A Decision Framework for AI/MLOps Toolchain Selection in Manufacturing

Martin Bischof[a] and Florian Wahl[b]
Faculty of Applied Computer Science, Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany
fl

Keywords: Computer Vision, Quality Assurance, Industrial Implementation, Deep Learning, Defect Detection, Platform Selection.

Abstract: This paper addresses the growing challenge of implementing and selecting appropriate Machine Learning Operations toolchains in manufacturing environments, where computer vision applications are becoming increasingly prevalent. We introduce a comprehensive framework that uniquely combines MLOps platform evaluation criteria with a practical workflow methodology tailored for manufacturing settings. To validate our framework, we conducted experiments using the MVTec Anomaly Detection dataset, achieving 77.78 % accuracy in granular defect-type classification when deployed through a commercial MLOps platform. Our framework effectively bridges the gap between theoretical principles and real-world manufacturing constraints by emphasizing both technical requirements and workflow considerations. This research advances industrial AI implementation by providing a systematic methodology that transcends conventional data mining approaches while specifically addressing manufacturing-sector challenges. Our findings demonstrate that successful MLOps toolchain selection necessitates a balanced evaluation of both functional capabilities and implementation workflows.

1 INTRODUCTION

The manufacturing sector is at a pivotal juncture, where traditional industrial processes intersect with modern artificial intelligence (AI) capabilities. Since the breakthrough in deep learning architectures (Krizhevsky et al., 2012), computer vision applications have become increasingly viable in manufacturing environments, driven by the widespread availability of high-resolution cameras and enhanced computational resources. However, the practical deployment of these systems remains challenging. Studies indicate that 75–85 % of machine learning projects fail to meet sponsor expectations (Studer et al., 2021), often due to difficulties in scaling from proof-of-concept to sustained operational implementation.

While traditional methodologies, such as the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework (Wirth and Hipp, 2000), have served as foundational guidelines for data mining projects, they were not explicitly designed to address the complexities of modern Machine Learning (ML) applications within manufacturing environments. Manufacturers encounter numerous challenges in operationalizing ML systems, including the need to ensure high-quality data, monitor model performance in dynamic and evolving settings, and integrate these systems with legacy industrial control infrastructures. Additionally, standardizing interfaces across diverse sensor manufacturers and sustaining model reliability amidst fluctuating production conditions remain persistent obstacles. The advent of specialized Machine Learning Operations (MLOps) platforms presents promising solutions to these challenges. However, manufacturers often struggle with selecting and implementing the most suitable toolchains for their unique requirements. This difficulty is compounded by the absence of a systematic decision-making framework tailored to the nuanced operational demands of manufacturing contexts.

In response to these challenges, this paper provides two key contributions:

1. A comprehensive requirements for evaluating MLOps platforms in manufacturing environments
2. A practical workflow methodology for computer

[a] https://orcid.org/0009-0005-1819-3976
[b] https://orcid.org/0000-0002-1163-1399

Bischof, M., Wahl and F.
A Decision Framework for AI/MLOps Toolchain Selection in Manufacturing.
DOI: 10.5220/0013523400003967
In *Proceedings of the 14th International Conference on Data Science, Technology and Applications (DATA 2025)*, pages 439-446
ISBN: 978-989-758-758-0; ISSN: 2184-285X

