

13th CIRP Global Web Conference (CIRPe 2025)

Automating Industrial Quality Assurance: A Zero-Shot MLLM Framework for Defect Detection and 3R

Shuchao Liu^a, Dominik Koch^a, Jan-Philipp Kaiser^a, Florian Stamer^{b*}

^a *wbk Institut für Produktionstechnik, Kaiserstraße 12, 76131 Karlsruhe, Germany*

^b *Leuphana University, Universitätsallee 1, 21335 Lüneburg, Germany*

* Corresponding author. E-mail address: florian.stamer@leuphana.de

Abstract

In modern industrial production, particularly within the circular economy, automated defect detection is crucial for quality control and resource efficiency. Traditional inspection methods are often inconsistent and inefficient, while existing machine learning approaches are constrained by data scarcity, poor generalization, and a lack of interpretability. This paper proposes a zero-training visual inspection framework that leverages Multi-Modal Large Language Models (MLLMs) for automated defect detection and decision-making in industrial components, using starter motors as a case study. The framework integrates the zero-shot reasoning capabilities of an MLLM with an Elasticsearch-powered knowledge base to create an end-to-end automated system for "defect identification, standard matching, and decision-making." By processing images and relevant technical documentation, the system identifies defects and provides handling recommendations based on the 3R principles (Reuse, Remanufacture, Recycle) without requiring task-specific training. Experimental results on 584 starter motor samples show that our optimized system achieves an overall accuracy of 81.4% and an F1 score of 85.3%. This work validates the potential of MLLMs for industrial visual inspection and provides a replicable framework for advancing intelligent automation in manufacturing.

© 2025 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Peer review under the responsibility of the scientific committee of the CIRPe 2025

Keywords: Multi-Modal Large Language Model, Visual Defect Detection, Zero-Shot Learning, 3R Decision-Making, Quality Assurance, Industrial Automation.

1. Introduction

The transition towards a circular economy necessitates advanced methods for extending product lifecycles and maximizing resource utilization [1]. A key enabler is efficient and accurate quality assurance (QA), particularly defect detection in used components to determine their suitability for Reuse, Remanufacture, or Recycling (3R). However, traditional manual inspection is prone to human error and low throughput [2], while existing automated solutions based on computer vision (CV) and deep learning (DL) face significant hurdles.

Classical CV methods rely on hand-crafted features and are often too rigid for complex industrial environments [3]. While DL models like Convolutional Neural Networks (CNNs) have

shown great promise, they typically require large, manually annotated datasets of defect samples, which are expensive and difficult to obtain in industrial settings [4]. Furthermore, these models often act as "black boxes," lacking the interpretability needed for critical decision-making and failing to generalize to novel, unseen defect types [5].

Single-modality methods, such as those based solely on RGB images, face inherent challenges in industrial defect detection. They often fail to capture the diverse range of industrial defects, are susceptible to noise and occlusion, and struggle to distinguish visually similar anomalies, leading to false positives and missed detections [15]. These limitations underscore the need for a more robust approach that integrates

complementary data modalities to enhance defect detection and decision-making.

To overcome these limitations, leveraging multi-modal inputs has emerged as a promising direction. By integrating visual and textual information, multi-modal methods can address the shortcomings of single-modality approaches and enhance defect detection in industrial settings. The recent emergence of MLLMs offers a paradigm shift. MLLMs, pre-trained on vast amounts of text and image data, possess remarkable zero-shot reasoning capabilities, allowing them to understand and analyze images without task-specific fine-tuning [6]. This enables them to perform complex tasks like identifying defects, understanding their context, and generating human-readable explanations.

This paper introduces a framework that harnesses these capabilities for industrial QA. Our main contributions are:

- A zero-training framework for visual defect detection using MLLMs, eliminating the need for extensive labeled data.
- An end-to-end automated pipeline that integrates visual analysis with a knowledge base of technical standards to provide justified 3R-based decisions.
- A comparative analysis of leading open-source MLLMs to identify a suitable model for industrial deployment.
- Empirical validation of the framework on a real-world dataset of starter motors, demonstrating its effectiveness and practical value.

