are categorized.

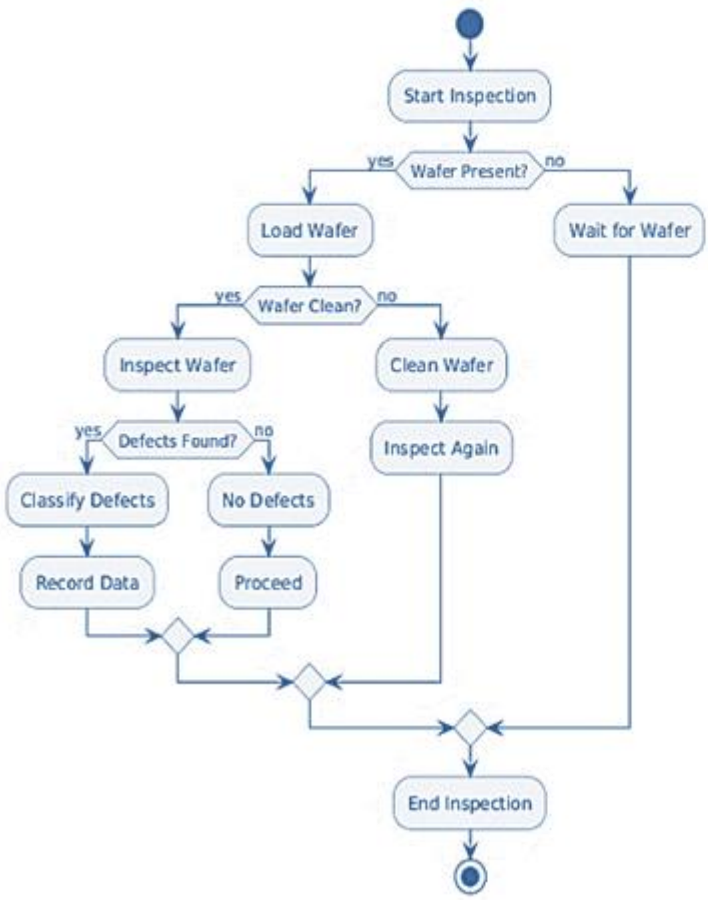

Figure 8: Wafer Inspection Process

1. Start Inspection: The inspection process begins.
2. Wafer Present: Ensure that there is a wafer.
   - Yes: Go on and load the wafer.
   - No: The next thing that Isaac would hear would be the wafer, so he waited for that signal.
3. Load Wafer: Put the loaded wafer into the inspection system.
4. Wafer Clean: Ensure there is no debris on the wafer.
   - Yes: Inspect the wafer.
   - No: Clean the wafer.
5. Clean Wafer: Do not allow any peculiar smell, stain, dust or grease to remain on the surface of the wafer.
6. Inspect Wafer: Perform optical inspection of the wafer for any defects in the wafers.
7. Defects Found: See whether there are any defects noted in the inspection done by the counselor.
   - Yes: Classify defects.
   - No: Proceed without defects.
8. Classify Defects: Determine the sort of defects observed on the respective waver and classify the defects into categories.
9. Record Data: Store the data; this will include the records of the different defects noted for analysis and reporting purposes.
10. Proceed: If none of the above areas depict defects, then it is safe to continue with the process.
11. Inspect Again: In this case after a wafer has been cleaned, ensure that it has no defects by again inspecting it.
12. End Inspection: Conclude the inspection process.

E. Inspection Speed

This is the time taken for the inspection of a wafer; the faster time taken has a direct impact on the throughput rate of manufacturing. This metric is calculated as the mean of the amount of time it takes to capture all the wafers' images, the amount of time it takes to preprocess them and the amount of time it takes to analyze them.

F. Detection Accuracy

Sensitivity is the proportion of defect actually present that was correctly identified. It depicts the skills of the model's call on the number of defective and non-defective areas of a given wafer. Operational detection accuracy means that a large number of defective wafers are not passed through the inspection process.

