

# Cloud-based ML Technologies for Visual Inspection: A Case Study in Manufacturing

Timo Koppe  
Technical University of Darmstadt  
[koppe@is.tu-darmstadt.de](mailto:koppe@is.tu-darmstadt.de)

Jonas Schatz  
Technical University of Darmstadt  
[schatz@is.tu-darmstadt.de](mailto:schatz@is.tu-darmstadt.de)

## Abstract

*In recent years, cloud-based Machine Learning services have received much attention for promising fast and cost-effective deployment. At the same time, manufacturing companies are beginning to evaluate and implement these new technologies in their production processes. This paper adopts the design science research approach to demonstrate the use of cloud-based Machine Learning services to implement a visual inspection system in the manufacturing industry. As a result, our developed IT artifact can correctly classify all of the given parts in a dataset consisting of 363 images, outperforming the current manual inspection. Thereby, it addresses the various challenges faced by the industry when introducing cloud-based Machine Learning technologies, evaluating return on investment (ROI), and how this can facilitate further digital transformation in production.*

## 1. Introduction

Artificial intelligence (AI) and specifically Machine Learning (ML) approaches are on their way to becoming the key technology for numerous business applications. One area where ML methods are particularly successful is the field of computer vision [1]. Since the advent of deep learning several years ago, those methods have dominated all important computer vision benchmarks [2]. On top of that, they have also become much more efficient in regards to system resources [3]. Among the best-known examples in everyday life are self-driving cars and the face unlock feature on smartphones. Beyond that, computer vision has the potential to have a high impact on industry and manufacturing. Potential applications in industrial production lines include identifying and localizing parts, detecting material defects on conveyor belts, and real-time analysis of materials [4]. Although visual monitoring systems using conventional image processing techniques have already been used for

several decades, conventional techniques usually come with several restrictions which could not be automated yet, especially when it comes to human flexibility, background knowledge and intuition.

In the course of global market trends, individualization of products, the need for more efficiency, shorter time-to-market cycles [5], and growing customer specifications, demands exceed conventional systems' capacities and capabilities. There is a growing need for more flexibility in these industrial processes.

Simultaneously, advances in computer vision promise to ease some of the restrictions of conventional visual inspection systems. Vendor cloud platforms promise a distributed adoption and use of ML and AI using Machine Learning as a Service (MLaaS). MLaaS simplifies the development and use of cloud-based off-the-shelf AI technologies and the adoption of pre-trained neural networks [6, 7]. It eventually supports the entire development process, from the raw data set to the operational machine model, even with smaller training data sets. On top of that, cloud technologies continue to promise payoffs in reduced costs, increased efficiencies and provide business model transformation opportunities [8].

Despite advances in computing, there is still much speculation about the use of ML for real applications, and the published benchmarks are often not realized [9]. In this respect, the literature criticizes that the current ML research has lost touch with the "real world". It rarely includes an assessment of whether quantitative performance improvements are relevant in practice. More research is needed to evaluate the solutions developed under real conditions [10] and the corresponding economic implications [9].

This paper addresses this gap by developing and evaluating an ML-based visual inspection solution in the manufacturing industry using cloud-based ML technologies. Specifically, we address the following research questions (RQ): How can MLaaS facilitate the development of ML-based visual inspection in a production process? What are further potentials

and challenges when introducing ML-based visual inspection into a production environment?

The remainder of the paper is structured as follows: The second section provides an overview of visual inspection techniques for quality control in a production environment. The research approach based on pertinent Design Science literature is introduced in Section 3. It is followed by the description of the designed artifact and design search process. Section 5 discusses the results and contributions to research and practice followed by concluding remarks and a description of future research directions in Section 6.

## **2. Theoretical Background**

Product quality inspection in manufacturing is about comparing the properties of a product with its specifications [11]. If we specifically refer to visual inspection, we imply visually perceptible criteria, which is also called an optical inspection. Visual inspection is used in many areas such as industrial manufacturing processes [12], food production [13], transport infrastructure maintenance [14], or the production of nuclear weapon parts [15]. As part of quality control, the aim is to ensure that faultless products with reliably consistent quality reach the customer. Products that are incorrectly sorted out, although they are acceptable, increase the scrap rate. Conversely, if a defective product is not recognized as such, it impacts customer complaints. Directly related to this are, among other things, customer satisfaction and loyalty [16], a strong corporate brand [17], and ultimately the competitiveness of the company. Due to its great importance, quality control is one of the four major management functions alongside planning, organizing, and leading [18].