2. Theory & Metrics

MLLMs are capable of integrating and processing diverse data modalities such as text, images, and audio, enabling cross-modal understanding and reasoning. Their general architecture consists of specialized modality encoders (which convert raw data from each modality into unified vector representations), an input projector (which aligns features from different modalities into the textual feature space for prompt construction), a core language model (which serves as the semantic engine for multimodal understanding, reasoning, and generation—typically using encoder-decoder or decoder-only frameworks), and output projectors and modality generators (which transform the LLM’s output tokens back into content in the target modality). This architecture allows MLLMs to efficiently handle complex cross-modal tasks—such as image-text retrieval, image captioning, and visual question answering—demonstrating strong capabilities in multimodal perception and reasoning for applications like defect detection [6].

The McNemar test rigorously evaluates whether the difference in binary outcomes between paired conditions—such as a large language model versus random selection—is statistically significant, by focusing on cases where individual outcomes change between methods [7].

3. Related Work

Traditional Computer Vision (CV): Early automated inspection systems relied on methods like threshold segmentation [8], morphological operations [8], and template matching [9]. These traditional computer vision (CV)

methodologies primarily rely on handcrafted features and explicit rules for image processing and defect detection, focusing on analyzing and interpreting image data through mathematical and algorithmic means. While these techniques are computationally efficient, they are highly sensitive to variations in lighting, orientation, and background, which limits their robustness in real-world scenarios.

Machine and Deep Learning (DL): The advent of machine learning introduced more robust classifiers like Support Vector Machines (SVM) and Random Forests [10], which improved performance but still depended on manual feature engineering. Deep learning, particularly CNNs, revolutionized the field by enabling end-to-end feature learning [4]. Models like YOLO [11] and others have achieved high accuracy in detecting various defects. However, their reliance on large, annotated datasets and their limited ability to handle unseen defect types remain major barriers to widespread adoption in dynamic industrial environments.

Multi-Modal Approaches: To overcome the limitations of single-modality systems, researchers have explored fusing data from multiple sources (e.g., RGB and depth images) [12]. Fusion can occur at early, intermediate, or late stages, each with its own trade-offs: early fusion integrates raw features to reduce parameters but risks modality imbalance [15][16]; intermediate fusion allows independent processing before aggregation, improving feature complementarity [15]; and late fusion combines predictions at the decision level, though it often underutilizes cross-modal correlations [17]. While these methods enhance robustness—for instance, EasyNet achieves an impressive 92.6% I-AUROC on the MVTec 3D-AD dataset through fusion techniques [18]—they rely on unsupervised learning, leveraging modality-specific datasets for training. Although this approach reduces dependency on labeled data, it introduces architectural complexity and fails to fundamentally address challenges such as data scarcity and zero-shot generalization.

In summary, traditional CV-based defect detection methods are sensitive to environmental variations, deep learning approaches require large labeled datasets and struggle with generalizing to novel classes, and multimodal fusion techniques—while more robust—add complexity without fully resolving data scarcity or adaptability. Recently, zero-shot MLLMs have enabled interpretation of both visual and textual data without task-specific training. Our work builds on these advances by leveraging the inherent zero-shot, cross-modal reasoning power of MLLMs. Instead of training a model on specific defects, we prompt a general-purpose MLLM to act as a domain expert, providing it with visual evidence and relevant technical context to make an informed, interpretable decision.

4. Methodology

Our proposed framework automates the quality assurance process through a three-stage pipeline, as illustrated in Figure 1.

