Dean&Francis
ISSN 2959-6157

The Advancements of Fault Detection in Microelectronics Based on Machine Learning Models

Zhongxu Huang

Department of Automation Academy, Nanjing Institute of Technology, Nanjing, China
cher63044@gmail.com

Abstract:

Fault detection in microelectronics constitutes an indispensable component of microelectronic applications, routinely employed to ensure the normal operation of electronic components. However, due to the inherent characteristics of microelectronic devices, such as their miniature size and high precision, conventional detection methods often struggle to meet the stringent requirements for their inspection. Machine Learning (ML) recently has gained prominence as a result of the development of Artificial Intelligence (AI) technologies, and it offers new methods for microelectronics. Convolutional structures are specifically used in Convolutional Neural Networks (CNNs), a type of deep neural network that helps mitigate model overfitting problems and lowers the memory footprint of deep networks. Recurrent Neural Networks (RNNs) are an important class of artificial neural networks that are specifically made to process sequential data. They can handle any length of sequence. Transformers are characterized by the introduction of the self-attention mechanism, enabling exceptional performance when dealing with sequential data. The application of models trained on these various architectures in the context of microelectronics inspection is the main focus of this paper, with an emphasis on how they affect fault detection accuracy. Furthermore, the concluding section of this paper discusses the current challenges associated with employing machine learning for microelectronics fault detection and explores potential future solutions. The process of addressing these challenges will not only drive the advancement of ML itself but is also expected to lead to more efficient and diverse methodologies for fault detection in microelectronics.

Keywords: Microelectronic; artificial intelligence; fault detection; deep learning.

ISSN 2959-6157

1. Introduction

Microelectronics, a subdiscipline of electronics, focuses on designing, fabricating, and applying microscopic electronic components and circuits. These elements are typically fabricated on semiconductor materials. Its foundational innovation lies in integrated circuits—miniaturized interconnections of millions to billions of microscopic transistors and other components on a single semiconductor substrate.

Positioned at the technological frontier, microelectronics provides critical components for aerospace [1], defense, and industrial applications. In biomedicine, S. Zhou et al. integrated microneedle technology with microelectronics to enhance drug delivery efficiency and physiological substance extraction [2]. Similarly, plasma etching technology constitutes an indispensable semiconductor manufacturing process for contemporary microelectronics, confronting significant societal challenges including—but not limited to—the transition toward three-dimensional device architectures and atomic-scale precision in process control [3].

Beyond manufacturing constraints, microelectronics faces inherent security vulnerabilities as essential constituents of communication devices (e.g., smartphones, computers) [4], alongside reliability concerns regarding storage component longevity. These imperatives demand exceptional integration density, performance, and reliability in electronic components. Consequently, microelectronics testing has become paramount. Conventional methodologies often prove inadequate due to structural complexity and extreme miniaturization. Recent advances in machine learning, however, have catalyzed novel solutions for microelectronics inspection.

Machine learning is Artificial Intelligence (AI) technology applied across various domains. For instance, in banking and finance, machine learning is utilized for detecting fraudulent transactions, providing risk assessments, and enhancing the security of fund transfers [5]. Biomedicine is another field where machine learning is extensively applied. In recent years, scholars have proposed its use in assisting pathologists to analyze blood samples, extracting information during the data integration stage to facilitate diagnostic reporting. Concurrently, the adoption of machine learning is driving the evolution of pathology towards digital pathology. In industrial production, machine learning plays a pivotal role by enhancing innovation, efficiency, and sustainability. However, as manual controls in the process are prone to issues, research into automating machine learning has been conducted, aiming to optimize the management of its workflows [6]. These applications involve diverse models, each with distinct emphases yielding different outcomes.

This article is divided into three sections, with the second section elucidating the application of various models in microelectronics fault detection, and the third section presenting an outlook for future work.