### **2.1. Manual and Automated Optical Inspection Systems**

Compared to other inspection methods, visual inspection is characterized by the fact that it is not only contact-less but also completely non-destructive, allowing 100% of manufactured products to be tested [19]. In the simplest case of manual optical inspection (MOI), it is often possible to operate without special equipment and thus without a high initial investment. Nevertheless, a higher resolution compared to most other non-destructive testing methods can be achieved. Apart from a few technical gadgets (microscopes, better lighting and more ergonomic workplaces), MOI's core has hardly changed over time. Most efforts in recent years have been aimed at

making MOI more reliable: With the advent of quality management procedures such as Six Sigma, the quality of human inspection was put under the microscope for the first time. It showed the error rate can range from 2-10% for simple inspection tasks to up to 30% for more complex tasks [20, 18]. This outcome is also difficult to reduce by multiple inspections, a fact known as the "Two-Inspector Problem" [21]. The error rate depends not only on the difficulty of the inspection task: Work results are influenced by numerous other factors. In addition to psycho-physical factors, such as the age and gender of the inspectors, the design of the workplace and social and organisational factors also have an influence [18].

At the end of the 1990s, a quantum leap in quality control took place with the advent of digital sensors and the associated automation of visual inspection. Automated optical inspection (AOI) systems operate using conventional image processing techniques, such as a reference comparison [22]. In order to be able to install systems for AOI, several necessary restrictions have to be met [23]. The inspection task to be solved must be described in detail and precisely in a technical manner, using the "language of image processing" [24], i.e., signal processing, edge detection, morphological image processing, and segmentation. Complex tasks are solved by combining these building blocks. For example, to count the number of pores in an object, the image is pre-processed so that the pores can be clearly distinguished from their background. Afterwards, the edges of the approximately circular pores are detected, fitted by circles, and counted in the last step. Furthermore, AOI systems require that all possible variants of test parts and defects must be known. In addition, the environmental conditions must be designed so that defects and objects can be detected at all. These environmental conditions must be kept as stable as possible [23].

If these prerequisites are given, automated visual inspection has some distinct advantages. It is efficient to a high degree and, once installed, has low operational costs. The inspection process is highly objective and detailed logs can be generated, from which the production process can be improved [25]. However, some aspects of visual inspection can not be automated, especially in terms of human flexibility, background knowledge, and intuition. Therefore, most literature suggests using so-called hybrid approaches, where the bulk of the task is automated, while humans perform mostly supervision tasks and decide in ambiguous cases [24].

## 2.2. ML-Based Visual Inspection

Machine Learning eases some restrictions for the use of an AOI system using conventional image processing techniques. First, the inspection task no longer needs to be described in a detailed and precise technical manner. Instead, ML automatically detects meaningful patterns in data [26]. For instance, surface defects are not formally described, but only provided as labeled example images for ML algorithms. By that, the implicit knowledge of the inspectors is contained in the training data. Second, not all variations of components and errors need to be known in advance. ML systems are, to a certain extent, able to react to previously unknown or partly challenging to describe or parameterizable error types. ML systems can also exhibit greater robustness to changing environmental conditions, e.g., by simulating such fluctuations in training data during data augmentation [27, 28]. In general, ML algorithms can perform various tasks during visual inspection. These include image classification, which classifies a test image into several categories, and object detection, which locates objects of interest in an image [29]. For this purpose, early ML algorithms used a two-stage approach: Feature descriptors were used to extract useful features from images. These were then used as input to a classification algorithm. However, the quality of the output hinges on the quality of those hand-crafted features [30]. With the rise of deep learning in the last decade, manual feature extraction is no longer necessary, as it is also learned [31]. Among those deep learning models, Convolutional Neural Networks (CNNs) are now used almost exclusively for image classification tasks. They work best when provided with data in the form of multiple arrays, such as images or linear signals [32].

## 2.3. Machine Learning as a Service

In addition to the progress in the field of computer vision, new tools and libraries, standard algorithms, and pre-trained neural networks facilitate the development process from raw data to an operational ML model [6] - even with smaller training data sets, e.g., by using transfer learning techniques [33]. Finally, these techniques and technologies have already been provided via several cloud platforms as Machine Learning as a Service. Cloud technologies enable rapid on-demand deployment of single use cases up to large-scale deployment throughout the company. Its advantages include on-demand computing power with quick implementation, fewer IT staff, low maintenance,

and lower cost [34]. Despite these promising conditions, cloud computing is subject to some limitations in an industrial context. These comprise, among other things, the requirements for real-time processing and interactions, the comparatively limited computing power of industrial sensors and controllers, and last but not least, security and privacy considerations [35]. These restrictions must be addressed to foster the use of cloud technology in the industrial environment.

## 3. Research Design and Case Description

The Design Science Research Methodology (DSRM) applied in this case study has been described by Peffers et al. [36]. They identified six activities performed in an iterative loop as the creation process of a Design Science artifact. Such an artifact can, for example, be a piece of software or an ML pipeline, as is the case in this research work. Additionally, the authors adhered to the guidelines for Design Science Research by Hevner et al. [37].