4.1. Framework Overview

This system employs a streamlined architecture with three main modules: part identification and description via a MLLM,

database retrieval using Elasticsearch, and comprehensive 3R (Reuse, Remanufacture, Recycle) decision-making based on integrated multimodal data, as shown in Figure 2. The workflow begins with image input, which is analyzed by the MLLM to identify the part and generate its description. Key details—such as part name, material, and surface condition—are then used to query a structured knowledge repository. This repository is powered by Elasticsearch, an open-source search and analytics engine optimized for fast, full-text and semantic retrieval across large document collections. It contains technical documentation such as maintenance manuals, inspection standards, and material data sheets, stored in indexed form to support rapid and flexible querying. The MLLM uses the retrieved documents—typically PDFs or structured texts—alongside the image input to deliver 3R recommendations based on a comprehensive multimodal assessment, while maintaining interfaces and mechanisms for human expert intervention.

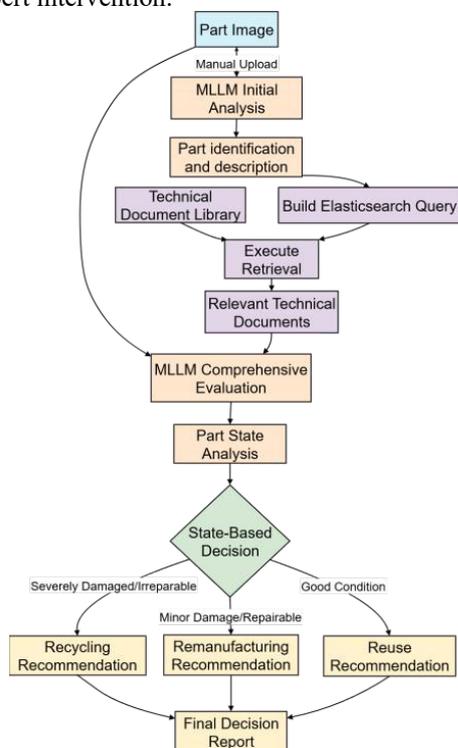


Figure 1: End-to-end system workflow, from image upload to final 3R decision report.

- Initial Part Analysis:** This module receives uploaded part images via a GUI, converts them to base64, and analyzes them through a remote MLLM API. Prompts perform two tasks: (a) part identification—matching against a predefined list using pretrained knowledge, and (b) appearance description—extracting material, surface features, and other attributes. Results are parsed, standardized, and stored in JSON for efficient downstream use. This process improves retrieval accuracy and supports database queries for technical documentation.
- Knowledge Retrieval:** The identified part name and descriptive keywords are used to build a query sent to an Elasticsearch-based knowledge base of technical documents (manuals, data sheets, 3R Standard Operating Procedures). This model aligns with recent efforts to use

structured data for autonomous QA, e.g., Asset Administration Shell concepts[14]. Retrieval proceeds in three stages: (1) Keyword-based search: queries from MLLM-identified terms are matched against Elasticsearch. This is fast and simple but depends on keyword quality and is vulnerable to misspellings or inconsistent wording. (2) Secondary filtering: preliminary results are checked so only documents matching all tokens remain. This improves precision but can be too strict, excluding valid documents with minor term variation. (3) Semantic retrieval: embedding models (e.g., all-mpnet-base-v2) encode queries and documents as semantic vectors, with cosine similarity ranking candidates. This boosts robustness and recall, handling natural language variation and recognition errors, though at higher computational cost. A hybrid approach combines keyword speed with semantic resilience, ensuring relevant documents are retrieved even with minor MLLM inaccuracies.

- Integrated Assessment & Decision-Making:** The MLLM Comprehensive Evaluation and Decision-Making Module serves as the system's core, integrating part images and technical documentation for analysis via a MLLM. The workflow includes data input, comprehensive evaluation, and decision generation. The system reads part images and corresponding technical texts, then submits them together as a consolidated prompt to the model, which acts as a QA expert. The model assesses the part's condition against provided standards, generating a structured JSON output for each inspection criterion (such as rust, cracks, deformation), a final 3R decision (Reuse, Remanufacture, or Recycle), a confidence score, and a detailed justification. These results are compiled into an Excel file for efficient analysis and traceability, enabling automated, transparent, and extensible inspection and decision-making. To enhance the system's robustness, the framework allows for a human intervention interface. During the evaluation process, if the AI confidence score falls below a predefined threshold, the system automatically triggers a mechanism for human expert intervention. Through an interactive interface, experts can review the AI's reasoning process and correct the results.