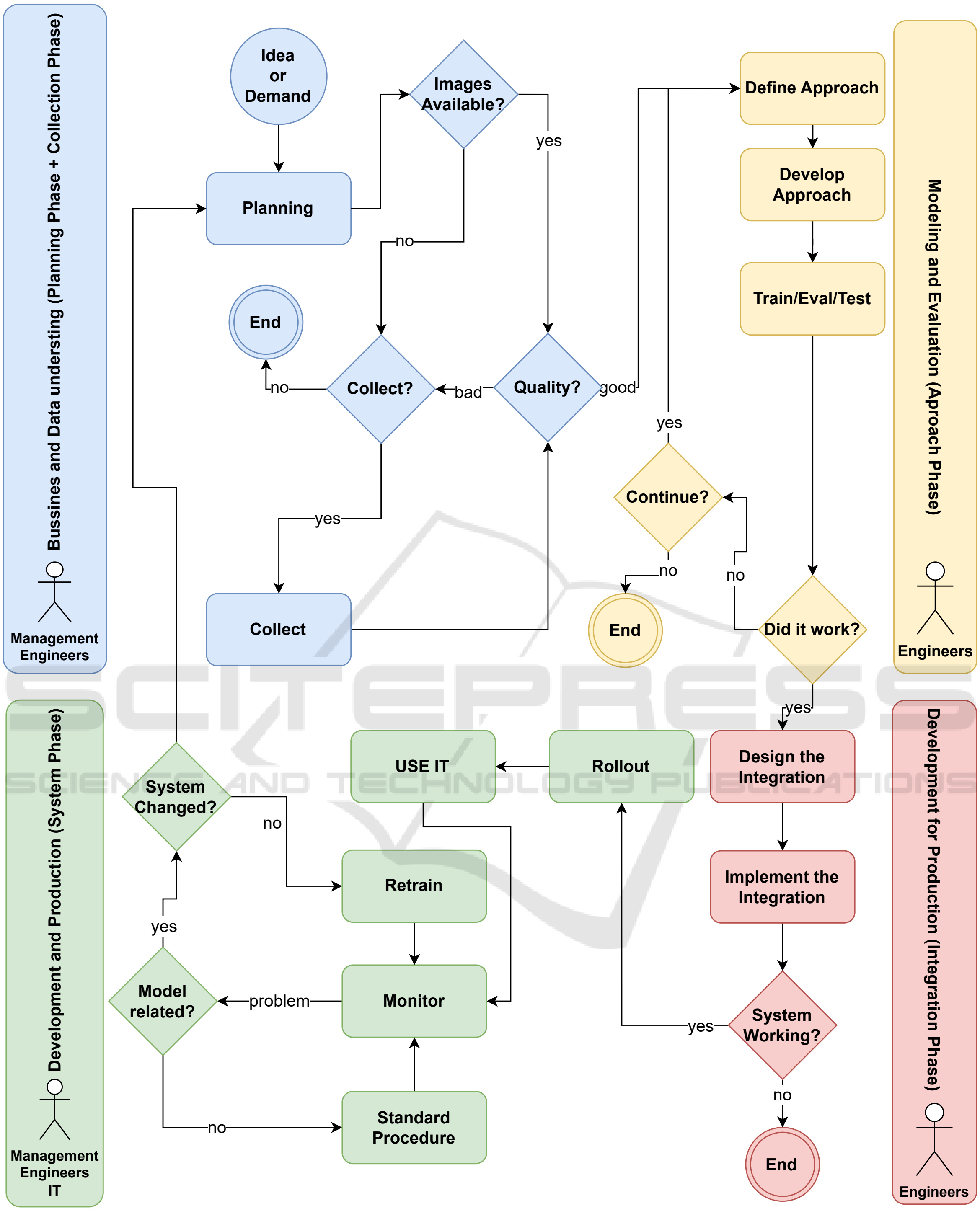

Figure 1: Our proposed workflow for computer vision projects in manufacturing environments. The workflow consists of five sequential phases: Planning, Collection, Approach, Integration, and System. Each phase incorporates specific decision points, documentation requirements, and stakeholder roles to ensure systematic progression and quality control throughout the implementation process.

vision implementation that bridges theoretical foundations with manufacturing constraints

Through these contributions, we aim to provide manufacturers with a structured approach to MLOps toolchain selection and implementation, specifically tailored to computer vision applications in industrial settings.

