International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 8 Issue 27 (December 2026) PP. 01-24



INTERNATIONAL JOURNAL OF INNOVATION AND INDUSTRIAL REVOLUTION (IJIREV)
www.gaexcellence.com/ijirev

GLOBAL ACADEMIC EXCELLENCE

EVOLUTION OF WAFER DEFECT DETECTION IN PHOTOLITHOGRAPHY: FROM CONVENTIONAL IMAGE PROCESSING TO DEEP LEARNING

Noraiza Aqilah Abdul Aziz[1], Zakaria Hussain[2*], Muhammad Khusairi Osman[3], Saiful Zaimy Yahaya[4]

[1]School of Electrical Engineering, Universiti Teknologi MARA (UiTM) Pulau Pinang, Malaysia
noraizaaqilah@gmail.com https://orcid.org/0009-0002-1159-250X
[2]School of Electrical Engineering, Universiti Teknologi MARA (UiTM) Pulau Pinang, Malaysia
zakaria183@uitm.edu.my https://orcid.org/0000-0002-2605-2621
[3]School of Electrical Engineering, Universiti Teknologi MARA (UiTM) Pulau Pinang, Malaysia
khusairi@uitm.edu.my https://orcid.org/0000-0002-5479-6409
[4]School of Electrical Engineering, Universiti Teknologi MARA (UiTM) Pulau Pinang, Malaysia
saiful053@uitm.edu.my https://orcid.org/0000-0002-6906-5086
*Corresponding Author

Article Info:

Article history:
Received date: 28.09.2026
Revised date: 13.10.2026
Accepted date: 02.11.2026
Published date: 01.12.2026

To cite this document:

Aziz, N. A. A., Hussain, Z., Osman, M. K., & Yahaya, S. Z. (2026). Evolution of Wafer Defect Detection in Photolithography: From Conventional Image Processing to Deep Learning. *International Journal of Innovation and Industrial Revolution, 8*(27), 01-24.

Abstract:

The increasing complexity of semiconductor manufacturing and the continuous miniaturisation of integrated circuits have intensified the need for accurate, automated, and robust wafer defect detection in photolithography. Conventional image processing techniques have provided the foundation for automated inspection; however, their dependence on handcrafted features and predefined rules limits their adaptability to complex defect patterns, process variations, and high-volume manufacturing conditions. This study presents a chronological review of the technological evolution of wafer defect detection from conventional image processing to advanced deep learning approaches. A systematic literature search was conducted using the Scopus database, applying predefined search terms, inclusion criteria, and exclusion criteria to identify relevant primary research articles. A total of 37 studies published between 1989 and 2026 were selected and analysed according to their chronological development, inspection methodologies, computational approaches, and technological contributions. The literature was organised into three temporal phases: Phase I: Pioneering and Foundational Research (1989–2019), characterised by knowledge-based inspection, conventional image processing, handcrafted features, and early machine learning; Phase II: Transition toward Intelligent Defect Detection (2021–2022), marked by

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 8 Issue 27 (December 2026) PP. 01-24

the increasing adoption of computer vision, machine learning, and convolutional neural networks; and Phase III: Rapid Advancement and Deep Learning Expansion (2023–2026), characterised by lightweight networks, attention mechanisms, transformer architectures, anomaly detection, synthetic data, multimodal learning, and real-time inspection. The findings demonstrate a clear shift from handcrafted feature engineering toward adaptive, data-driven, and integrated intelligent inspection systems. Nevertheless, challenges remain in dataset availability, model generalisation, false-positive reduction, computational efficiency, explainability, and high-volume manufacturing deployment. This review provides a chronological perspective of technological development and identifies key research gaps that can guide the development of robust, scalable, and industry-ready wafer defect detection systems for next-generation semiconductor manufacturing.

DOI: 10.35631/IJIREV.827001

Keyword:

Deep Learning, Photolithography, Wafer Defect

© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact ijirev@gaexcllence.com.

Introduction

The evolution of wafer defect detection in photolithography has moved from manual visual checks and simple image differencing to multi-modal optical platforms, electron-beam review, and AI-driven classification tightly linked to yield control. Photolithography sits at the core of semiconductor manufacturing, and shrinking device dimensions have made wafer defect detection a central problem for yield, reliability, and cost control (Basu et al., 2025; Magnanini et al., 2025; Ohri et al., 2024). As critical dimensions continue to decrease and circuit layouts become more intricate, inspection systems face increasing pressure to balance sensitivity, specificity, throughput, and cost of ownership (Reut et al., 2025; Zhu et al., 2022). Defects on wafers directly affect device performance and can propagate into yield loss if they are missed at early process stages (Jianhong et al., 2023; W. Liu et al., 2024),(Ohri et al., 2024). This pressure has intensified with sub-10 nm, 4 nm, and newer nodes, where even conventional high-resolution inspection methods operate close to sensitivity limits and face stronger noise, material, and pattern-complexity constraints (Jianhong et al., 2023; Lechien et al., 2023; Zhu et al., 2022). The historical development of defect detection is therefore inseparable from the broader scaling of lithography itself, including EUV, high-NA systems, and alternative patterning platforms such as nanoimprint lithography (Basu et al., 2025; Reut et al., 2025).

In the earliest stage of this evolution, wafer inspection depended heavily on human expertise and classical machine-vision methods, which were practical for larger and more visually distinct defects but struggled with speed, objectivity, and generalization (Ji et al., 2025; Jianhong et al., 2023; Yu et al., 2023). Traditional approaches relied on template matching, wavelet transforms, threshold segmentation, spatial filtering, principal component analysis,

and reference-image subtraction, making performance highly dependent on image quality and on assumptions about defect morphology (Jianhong et al., 2023; Yu et al., 2023). These methods were often tailored to specific defect classes and therefore generalized poorly across wafers, pattern types, and process conditions (Jianhong et al., 2023; Yu et al., 2023). Even so, this period established a lasting conceptual foundation: defect detection was framed as the localization of anomalous signal against a nominal background, whether on bare wafers dominated by particles and scratches or on patterned wafers complicated by topography and materials contrast (Yang et al., 2025; Zhu et al., 2022).

As pattern density increased, optical inspection became more sophisticated rather than obsolete. For bare wafers, dark-field scattering remained the workhorse because particles and scratches generate strong high-frequency scattering signatures (Zhou et al., 2025; Zhu et al., 2022). Recent spot-scanning multimodal imaging demonstrates how this hardware evolution improves practical sensitivity, sub-micron scratches, and stains below 20μm on unpatterned wafers (Zhou et al., 2025). For patterned wafers, inspection has become a multi-step and multi-tool exercise because double, triple, and quadruple patterning increase the number of inspection points while also increasing the risk that missed events transfer downstream (Zhu et al., 2022). At advanced EUV nodes, defectivity is no longer only a question of visible particles; bridging, line collapse, buried residue, and chemical-interaction defects demand combined optical inspection, e-beam review, and process-specific monitors rather than conventional silicon particle monitors alone (Ohri et al., 2024; Yang et al., 2025).

The next major transition was from hand-crafted rules to machine learning and then to deep learning. Reviews now describe the field as progressing through three broad stages: image signal processing, classical machine learning, and deep learning (Jianhong et al., 2023; Lechien et al., 2023). Classical machine learning improved robustness in some settings, with methods such as KNN, decision trees, and SVM used for wafer defect recognition and classification (Jianhong et al., 2023; M. Li et al., 2025). Deep learning then became attractive because it learns hierarchical image features directly, reducing reliance on manual feature engineering and improving objectivity and speed (Jianhong et al., 2023; Zhu et al., 2022). CNNs, segmentation networks, transformers, and YOLO-family detectors have all been applied to wafer or SEM inspection, with particular emphasis on small-defect recognition, lightweight deployment, and real-time inference (Ji et al., 2025; Samir et al., 2024; Shinde et al., 2025). However, the literature is equally clear that the hardest problems remain data scarcity, defect unpredictability, class imbalance, and the weak visual signatures of very small defects, which occupy only tiny image areas and are easily confused with nuisance noise (Shinde et al., 2025).