The case study was conducted within a German technology group. Although the individual divisions of the group pursue business in different markets, a central organisational unit enables cooperation across all areas concerned with digital transformation. One of the business units is active in the mass production of electronic sensors and actuators. For one of its products, surface inspection of production flaws and dirt particles is performed manually at a production throughput in the order of magnitude of several hundred thousand parts per month. In the context of this case study, this process should be automated by ML-based visual inspection. Further, the current manual inspection only could perform binary classification, i.e., sorting the produced actuators into two categories (OK and not OK). Thus, the case study's primary goal was to provide an algorithm that could outperform the human inspectors in that dimension. Beforehand, the performance of each human inspector was evaluated individually using the F1 metric, which resulted in the failure rates for simple inspection tasks (2-10%) [18, 20].

However, improvements in other dimensions were needed to justify the investment and the risk associated with adopting this new approach. So, a secondary goal was to provide new insights into the production process, which could improve the manufacturing process and detect deviations from the defined standard conditions before those deviations could manifest themselves in errors. Thus, investigation was not only focused on algorithms for binary classification, but also for multilabel classification (determining the kind of error) and detection (localizing the error).

Since this case study shall demonstrate how ML can be incorporated into an already established production environment within a company, particular focus lies in communicating the artifact to support decision-making for the stakeholder. Therefore, the demonstration and evaluation of the artifact include an analysis of the financial feasibility and visualizations of the performance and capabilities of the ML system. Lastly, proprietary data was used for this case study. Therefore, no original images can be shown. Figure 1 depicts an anonymized version of the device.

## 4. Results

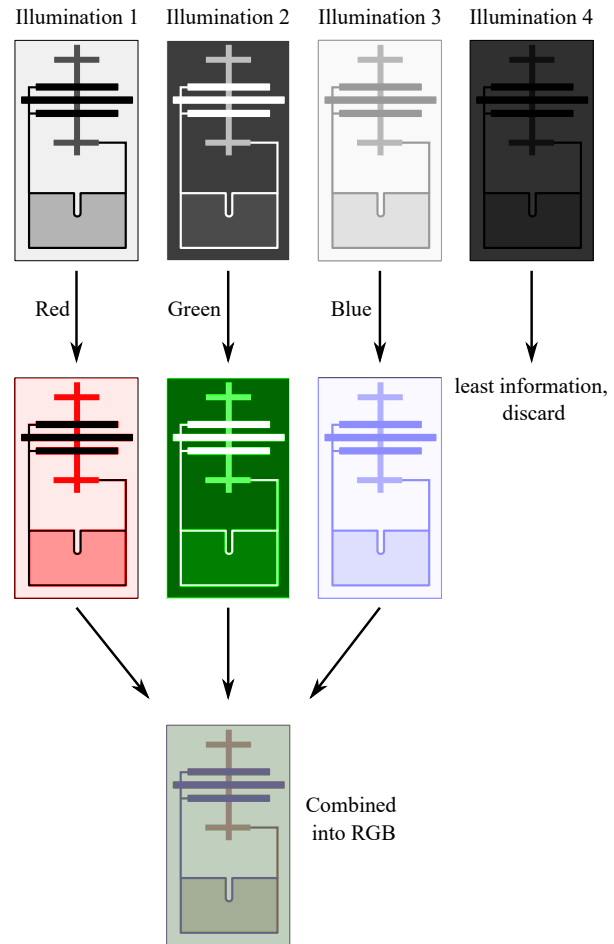
The following section describes the designed artifact and the design process. Since a large focus of the DSRM lies on communication and demonstration, the artifact will be explained extensively to the stakeholders of the production process, as shown in Section 4.2. In Section 4.3, the algorithms' performance will be evaluated and compared to the manual visual inspection, in the domains of prediction accuracy and financial feasibility.

### 4.1. Design and Development

The structure of the design and development phase follows the flowchart shown in Figure 2. In the following, the individual steps are outlined in greater detail.

**Image Capturing** The device under inspection is a small thermal sensor, which is shown in Figure 1 as a simplified version. It consists of multiple different precious metals, screen-printed on a ceramic substrate, and covered by a protective glass layer. Different errors can be introduced during the production process, e.g., dust particles underneath the glass layer, scratches on the surface, or cavities in the printed metal parts. Due to the variety of those errors, they can not be made visible within a single image. In order to capture the full breadth of errors, four images under different lighting conditions (incident light with different amounts of dark-field illumination as well as transmitted and coaxial light) are captured for each of the 363 specimens of the sample dataset. A simplified depiction of the resulting images is shown in the top row of Figure 1. Since position and orientation of the camera are kept constant during the capturing of the image series, the resulting images can be stacked on top of each other.

**Labeling** A production engineer labels each of the four illuminations for every image in the dataset, based



**Figure 1. Visualization of converting four illuminations to one RGB image.**

on the error catalogue that is also used for the current manual visual inspection. They mark each error with a rectangle and further annotate the category of the error. This allows to additionally deduct the labels for the simpler binary and multilabel cases, where the location detail is irrelevant.