4.2. MLLM Selection

Selecting an appropriate MLLM is critical. We conducted a comprehensive evaluation of five prominent models: gpt-4o (as



Figure 2: System Architecture Overview.

a closed-source benchmark), llama-3.2-90b-vision, minimax-01, qwen2.5-vl-32b, and pixtral-large-2411.

The evaluation utilized a dataset of 584 real-world starter motor images, including 386 defective (NOK) and 198 non-defective (OK) samples, resulting in a moderate class imbalance. Each image was simply labeled as either OK or NOK (cf. Figure 3).

To ensure a thorough and objective comparison, the evaluation was structured into three progressive rounds. The



Figure 3: Examples of OK starter motor (l.) and NOK starter motor (r.)

first round focused on simple binary classification, where each model was tasked with categorizing images as either OK or NOK, providing a direct measure of their baseline defect detection capabilities. In the second stage, the built-in prompts were optimized to provide more specific evaluation criteria, and models were required to output a confidence score between 0 and 1 for each prediction. This allowed for a more detailed analysis of the models' judgment confidence and decision boundaries. In this way, we not only assessed the accuracy of the predictions but also analyzed how confidently each model distinguished between defective and non-defective samples. The third round introduced threshold optimization, systematically adjusting the decision threshold to identify the optimal balance between precision and recall for each model. This process was crucial for uncovering each model's sensitivity to threshold selection and for maximizing overall detection performance in practical scenarios.

During the performance evaluation, we adopted a rigorous and comprehensive set of standard metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC), to assess the performance of each model in defect detection tasks from multiple dimensions and perspectives. These metrics not only reflect the overall classification capability of the models but also provide deeper insights into their recognition performance across different sample categories and their reliability in practical applications. To further ensure the scientific validity and statistical significance of the experimental results, we employed the McNemar test to compare the model predictions against random guessing. The test results indicated that all models performed significantly better than random guessing ($p < 0.00001$), fully demonstrating the effectiveness and practical value of the models in real-world defect detection tasks.

While gpt-4o achieved the highest overall performance (AUC 0.90), its closed-source nature imposes limitations on flexibility and customization for industrial deployment. Among the open-source models, pixtral-large-2411 demonstrated the best balance between identifying OK and NOK samples and maintained stable performance across different metrics (AUC 0.87). Its confidence score distributions revealed robust discriminative capability without excessive bias toward either class, which is particularly important in real-world scenarios characterized by class imbalance and diverse defect types. Moreover, Pixtral's open-source accessibility offers significant advantages in terms of scalability, adaptability, and cost-

effectiveness. Its architecture allows for further fine-tuning and integration within existing manufacturing systems, ensuring both autonomy and data security for enterprise applications.

In summary, the comprehensive multi-phase evaluation underscores Pixtral's suitability as the core MLLM for subsequent system development. Its balanced nature, robust performance across multiple metrics, and open-source flexibility make it the ideal choice for integration into our final starter motor visual defect detection system. Figure 4 shows the performance metrics for Pixtral at its optimal threshold, further illustrating its effectiveness and reliability in industrial inspection tasks.

5. Experiments and Results

This study presents the evaluation of a comprehensive end-to-end decision support system powered by the Pixtral MLLM. The system integrates image-based component recognition with automated retrieval of relevant technical standards and maintenance manuals. Given an input image of a starter motor component, the system first performs component identification, then retrieves associated documentation, and finally conducts a 3R (Reuse, Repair, Recycle) decision analysis. The model autonomously summarizes inspection items and criteria from both image and technical documents, generating a structured analysis report that includes visual assessment results, relevant standards, decision grades, and detailed justifications. Furthermore, the alignment mechanism between text and image was reinforced to ensure the reasoning process is strictly grounded in the reference documents, thus improving the consistency of multimodal information integration.