The current phase of wafer defect detection in photolithography is defined less by any single sensor or algorithm than by integrated inspection ecosystems. Synthetic-data training is being used to compensate for the shortage of label defect images, and in one recent EUV-related study YOLOv8 trained on artificial SEM data achieved strong small-defect detection and transferred reasonably to real images (Shinde et al., 2025). Unsupervised anomaly detection, supervised segmentation, and dual-stage fusion models are being explored to support high-throughput deployment while incorporating defect severity into pass-fail decisions (Façoco et al., 2025). At the fab level, AI is now being coupled to scanner topography maps for virtual metrology on every wafer, to automated hardware matching across inspection fleets, and to defect-aware process qualification frameworks for 2 nm, 1.4 nm, NIL, and EUV 3D integration (Bose, Pandey, Staib, Womach, Balla, Olmedo, Khandelwal, Ahmadi, et al., 2025; J.-B. Kim et al.,

2026; Reut et al., 2025; Shin et al., 2025; Yang et al., 2025). Across this trajectory, the field has evolved from detecting obvious surface flaws to building predictive, defect-aware control loops that connect inspection signals to root-cause analysis, model calibration, and yield learning (Bose, Pandey, Staib, Womach, Balla, Olmedo, Khandelwal, Lerner, et al., 2025; Magnanini et al., 2025; Samir et al., 2024). The introduction to the evolution of wafer defect detection in photolithography is the transformation of inspection itself: from isolated observation to data-driven process intelligence for advanced semiconductor manufacturing.

Literature Review

Evolution of Wafer Defect Detection Technologies

The increasing complexity of semiconductor manufacturing and the continuous reduction of feature dimensions have made defect inspection an essential component of process control and yield management. In photolithography, defects can originate from patterning errors, particles, process variation, resist-related abnormalities, and imperfections introduced during subsequent manufacturing stages. As technology nodes continue to shrink, inspection systems must detect increasingly small and complex defects while maintaining high throughput and low false-positive rates. Recent developments in lithography, particularly extreme ultraviolet (EUV) lithography, have further intensified these requirements because defectivity and process-window control become increasingly critical at advanced technology nodes (Basu et al., 2025; Ohri et al., 2024; Yang et al., 2025; Zhu et al., 2022) . Consequently, wafer inspection has progressively evolved from conventional image-processing techniques toward machine learning, deep learning, multimodal inspection, and intelligent manufacturing systems.

The early development of automated wafer inspection was primarily based on knowledge-driven and image-processing approaches. One of the earliest studies identified in this review was conducted by Ejiri et al. (Ejiri et al., 1989), who proposed a knowledge-directed inspection approach for complex multilayered patterns. This work represents an important foundation for automated semiconductor inspection because the inspection process relied on predefined knowledge and image characteristics to identify abnormalities in patterned structures. At this stage, defect detection was strongly dependent on human-defined rules, image characteristics, and domain knowledge. Such approaches were suitable for relatively controlled inspection environments but became increasingly difficult to maintain as pattern structures and defect morphologies became more complicated.

The subsequent development of machine learning introduced a transition from explicitly programmed inspection rules toward data-driven defect identification. (Chang et al., 2009) proposed an unsupervised neural network approach for automatic semiconductor wafer defect inspection, demonstrating the potential of neural computation to reduce dependence on manually designed inspection rules. Similarly, (Y.-H. Liu & Chen, 2011) investigated automatic defect detection using a support vector data description approach, demonstrating the application of machine learning to automated defect identification. These studies represented an important intermediate stage between conventional image processing and modern deep learning because the inspection system could learn patterns from data rather than relying exclusively on manually specified rules.

By 2019, data-driven approaches had become more prominent in optical patterned defect detection. (Henn et al., 2019) investigated data-driven methods for optical patterned defect

detection, reflecting the increasing use of statistical learning to improve automated inspection. This development was further strengthened by the emergence of convolutional neural networks (CNNs), which demonstrated strong capabilities in learning visual features directly from images. Although some early CNN applications were developed outside semiconductor wafer inspection, they provided methodological foundations for later semiconductor applications. For example, (Westphal & Seitz, 2021) demonstrated CNN-based defect detection and visualization in additive manufacturing, while (Zheng et al., 2021) reviewed the growing application of deep learning to surface defect inspection in industrial products. These developments established deep learning as a promising solution to the limitations associated with handcrafted features and conventional machine learning.

Transition from Conventional Inspection to Intelligent Defect Detection

The period from approximately 2020 to 2022 can be considered a transition toward intelligent wafer defect detection. During this period, CNN architectures became increasingly integrated into semiconductor inspection, while research also began addressing mixed defect patterns, feature selection, and the need for computationally efficient models. (Wang et al., 2020), for example, developed deformable convolutional networks for mixed-type wafer defect pattern recognition. The use of deformable convolution demonstrated an important methodological advancement because conventional convolution operations may have difficulty representing irregular or spatially variable defect patterns. Adaptive feature extraction therefore provided greater flexibility for handling complex wafer defect structures.

Research also increasingly focused on improving the quality and relevance of features used for classification. (Gomez-Sirvent et al., 2022) investigated optimal feature selection for semiconductor wafer defect classification, indicating that feature representation remained an important consideration even as machine learning techniques became more advanced. (Lin et al., 2021) applied deep learning and pseudo-colorization to defect identification in photovoltaic modules, while (Zheng et al., 2021) documented the rapid expansion of deep learning methods for industrial surface inspection. These studies collectively indicate that the transition to deep learning was not an isolated development within semiconductor manufacturing but part of a broader movement toward data-driven visual inspection.

The emergence of deep learning also encouraged researchers to consider architectures beyond conventional CNNs. Knowledge-based and hybrid approaches became particularly relevant because purely data-driven models can require large amounts of labelled data and may lack explicit process knowledge. (Wang et al., 2023) proposed a knowledge-augmented broad learning system for mixed-type defect detection in semiconductor manufacturing, demonstrating an alternative direction that combines domain knowledge with machine learning. Such hybrid approaches are particularly relevant to semiconductor manufacturing because inspection systems operate under strict process constraints and require reliable decisions rather than accuracy alone.

Rapid Expansion of Deep Learning-Based Inspection

From 2023 onward, the literature demonstrates a marked acceleration in the application of deep learning to semiconductor and wafer defect detection. Research during this period moved beyond basic classification toward defect localization, segmentation, mixed-defect recognition, anomaly detection, and real-time inspection. (Jianhong et al., 2023) reviewed wafer surface

defect detection methods and highlighted the increasing importance of intelligent approaches for semiconductor manufacturing. Similarly, (Dehaerne et al., 2025) and (Lechien et al., 2023) systematically reviewed automated semiconductor defect inspection using scanning electron microscopy, demonstrating the growing importance of automated image interpretation for high-resolution semiconductor inspection.

One important development was the use of lightweight deep learning models to address the computational requirements of industrial inspection. (López de la Rosa et al., 2023) combined geometric transformation-based data augmentation with SqueezeNet for semiconductor wafer defect detection and classification. The use of a lightweight CNN demonstrates the shift from pursuing detection accuracy alone toward balancing accuracy with computational efficiency. Lightweight architectures are particularly important for high-throughput manufacturing environments where inspection must be performed rapidly and potentially close to the production line.

Other studies focused on improving feature extraction and segmentation capabilities. (Y. Kim et al., 2023) employed a pre-trained CNN-based TransUNet architecture for mixed-type defects in wafer maps, reflecting the emergence of transformer-based and hybrid CNN-transformer architectures. (Wang et al., 2023) investigated knowledge-augmented learning for mixed-type semiconductor defects, while (Solorzano & Tsai, 2023) explored cosine-convolutional semantic networks. These approaches illustrate the diversification of deep learning architectures beyond conventional CNNs.

The same period also saw increasing attention to specialized inspection configurations. (Fukuda et al., 2023) investigated trainable die-to-database inspection for large-field-of-view electron-beam inspection, demonstrating the application of trainable approaches to high-resolution semiconductor inspection. (Yu et al., 2023) roposed a full-flow machine-vision inspection method for wafer surface defects, indicating a movement toward integrated inspection pipelines rather than isolated image-classification tasks. This development is significant because practical semiconductor inspection requires the integration of image acquisition, preprocessing, defect detection, classification, and decision-making.

Expansion toward Advanced and Real-Time Inspection

Research published during 2024–2026 demonstrates a further shift toward optimized, lightweight, high-resolution, and real-time inspection systems. (D. Kim, 2024) investigated optimized operation methods for wafer surface defect detection, while (W. Liu et al., 2024) incorporated multimodal feature fusion for semiconductor wafer particle defect classification. The integration of multiple sources of information reflects the increasing recognition that complex semiconductor defects may not always be adequately represented by a single image modality.

Transformer-based approaches have also become increasingly relevant. (Samir et al., 2024) investigated transformer-based defect detection and classification using CD-SEM images, reflecting the growing use of attention-based architectures for high-resolution semiconductor inspection. At the same time, (He et al., 2025) proposed an improved coordinate attention mechanism for semiconductor wafer defect recognition, demonstrating continued efforts to enhance feature representation within CNN-based architectures. (M. Li et al., 2025) further

explored an improved YOLOv7-Tiny architecture, emphasizing the importance of real-time detection and computational efficiency.

Segmentation and anomaly detection have similarly gained attention. (Façoco et al., 2025) explored lightweight approaches for wafer visual inspection, particularly anomaly detection and defect segmentation, while (Ozcelik & Alimli, 2025) investigated wafer defect segmentation using U-Net++. These studies demonstrate a shift from simply determining whether a defect exists toward identifying its precise location and morphology. Such capabilities are increasingly important for process diagnosis and root-cause analysis.