Table 3: Detection Accuracy Metric

Model	Accuracy (%)
SVM	85
Decision Tree	82
CNN (ResNet)	95

G. False Positive Rate

The false positive rate, on the other hand, is the percent of non-defective regions that are classified as defective. If false positive rates are improved by reducing them, this, in turn, leads to less rework and scrap and therefore, the manufacturing process is made more efficient.

H. Yield Improvement

Yield improvement is related to the improvement of the AI-based inspection system by outlining the percentage of wafers free of defects after the implementation of the system. These signify the extent to which the system has transformed the overall manufacturing quality as well as efficiency.

IV. RESULT AND DISCUSSION

A. Experimental Setup

a) Dataset Description

Let the dataset be a clear image dataset with 10,000 high-quality wafer images from where the experimental setting will be enacted. They are labeled by professionals pointing at different defects, such as scratches, particles, patterns, and even cracks. The dataset is divided into three subsets: They split them into a training set at 70%, a validation set at 20% and a test set at 10%. This split implies that the models will have enough data to learn from as well as validate their performance during their training process; they are also so designed that once trained, they can be tested for their accuracy using data that was not used in their training process.

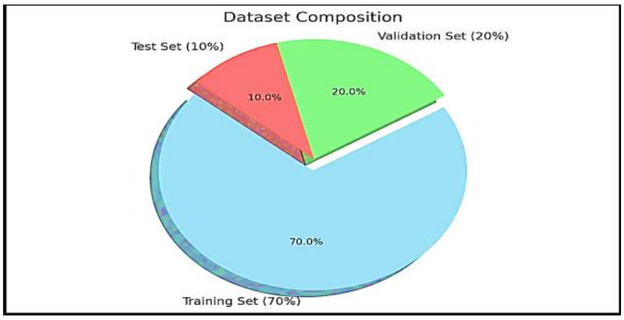

Figure 9: Dataset Composition

b) Data Splitting

The dataset is split in a 70: Proper splitting of this data is 20 images for validation, 7 images for testing as the given split is a 20:10 ratio, so 7000 images for training, 2000 images for validation and 1000 for testing purposes. This stratified split also confirms that all defect types come to different subsets so there is a division of defects in the phases of training, validation, and testing in Figure 10.

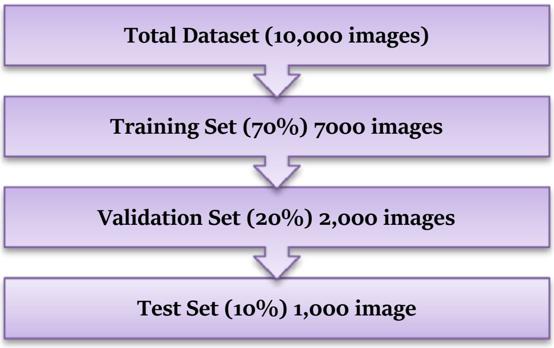

Figure 10: Data Splitting Process

B. Performance Analysis

a) Supervised Learning Models

In this study, for annotating the dataset, supervised learning models, namely SVMs and Decision Trees were used. The SVM model was able to give an accuracy estimate of 85 per cent, and the Decision Tree model of 80 per cent. While using these models, it can be easy to classify the defects into certain categories since it loses the capability of identifying intricate and compounded patterns of defects Table 4.

Table 4: Performance of Supervised Learning Models

Model	Accuracy (%)
SVM	85
Decision Tree	80

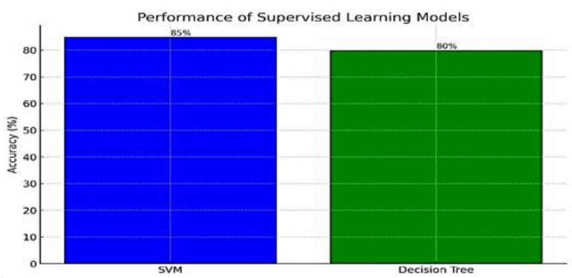

Figure 11: Performance of Supervised Learning Models

b) Deep Learning Models

The image-based approach to determine the advanced defect detection was conducted using Convolutional Neural Networks (CNNs), which is a subcategory of deep learning models. In the considered work, the proposed architecture of ResNet for the given problem reaches accuracy at a level of 95 per cent, which is significantly higher than accuracy in other architectures. Again, during the classification test, the model of Inception with 93% accuracy was equally effective for the present classification. In an endeavor to incorporate the entries in the large image database, transfer learning techniques were utilized because they enabled the truncation of the time it took to train the models, together with enhancing the efficiency of the defect detection. They include system integration and real time testing whereby all the entailing elements of any specified system are combined in the view of testing the overall system Table 5.