A total of 584 starter motor images were used for full workflow validation. The component recognition module achieved a correct identification rate of 79.5%. During the document retrieval stage, the system reached a relevant document retrieval rate of 86.1%. The final classification task was performed on the 503 samples where documents were successfully matched, allowing comparison between image-only and image-plus-document modes, as well as further optimization through decision threshold adjustment.

Table 1 summarizes the core results. When using the MLLM with **image only**, the model achieved a baseline F1 score of 0.8614, with a precision of 0.8246 and a recall of 0.9016.

However, when we introduced the **image + technical documents**, the model's performance profile changed dramatically. Precision surged to 0.9677, while recall dropped sharply to 0.4286. This pronounced shift reflects the influence of the technical documentation, which imposed strict and explicit criteria on the model's decision-making process. The presence of these detailed standards made the model highly confident in its positive predictions, resulting in a substantial reduction in false positives and thus a notable improvement in precision. At the same time, the model became considerably more conservative, tending to classify borderline or incomplete cases as non-defective, which led to a significant increase in false negatives and a corresponding decline in recall. This phenomenon can be described as a dual constraint effect: while the technical documentation enhances the model's ability to

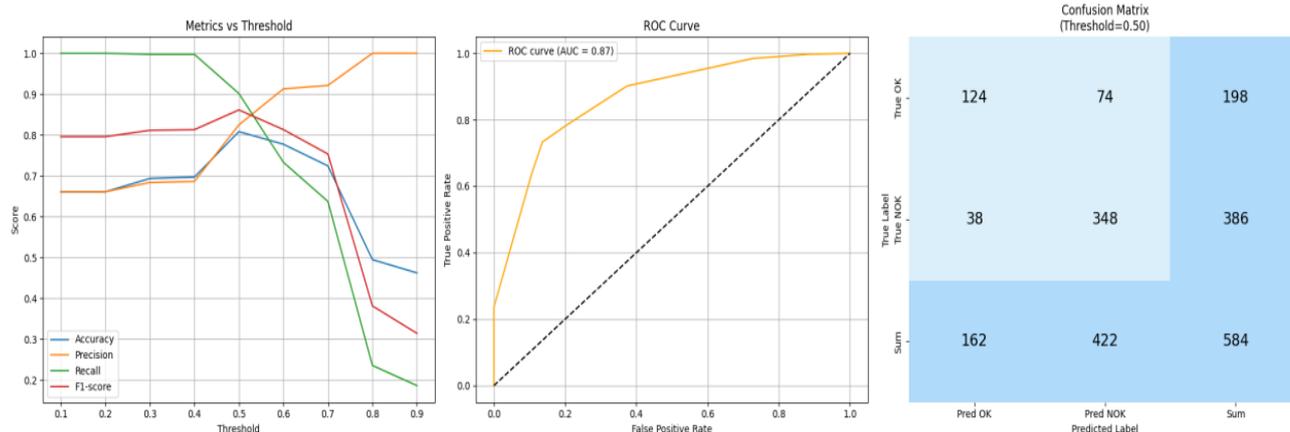


Figure 4: Performance metrics for Pixtral at the optimal threshold (0.50), including ROC curve (AUC=0.87) and confusion matrix.

confirm defects with certainty, it also raises the threshold for what is considered defective, making the system less tolerant of ambiguity and more likely to overlook subtle or incomplete defects.

To address this trade-off between precision and recall, we performed **threshold optimization** on the model's confidence scores with the goal to check what is theoretically possible meaning this is not an option for a real world application as the dataset is not known a priori. By lowering the decision threshold from the default 0.5 to 0.3, we were able to achieve a more balanced performance. The final optimized system reached an **accuracy of 0.8140**, a **precision of 0.9010**, a **recall of 0.8100**, and a significantly improved **F1 score of 0.8530**. This systematic threshold adjustment demonstrates that it is possible to alleviate the rigid trade-off imposed by the technical documentation, enabling the system to maintain high precision while recovering much of the lost recall. In practical terms, this means the model can effectively identify more true defects without substantially increasing the risk of false positives, thus providing a more reliable and robust solution in real-world applications.