Another important development is the increasing use of synthetic data and data augmentation to address limitations in defect datasets. (Shinde et al., 2025) investigated deep learning models for photolithographic pattern defect detection using synthetic data, while (Friedrich et al., 2026) examined simulated wafer-dicing faults as an augmentation strategy for deep learning-based visual inspection. These studies address one of the major challenges in semiconductor inspection: the difficulty of obtaining sufficiently large and representative datasets containing rare defect types. Synthetic data can therefore improve model training while reducing dependence on costly manual defect collection and annotation.

Recent research has also expanded from image-level defect recognition toward system-level integration. (Fu et al., 2025) investigated the application of artificial intelligence to improve chip defect detection using semiconductor equipment, while (Noh, 2025) proposed a real-time operator-centric patch extraction and analysis platform for semiconductor wafer processes. (Chen et al., 2026) further demonstrated system-level integration of deep learning and computer vision for semiconductor manufacturing. These developments indicate that current research is increasingly concerned not only with model performance but also with how intelligent inspection technologies can be integrated into manufacturing operations.

At the process level, virtual metrology and inspection have also emerged as complementary technologies. (Bose, Pandey, Staib, Womach, Balla, Olmedo, Khandelwal, Ahmadi, et al., 2025) proposed a virtual defect metrology system using photolithography scanner topography maps to detect gross defects, predict inline yield, and identify defect-generating tools. Similarly, (Shin et al., 2025) investigated the matching of optical metrology and inspection equipment for yield enhancement. These studies demonstrate a broader transformation in semiconductor inspection from isolated defect detection toward predictive and process-aware quality management.

The importance of advanced inspection is particularly evident at leading-edge technology nodes. (Zhu et al., 2022) investigated optical wafer defect inspection at the 10 nm technology node and beyond, while (J.-B. Kim et al., 2026) addressed AI-based optical inspection for critical defects at 2 nm and 1.4 nm technology nodes. (Reut et al., 2025) examined inspection sensitivity and cost-of-ownership considerations in nanoimprint lithography, while (Yang et al., 2025) investigated defectivity-aware EUV process-window characterization. These studies indicate that future inspection systems must simultaneously achieve high sensitivity, high throughput, low false-positive rates, and economic feasibility.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 8 Issue 27 (December 2026) PP. 01-24

Current Research Direction and Emerging Challenges

The literature from 1989 to 2026 demonstrates a clear technological progression from knowledge-directed image processing toward increasingly autonomous, data-driven, and intelligent inspection systems. The earliest approaches relied heavily on predefined rules and domain knowledge (Ejiri et al., 1989), followed by unsupervised neural networks and conventional machine learning (Chang et al., 2009; Y.-H. Liu & Chen, 2011). The introduction of data-driven optical inspection (Henn et al., 2019) and the subsequent adoption of CNNs and deep learning established a new paradigm in which visual features could be learned automatically from inspection data. More recent research has expanded this paradigm through deformable convolutions, transformer-based models, attention mechanisms, multimodal feature fusion, lightweight networks, synthetic data, anomaly detection, and real-time industrial deployment (Façoco et al., 2025; Friedrich et al., 2026; He et al., 2025; Y. Kim et al., 2023; M. Li et al., 2025; W. Liu et al., 2024; López de la Rosa et al., 2023; Samir et al., 2024; Shinde et al., 2025).

Despite these advances, several challenges remain. First, the increasing complexity of semiconductor patterns creates difficulties in detecting extremely small, irregular, and mixed-type defects. Second, labelled defect datasets are often limited because critical defects can be rare and expensive to collect. Third, highly accurate deep learning models may impose computational and deployment constraints in high-throughput manufacturing environments. Fourth, false positives and over-detection remain important concerns because unnecessary inspection alarms can reduce manufacturing efficiency. Finally, the transition from laboratory-level model validation to robust industrial deployment requires consideration of process variation, equipment differences, changing technology nodes, and system-level integration.

Recent review studies reinforce the need to address these issues systematically. (Akele & Gurau, 2026) highlighted the broader development of machine learning-powered vision inspection in semiconductor manufacturing, while (J. Li et al., 2026) examined wafer defect detection from the perspectives of algorithms, systems, and data. (Shi, 2026) further reviewed deep learning-based surface defect detection for microdevices. These recent reviews demonstrate that the field has reached a level of maturity in which algorithmic performance must increasingly be considered together with data availability, system architecture, inspection equipment, and manufacturing requirements.

Overall, the literature reveals a fundamental transformation in wafer defect inspection: from rule-based and handcrafted image analysis to machine-learning-assisted recognition, and finally to deep learning-enabled intelligent inspection and process integration. However, the progression has not been uniform. Different technologies have continued to coexist because conventional image processing remains useful for structured and well-defined defects, whereas deep learning provides greater flexibility for complex and variable defect patterns. Therefore, the current research direction is not simply the replacement of conventional methods by deep learning, but the development of integrated inspection frameworks that combine image processing, machine learning, deep learning, domain knowledge, multimodal sensing, virtual metrology, and real-time manufacturing control. This technological trajectory provides the basis for a chronological analysis of wafer defect detection and highlights the continuing need for robust, explainable, computationally efficient, and industrially deployable inspection systems.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 8 Issue 27 (December 2026) PP. 01-24

Material and Method

Research Design

This study employed a structured chronological literature review to examine the evolution of wafer defect detection in photolithography from conventional image processing to machine learning and deep learning-based inspection. The review was designed to identify not only changes in publication activity but also technological and methodological transitions across different periods. In contrast to a conventional literature review that primarily summarizes existing knowledge, the present study adopts a temporal perspective to reconstruct the development of wafer defect detection technologies and to establish relationships between earlier inspection approaches and subsequent technological innovations.

The review process consisted of four main stages: (1) Data source and Literature Search Strategy, (2) Study Selection and Data Extraction, (3) Chronological Clustering of the Literature and (4) Analytical Framework. The literature review was used as the conceptual basis for identifying the major technological dimensions of wafer defect detection, including conventional image processing, machine learning, convolutional neural networks (CNNs), advanced deep learning architectures, defect segmentation, anomaly detection, multimodal inspection, and real-time industrial inspection.

Data Source and Literature Search Strategy

Scopus was selected as the primary bibliographic database because of its broad coverage of peer-reviewed literature in engineering, computer science, semiconductor manufacturing, materials science, and related technological disciplines. The database was considered suitable for this review because wafer defect detection is an interdisciplinary research area involving image processing, computer vision, artificial intelligence, manufacturing systems, and semiconductor process control.

A structured single-step search strategy was conducted using the **TITLE-ABS-KEY** field in Scopus. The search strategy incorporated four conceptual groups connected using the Boolean operator **AND**. The first group represented the wafer inspection domain and included *chronology*, *wafer defect detection*, *defect detection*, *wafer inspection*, and *wafer quality*. The second group represented the semiconductor manufacturing context and included *semiconductor*, *semiconductor industry*, and *semiconductor manufacturing*. The third group represented intelligent computational techniques and included *deep learning*, *machine learning*, *AI*, *neural network*, and *artificial intelligence*. The fourth group represented image-based inspection approaches and included *image processing*, *traditional image processing*, *conventional image processing*, and *computer vision*.

The search strategy was designed to capture literature representing both the earlier foundations of automated inspection and the subsequent transition toward intelligent defect detection. This approach was particularly important for a chronological review because restricting the search exclusively to contemporary deep learning terminology could overlook earlier research based on knowledge-directed inspection, conventional image processing, and traditional machine learning.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 8 Issue 27 (December 2026) PP. 01-24

The final Scopus search string was:
TITLE-ABS-KEY ((chronology OR "wafer defect detection" OR "defect detection" OR "wafer inspection" OR "wafer quality") AND (semiconductor OR "semiconductor industry" OR "semiconductor manufacturing") AND ("deep learning" OR "machine learning" OR AI OR "neural network" OR "artificial intelligence") AND ("image processing" OR "traditional image processing" OR "conventional image processing" OR "computer vision")) AND LIMIT-TO(LANGUAGE, "English") AND LIMIT-TO(SUBJAREA, "ENGI") AND LIMIT-TO(SRCTYPE, "j")**

The search was restricted to English-language journal articles within the Engineering subject area. Following the application of the eligibility criteria, **37 primary research articles** were retained for the chronological analysis.

Study Selection and Data Extraction

The selected publications were organized according to publication year and examined individually to identify their technological and methodological contributions. Information extracted from each study included publication year, inspection objective, inspection modality, image-processing technique, machine learning or deep learning approach, model architecture, feature-extraction strategy, defect classification or detection method, segmentation or localization method, data-generation or augmentation strategy, and reported application context.