2 RELATED WORK

Our analysis focuses on three key research domains: Computer Vision methodologies, manufacturing AI integration, and MLOps. The field of AI has established foundational concepts and methodologies, as fully documented in the standard reference work (Russell and Norvig, 2020). Similarly, the domains of data mining and ML have undergone a significant evolution since the introduction of standardized methodologies. The CRISP-DM framework (Wirth and Hipp, 2000), for instance, emerged as a hierarchical process model structured across four levels of abstraction, ranging from general phases to specific process instances. This methodology provided a robust and comprehensive approach for executing data mining projects, remaining independent of both industry sectors and the technologies employed. However, recent research has highlighted limitations in the original CRISP-DM framework, particularly in its applicability to modern ML use cases. The traditional model lacks explicit guidance on quality assurance methodologies and does not adequately address scenarios where ML models must make real-time decisions over extended periods. To overcome these deficiencies, Cross-Industry Standard Process for Machine Learning with Quality Assurance (CRISP-ML(Q)) (Studer et al., 2021) was introduced, incorporating quality assurance practices across six well-defined phases while preserving its neutrality with respect to industries and applications. This evolution has proven critical, as surveys indicate that 75-85 % of practical ML projects fail to meet sponsor expectations. In manufacturing contexts, the adoption of data-driven approaches presents distinct challenges. For example, (Tripathi et al., 2020) demonstrated that applying robust, industry-specific knowledge discovery models often encounters numerous obstacles related to data and model development. These challenges include experimental design considerations, managing model complexity, addressing class imbalance issues, and mitigating concerns related to data dimensionality. Moreover, the manufacturing sector requires systematic and efficient coordination between different phases of the knowledge discovery process to ensure success. The emergence of MLOps as a discipline has introduced new paradigms for implementing ML systems in manufacturing environments. For instance, (Beck et al., 2020) examines processes for developing, integrating, and operating ML systems effectively. In addition, (Faubel and Schmid, 2024) conducted multiple case studies on implementing MLOps within Industry 4.0 contexts, emphasizing the processes, tools, and organizational structures necessary for reliable model deployment. Recent advancements (Jon Bokrantz and Skoogh, 2024) have further extended the CRISP-DM framework specifically for manufacturing applications by introducing an “Operation and Maintenance” phase. This extension underscores the importance of managing AI drift while ensuring that domain expertise, data science proficiency, and data engineering competency are maintained throughout all process phases. In particular, it highlights the critical role of data engineering, a component often overlooked in conventional AI workflows. In the realm of computer vision, significant progress has been made since (Krizhevsky et al., 2012) demonstrated breakthrough performance in image classification using deep convolutional neural networks. Current industrial implementations focus on practical considerations such as build versus buy decisions for vision-based AI software in manufacturing environments (Robovision, 2024). Additionally, (Schneider et al., 2024) explored integration challenges within Industry 4.0 ecosystems, identifying four key areas: system integration, data-related issues, workforce adaptation concerns, and ensuring trustworthy AI implementation.

3 PROPOSED WORKFLOW

We introduce a structured framework designed to guide computer vision projects in manufacturing environments. Divided into five distinct phases, the workflow addresses critical aspects of project development and implementation.

3.1 The Five Phases

Our proposed workflow is structured in five phases as follows:

Phase 1: Planning Phase The foundational planning phase establishes the groundwork necessary for achieving project success. During this stage, stakeholders engage in comprehensive requirements engineering to clearly define the scope and objectives. Collaborative sessions facilitate the creation of detailed documentation outlining resource allocation, timeline constraints, and key success factors. Quantifiable quality metrics and well-defined acceptance criteria are developed to serve as benchmarks for subsequent phases.

Phase 2: Collection Phase Building on the planning stage, the collection phase emphasizes systematic data acquisition and preparation activities. Rigorous protocols ensure consistency and reliability in image acquisition processes. To maintain high data quality throughout the project life cycle, standardized and robust validation procedures are implemented. Special focus is placed on verifying image quality parameters and ensuring dataset completeness before advancing to later stages.

Phase 3: Approach The approach phase focuses on designing technical solutions and crafting an implementation strategy tailored to manufacturing constraints. This pivotal stage involves algorithm selection and architectural decisions aligned with production requirements. Model development proceeds iteratively, with each cycle incorporating optimization strategies informed by manufacturing-specific performance metrics. Validation processes are carefully designed to address metrics relevant to industrial applications.

Phase 4: Integration Phase Integration serves as a critical link between development and production deployment. Seamless compatibility with existing manufacturing infrastructure is prioritized to ensure system performance under real-world conditions. Interfaces are developed alongside comprehensive testing protocols to guarantee system reliability. Deployment documentation becomes increasingly detailed, incorporating practical insights from the manufacturing environment to streamline implementation efforts.

Phase 5: System Phase A final system phase addresses challenges associated with production deployment and long-term operation. Continuous assessment of system health is enabled through sophisticated monitoring mechanisms. Clear guidelines for system upkeep are established through maintenance protocols, while iterative improvement mechanisms drive ongoing optimization efforts. Performance in dynamic manufacturing environments remains a central focus during this phase.

3.2 Implementation

The effectiveness comes from the structured progression of the workflows through clearly defined transitions between phases. Each transition is validated against predefined criteria, ensuring systematic advancement while upholding quality standards throughout the process. Empirical observations have informed specific cycle limitations that prevent excessive iteration while maintaining thorough development practices. This balanced approach ensures refinement without compromising practical constraints, resulting in a robust framework that effectively guides computer vision projects from initial conception to full-scale production deployment.