As a labeling software, IBM's *Maximo Visual Inspection* (formerly *PowerAI Vision*) [38] is used. The images of the different illuminations are labeled independently from one another. This way, only errors that are actually visible in a certain illumination are marked as such. This procedure will become important when combining these information in the red, green, and blue channels of an RGB image.

**Preprocessing** For preprocessing, we use OpenCV's Template Matching to crop the labeled microscopic images to only show the device [39]. Next, we perform image augmentation in accordance with the device's

actual properties: Since it shows mirror symmetry, the size of the dataset can be doubled by creating vertically flipped copies. Since the substrate may come from different suppliers and may have a slightly deviating color, thickness and density, we can create additional images for training by changing the colors for the substrates.

**Conversion to RGB** One of the strengths of MLaaS is its ability to work with pre-trained models. However, there is no out-of-the-box solution to handle the information contained in the four different images we have of each specimen in this use case. To overcome this limitation, we selected the three illuminations containing the most information about errors. We copied each illumination to the red, green, and blue channels of a new color image in the next step. Further, we discarded the illumination with the least information content. This newly created color image is then used as input to the pre-trained models.

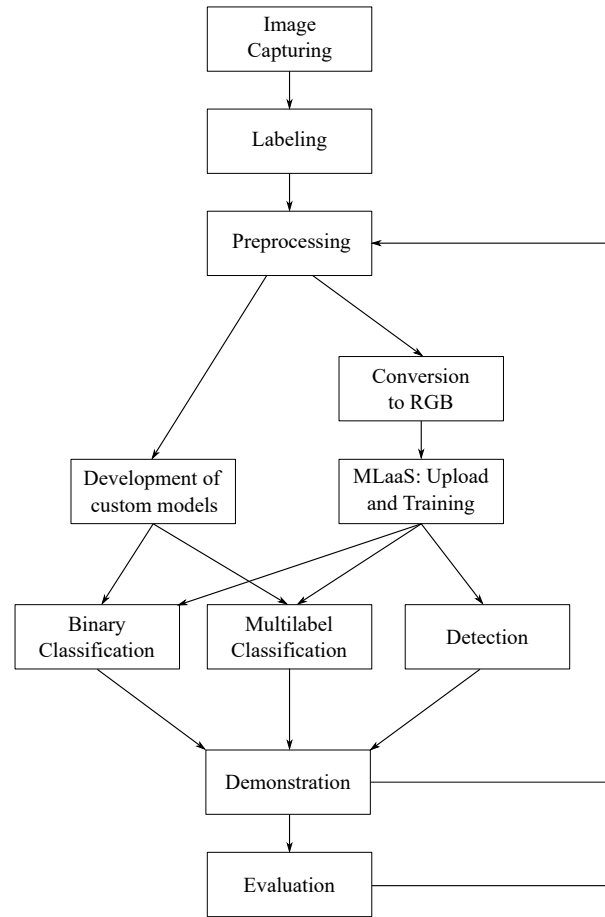
In order to determine the information content of an illumination, three metrics are computed for each image and then aggregated over all images of one illumination:

- Total number of labels
- Unique labels: As mentioned above, one actuator's four illuminations are labeled independently from one another. Often, errors can be detected in more than one illumination, resulting in bounding boxes at similar positions in these images. A label is considered unique if its bounding box does not overlap with another bounding box on another illumination.
- Essential labels: If an actuator only has errors detectable in a single illumination, the removal of this illumination would make an otherwise rejected part undetectable.

Taking these three metrics into account, we are able to keep most of the information while removing one illumination. This process is visualized in Figure 1.

**Upload and Training** We use two MLaaS providers in this case study: IBM's *Maximo Visual Inspection* and Google's *Cloud AutoML Vision* [40]. Both work in a similar manner: The images for training are uploaded, in addition with a specification file describing the labels. Afterwards, we can train a ML model in the cloud. After the training is completed, we can perform online predictions using the vendor-supplied APIs or download the model for use on an edge device.

For all training processes, a split into training (70%),



**Figure 2. Flowchart showing the development process of the artifact.**

validation (15%) and test set (15%) is performed. In all cases, the test set is never uploaded into the used tools to preserve its independence. IBM's tool allows for some adjustments to be made: For the first task (binary classification), we choose GoogLeNet [41], pre-trained on the ImageNet dataset, as the model architecture. Multiple other pre-trained models are available within the tool [42], none however were fit for the specialised use case at hand. Next, we perform a grid search over learning rate and weight decay to find the best model based on the F1 metric. At the time of the case study, the service did not offer a feature to solve the second task (multilabel classification). However, we could exploit a property of the device to reduce the task of multilabel classification to a binary classification task for each label: The different errors labels correspond to the material classes on the device, e.g., "Substrate Error", "Gold Error", and "Glass Error". With one small exception, these areas do not overlap. We can then create one binary classification task for each class

of material, by cropping out the rest of the device and training only on the area of the selected material. Again, a pre-trained GoogLeNet is chosen, grid search over learning rate and weight decay is performed and the best model for each error class is determined. For the third task (detection), the faster R-CNN algorithm [1] is available through the platform.