Where available, information concerning detection accuracy, computational efficiency, real-time capability, robustness, data requirements, and industrial applicability was also recorded. These characteristics were not treated as directly comparable quantitative measures because the studies differed substantially in datasets, defect types, imaging systems, experimental conditions, and evaluation metrics. Instead, they were used to identify methodological improvements and changes in research priorities within their respective technological contexts.

Chronological Clustering of the Literature

The selected studies were grouped according to publication year and technological characteristics. Chronological clustering was performed to transform individual publication records into broader periods representing meaningful stages of technological development. The phase boundaries were determined by considering both temporal publication patterns and methodological transitions identified through the literature analysis.

The selected studies were grouped according to publication year and technological characteristics. Chronological clustering was performed to transform individual publication records into broader periods representing meaningful stages of technological development. The phase boundaries were determined by considering both temporal publication patterns and methodological transitions identified through the literature analysis.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 8 Issue 27 (December 2026) PP. 01-24

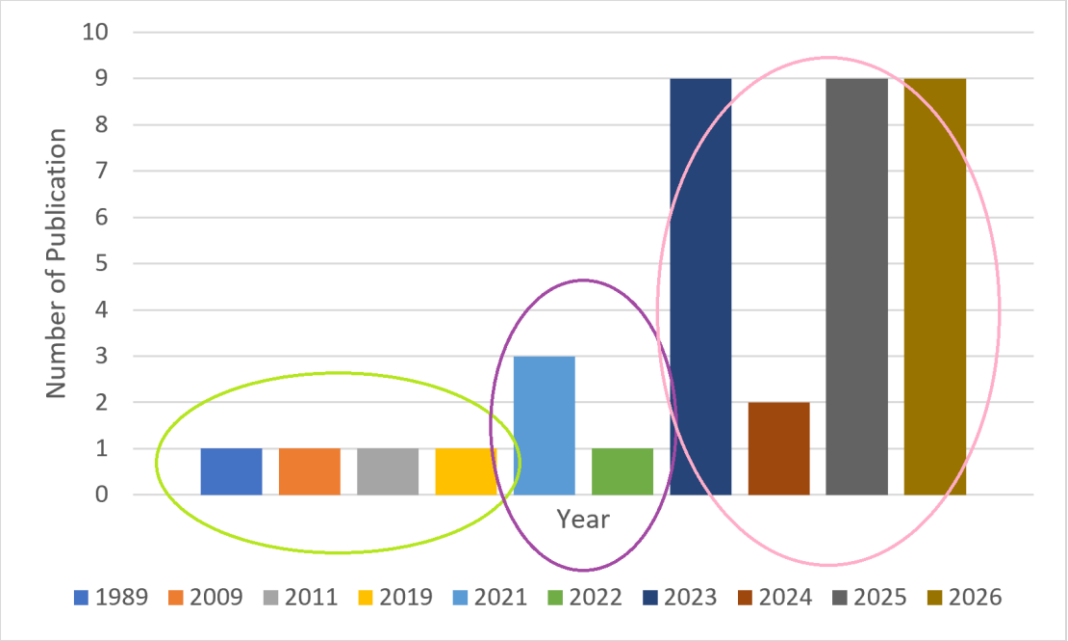


Figure 1: Number of Publication from year 1989 to August 2026

Phase I: Pioneering and Foundational Research (1989–2019)

This phase represents the earliest stage in the evolution of wafer defect detection research. Scientific output remained limited and sporadic, reflecting the exploratory nature of the field and the technological constraints of the period. Research primarily focused on conventional image processing techniques, including thresholding, edge detection, morphological operations, and handcrafted feature extraction, to identify defects in semiconductor wafers. Although these methods established the fundamental principles of automated optical inspection, they were generally sensitive to variations in illumination, noise, and increasingly complex defect patterns, thereby limiting their applicability to next-generation semiconductor manufacturing.

Phase II: Transition toward Intelligent Defect Detection (2021–2022)

The second phase marks the beginning of a technological transition from conventional image processing toward intelligent data-driven approaches. The gradual increase in publication activity reflects growing awareness of the limitations of handcrafted feature engineering and the emergence of machine learning as a promising alternative. Advances in computing hardware, increased data availability, and the expanding adoption of artificial intelligence in industrial manufacturing enabled researchers to investigate learning-based models capable of improving defect recognition accuracy, robustness, and automation. This period laid the methodological foundation for the widespread adoption of deep learning in subsequent years.

Phase III: Rapid Advancement and Deep Learning Expansion (2023–2026)

The final phase is characterized by an exponential increase in research productivity, demonstrating the rapid maturation of deep learning for wafer defect detection. The substantial growth in publications reflects the increasing demand for intelligent inspection systems capable of addressing the complexity of modern photolithography processes and advanced

semiconductor nodes. Research during this period extends beyond conventional convolutional neural networks to encompass transfer learning, attention mechanisms, transformer architectures, explainable artificial intelligence, and real-time industrial deployment. This sustained growth indicates that deep learning has become the dominant technological paradigm, with current research emphasizing robust, scalable, and highly automated inspection systems to support the next generation of semiconductor manufacturing.

The chronological phases were therefore established through a combination of publication chronology and technological characteristics, rather than solely according to the number of publications in each year.

Analytical Framework

A technology-oriented analytical framework was developed from the literature review to guide the chronological synthesis. Four principal dimensions were used.

First, inspection technology considered the evolution of inspection modalities, including conventional optical inspection, machine vision, electron-beam and SEM-based inspection, multimodal inspection, and virtual metrology. This dimension was used to determine how advances in image acquisition and inspection systems influenced defect-detection capabilities.

Second, computational methodology examined the progression from knowledge-based rules and handcrafted image features to statistical learning, neural networks, CNNs, deformable convolutions, attention mechanisms, transformer architectures, hybrid models, and other deep learning approaches. This dimension provided the principal basis for evaluating the transition from conventional image processing to intelligent inspection.

Third, defect-analysis objective considered the evolution from basic defect detection toward classification, localization, segmentation, anomaly detection, mixed-defect recognition, defect prediction, and process diagnosis. This dimension was used to determine whether research objectives became increasingly comprehensive as computational methods developed.

Fourth, industrial applicability examined computational efficiency, real-time capability, robustness, scalability, equipment integration, yield improvement, and applicability to advanced technology nodes. This dimension was included because high detection accuracy alone does not guarantee suitability for semiconductor manufacturing environments.

The four dimensions were analysed across the three chronological phases to identify relationships between technological development and changes in research objectives. The resulting synthesis was then used to explain how limitations of earlier approaches contributed to the development and adoption of subsequent methods.

Result and Discussion

Phase I: Pioneering and Foundational Research (1989–2019)

The pioneering phase of wafer defect detection research was characterized by the gradual transition from conventional image processing toward intelligent computational approaches capable of addressing the increasing complexity of semiconductor manufacturing. Although

research activity remained limited throughout this period, the published studies collectively established the conceptual and methodological foundation that later enabled the adoption of advanced artificial intelligence techniques. Early investigations primarily concentrated on improving inspection efficiency through deterministic image processing algorithms, whereas subsequent studies progressively incorporated machine learning to enhance adaptability, detection accuracy, and automation. This gradual methodological evolution illustrates how technological advances in computing and pattern recognition influenced the development of wafer inspection systems over three decades.

The earliest contribution during this phase was presented by (Ejiri et al., 1989), whose findings demonstrated that conventional image processing alone was insufficient for inspecting increasingly complex multilayer semiconductor wafer patterns. The proposed knowledge-directed image processing architecture integrated design knowledge directly into the inspection pipeline, allowing image processing operations to be dynamically controlled through predefined parametric information. The study highlighted that embedding domain-specific knowledge within dedicated hardware processing units substantially reduced processing delays that traditionally hindered knowledge-based inspection systems. The discussion emphasized that the pipelined architecture improved inspection efficiency by enabling real-time operation while maintaining consistent processing performance across complex wafer structures. Rather than relying solely on sequential image analysis, the inspection strategy combined prior design information with image interpretation, representing one of the earliest attempts to incorporate intelligent decision-making into automated wafer inspection. Although the approach remained dependent on expert-defined rules, the study established an important technological milestone by demonstrating that effective defect detection required both image analysis and contextual manufacturing knowledge.

Approximately two decades later, (Chang et al., 2009) introduced a significant methodological advancement through the application of self-organizing neural networks for automated wafer inspection. Experimental findings indicated that the proposed unsupervised learning framework successfully identified defective wafer regions using industrial inspection data while eliminating the dependence on manually labelled training samples. The discussion highlighted a notable improvement in product flexibility compared with earlier supervised neural network approaches because the clustering mechanism automatically adapted to different defect distributions without requiring application-specific retraining. The proposed inspection framework also addressed practical manufacturing challenges associated with manual microscopic inspection, including operator fatigue, inconsistent judgement, and high labour requirements. Consequently, the study represented a shift away from rigid rule-based inspection toward adaptive computational intelligence capable of learning defect characteristics directly from production data. This transition reflected an increasing recognition that learning-based models could overcome many limitations associated with handcrafted inspection algorithms.