2. Method

2.1 CNN-based Detection

Convolutional Neural Networks (CNNs), incorporate convolutional computations that reduce the memory footprint of deep networks while maintaining a deep structure. Through local receptive fields, weight sharing, and hierarchical feature extraction, CNNs effectively minimize the number of network parameters and mitigate overfitting, establishing themselves as the cornerstone architecture for processing grid-like topology data (images/video/speech).

The Calabrese et al. employed an open-source Deep Learning (DL) framework to address technical challenges in the Quality Control (QC) process associated with printed circuit board (PCB) manufacturing. Given that the identified components and features may be at a sub-millimeter scale with ill-defined boundaries, ensuring PCB manufacturing quality while considering the feasibility of implementing automated solutions for QC presents a significant challenge. As a result, the proposed study looked into the possibility of using a DL algorithm—Mask R-CNN in particular—to create a tool that would support the PCB fabrication QC process. Two DL algorithms—Mask R-CNN implemented via Detectron2 and YOLOv8—were the main focus of the study. A dataset of open-source images was used to test the algorithms. With curated subsets from the open-source database, the study trained two different models (Mask R-CNN and YOLOv8) that target two specific types of PCB defects: missing holes and short circuits. Finally, the Mask R-CNN model outperformed YOLOv8 in terms of defect segmentation and detection [7].

To precisely identify and categorize manufacturing flaws on printed circuit boards (PCBs), Bhattacharya et al. used a Fast Region-based Convolutional Neural Network (FRCNN). They utilized the CNN backbone to extract underlying geometric features from images. To achieve a more comprehensive representation, Bhattacharya et al. incorporated considerations for global dependencies alongside local modeling, with their model operating directly on the feature maps generated by the convolutional network. This approach leverages the complementary strengths of CNNs and Transformers: modeling long-range dependencies while combining shift-invariant local representations to learn scale-invariant features. Thus, the model was able to locate, identify, and categorize a number of flaws in low-resolution bare-board PCB pictures. Benefiting from the favorable properties of the CNN architecture, Bhattacharya's model is lightweight and compatible with low-resolution inputs. Compared to the standard YOLOv5m model, it achieved an overall improvement of 3.2% in mAP at an IoU threshold of 0.5. This advancement holds significant potential for enhancing production efficiency and substantially reducing costs [8].

2.2 RNN-based Detection

Neural networks that are specifically made for processing sequential data are called recurrent neural networks, or RNNs. They excel at capturing temporal information and semantic information within such data. The shortcomings of conventional feedforward neural networks, like Multilayer Perceptrons (MLPs) and CNNs, in processing data with dependencies spanning multiple time steps are addressed by RNNs, an essential class of Artificial Neural Networks (ANNs) designed for sequence processing [9].

Liu et al. proposed a model based on a RNN as an alternative to Simulink models. This approach aims to achieve accurate assessment of power semiconductor lifetime in permanent magnet synchronous coupling converters while meeting acceptable simulation time requirements to satisfy whole-lifecycle design criteria. By utilizing the RNN model, the team derived chip temperature estimates, encompassing cumulative damage calculations for both long-term and short-term thermal dynamics. The approximations obtained by this model proved significantly faster and more accurate, achieving a error of only 0.51% for a 1-hour extreme mission profile [10].

This PI-LSTM, which is a variation of RNN models, was trained with data from simulations using the Finite Element Method (FEM), enhanced by adding flow rules to its loss function. Despite having little training data, the physics-informed LSTM showed strong stability and high accuracy.The equivalent plastic strain and stress components within solder joints were predicted by the researchers using PI-LSTM. By establishing a correlation with failure cycles, these anticipated plastic strains make reliability determination possible. Multiple reliability models confirmed that failure cycles predicted by PI-LSTM closely aligned with FEM simulation results, highlighting its potential for reliability assessment applications. Comparative analyses revealed that PI-LSTM achieved comparable or superior R^2 scores to alternative methods, despite utilizing significantly smaller datasets. This approach substantially reduces the time required to create training databases for AI models. Moreover, PI-LSTM remains applicable in data-scarce scenarios where other models fail to achieve sufficient accuracy [11].