Google's service allows only to set a budget of node hours for training. We train the models with the same images as on IBM's platform. Here, the only parameter to choose from is the amount of node hours the training should be performed with. This value is set to 100 hours. This way it is high enough, so that convergence is reached in every training. Again, the multilabel classification is reduced to a binary classification task using the same blacked-out data. Even though this service is able to perform multilabel classification, this approach lead to better results and comparison between platforms.

The training duration ranges from a few minutes for the classification tasks to one hour for the detection models on both platforms. The training process is performed completely transparent for the user, who will have no further influence on the process until the completion of training. Furthermore, cross-validation can not be performed. At the time of experimentation, the used tools did not have the option for cross-validation, instead it would have had to be performed manually.

**Development of Custom Models** To further evaluate the performance and use of MLaaS we decided to develop a custom model.

MLaaS, especially in the field of computer vision, uses powerful, but very generic algorithms to solve a wide variety of use cases. To potentially improve upon the performance of those MLaaS providers, we develop a custom model that is free of the restriction of the cloud platforms. This approach took advantage of two properties from our special application:

First, our custom model will use the information of all four illuminations. This is an improvement over the MLaaS platforms, where only three illuminations could be used in the three channels of an RGB image. Second, each image is taken under the same orientation and lighting so that each time there is a fixed structure that does not change drastically from image to image. The only differences are potential production flaws.

Both of these facts can be exploited by implementing a variation of an established facial recognition algorithm [43]. The images within each illumination are normalized by subtracting the mean of the images and then scaling the data to unit variance. Each normalized image is then flattened into a

vector and arranged into a feature matrix. Of this feature matrix, the eigenvalues and eigenvectors are calculated using principal component analysis (PCA). This decomposition has some interesting properties: First, the mean of the images which was subtracted in the normalization step shows a device without flaws. Second, the eigenvectors belonging to the biggest eigenvalues represent the most significant flaws present in the test data.

This eigendecomposition can now be used to classify the devices. We could show that the eigenvectors belonging to the 800 biggest eigenvalues contain 99% of the variance of the original data. We can now project the images into this lower-dimensional space, spanned by the corresponding eigenvalues, and use these coordinates as new feature vectors. Classification is performed using a Support Vector Classifier (SVC). This approach is capable of performing binary and multilabel classification. However, detecting the location of production flaws is not possible.

## 4.2. Demonstration

Both MLaaS providers allow for online as well as offline predictions. For this case study, the evaluation was performed using online prediction, since no additional setup was required. However, when used in production, the models need to be downloaded onto edge devices: Preprocessing, creating the RGB image, upload and inference take around 270 ms in our calculations. This additional latency would already impact the throughput of the production line. Additionally, online predictions would make the whole production process reliant on an internet connection.

For the further demonstration of the design artifact and showing how ML can be used in our case study, we designed two ML-based visual inspection systems aligned to our business environment. In the first system design, we only replaced the MOI by the ML component and appropriate hardware. However in this design, product handling is still done by the production staff, as it had been done when parts were taken to MOI and back. In the second design, the handling of the product parts is automated by a pick-and-place machine. This system design therefore requires more modifications to the production process by introducing a new machine into the process. These two new system designs were presented to the business.

As a further communication instrument, we presented some exemplary visualizations. These visualizations show how additional quantifiable information obtained through the ML model can be used in the production process. The first visualization shows a Pareto chart,

which depicts the frequency of different error types as a bar chart. The diagram allows the viewer to quickly compare the current AI predictions with historical predictions to detect changes. Secondly, we designed a heat map of the product. The heat map uses color concentration to indicate a concentration of defects in certain areas of the product within a certain time. This tool can also be used to monitor the running production and to react to deviations.

Lastly, a graphical user interface (GUI) demo was implemented, which served to communicate the ML model's functionality. The GUI allows users to use the AI to analyze their sample images of products. The output then displays the analysis (error classification) and a detection as a bounding box over the corresponding area of the image.

### 4.3. Evaluation

The created artifact, i.e., the trained Machine Learning model, needs to be evaluated in terms of several criteria:

First, the performance of the different models in the testing environment is determined. For this, a suitable metric needs to be found. Accuracy, which is the percentage of correctly classified parts, can hide bad prediction performance on underrepresented classes. When the production process is running as intended, the fraction of faulty parts is going to be small, making accuracy an ill-suited metric. From a business standpoint, high precision (a small number of undetected faults) and high recall (a small number of erroneously discarded non-faulty parts) were decided to be equally important. Thus, the F1 score, which is the harmonic mean of precision and recall, is chosen to compare the different models. Table 1 shows the best performances under this metric for the two MLaaS providers and the custom model. In multilabel classification and detection, the average F1 score over the types of errors is reported. In all cases, the test set that was not used during training is used to determine these metrics.