4 EXPERIMENT

To validate our proposed workflow and derive platform requirements, we utilized the MVTec Anomaly Detection (MVTec AD) dataset (Bergmann et al., 2019) of which some examples can be seen in Figure 2. Chosen for its extensive coverage of industrial defect scenarios and established relevance in manufacturing computer vision applications, this data set proved particularly well suited for workflow validation, because of its real world comparability.

4.1 Dataset Structure and Preparation

Comprising 15 distinct object categories, the MVTec AD dataset represents a diverse array of industrial products and textures. Each category includes defect-free samples alongside various defect manifestations, offering a robust foundation for validation efforts. To align with our objectives, manual reorganization of the dataset resulted in two distinct classification configurations. The first configuration consolidated all the defect variants within each product category into a unified defect class. For example, images depicting contamination, broken seals, and surface scratches in the bottle category were aggregated into a single comprehensive defect class. This approach facilitated validation focused on broad-spectrum defect detection capabilities. In contrast, the second configuration retained the granular classifications, preserving the detailed structure of the original MVTec dataset. This arrangement enabled validation of fine-grained defect discrimination capabilities. Original image distribution patterns were maintained during preparation; for example, the bottle category contained 209 defects-free images and 63 defective samples spanning various types of anomaly.

4.2 Validation Process

Validation involved a comprehensive implementation of each phase using an ML Ops platform. The systematicatic data ingestion and preparation procedures established processingssing pipelines for both dataset configurations. Subsequent phases included model development and training, with hyper-parameter settings and training protocols meticulously documented through platform support.

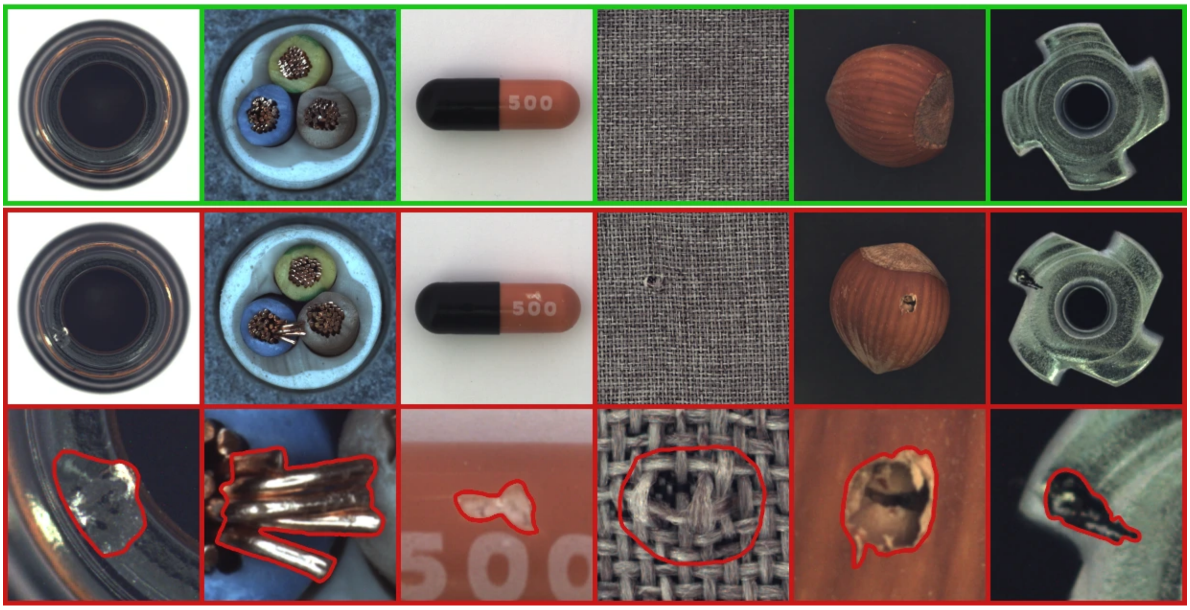

Figure 2: The used dataset example. The figure shows examples from the MVTec Anomaly Detection dataset, displaying normal samples (top row), defective samples (middle row), and ground truth defect annotations (bottom row). From left to right: metal nut, cable, capsule, fabric, hazelnut, and metal part categories (Bergmann et al., 2019).

4.2.1 Workflow Implementation

A full execution cycle of the proposed workflow was carried out during this phase. Data preparation and ingestion procedures initiated the process, followed by iterative model development stages. Rigorous integration testing and production environment simulations validated real-world applicability. Throughout this cycle, emerging requirements were systematically documented as part of practical workflow execution.