Table 1: Overall Performance Comparison

Metric	Unimodal (Image Only)	Multimodal (Image + Text)	Multimodal (Optimal Threshold (0.3))
Accuracy	0.8082	0.5900	0.8140
Precision	0.8246	0.9677	0.9010
Recall	0.9016	0.4286	0.8100
F1 Score	0.8614	0.5941	0.8530

In summary, the observed changes in model performance highlight the nuanced impact of integrating technical documentation into the defect detection pipeline. While explicit standards can greatly improve precision by minimizing false positives, they may also inadvertently suppress recall by causing the model to disregard ambiguous or marginal cases. Overall, it is important to note, that the baseline OK/NOK labels used in evaluation are based on subjective human judgment. In practice, the decision whether a component is "OK" or "NOK" is often qualitative, relying on experience-based assessments rather than clearly defined, objective

thresholds. This subjectivity can introduce inconsistencies into the ground truth and limits the interpretability of absolute performance scores, especially in borderline cases.

6. Conclusion and Future Work

This paper has presented a novel framework for automated quality assurance in industrial settings using MLLMs. By combining zero-shot visual analysis with a knowledge base of technical standards, our system provides an end-to-end, interpretable solution for 3R decision-making without the need for task-specific training data. Experiments show that part identification accuracy is 79.5%, which limits the reliability of subsequent knowledge retrieval (accuracy ~86.1%). A closer analysis of these identification errors reveals that they primarily stem from high inter-class similarity (e.g., visually similar), high intra-class variation (e.g., wear and tear), inconsistent image quality (lighting, focus), and data imbalance. These results highlight ongoing challenges with misidentification and retrieval errors in complex industrial settings, emphasizing the need for greater robustness and generalizability. Analysis further reveals that incorporating technical documentation improves precision but significantly lowers recall, thereby reducing the F1 score and overall performance. Adjusting the decision threshold helps balance precision and recall, raising accuracy to 0.8140, precision to 0.9010, recall to 0.8100, and F1 score to 0.8530. This suggests that the benefits of multimodal information integration depend heavily on appropriate parameter tuning and fusion strategies.

This study demonstrates the application potential of open-source MLLMs in 3R (Remanufacturing, Reuse, Recycling) decision support systems for industrial parts, while highlighting ongoing challenges in part identification, information fusion, and practical deployment. Building on the strengths of MLLMs, future optimizations can be pursued from several key perspectives.

- Enhancing Part Identification:** The accuracy and robustness of part identification remain the main bottlenecks. In complex industrial environments with diverse and morphologically varied parts, constructing larger, more representative image datasets is crucial to fully leverage MLLMs' visual and semantic capabilities. Enhancing model generalization under small-sample and long-tail scenarios will be critical for real-world utility.

2. **Optimizing Multimodal Information Fusion:** Optimizing multimodal information fusion also requires further exploration. The current system uses prompt-level serial fusion, inputting images and text separately for unified processing—an approach that, while simple, leaves room for improvement in complementarity and alignment. Future work should explore more flexible fusion strategies, such as adaptive weighting to dynamically balance modalities or multi-level attention structures for deeper modeling of image-text associations.
3. **Improving Utilization of Technical Documentation:** Finally, the utilization of technical documents remains an important challenge. Although integrating such documents improves precision, recall declines, indicating limited knowledge integration and handling of borderline cases. For the future, an impact analysis of the documents covering all technical aspects is a first step. In addition, future research should develop more intelligent document parsing methods to improve the system's interpretation and application of complex technical standards.