2.3 Transformer-based Detection

The Transformer, a novel and simplified network architecture proposed by the Vaswani et al, relies entirely on the attention mechanism, dispensing with recurrence and convolutions. Experiments demonstrated that these models achieve superior quality, are significantly more parallelizable, and require substantially less training time [12]. Thus, Transformers address three major limitations of Recurrent Neural Networks (RNNs): failure in handling long-range dependencies, lack of parallelizability, and the information bottleneck.

To improve biomedical data analysis, Suresh et al. created a novel methodology by combining the Transformer algorithm with sensors based on nanoelectronics. Compared to conventional approaches, the proposed method improved diagnostic accuracy by 25% and reduced processing time by 30%. Transformer model with smaller error margins and greater prediction precision could produced lower error values for the analysis of nano-circuitry data. In detecting and diagnosing anomalies in nano-circuitry data, its high accuracy and detection rates show its efficacy [13].

Consumer electronics often incorporate complex sensors susceptible to malfunctions induced by temperature fluctuations, humidity variations, vibration, and mechanical shock. To ensure reliable sensor operation, Lin et al. developed a CESFDNet. After extracting local correlation features from neighboring data points using multi-layer convolutional processes, CESFDNet combines these features with global dependencies that the Transformer architecture has detected. The researchers further refined the self-attention mechanism to mitigate noise interference in temporal sequences that compromise diagnostic accuracy. Comprehensive experimental results conclusively demonstrate the superiority and reliability of the Transformer-based CESFDNet framework [14].

3. Discussion

Despite increasing methods and advancing technologies in microelectronics fault detection—where machine learning has demonstrated significant practical value—challenges persist. In counterfeit integrated circuit (IC) detection, techniques like SIFT keypoint extraction, template matching facilitate identification and analysis of PCB and IC components. Hardware security still faces significant challenges, despite the fact that deep learning and ANNs have improved feature extraction efficiency, especially through models like AlexNet, and Inception-v3, including high-density PCBs' inherent image clutter and the lack of

ISSN 2959-6157

extensive annotated datasets. Addressing these challenges requires future exploration of multimodal imaging approaches, development of publicly accessible benchmark datasets, and earlier integration of deep learning within computer vision pipelines. For complete solutions, it is still crucial to combine knowledge from computer vision, imaging technology, hardware design, and machine learning [6].

In machine learning, while ML algorithms exhibit significant predictive capabilities, they often fail to provide causal explanations for their predictions. Determining which particular factors, or combinations of them, influence model decisions is nearly impossible due to the multilayer architecture and inherent complexity of neural networks. This opacity can be problematic—the 'black-box' nature of such models precludes clear articulation of how various factors are weighted during decision-making processes. Moreover, in certain application scenarios, whether in supervised or unsupervised settings, it is necessary to fully comprehend the capabilities or skills being learned. This requirement significantly complicates the learning process [15].

To address these challenges, future research should explore distinct pathways. In order to reduce bias and guarantee fairness, research could primarily concentrate on creating more efficient algorithms and methods [16]. Investigating bias reduction strategies, like data preprocessing approaches and algorithmic fairness techniques, is necessary to achieve equitable results and avoid discrimination. Understanding and reducing biases in supervised learning algorithms requires the development of models that are both transparent and understandable [17].

The performance and generalization capabilities of supervised learning may also be enhanced by transfer learning techniques, particularly when obtaining labeled data is costly or time-consuming. The creation of strong transfer learning frameworks that can adapt to various data distributions and transfer knowledge across domains efficiently while reducing domain shift could be the main focus of future research [18].