Several clear trends can be seen:

In the two classification tasks, the custom model has the worst performance of the three approaches, even though it is able to exploit the most information from the data. This shows the supremacy of the pre-trained models used by the MLaaS providers, especially on such a small dataset.

Of the two MLaaS platforms, *Maximo Visual Inspection* performs better in all three categories. In the case of binary classification, it can correctly predict all samples

from previously unseen images, even with different data splits. However, since training and evaluation are based only on a small labeled dataset, it remains to be seen how this model will behave in a production environment. The higher performance of the *Maximo Visual Inspection* platform can be attributed to two factors: First of all, it allows the user to adjust some settings and parameters. This is in contrast to Google's platform, which just allows the user to only define a training budget and leaves the rest to the automated ML system running in the background. Additionally, further insights and metrics from the training process are provided, making a continuous improvement possible.

**Table 1. Comparison of F1 scores of the two MLaaS providers and the custom model. Note that the latter is not able to perform detection tasks.**

	AutoML Vision	Maximo Vis. Insp.	Custom Model
Binary Class.	0.90	1.00	0.86
Multilabel Class.	0.85	0.90	0.71
Detection	0.75	0.78	-

Another goal of the Case Study is to analyze the economic implications of switching to the new ML-based visual inspection system. First, the costs of the current manual inspection are evaluated. For most factors, numbers are available from the business unit or through interviews, e.g., labor costs, office space, and equipment. Furthermore, manual inspection performance is taken into account for the cost calculation, which is derived from a measurement system analysis. This leads to the cost of "false negatives" classifications in quality inspection. False negatives are products that are incorrectly classified as defective. Second, the costs of the two different proposed automatic visual inspection system designs (manual handling and fully automated) are estimated. The costs divide up into three categories: Development costs, the initial investment into hardware, and running costs. Development costs include data labeling costs, labor costs, and cloud infrastructure costs. Next, initial investments include the costs for product handling, image sensors, and ML hardware. Lastly, the operating costs for the AI solution were calculated. This includes the costs for the False Negatives and costs for the operation and management.

Finally, the variable and fixed costs of different solutions can be compared. The case study showed that the costs amortize already within the second month. This analysis also helps to compare scenarios that may involve changes in the existing production process. Although the AI solution has a high initial investment,

it has the lowest variable costs and is, therefore, the preferred option in the long run. This also satisfies the condition given by the business unit that a visual inspection system should be profitable within two years. A postimplementation review is nonetheless necessary and open for further research.

The above considerations are purposefully kept very simple. It is expected that higher costs and unforeseen difficulties arise during the actual implementation. However, in the long term, the automatic inspection based on AI is expected to have a positive return on investment, taking all previous considerations into account.

Finally, the developed artifact is evaluated in its business environment [37]. For this purpose, the solution is presented to the decision-makers in the company. The observations on the presentation show that the demonstration of the solution is well received. The functionality and decisions of the AI can be replicated and visualized through the GUI. In addition, the dashboards for statistical monitoring of the solution are considered to be a valuable tool. It was decided to use the ML solution in production.

## 5. Discussion

Commencing with the *theoretical implications*, our study contributes to the current “Develop” research stream of digital innovation [44] by showing how Information systems (IS) artifacts are developed and what adoption antecedents are. We showed that MLaaS can be used to implement a ML-based visual inspection system in the manufacturing industry. Further, we demonstrated which steps need to be considered during ML development (e.g., Labeling, Conversion to RGB, Upload and Training) in order to use MLaaS. The outcome of the system design relies on attention to all steps and new work routines, starting from capturing the training data up to addressing the challenge of introducing a ML black-box solution into a production process, embracing the central idea of socio-technical theory [45].

Furthermore, our study contributes to the discussion that the “best practice” design of a vendor-supplied IS may be a myth [46]. On our way to find an optimal ML model for the different prediction tasks (binary classification, multilabel classification, and detection) the two vendor platforms performed quite differently in terms of configuration process (provided configuration and optimization options) and performance (overall F1-Score). Different vendors provided the best results for different parts of the solution. Therefore, this case study cannot be considered as an overall comparison

between the providers, and it was not our intention to highlight a better cloud platform. However, the different performances can have a substantial financial impact, especially when it comes to high volume manufacturing. The advantages of cloud computing, among others, are rapid on-demand deployment, quick implementation, and fewer IT staff [34]. These advantages also became evident when using AI cloud technologies. MLaaS helps non-IT companies in the subject matter to build expertise in certain AI areas, like computer vision. A comparison with our custom-developed AI approach shows that the performance of the providers’ black box solutions cannot easily be beaten by in-house developments. Therefore, these solutions provide a high potential for speeding up the development of ML products. Simultaneously, it also becomes apparent that MLaaS services do not (yet) make data science expert knowledge obsolete since challenges with the data, training or evaluation of the models remain. One example is the preprocessing of the available data into the three color channels of an RGB image explained in Section 4.1.