4.2.2 Requirements Documentation

Functional and non-functional requirements were identified and documented during systematic implementation. These requirements spanned technical specifications, integration prerequisites, and operational constraints essential for successful deployment in manufacturing environment. Special attention was given to requirements critical for production-grade computer vision implementations in industrial contexts. Concrete evidence of both the workflow's practical applicability and the platform's ability to support advanced manufacturing computer vision workflows emerged from this experimental approach. Findings from this validation process formed the basis for a comprehensive analysis of platform requirements and an assessment of its effectiveness.

5 CONTRIBUTIONS

Two key contributions were developed to aid manufacturers in selecting and implementing MLOps platforms. These contributions underwent validation through practical application using the MVTec AD dataset (Bergmann et al., 2019). First, a set of essential requirements was compiled for evaluating MLOps platforms in manufacturing environments. Second, a practical workflow was created to guide organizations through computer vision implementation from start to finish.

Experimental validation delivered compelling results across both dataset configurations. Training utilized 1132 samples for training and 126 samples for validation, maintaining a consistent 90/10 split ratio because of the limited data. Defect-specific classification models (MV TEC defect types) exhibited promising yet more nuanced performance. EfficientNet implementations achieved validation accuracies of 77.78 % and 78.57 %, reflecting the increased complexity of distinguishing specific defect types. These models were trained using the same data distribution of 1132 training samples and 126 validation samples, ensuring consistency across experiments.

Both contributions, the requirements and practical workflow, are intrinsically linked; neither can function effectively without the other. Requirements facilitate platform selection, while the workflow optimizes utilization.

5.1 Software Tool Requirements

Analysis revealed that distinguishing between functional and non-functional requirements is crucial for MLOps platform selection. Operational needs diverge from systemic constraints in industrial computer vision deployments, empowering manufacturers to evaluate toolchain capabilities against both technical specifications and quality attributes.

5.2 Functional Requirements

5.2.1 Must Haves

Operations on Data: Systematic processes are essential for managing visual data throughout the computer vision lifecycle. Image data ingestion into centralized repositories like data lake architectures, configured to store raw images while maintaining metadata, serves as a successful starting point. Critical sub-operations include automated data preparation pipelines, annotation tools, and domain-specific augmentation strategies. Rigorous quality assurance protocols leverage automated analysis and outlier detection to ensure dataset integrity. Annotation workflows perform optimally when integrating human-in-the-loop validation and active learning mechanisms. Dataset management requires splitting methodologies aligned with data science standards to prevent model overfitting.

Model Development and Training Capabilities: Structured methodologies are vital for creating industrial-grade computer vision models. Development pipelines must implement multi-stage training protocols, combining transfer learning from domain-specific pretrained models with automated hyperparameter optimization. Modern validation require stratified evaluation across manufacturing edge cases, supported by explainability techniques such as Grad-CAM heatmaps to audit model decisions. Version-controlled experimentation tracking ensures reproducibility of architecture variants and training configurations.

Deployment Capabilities: Practical deployment is a critical functional requirement for MLOps platforms in manufacturing environments. Models must be seamlessly deployed on local servers or in the cloud, with version tracking enabling precise identification of model versions in use across facilities. Flexibility is essential, some factories require simultaneous updates across systems, while others prefer gradual rollouts. Container support ensures consistent model performance across diverse systems, while rollback capabilities minimize downtime during failures. Integration with existing factory systems eliminates isolated solutions, aligning deployments with modern manufacturing practices.

5.2.2 Should Haves

Hyperparameter Tuning and Comparability: Systematic optimization of model configurations is critical for efficient experimentation workflows. Automated hyperparameter search capabilities eliminate manual trial-and-error processes, while concurrent experiments on distributed compute resources accelerate development timelines.

Extensibility: Adapting AI/MLOps platforms to unique workflows requires structured plugin architectures and SDKs with well-defined extension points. Extensibility preserves institutional knowledge by allowing engineers to embed custom scripts that reflect factory-specific expertise.

5.3 Nonfunctional Requirements

5.3.1 Must Haves

Usability: Intuitive UI/UX design is essential for manufacturing environments where operators may lack specialized expertise in machine learning workflows. Interfaces must abstract complex operations into domain-specific workflows that facilitate rapid execution by manufacturing engineers.

Deployment Monitoring: Robust monitoring capabilities are necessary to maintain smooth operation of computer vision systems over time. Engineers require clear insights into prediction accuracy, response times, and model confidence levels, with alerts triggered when metrics deviate from acceptable ranges.