Further progress toward intelligent defect detection was demonstrated by (Y.-H. Liu & Chen, 2011) through the development of the quasiconformal kernel support vector data description (QK-SVDD) model for industrial surface inspection. Experimental evaluation showed that the proposed algorithm achieved a defect detection rate of 96% while improving the generalization capability of the conventional support vector data description framework by more than 30%. The reported inspection time of approximately 60 milliseconds per image also demonstrated that high detection performance could be achieved without sacrificing computational

efficiency. The discussion emphasized that incorporating kernel transformation substantially enhanced the classifier's ability to distinguish defective and defect-free samples under varying inspection conditions, thereby overcoming one of the principal weaknesses of conventional kernel-based methods. Although the study focused on thin-film transistor liquid crystal display manufacturing, the methodological contribution reflected broader developments in industrial visual inspection, where machine learning increasingly replaced manually designed feature extraction procedures with statistically optimized classification models capable of adapting to complex image characteristics.

The culmination of the pioneering phase was observed in (Henn et al., 2019), whose findings reflected the emergence of deep learning as a transformative approach for semiconductor defect inspection. The study demonstrated that convolutional neural networks substantially improved defect detectability and classification performance, particularly for defects considerably smaller than the optical inspection wavelength and in the presence of line edge roughness, which traditionally reduced inspection reliability. Simulation results showed that deep learning models successfully extracted discriminative image representations from complex scattering patterns where conventional intensity thresholding methods were no longer effective. The discussion further indicated that increasing computational capability enabled data-driven learning algorithms to exploit subtle image variations that were difficult to capture using handcrafted feature engineering. This investigation represented a major technological transition from conventional image processing and classical machine learning toward end-to-end feature learning, laying the foundation for the rapid expansion of deep learning research observed in subsequent years. Collectively, the studies published between 1989 and 2019 illustrate a clear chronological progression from knowledge-guided image processing, to unsupervised neural learning, to advanced kernel-based machine learning, and finally to convolutional deep learning, demonstrating the gradual maturation of wafer defect detection technologies in response to increasing manufacturing complexity and inspection requirements.

Phase II: Transition toward Intelligent Defect Detection (2021-2022)

The transition phase represents a critical period in the evolution of wafer defect detection, during which research emphasis shifted from validating individual machine learning algorithms to developing comprehensive intelligent inspection frameworks capable of supporting industrial-scale manufacturing. Compared with the pioneering stage, studies published during 2021–2022 exhibited a broader research perspective by integrating computer vision, machine learning, and deep learning into complete inspection pipelines rather than treating defect detection as an isolated classification task. Collectively, this period illustrates the gradual movement from conventional feature engineering toward data-driven learning while acknowledging that hybrid approaches combining image processing with artificial intelligence remained highly competitive for practical semiconductor inspection.

Research published in 2021 (Westphal & Seitz, 2021) highlighted the growing maturity of intelligent inspection technologies through both methodological development and comprehensive evidence synthesis. (López de la Rosa et al., 2023) consolidated findings from previous studies and reported that machine learning and deep learning approaches consistently outperformed conventional manual inspection by improving defect detection and classification performance in scanning electron microscope images. The comparative analysis demonstrated considerable variation among learning architectures and parameter configurations, indicating that inspection accuracy depended not only on the selected algorithm but also on feature

representation, network design, and training strategy. The discussion emphasized that systematic comparison of existing techniques provides an important basis for identifying high-performing solutions and directing future investigations toward more robust and efficient semiconductor inspection systems.

The practical implementation of deep learning was further demonstrated by (Lin et al., 2021) through the development of an automated inspection framework that integrated cell segmentation, convolutional neural network-based defect recognition, and pseudo-colour visualization (Lin et al., 2021). Experimental findings showed that the proposed segmentation procedure accurately extracted individual cells with minimal localization error while maintaining a processing time suitable for industrial applications. The convolutional neural network achieved a defect detection accuracy of 99.8%, indicating that deep learning had reached a level of performance capable of supporting reliable automated visual inspection. The addition of pseudo-colour visualization further improved defect interpretation by clearly highlighting defective regions after classification. The discussion suggested that combining accurate segmentation with deep feature learning significantly enhanced inspection consistency while reducing dependence on subjective human judgement. Although the application focused on photovoltaic manufacturing, the methodological framework demonstrated characteristics directly applicable to wafer inspection environments requiring high-throughput and high-precision defect analysis.

A broader perspective on intelligent surface inspection was provided by Zheng et al. (2021) (Zheng et al., 2021) , whose review identified a clear technological progression from traditional statistical, spectral, model-based, and learning-based inspection techniques toward deep learning architectures. The findings indicated that recent advances in automated visual inspection were supported not only by increasingly sophisticated neural network models but also by improvements in hardware platforms, imaging technologies, and publicly available datasets. The discussion highlighted that deep learning substantially expanded the capability of automated inspection systems to recognise complex defect characteristics across multiple manufacturing industries, including semiconductor production. Furthermore, the availability of benchmark datasets was identified as an important factor supporting algorithm development and objective performance comparison. These observations suggest that intelligent defect detection had evolved beyond isolated algorithm design into a broader research ecosystem encompassing data resources, computational infrastructure, and standardized evaluation practices.

The methodological progression continued in 2022, when (Gomez-Sirvent et al., 2022) proposed a hybrid inspection framework that combined conventional computer vision with machine learning to improve semiconductor wafer defect classification. Experimental results demonstrated that the integration of image preprocessing, multiple feature descriptors, exhaustive feature selection, and support vector machine classification produced superior F1-score performance compared with several widely used feature selection techniques. The discussion indicated that carefully selecting complementary descriptors representing shape, texture, histogram characteristics, and key-point information substantially enhanced defect characterization before classification. Unlike purely end-to-end deep learning models, this approach illustrated that conventional image processing continued to provide valuable contributions when integrated with intelligent classification algorithms. Consequently, the findings reinforce the view that the transition period was characterised by methodological diversification rather than complete replacement of traditional techniques, with hybrid systems

emerging as practical solutions capable of balancing classification accuracy, computational efficiency, and interpretability.

Overall, publications during 2021–2022 demonstrate a clear transformation in wafer defect detection research from isolated algorithm development toward comprehensive intelligent inspection systems. Comparative reviews established the relative strengths of machine learning and deep learning approaches, deep convolutional networks achieved near-human inspection performance with automated visualization capabilities, broad technological surveys confirmed the increasing dominance of deep learning across industrial inspection applications, and hybrid computer vision frameworks demonstrated that conventional feature engineering continued to complement artificial intelligence under specific manufacturing conditions. These developments collectively established the technological and methodological foundation for the subsequent phase, during which deep learning became the predominant paradigm driving rapid research expansion and innovation in semiconductor wafer defect detection.

Phase III: Rapid Advancement and Deep Learning Expansion (2023-2026)

The first stage of rapid expansion (2023) demonstrated a decisive transition from conventional deep learning applications toward intelligent, high-performance inspection systems designed for industrial semiconductor manufacturing. Research emphasis shifted from improving detection accuracy alone to developing complete inspection pipelines capable of addressing segmentation, localization, classification, computational efficiency, and robustness simultaneously. (Yu et al., 2023) demonstrated that integrating die segmentation with region-based anomaly clustering achieved defect detection accuracy exceeding 97%, indicating the effectiveness of combining machine vision with anomaly-based analysis for industrial inspection. Likewise, (López de la Rosa et al., 2023) reported that a lightweight SqueezeNet architecture produced classification performance comparable to heavyweight convolutional neural networks while substantially reducing computational time, highlighting the increasing importance of deployment efficiency in production environments. Improvements in segmentation capability were further demonstrated by (Solorzano & Tsai, 2023) through a modified encoder–decoder architecture capable of accurately identifying low-contrast defects with minimal computational overhead, while (Y. Kim et al., 2023) showed that a TransUNet incorporating contrastive learning significantly improved mixed-defect segmentation, achieving an accuracy and F1-score of 0.995. Similarly, (Wang et al., 2023) proposed a knowledge-augmented broad learning framework capable of decoupling mixed defect patterns more effectively than conventional deep learning models. Complementing these algorithmic developments, review studies by (Ling & Isa, 2023), (Chisti et al., 2023), and (Jianhong et al., 2023) consistently identified a clear technological evolution from traditional image processing and machine learning toward deep learning, while also recognizing persistent challenges associated with data availability, model interpretability, computational complexity, and industrial deployment. At the manufacturing level, (Fukuda et al., 2023) demonstrated that trainable die-to-database inspection substantially improved inspection robustness under low-resolution imaging, image distortion, and high-noise conditions encountered in extreme ultraviolet lithography, illustrating the growing maturity of trainable inspection systems.