MLaaS helps especially with hyperparameter tuning or the search for an optimal architecture. Choosing the right tool for the right task, however, is still up to the data scientist. In addition, proprietary MLaaS aggravates the black box dilemma that is already known in the context of cloud computing services [34], since the developed solutions are often black boxes again. This dilemma can lead to several problems, e.g., lower acceptance and user’s trust of model predictions [47] as well as challenges in clarifying the accountability for such a system [48]. In turn, MLaaS makes the development of such AI solutions more accessible and promote their widespread use within a company, contributing to the democratization of AI tools in the company [7].

Our final IT Artefact consisted of the ML model and a software (including a user interface) that allows the production staff to upload sample images of the products to review the AI’s decisions, as explained in Section 4.2. Its use indicates that the models are able to generalize even to previously unseen error patterns. In addition to the demonstration of several graphical ways to visualize the resulting data on dashboards, this led to a high acceptance from the responsible business unit. These dashboards, based on the AI output information, provide an outlook on how such a solution can create useful additional information for new process improvements (e.g., improved error detection) to facilitate further innovation. At the same time, the additional representations and feedback from the business unit lead to a better insight into the mechanisms and operation of the model, which also



contributed to the question of the responsibility and ethical implications of such a system [48].

Moving beyond theoretical implications, our study also has several *practical implications*. At first, all three technical challenges (binary classification, multilabel classification, and detection) could be solved with higher than expected performance. By combining the information contained in four differently illuminated images into one RGB image, high-performing models for each of these tasks were trained using standard architectures within cloud AI services. Furthermore, they outperformed a custom algorithm based on PCA, which was specifically developed to handle the information contained in all illuminations. In addition to technical feasibility, financial viability was shown as part of a case study. Estimates about the costs of the automated inspection were compared to the costs of the current manual inspection, showing a positive return on investment and, based on the concrete implementation, possible cost savings of several hundred thousand Euros per year.

## 6. Conclusion

In many manufacturing companies, the process of manual optical inspection requires considerable human and financial resources. Also, it involves various challenges that influence inspection quality. Here, ML-based visual inspection has emerged as a powerful technology to automatically monitor product quality and thus reduce the number of defective products. Simultaneously, cloud services and the use of off-the-shelf AI technologies simplify the development and the adoption of pre-trained neural networks. In this paper, we demonstrated different approaches to the development of an ML-based visual inspection system in the manufacturing industry utilizing MLaaS platforms. Especially IBM's *Maximo Visual Inspection* could achieve excellent results in binary classification, multilabel classification and detection tasks, performing even better than a custom model based on PCA. We demonstrated the value of this artifact to the involved business unit, fostering understanding in AI and showing how they can use the data from these models to improve their production processes. Further research can combine the findings from this and several similar use cases into a framework for automating visual inspection problems using Machine Learning.

## References

- [1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.
- [3] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [4] Y. Yang, C. Miao, X. Li, and X. Mei, "On-line conveyor belts inspection based on machine vision," *Optik*, vol. 125, no. 19, pp. 5803–5807, 2014.
- [5] J. F. Lachenmaier, H. Lasi, and H.-G. Kemper, "Entwicklung und evaluation eines informationsversorgungskonzepts für die prozess- und produktionsplanung im kontext von industrie 4.0.," in *Wirtschaftsinformatik*, pp. 1–15, 2015.
- [6] A. Carvalho, A. Levitt, S. Levitt, E. Khaddam, and J. Benamati, "Off-the-shelf artificial intelligence technologies for sentiment and emotion analysis: A tutorial on using IBM natural language processing," *Communications of the Association for Information Systems*, vol. 44, no. 1, pp. 918–943, 2019.
- [7] A. Rai, P. Constantinides, and S. Sarker, "Editor's Comments: Next-Generation Digital Platforms: Toward Human-AI Hybrids," *Management Information Systems Quarterly*, vol. 43, mar 2019.
- [8] P. J. Stamas, M. L. Kaarst-Brown, and S. A. Bernard, "The business transformation payoffs of cloud services at mhawk.," *MIS Quarterly Executive*, vol. 13, no. 4, 2014.
- [9] L. Baier, F. Jöhren, and S. Seebacher, "Challenges in the Deployment and Operation of Machine Learning in Practice," *Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm and Uppsala, Sweden, June 8 - 14, 2019. Research Papers.*, 2019.
- [10] K. Wagstaff, "Machine learning that matters," in *Proceedings of the 29th International Conference on Machine Learning*, 2012, pp. 529–536, 2012.
- [11] T. S. Newman and A. K. Jain, "A survey of automated visual inspection," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 231–262, 1995.
- [12] R. T. Chin, "Automated visual inspection: 1981 to 1987," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 3, pp. 346–381, 1988.
- [13] J.-L. Maire, M. Pillet, and N. Baudet, "Toward a better visual inspection of products," *The TQM Journal*, vol. 28, no. 3, pp. 411–430, 2016.
- [14] Y. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 731–747, 2018.
- [15] J. E. See, "Visual inspection reliability for precision manufactured parts," *Human factors*, vol. 57, no. 8, pp. 1427–1442, 2015.
- [16] T. Čater and B. Čater, "Product and relationship quality influence on customer commitment and loyalty in b2b manufacturing relationships," *Industrial Marketing Management*, vol. 39, no. 8, pp. 1321–1333, 2010.