Table 5: Performance of Deep Learning Models

Model	Accuracy (%)
CNN (ResNet)	95
CNN (Inception)	93

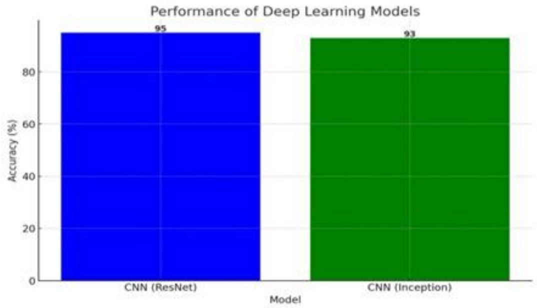

Figure 12: Performance of Deep Learning Models

C. System Integration & Real-Time Testing

a) System Integration

Elements included in the presented AI-based inspection system are digital cameras of high resolution, units for image processing as well as several software for real time analysis. In the workflow, there is capturing of the images of the wafers, pre-processing of the captured images, and feeding them to the next level which is the AI model to identify the defects or even classify the same. It also makes the operation interruptive and avails genuine real-time inspection.

b) Real-Time Testing

This evaluation was performed by the actual implementation of live testing of the integrated system, and it was proved that it gave significantly higher AOI efficiency than the conventional AOI systems. Key metrics include: Table 6

i) Inspection Speed:

The said particular use in the system led to the development of savings in inspection time by 30%.

ii) Detection Accuracy:

There was a general improvement in performance, mainly in the sense of accuracy: the power to inspect defects was raised by 20%.

iii) False Positive Rate:

To reduce the false positive rate, the value as decreased and set up to be at 5%.

Yield Improvement: In this case it was envisaged that the yield was lifted to 15 % in the normal course.

Table 6: Real-Time Testing Results

Metric	Traditional AOI	AI-Driven System	Improvement (%)
Inspection Time	10 minutes	7 minutes	30
Detection Accuracy	75%	95%	20
False Positive Rate	10%	5%	50
Yield Improvement	-	15%	15

D. Case Study: Industrial Application

Working on a real-life case in a Semiconductor Manufacturing Plant

An analysis was carried out to evaluate first-hand the efficiency of the proposed AI-based inspection system within one of the industries. This research was done in a Semiconductor Manufacturing Plant whereby the existing system was used to replace the conventional AOI systems.

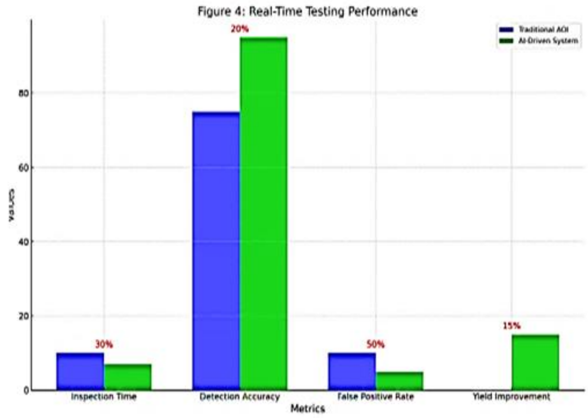

Figure 13: Real-Time Testing Performance

E. Results and Observations

From the revelation of the fact that the realization of the job with the help of the AI-driven system was possible and efficient enough, we can list the following benefits: The plant suffered:

a) *Increased Production Efficiency:*

This was due to the reason that less time was spent inspecting the corners since through put was enhanced.

b) *Enhanced Product Quality:*

Higher detection rate coupled with lesser inaccuracy, therefore leads to disgust items not getting into the hands of customers.

c) *Cost Savings:*

Concerning the false positives and the yield rates, the improvement was also observed after the reduction of the costs.