The second stage (2024–2025) reflected the consolidation of deep learning into practical semiconductor manufacturing by emphasizing deployment efficiency, industrial scalability, and intelligent decision support. Although (D. Kim, 2024) primarily introduced a mathematical optimisation framework for analysing wafer surface inspection systems, subsequent studies

increasingly focused on addressing practical manufacturing constraints rather than solely improving algorithmic accuracy. (Sun et al., 2024) proposed a cascaded detection framework capable of processing high-resolution industrial inspection images while maintaining high localisation accuracy and controlling false-negative rates, demonstrating improved suitability for production environments. During 2025, research diversified considerably to address equipment inspection, automatic root cause analysis, few-shot learning, virtual metrology, lightweight artificial intelligence, and large-scale inspection systems. (Fu et al., 2025) demonstrated that integrating artificial intelligence with automated optical inspection reduced false alarms and manual intervention while improving production efficiency. (Bose, Pandey, Staib, Womach, Balla, Olmedo, Khandelwal, Lerner, et al., 2025) extended intelligent inspection by introducing an inline virtual metrology framework capable of monitoring every wafer throughout photolithography and performing automated root cause analysis using deep learning and probabilistic reasoning. Model robustness under limited training data received increasing attention through the few-shot learning strategy proposed by (He et al., 2025), whereas (Yang et al., 2025) improved recognition of single and mixed wafer defects using an attention spatial pyramid pooling network that outperformed previous deep learning models. Comparative investigations by (Ozcelik & Alimli, 2025) revealed that classical image processing remained advantageous for rapid inspection, whereas deep learning achieved superior segmentation accuracy in demanding inspection scenarios, suggesting that hybrid methodologies continued to retain industrial relevance. Meanwhile, (Façoco et al., 2025) demonstrated the potential of combining unsupervised anomaly detection with supervised segmentation for balanced inspection accuracy and computational efficiency, while (Noh, 2025) and (Htay et al., 2025) illustrated the growing integration of Vision Transformers, automated annotation, and customised optimisation strategies into intelligent manufacturing systems. In parallel, the systematic review conducted by (Dehaerne et al., 2025) confirmed that deep learning became the dominant approach for semiconductor inspection after 2020 while identifying throughput, automation, and reduced manual configuration as continuing research priorities.

The most recent stage (2026) illustrates the emergence of intelligent, scalable, and highly adaptive inspection frameworks that extend beyond conventional convolutional neural networks toward transformer architectures, vision-language learning, synthetic data generation, unsupervised learning, and multimodal artificial intelligence. (Friedrich et al., 2026) demonstrated that synthetic image generation substantially improved defect classification when integrated with residual neural networks, highlighting data augmentation as an effective solution for scarce industrial datasets. Similarly, (An et al., 2026) proposed an unsupervised framework integrating autoencoders, generative adversarial networks, and attention mechanisms that achieved high precision while maintaining robustness under limited labelled data. Comprehensive reviews by (Akele & Gurau, 2026) and (J. Li et al., 2026) concluded that although defect detection accuracies frequently exceeded 98%, challenges relating to scalability, dataset quality, false-positive reduction, and industrial deployment remained unresolved, with only a limited proportion of proposed systems reaching high-volume manufacturing. Algorithmic innovation continued through lightweight transformer architectures (Shi, 2026), dynamic artificial intelligence-assisted inspection systems (Chen et al., 2026) enhanced YOLO-based real-time detection networks (Han et al., 2026), prototype-based vision-language alignment frameworks (Wang et al., 2020), and multimodal thermography-assisted defect identification for three-dimensional semiconductor packaging (Zhao et al., 2026). Collectively, these studies demonstrate that research has progressed beyond improving detection accuracy toward developing intelligent inspection ecosystems capable of

integrating multiple sensing modalities, semantic knowledge, synthetic data, lightweight computation, and real-time decision support. This progression represents the maturation of wafer defect detection from isolated deep learning algorithms into comprehensive artificial intelligence-driven manufacturing solutions capable of supporting next-generation semiconductor production.

Conclusion

This chronological review examined the evolution of wafer defect detection in photolithography from 1989 to 2026, with the objective of providing a structured understanding of how inspection technologies have developed from conventional image processing to modern deep learning-based approaches. A total of 37 primary studies were selected and analysed according to their publication period, methodological approaches, technological characteristics, and contributions to wafer defect inspection. Unlike a purely thematic review, this study organised the selected literature into three temporal phases: Phase I: Pioneering and Foundational Research (1989–2019), Phase II: Transition toward Intelligent Defect Detection (2021–2022), and Phase III: Rapid Advancement and Deep Learning Expansion (2023–2026). This chronological structure provides a historical perspective on how research priorities and inspection technologies have progressively changed in response to the increasing complexity of semiconductor manufacturing.

The findings demonstrate a clear technological transition across the reviewed period. Phase I represents the foundation of automated wafer inspection, where knowledge-based approaches, conventional image processing, handcrafted features, and early learning algorithms were used to identify and classify defects. These approaches established important principles for automated inspection but were generally dependent on predefined rules and manually designed features. Their effectiveness could be affected by image quality, illumination conditions, noise, process variation, and the increasing complexity of semiconductor patterns. Phase II marked an important transition toward intelligent defect detection, as machine learning and early deep learning approaches increasingly reduced dependence on handcrafted feature engineering. The development of convolutional neural networks and other learning-based techniques enabled automatic feature representation and improved the ability of inspection systems to recognise more complex defect patterns. This period therefore served as an important bridge between conventional image analysis and the more advanced deep learning approaches observed in recent research.

The most significant development was observed in Phase III (2023–2026), which demonstrated rapid expansion and diversification of deep learning applications. Research increasingly moved beyond conventional defect classification toward integrated capabilities involving detection, localisation, segmentation, anomaly identification, and process-level analysis. Advanced CNN architectures, lightweight networks, attention mechanisms, transformer-based models, hybrid approaches, and knowledge-augmented frameworks were increasingly investigated. At the same time, researchers addressed practical industrial requirements such as computational efficiency, real-time processing, robustness under low-quality imaging conditions, and reduced dependence on labelled data. The emergence of synthetic data generation, unsupervised learning, multimodal inspection, virtual metrology, and intelligent process monitoring further demonstrates that wafer defect detection is evolving from an isolated image-analysis task into a broader intelligent manufacturing function.

An important finding from the chronological analysis is that detection accuracy is no longer the only primary objective of research. Although high-performing models have demonstrated substantial improvements in recognition and segmentation, recent studies increasingly consider factors such as computational cost, inference speed, data availability, model robustness, false-positive reduction, scalability, and industrial integration. This indicates a shift from developing models that perform well under controlled experimental conditions toward developing systems that can operate reliably within complex semiconductor production environments. Interestingly, conventional image processing has not become entirely obsolete. Recent comparative studies indicate that traditional approaches may still provide advantages in terms of speed and simplicity for certain inspection tasks, whereas deep learning is more suitable for complex and variable defect patterns. Therefore, future inspection systems may increasingly adopt hybrid architectures that combine conventional image processing with machine learning and deep learning rather than relying exclusively on a single technological paradigm.

The chronological approach also provides an important contribution to understanding the research trajectory. By connecting earlier foundational studies with subsequent technological developments, the review demonstrates that current deep learning systems did not emerge independently but resulted from a gradual accumulation of knowledge in image processing, machine vision, machine learning, and semiconductor inspection. The three-phase framework therefore provides a clearer interpretation of technological maturity and research transformation than an analysis based only on individual technical categories. It also illustrates how increasing semiconductor miniaturisation and manufacturing complexity have progressively driven the demand for more adaptive, automated, and intelligent inspection technologies.

Despite the substantial progress, several research gaps remain. First, many existing studies continue to rely on specific datasets, defect categories, imaging conditions, or controlled experimental environments, which may limit the generalisation of models across different semiconductor processes and manufacturing conditions. Second, the availability of sufficiently large, balanced, high-quality, and accurately labelled defect datasets remains a major challenge because critical defects are often rare and industrial datasets may contain sensitive or proprietary information. Third, false positives, model interpretability, computational requirements, and robustness under changing process conditions remain important barriers to large-scale deployment. Fourth, although recent research has introduced transformers, synthetic data, multimodal learning, and virtual metrology, these technologies are still relatively fragmented and are not yet consistently integrated into unified inspection frameworks. Most importantly, there remains a gap between high experimental detection performance and validated high-volume manufacturing deployment. These gaps indicate the need for future research to focus not only on developing more accurate algorithms but also on building robust, explainable, computationally efficient, and adaptable inspection systems that can operate under real manufacturing constraints.