- [17] A. C. Van Riel, C. P. De Mortanges, and S. Streukens, "Marketing antecedents of industrial brand equity: An empirical investigation in specialty chemicals," *Industrial Marketing Management*, vol. 34, no. 8, pp. 841–847, 2005.
- [18] A. Kujawińska and K. Vogt, "Human factors in visual quality control," *Management and Production Engineering Review*, vol. 6, 2015.
- [19] Q. Luo and Y. He, "A cost-effective and automatic surface defect inspection system for hot-rolled flat steel," *Robotics and Computer-Integrated Manufacturing*, vol. 38, pp. 16–30, 2016.
- [20] A. D. Swain and H. E. Guttman, "Handbook of human-reliability analysis with emphasis on nuclear power plant applications. final report," report, Sandia National Labs., 1983.
- [21] C. G. Drury, M. H. Karwan, and D. R. Vanderwarker, "The two-inspector problem," *IIE Transactions*, vol. 18, no. 2, pp. 174–181, 1986.
- [22] A. D. Thomas, M. G. Rodd, J. D. Holt, and C. Neill, "Real-time industrial visual inspection: A review," *Real-Time Imaging*, vol. 1, no. 2, pp. 139–158, 1995.
- [23] C. Demant, B. Streicher-Abel, and A. Springhoff, *Industrielle Bildverarbeitung: wie optische Qualitätskontrolle wirklich funktioniert*. Springer-Verlag, 2011.
- [24] J. Beyerer, F. P. León, and C. Frese, *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. Springer-Verlag, 2016.
- [25] Y. Yang, Z.-J. Zha, M. Gao, and Z. He, "A robust vision inspection system for detecting surface defects of film capacitors," *Signal Processing*, vol. 124, pp. 54–62, 2016.
- [26] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [27] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard, "Adaptive data augmentation for image classification," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3688–3692, Ieee, 2016.
- [28] D. Yin, R. G. Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, "A fourier perspective on model robustness in computer vision," in *Advances in Neural Information Processing Systems*, pp. 13276–13286, 2019.
- [29] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Pearson, 2011.
- [30] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [31] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [33] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, pp. 322–330, 2017.
- [34] H. Yang and M. Tate, "A descriptive literature review and classification of cloud computing research," *Communications of the Association for Information Systems*, vol. 31, pp. 35–60, 2012.
- [35] C. Matt, "Fog computing," *Business & information systems engineering*, vol. 60, no. 4, pp. 351–355, 2018.
- [36] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.
- [37] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
- [38] IBM, "IBM Maximo Visual Inspection - Overview," [www.ibm.com/products/ibm-maximo-visual-inspection](http://www.ibm.com/products/ibm-maximo-visual-inspection), 2020-07-15.
- [39] OpenCV, "Object Detection - OpenCV 2.4.13.7 documentation," [docs.opencv.org/2.4/modules/imgproc/doc/object\\_detection.html](https://docs.opencv.org/2.4/modules/imgproc/doc/object_detection.html), 2020-07-15.
- [40] Google, "Cloud AutoML - Custom Machine Learning Models," [cloud.google.com/automl](https://cloud.google.com/automl), 2020-07-15.
- [41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.
- [42] IBM, "Version 1.3.0: IBM Maximo Visual Inspection Guide," [ibm.com/support/knowledgecenter/SSRU69](http://ibm.com/support/knowledgecenter/SSRU69), 2020-07-15.
- [43] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [44] R. Kohli and N. P. Melville, "Digital innovation: A review and synthesis," *Information Systems Journal*, vol. 29, no. 1, pp. 200–223, 2019.
- [45] E. Mumford, *Redesigning human systems*. IGI Global, 2003.
- [46] J. Swan, S. Newell, and M. Robertson, "The illusion of 'best practice' in information systems for operations management," *European Journal of Information Systems*, vol. 8, no. 4, pp. 284–293, 1999.
- [47] J. Wanner, L.-V. Herm, and C. Janiesch, "How much is the black box? the value of explainability in machine learning models.," in *ECIS*, 2020.
- [48] K. Martin, "Designing ethical algorithms," *MIS Quarterly Executive*, vol. 18, no. 2, 2019.