5.3.2 Should Haves

Standard Algorithm Library: Foundational libraries offering pre-implemented algorithms save significant development time for common tasks such as classification, anomaly detection, segmentation, and optical character recognition (OCR). These out-of-the-box solutions reduce initial development overhead while enabling engineers to focus on task-specific fine-tuning.

User Management Integration: Native LDAP/AD compatibility simplifies deployment by integrating seamlessly with existing user authentication systems prevalent in manufacturing IT environments. When unavailable, built-in user management systems provide secure access control without requiring extensive custom coding or external dependencies.

6 DISCUSSION

Experimental results reveal both strengths and limitations of the proposed framework. Performance on defect-specific classifications (77.78 % and 78.57 %) underscores the inherent complexity of fine-grained defect categorization in manufacturing environments. Distinguishing between similar defect types remains a significant challenge, particularly when working with limited data diversity and class imbalance.

These findings offer valuable insights into MLOps platform selection for manufacturing applications. First, successful implementation of both classification approaches using MLOps platforms validates the requirements ability to identify appropriate toolchain capabilities. Supporting workflows from data preparation to model deployment, the platform demonstrates practical utility as part of an integrated management process approach.

Certain limitations in the validation approach must be acknowledged. Relying on a 90/10 split ratio, while practical for initial validation, limits the robustness of performance evaluation compared to a three-way split (training/validation/test). This limitation stems from platform constraints, as Robovision currently lacks native support for advanced dataset splitting strategies. Additionally, exclusive reliance on the MVTec AD dataset restricts validation to a narrow subset of manufacturing defect scenarios, leaving broader applicability unexplored.

Conducted through February 2025, the experimental timeline illustrates the relevance with current MLOps platforms and modern deep learning architectures. Using a consistent training set of 1132 images and validation set of 126 samples across all experiments ensured stable comparisons between category-level and defect-specific classification tasks. Despite these constraints, results demonstrate the framework's potential for guiding effective MLOps platform selection and implementation in manufacturing contexts.

7 CONCLUSION

This paper presents two significant contributions to manufacturing AI implementation: a comprehensive requirements for MLOps platform selection and a practical workflow for computer vision deployment. Experimental validation using the MVTec AD dataset demonstrates the effectiveness in guiding both platform selection and implementation decisions.

Findings suggest that successful MLOps toolchain selection in manufacturing requires careful consideration of functional and non-functional requirements alongside a structured implementation approach. The ability to support both broad category-level classification and fine-grained defect-type discrimination highlights its flexibility in addressing diverse manufacturing needs. Achieving perfect accuracy in category-level classification while maintaining reasonable performance in more complex tasks validates its practical utility for real-world applications.

Future research could expand this work by exploring applicability across other manufacturing domains beyond computer vision and testing its effectiveness with alternative MLOps platforms. Extending validation to real-time production environments and incorporating more diverse defect scenarios would further strengthen practical relevance.

ACKNOWLEDGEMENTS

This research was supported by the Hightech Agenda Bavaria.

REFERENCES

Beck, N., Martens, C., Sylla, K.-H., Wegener, D., and Zimmermann, A. (2020). Machine learning operations (MLOps): Prozesse für entwicklung, integration und betrieb. Whitepaper, Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Sankt Augustin. Zukunftssichere Lösungen für Maschinelles Lernen.

Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. (2019). Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9592–9600.

Faubel, F. and Schmid, U. (2024). MLOps: A multiple case study in Industry 4.0. In *IEEE ETFA*. IEEE. To appear.

Jon Bokrantz, M. S. and Skoogh, A. (2024). Realising the promises of artificial intelligence in manufacturing by enhancing crisp-dm. *Production Planning & Control*, 35(16):2234–2254.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105.

Robovision (2024). Build vs. buy: Choosing between in-house development and off-the-shelf vision AI. Whitepaper, Robovision. Includes McKinsey AI implementation archetypes and PackCheck case study.

Russell, S. and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition. Global Edition equivalent to 2021 US Edition (ISBN 978-0-13-461099-3).

Schneider, J., Fischer, L., and Voelter, M. (2024). Artificial intelligence in industry 4.0: A review of integration challenges. *arXiv preprint arXiv:2405.18580*. intro + motivation.

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., and Müller, K.-R. (2021). Towards crisp-ml(q): A machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2):392–413.

Tripathi, S., Muhr, D., Brunner, M., Emmert-Streib, F., Jodlbauer, H., and Dehmer, M. (2020). Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing. *CoRR*, abs/2007.14791.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining (PAKDD 2000).