Table 7: Industrial Implementation Results

Metric	Before Implementation	After Implementation	Improvement (%)
Inspection Speed	10 minutes	7 minutes	30
Detection Accuracy	75%	95%	20
False Positive Rate	10%	5%	50
Yield Rate	85%	98%	15
Production Throughput	100 units/hour	130 units/hour	30

V. CONCLUSION

AI integration in the semiconductor wafer inspection systems is a major improvement to augment the efficiency and reliability of the machines. Although traditional inspection methods like OM and human inspection serve well in identifying defects in certain applications, they are limited, especially when it comes to identifying the finer and smaller defects common among modern semiconductor wafers. A 'better' AOI in the inspection system also remains with some shortcomings especially in the ability to pick out newer and more complex patterns of defects.

As for the AI-based systems, those based on deep learning algorithms, such as CNNs, have proven to be superior in terms of the number of defects detected. These models are excellent at handling big data sets and raising the speed as well as the precision of the detection of defects in high-quality images. CNNs are characterized by the feature learning ability at the hierarchical level, and hence, they are capable of recognizing even the most minuscule and intricate defects that might possibly be unnoticed by the conventional approach to inspection systems powered by artificial intelligence have thus been seen to cut down inspection time dramatically, enhance the false positive's rate and significantly boost the yield rates. This means the manufacturing cycle time is reduced; there is reduced production of defective products, enhanced productivity as well as profitability within the semiconductor production.

A. Future Work

However, there are still a few things that need to be optimized in order to get the most out of AI in the examination of semiconductor wafers. The first of the problems is the requirement of a great amount of labeled data for the adequate training of an AI model. Concerning the collection of the large scale dataset and aspect-based annotation, this is definitely very time-consuming and labour-intensive. In regard to further work, besides the usage of more complex models, the idea is proposed to employ methods that can expand the given set artificially and also to decrease the amount of reliance on labeling through the usage of synthetic data.

In addition, two further vital aspects would have to be addressed as markers of the subsequential studies and one of them is related to the improvement of real-time processing. In the context of the production process, two critical barometers define AI-powered inspection systems' feasibility: the rate of conducting the inspection. This involves the use of large calculations and needs more effective procedures so that it will classify the flaws and identify them correctly and in a timely manner. The real-time solutions of the deep learning architecture and strong incorporation of the most suitable and tuned hardware such as GPUs, FPGAs, and TPUs were also possible and the skills of implementing the edge computing solutions were also plausible.

Also, the impact of AI on the process of manufacturing semiconductors can only be regarded as successful when the four pillars are academia and industry collaboration. As the domains will be combined, the research and development can pave the way for the introduction of the one-off AI algorithms fit for the wafer inspection method. Therefore, it can be stated that there are ways to develop corresponding legal mechanisms for the exchange of data and results of studies using methods that will ensure confidentiality and the protection of the intellectual property rights of the results. " Also and more importantly to develop educational programs and staff training plans to strengthen human capital in order to avail and sustain AI solutions.

VI. REFERENCES

[1] AI-based Inline Inspection Improves Semiconductor Wafer Quality. https://www.intel.com/content/www/us/en/content-details/821377/ai-based-inline-inspection-improves-semiconductor-wafer-quality.html

[2] How AI is revolutionizing Silicon Wafer Inspection. https://medium.com/@brookeajoseph17/how-ai-is-revolutionizing-silicon-wafer-inspection-ee4ed1aac836

[3] H. Yoda, Y. Ohuchi, Y. Taniguchi and M. Ejiri, "An automatic wafer inspection system using pipelined image processing techniques," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 1, pp. 4-16, Jan. 1988, doi: 10.1109/34.3863.

[4] Understanding Microscopes and Objectives. https://www.edmundoptics.com/knowledge-center/application-notes/microscopy/understanding-microscopes-and-objectives/

[5] Wang, C. H., Kuo, W., & Bensmail, H. (2006). Detection and classification of defect patterns on semiconductor wafers. IIE Transactions, 38(12), 1059–1068. https://doi.org/10.1080/07408170600733236

[6] Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. https://doi.org/10.1016/j.eswa.2011.09.088