References

- [1] K.-Y. Lin, 'Circular supply chain for smart production in Industry 4.0', *Computers & Industrial Engineering*, vol. 198, p. 110682, Dec. 2024, doi: 10.1016/j.cie.2024.110682.
- [2] P. Kopardekar, A. Mital, and S. Anand, 'Manual, Hybrid and Automated Inspection Literature and Current Research Current Research', *Integrated Manufacturing Systems*, vol. 4, no. 1, pp. 18–29, Jan. 1993, doi: 10.1108/09576069310023838.
- [3] D. Ziou and S. Tabbone, "'Edge detection techniques: An overview'", *International Journal of Pattern Recognition and Image Analysis*, vol. 4, pp. 537–559, Jan. 1998.
- [4] R. Khanam, M. Hussain, R. Hill, and P. Allen, 'A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications', *IEEE Access*, vol. 12, pp. 94250–94295, 2024, doi: 10.1109/ACCESS.2024.3425166.
- [5] F. Stamer, R. Jachemich, S. Puttero, E. Verna, and M. Galetto, 'Integrative inspection methodology for enhanced PCB remanufacturing using artificial intelligence', *Procedia CIRP*, vol. 132, pp. 227–232, Jan. 2025, doi: 10.1016/j.procir.2025.01.038.
- [6] U. Kamath, K. Keenan, G. Somers, and S. Sorenson, *Large Language Models: A Deep Dive: Bridging Theory and Practice*. Cham: Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-65647-7.
- [7] Q. McNemar, 'Note on the sampling error of the difference between correlated proportions or percentages', *Psychometrika*, vol. 12, no. 2, pp. 153–157, June 1947, doi: 10.1007/BF02295996.
- [8] N. Otsu, 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
- [9] A.-A. I. M. Hassanin, F. E. Abd El-Samie, and G. M. El Banby, 'A real-time approach for automatic defect detection from PCBs based on SURF features and morphological operations', *Multimed Tools Appl*, vol. 78, no. 24, pp. 34437–34457, Dec. 2019, doi: 10.1007/s11042-019-08097-9.
- [10] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You Only Look Once: Unified, Real-Time Object Detection', May 09, 2016, *arXiv*: arXiv:1506.02640. doi: 10.48550/arXiv.1506.02640.
- [12] R. Chen et al., 'EasyNet: An Easy Network for 3D Industrial Anomaly Detection', Sept. 01, 2023, *arXiv*: arXiv:2307.13925. doi: 10.48550/arXiv.2307.13925.
- [13] D. Koch, J.-P. Kaiser, F. Stamer, R. Stark, and G. Lanza, 'Enhancing Visual Inspection in Remanufacturing: A Reinforcement Learning Approach with Integrated Robot Simulation', *Procedia CIRP*, vol. 134, pp. 939–944, Jan. 2025, doi: 10.1016/j.procir.2025.02.228.
- [14] A. Bilen, F. Stamer, S. Behrendt, and G. Lanza, 'A Quality Data Model Based on Asset Administration Shell Technology to Enable Autonomous Quality Control Loops', in *Production at the Leading Edge of Technology*, T. Bauernhansl, A. Verl, M. Liewald, and H.-C. Möhring, Eds, Cham: Springer Nature Switzerland, 2024, pp. 195–204. doi: 10.1007/978-3-031-47394-4_20.
- [15] Y. Lin, Y. Chang, X. Tong, J. Yu, A. Liotta, G. Huang, W. Song, D. Zeng, Z. Wu, Y. Wang, and W. Zhang, 'A survey on RGB, 3D, and multimodal approaches for unsupervised industrial image anomaly detection', *Information Fusion*, vol. 121, p. 103139, Sept. 2025. doi: 10.1016/j.inffus.2025.103139.
- [16] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, 'Early versus late fusion in semantic video analysis', in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, Nov. 2005, pp. 399–402. doi: 10.1145/1101149.1101236.
- [17] K. Gadzicki, R. Khamsehashari, and C. Zetsche, 'Early vs late fusion in multimodal convolutional neural networks', in *Proceedings of the 2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, Jul. 2020, pp. 1–8. doi: 10.23919/FUSION45008.2020.9190246.
- [18] R. Chen, G. Xie, J. Liu, J. Wang, Z. Luo, J. Wang, and F. Zheng, 'Easynet: an easy network for 3D industrial anomaly detection', in *arXiv preprint arXiv:2307.13925*, Jul. 2023. doi: 10.48550/arXiv.2307.13925.