Future research should therefore move toward integrated and industry-oriented intelligent inspection frameworks. Greater attention should be given to combining deep learning with multimodal sensing, synthetic data, domain knowledge, process monitoring, digital twins, edge computing, and explainable artificial intelligence. Cross-domain collaboration between semiconductor manufacturing, computer vision, artificial intelligence, materials science, and industrial engineering will also be important for translating algorithmic developments into practical manufacturing solutions.

Overall, the evolution from conventional image processing to deep learning reflects a fundamental transformation in wafer defect inspection, from rule-based and feature-engineered analysis toward adaptive, data-driven, and intelligent manufacturing systems. The chronological synthesis of 37 studies provides evidence of this long-term transformation while identifying the remaining gap between technological innovation and reliable industrial deployment. Addressing this gap represents an important direction for future research and can contribute to the development of more accurate, efficient, scalable, and reliable wafer defect detection systems for next-generation semiconductor manufacturing.

Acknowledgements:	The authors would like to express their sincere gratitude to Universiti Teknologi MARA (UiTM) Pulau Pinang for providing the necessary resources and support throughout the course of this research. Special appreciation is extended to colleagues and peers who contributed valuable insights and constructive feedback, which greatly enhanced the quality of this paper.
Funding Statement:	No Funding
Conflict of Interest Statement:	The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Innovation and Industrial Revolution (IJIREV).
Ethics Statement:	This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards.
Author Contribution Statement:	All authors contributed significantly to the development of this manuscript. Author 1 was responsible for the conceptualization, methodology, and overall supervision of the study. Author 2 handled data collection, analysis, and interpretation of results. Author 3,4 contributed to the literature review, drafting, and critical revision of the manuscript. All authors read and approved the final version of the manuscript prior to submission.

References

Akele, Y. K., & Gurau, V. (2026). A critical survey of machine learning - powered vision inspection in semiconductors manufacturing. *Array*, *30*. https://doi.org/10.1016/j.array.2026.100888

An, S., Wu, J., & Li, J. (2026). Unsupervised industrial image defect detection based on autoencoder and GANs. *PLOS ONE*, *21*(4 April). https://doi.org/10.1371/journal.pone.0346637

Basu, P., Verma, J., Abhinav, V., Ratnesh, R., Singla, Y., & Kumar, V. (2025). Advancements in Lithography Techniques and Emerging Molecular Strategies for Nanostructure Fabrication. *International Journal of Molecular Sciences*, *26*. https://doi.org/10.3390/ijms26073027

Bose, P., Pandey, V. K., Staib, M., Womach, B., Balla, I., Olmedo, M., Khandelwal, A., Ahmadi, A., & Bisht, G. (2025). Virtual defect metrology system utilizing photolithography scanner topography maps to quantitatively detect gross defects, perform inline yield prediction, and identify defect-generating tools. *Journal of Micro/Nanopatterning, Materials and Metrology*, *24*(4). https://doi.org/10.1117/1.JMM.24.4.041404

Bose, P., Pandey, V. K., Staib, M., Womach, B., Balla, I., Olmedo, M., Khandelwal, A., Lerner, T., Ahmadi, A., & Bisht, G. (2025). *A virtual defect metrology system utilizing photolithography scanner topography maps to quantitatively detect gross defects, perform in-line yield prediction, and identify defect generating tools*. *13426*, 134260. https://doi.org/10.1117/12.3050587

Chang, C.-Y., Li, C., Chang, J.-W., & Jeng, M. (2009). An unsupervised neural network approach for automatic semiconductor wafer defect inspection. *Expert Systems with Applications*, *36*(1), 950–958. https://doi.org/10.1016/j.eswa.2007.10.033

Chen, T.-H., Chou, H.-H., Zou, S., Chen, Y.-H., & Hsieh, S.-Y. (2026). System-level integration of deep learning and computer vision for contact ring seal defect detection in semiconductor manufacturing. *Expert Systems with Applications*, *298*. https://doi.org/10.1016/j.eswa.2025.129551

Chisti, M. K. M., Srinivas Kumar, S., & Prasad, G. (2023). Defects Identification, Localization, and Classification Approaches: A Review. *IETE Journal of Research*, *69*(7), 4323–4336. https://doi.org/10.1080/03772063.2021.1958073

Dehaerne, E., Dey, B., Blanco, V., & Davis, J. (2025). Scanning electron microscopy-based automatic defect inspection for semiconductor manufacturing: a systematic review. *Journal of Micro/Nanopatterning, Materials and Metrology*, *24*(2). https://doi.org/10.1117/1.JMM.24.2.020901

Ejiri, M., Yoda, H., Sakou, H., & Sakamoto, Y. (1989). Knowledge-directed inspection for complex multilayered patterns. *Machine Vision and Applications*, *2*(3), 155–166. https://doi.org/10.1007/BF01212456

Façoco, I., Carvalho, R., & Rosado, L. (2025). Towards efficient wafer visual inspection: Exploring novel lightweight approaches for anomaly detection and defect segmentation. *Intell. Syst. Appl.*, *28*, 200576. https://doi.org/10.1016/j.iswa.2025.200576

Friedrich, M., Schlosser, T., & Kowerko, D. (2026). Simulation of semiconductor wafer dicing induced faults on chips and their application as augmentation method for a deep learning based visual inspection system. *Journal of Intelligent Manufacturing*, *37*(2), 573–596. https://doi.org/10.1007/s10845-024-02559-0

Fu, C.-J., Chen, H.-L., & Tseng, H.-Y. (2025). Application of Artificial Intelligence to Improve Chip Defect Detection Using Semiconductor Equipment †. *Engineering Proceedings*, *98*(1). https://doi.org/10.3390/engproc2025098026

Fukuda, K., Ouchi, M., Ishikawa, M., Yoshida, Y., Fukaya, K., Kagetani, R., & Shindo, H. (2023). Trainable die-to-database for large field of view e-beam inspection. *Journal of Micro/Nanopatterning, Materials and Metrology*, *22*(2). https://doi.org/10.1117/1.JMM.22.2.021004

Gomez-Sirvent, J. L., De La Rosa, F. L., Sanchez-Reolid, R., Fernandez-Caballero, A., & Morales, R. (2022). Optimal Feature Selection for Defect Classification in Semiconductor Wafers. *IEEE Transactions on Semiconductor Manufacturing*, *35*(2), 324–331. https://doi.org/10.1109/TSM.2022.3146849

Han, T., Wang, X., Dong, Q., Jia, J., Zhang, F., & Huang, M. (2026). CDNet: Efficient Real-Time Chip Defect Detection for Consumer Electronics. *IEEE Transactions on Consumer Electronics*, *72*(2), 2827–2839. https://doi.org/10.1109/TCE.2025.3608075

He, H., Wei, Y.-C., Lin, X., Zhu, M., & Zhang, H. (2025). Semiconductor Wafer Defect Recognition Based on Improved Coordinate Attention Mechanism. *IEEE Access*, *13*, 46856–46864. https://doi.org/10.1109/access.2025.3546781

Henn, M.-A., Zhou, H., & Barnes, B. M. (2019). Data-driven approaches to optical patterned defect detection. *OSA Continuum*, *2*(9), 2683–2693. https://doi.org/10.1364/OSAC.2.002683

Htay, T. S. S., Huang, C.-C., Thi, V. D., Tsai, Y.-C., & Su, C.-Y. (2025). High-precision and rapid detection of complex defects in transferred CVD graphene enabled by machine learning algorithms. *Carbon*, *244*. https://doi.org/10.1016/j.carbon.2025.120669

Ji, P., He, Z., Yang, W., Du, J., Ye, G., & Lu, X. (2025). Wafer Defect Detection Technology Based on CTM-IYOLOv10 Network. *Journal of Imaging*, *11*. https://doi.org/10.3390/jimaging11110408

Jianhong, Zhang, T., Yang, C., Cao, Y., Xie, L., Tian, H., & Li, X. (2023). Review of Wafer Surface Defect Detection Methods. *Electronics*. https://doi.org/10.3390/electronics12081787

Kim, D. (2024). Optimized Operation Methods of the Wafer Surface Defect Detection. *SN Computer Science*, *5*(7). https://doi.org/10.1007/s42979-024-03076-w

Kim, J.-B., Chung, H., Moon, S., Ryu, G., Kim, T., Lee, W., Cho, D., Kang, C., Kim, H., Park, H.-B., Cha, B., Koushnir, L., Shatsberg, Z., Sreenivas, A., Pandey, A. K., & Elias, N. (2026). *AI-based detection algorithm for optical inspection of critical defects on 2nm and 1.4nm technology nodes*. *13981*, 139810. https://doi.org/10.1117/12.3090867