Additionally, creating scalable and reliable algorithms to handle massive datasets is essential. Traditional supervised learning algorithms might find it difficult to effectively process and extract knowledge from enormous volumes of data as big data becomes more and more available. Future research should prioritize creating algorithms capable of handling large-scale datasets effectively, such as distributed learning methodologies [19].

Lastly, there may be a lot of advantages to combining Supervised Learning (SL) algorithms with other machine learning techniques. Robust models that can handle a variety of tasks can be created by combining different learning paradigms. Models may be pre-trained on unlabeled data by unsupervised learning algorithms, for example, and then labeled data will be refined using through supervised learning [20].

4. Conclusion

Machine learning has been extensively applied in microelectronics fault detection. This work provides a systematic exposition of microelectronics and machine learning concepts, summarizes key applications of machine learning in microelectronics, and offers detailed examination of fault detection case studies. It specifically analyzes the implementation of various models—including CNNs, RNNs and Transformer—in fault detection scenarios. The paper comprehensively addresses prevalent algorithms, current challenges, and future research directions. While encouraging further research to improve their efficacy, a basic overview of machine learning applications in microelectronics is introduced in the article. It is anticipated that this work will be helpful to scholars, practitioners, and students in related fields.

References

[1] Kreuzberger D, Kühl N, Hirschl S. Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 2023, 11: 31866–31879.

[2] Zhou S, Li X, Gao B. Emerging microelectronic microneedles (eMN) for biomedical applications. Journal of Materials Chemistry C, 2024.

[3] Guarascio D, Pianta M. Digital technologies: civilian vs. military trajectories. LEM Working Paper Series, 2025, No. 2025/08.

[4] Dehkharghanian T, et al. Applied machine learning in hematopathology. International Journal of Laboratory Hematology, 2023, 45: 87–94.

[5] Nazareth N, Reddy YVR. Financial applications of machine learning: A literature review. Expert Systems with Applications, 2023, 219: 119640.

[6] Nyako K, et al. Building trust in microelectronics: A comprehensive review of current techniques and adoption challenges. Electronics, 2023, 12(22): 4618.

[7] Calabrese M, et al. Application of Mask R-CNN and YOLOv8 algorithms for defect detection in printed circuit board manufacturing. Discover Applied Sciences, 2025, 7(4): 257.

[8] Bhattacharya A, Cloutier SG. End-to-end deep learning framework for printed circuit board manufacturing defect classification. Scientific Reports, 2022, 12(1): 12559.

[9] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 2020, 404: 132306.

[10] Liu S, et al. Recurrent neural networks model based reliability assessment of power semiconductors in PMSG converter. Microelectronics Reliability, 2021, 126: 114314.
[11] de Jong SDM, Ghezeljehmeidan AG, van Driel WD. Solder joint reliability predictions using physics-informed machine learning. Microelectronics Reliability, 2025, 172: 115797.
[12] Vaswani A, et al. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30.
[13] Suresh G, et al. Advancements in nanoelectronics leveraging transformer algorithms for enhanced biomedical data science applications.
[14] Lin W. Intelligent fault diagnosis of consumer electronics sensor in IoE via transformer. IEEE Transactions on Consumer Electronics, 2023, 70(1): 1259–1267.
[15] Barbierato E, Gatti A. The challenges of machine learning: A critical review. Electronics, 2024, 13(2): 416.
[16] Obaido G, et al. Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. Machine Learning with Applications, 2024, 17: 100576.
[17] Thulasiram PP. Explainable Artificial Intelligence (XAI): Enhancing transparency and trust in machine learning models, 2025.
[18] Zhao Z, Li Z, Jiang J, et al. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. Expert Systems with Applications, 2024, 242: 122807.
[19] Bian K, Priyadarshi R. Machine learning optimization techniques: A survey, classification, challenges, and future research issues. Archives of Computational Methods in Engineering, 2024, 31(7): 4209–4233.
[20] Bao W, Ning L, Kong Y. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems with Applications, 2019, 128: 301–315.