Kim, Y., Lee, J.-H., & Jeong, J. (2023). Pre-trained CNN-based TransUNet Model for Mixed-Type Defects in Wafer Maps. *WSEAS Transactions on Information Science and Applications*, *20*, 238–244. https://doi.org/10.37394/23209.2023.20.27

Lechien, T., Dehaerne, E., Dey, B., Blanco, V., Gendt, S., & Meert, W. (2023). Automated Semiconductor Defect Inspection in Scanning Electron Microscope Images: a Systematic Review. *ArXiv*, *abs/2308.0*. https://doi.org/10.48550/arxiv.2308.08376

Li, J., Zhang, X., Liang, X., Yin, G., Yan, L., & Wei, S. (2026). Review of wafer defect detection in semiconductor manufacturing: Algorithms, systems, and data. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-026-02845-z

Li, M., Wang, X., Zhang, H., & Hu, X. (2025). An Improved YOLOv7-Tiny-Based Algorithm for Wafer Surface Defect Detection. *IEEE Access*, *13*, 10724–10734. https://doi.org/10.1109/access.2025.3528242

Lin, H.-H., Dandage, H. K., Lin, K.-M., Lin, Y.-T., & Chen, Y.-J. (2021). Efficient cell segmentation from electroluminescent images of single-crystalline silicon photovoltaic

modules and cell-based defect identification using deep learning with pseudo-colorization. *Sensors*, *21*(13). https://doi.org/10.3390/s21134292

Ling, Q., & Isa, N. A. M. (2023). Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. *IEEE Access*, *11*, 15921–15944. https://doi.org/10.1109/ACCESS.2023.3245093

Liu, W., Li, G., Cui, Q., Zhang, Z., Wang, X., & Bai, C. (2024). Semiconductor Wafer Particle Defect Classification Based on Deep Learning and Multimodal Feature Fusion. *2024 IEEE 12th International Conference on Computer Science and Network Technology (ICCSNT)*, 29–34. https://doi.org/10.1109/iccsnt62291.2024.10776716

Liu, Y.-H., & Chen, Y.-J. (2011). Automatic defect detection for TFT-LCD array process using quasiconformal kernel support vector data description. *International Journal of Molecular Sciences*, *12*(9), 5762–5781. https://doi.org/10.3390/ijms12095762

López de la Rosa, F., Gómez-Sirvent, J. L., Morales, R., Sánchez-Reolid, R., & Fernández-Caballero, A. (2023). Defect detection and classification on semiconductor wafers using two-stage geometric transformation-based data augmentation and SqueezeNet lightweight convolutional neural network. *Computers and Industrial Engineering*, *183*. https://doi.org/10.1016/j.cie.2023.109549

Magnanini, M., Djurdjanović, D., Pomi, R., & Tolio, T. (2025). System-level evaluation of productivity and quality in semiconductor frontend fabrication integrating product and process models. *CIRP Journal of Manufacturing Science and Technology*. https://doi.org/10.1016/j.cirpj.2025.05.010

Noh, G. (2025). A Real-Time Operator-Centric Patch Extraction and Analysis Platform for Industrial Semiconductor Wafer C&C Processes. *IEEE Access*, *13*, 202340–202353. https://doi.org/10.1109/ACCESS.2025.3638806

Ohri, A., Snyder, S., Nair, V., Watson, B., Ahn, K., Ege, E., Singh, Y., Lee, E., Chen, H., Light, S., Muthuraj, J. J. J. J. J., Chen, L., Weich, K., & Hobbs, J. (2024). *Enhanced EUV photolithography control for overcoming defectivity challenges*. *13215*, 132150. https://doi.org/10.1117/12.3034557

Ozcelik, Z., & Alimli, D. (2025). Comparative Analysis of Wafer Defect Segmentation Based on Image Processing with U-Net++. In A. E. (Ed.), *International Conference on Computer Science and Engineering, UBMK* (Number 2025, pp. 58–62). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/UBMK67458.2025.11207000

Reut, G., Ovdat, O., Cohen, O., Yasharzade, S., Zacs, L., Kolsky, I., Ilan, H., Ishida, S., Saito, T., Sawasato, A., Kawabe, Y., & Mizuno, S. (2025). *Optimization of defect detection sensitivity and cost of ownership reduction in nano imprint lithography through design for inspection and advanced optical inspection*. *13426*, 134262. https://doi.org/10.1117/12.3051318

Samir, H., Amin, B., & Ghoneima, M. (2024). CD-SEM Image Defect Detection and Classification Using Transformers. *Journal of Electrical Systems*. https://doi.org/10.52783/jes.6062

Shi, Z. (2026). Deep Learning-Based Surface Defect Detection Technology for Microdevices. *HighTech and Innovation Journal*, *7*(2), 732–745. https://doi.org/10.28991/HIJ-2026-07-02-022

Shin, H., Lee, H., & Shin, D. (2025). Precision Matching Optimization of Optical Metrology and Inspection Equipment for Yield Enhancement in Semiconductor Manufacturing. *JOURNAL OF SEMICONDUCTOR TECHNOLOGY AND SCIENCE*. https://doi.org/10.5573/jsts.2025.25.6.623

Shinde, P., Pai, P., Adiga, S., Mayya, K., Seo, Y., Hwang, M., Go, H., & Park, C. (2025). Defect Detection in Photolithographic Patterns Using Deep Learning Models Trained on Synthetic Data. *ArXiv*, *abs/2505.1*. https://doi.org/10.1016/j.heliyon.2025.e43377

Solorzano, C., & Tsai, D.-M. (2023). Watermark Detection in CMOS Image Sensors Using Cosine-Convolutional Semantic Networks. *IEEE Transactions on Semiconductor Manufacturing*, *36*(2), 279–290. https://doi.org/10.1109/TSM.2023.3245606

Sun, T., Li, Z., Xiao, X., Guo, Z., Ning, W., & Ding, T. (2024). Cascaded detection method for surface defects of lead frame based on high-resolution detection images. *Journal of Manufacturing Systems*, *72*, 180–195. https://doi.org/10.1016/j.jmsy.2023.11.017

Wang, J., Gao, P., Zhang, J., Lu, C., & Shen, B. (2023). Knowledge augmented broad learning system for computer vision based mixed-type defect detection in semiconductor manufacturing. *Robotics and Computer-Integrated Manufacturing*, *81*. https://doi.org/10.1016/j.rcim.2022.102513

Wang, J., Xu, C., Yang, Z., Zhang, J., & Li, X. (2020). Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition. *IEEE Transactions on Semiconductor Manufacturing*, *33*(4), 587–596. https://doi.org/10.1109/TSM.2020.3020985

Westphal, E., & Seitz, H. (2021). A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. *Additive Manufacturing*, *41*, 101965. https://doi.org/10.1016/j.addma.2021.101965

Yang, H., Kim, M.-S., Zhou, X., Beral, C., Charley, A., Baskaran, B., Sah, K., Bouckou, L., Barbisan, L., Durbha, G., Paithankar, N., Chen, Z., & Gronheid, R. (2025). *Defectivity-aware EUV process window characterization for monolithic complementary FET HAR and 3D patterning*. *13686*, 136860. https://doi.org/10.1117/12.3071887

Yu, N., Li, H., & Xu, Q. (2023). A full-flow inspection method based on machine vision to detect wafer surface defects. *Mathematical Biosciences and Engineering : MBE*, *20 7*, 11821–11846. https://doi.org/10.3934/mbe.2023526

Zhao, C., Sim, J., Chang, S., Im, Y., & Won, Y. (2026). ThermoVision: Nondestructive Defect Mapping in 3-D Stacked Dies via Surface Thermography and Deep Learning. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *16*(4), 757–764. https://doi.org/10.1109/TCPMT.2026.3661910

Zheng, X., Zheng, S., Kong, Y., & Chen, J. (2021). Recent advances in surface defect inspection of industrial products using deep learning techniques. *International Journal of Advanced Manufacturing Technology*, *113*(1–2), 35–58. https://doi.org/10.1007/s00170-021-06592-8

Zhou, Z., Luo, H., Xiong, W., Qu, D., Ding, R., Li, Z., Jin, W., Ru, Y., Jia, S., & Hong, J. (2025). Characterization of Multimodal Spot Scanning Imaging System for Wafer Defect Inspection. *IEEE Transactions on Semiconductor Manufacturing*, *38*, 4–11. https://doi.org/10.1109/tsm.2024.3481291

Zhu, J., Liu, J., Xu, T., Yuan, S., Zhang, Z., Jiang, H., Gu, H., Zhou, R., & Liu, S. (2022). Optical wafer defect inspection at the 10 nm technology node and beyond. *International Journal of Extreme Manufacturing*, *4*(3). https://doi.org/10.1088/2631-7990/ac64d7