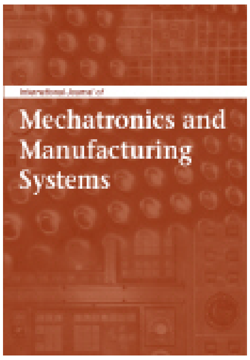



**International Journal of Mechatronics and Manufacturing Systems**

ISSN online: 1753-1047 - ISSN print: 1753-1039
https://www.inderscience.com/ijmms

---

# Enhancing power skiving tool longevity: the synergy of AI and robotics in manufacturing automation

Daniel Kiefer, Florian Grimm, Tim Straub, Günter Bitsch, Clemens Van Dinther

**DOI:** 10.1504/IJMMS.2024.10068102

**Article History:**

| | |
|---|---|
| Received: | 15 February 2024 |
| Last revised: | 21 April 2024 |
| Accepted: | 04 May 2024 |
| Published online: | 02 December 2024 |

# Enhancing power skiving tool longevity: the synergy of AI and robotics in manufacturing automation

Daniel Kiefer*, Florian Grimm, Tim Straub and Günter Bitsch

ESB Business School,
Reutlingen University,
Alteburgstraße 150, 72762, Reutlingen, Germany
Email: daniel.kiefer@reutlingen-university.de
Email: florian.grimm@reutlingen-university.de
Email: tim.straub@reutlingen-university.de
Email: guenter.bitsch@reutlingen-university.de
*Corresponding author

Clemens Van Dinther

Karlsruhe Institut of Technology (KIT),
Institute of Information Systems and Marketing (IISM),
Kaiserstraße 89, 76133, Karlsruhe, Germany
Email: c.v.dinther@gmx.de

**Abstract:** In gear manufacturing, the longevity and cost-effectiveness of *power skiving* tools are essential. This study presents an innovative approach that combines artificial intelligence and robotics in manufacturing automation to prevent tool breakage *to improve the remaining useful life* (RUL). Using a robotic cell, the system captures six images per tooth from different angles. An unsupervised *generative* deep learning model approach is used because it is more suitable for industrial application as it can be trained with a small number of defect-free images. It is used in a first step as a classifier and, in a second step, to segment tool wear. This approach promises economic benefits by reducing manual inspection and, through automated tool inspection, detecting wear earlier to prevent tool breakage.

**Keywords:** power skiving; RUL; remaining useful life; artificial intelligence; robotics; anomaly detection; deep learning; economic efficiency; industrial applications.

**Reference** to this paper should be made as follows: Kiefer, D., Grimm, F., Straub, T., Bitsch, G. and Van Dinther, C. (2024) 'Enhancing power skiving tool longevity: the synergy of AI and robotics in manufacturing automation', *Int. J. Mechatronics and Manufacturing Systems*, Vol. 17, No. 2, pp.201–224.

**Biographical notes:** Daniel Kiefer studied Mechanical Engineering at HTW Aalen and completed his Master's degree in Operations Management at ESB Business School. He is currently pursuing a PhD in Information Systems at the Karlsruhe Institute of Technology (KIT) while working as a research associate at ESB Business School.

Florian Grimm earned his Bachelor's and Master's degrees in Media Informatics from the University of Tübingen. He is pursuing a PhD in Information Systems at KIT and serves as a research associate at ESB Business School.

Tim Straub is a Professor of General Business Administration, with a focus on Management of Digital Technologies, since 2023. He has held various academic and professional roles, including acting professorships in Business Informatics and Business Analytics, and leadership positions in Information Management & Analytics at the FZI Research Center for Information Technology. He holds a Dr. rer. pol. from KIT and has extensive experience in research and project management.

Günter Bitsch is the Founder and Managing Partner of becos GmbH in Stuttgart, renowned for his expertise in Manufacturing Execution Systems (MES) and Business Intelligence (BI). He earned his doctorate from the University of Siegen in 2010 and has been a Professor at ESB Business School since 2013, becoming a permanent Faculty Member in the summer semester of 2020.

Clemens van Dinther was a Full Professor at ESB Business School, focusing on business informatics and analytics. He has served as the CEO of Vivavis and is currently a Senior Advisor to 3DSE Management Consultants, providing strategic guidance based on his extensive industry and academic experience.

This paper is a revised and expanded version of a paper entitled 'Anomaly detection in hobbing tool images: using an unsupervised deep learning approach in manufacturing industry' presented at *5th International Conference on Industry 4.0 and Smart Manufacturing (ISM 2023)*, Lisbon, Portugal, 22–24 November, 2023.

## 1 Introduction

This paper introduces an anomaly detection approach in gear manufacturing, leveraging the capability of artificial intelligence (AI) and robotics *to prevent tool breakage to improve the remaining useful life* (RUL) of power skiving tools.

This research is part of a project at a leading family-owned German company in hydrodynamic transmissions and provider of diverse drive and braking systems, with applications across energy, transportation and the automotive sector (henceforth referred to as the small and medium-sized enterprise (SME) project partner). Small and medium-sized enterprises (SMEs) often face significant challenges in adopting AI, primarily due to the complexity of AI technologies and the limited resources available to these businesses to keep pace with rapid technological advancements (Kiefer et al., 2022; Zamani, 2022). Neural network models in manufacturing automation offer distinct advantages and challenges. Their ability to process large, unorganised datasets significantly enhances system flexibility and accuracy, proving beneficial in defect detection and workflow acceleration (Dey and Yodo, 2022; D'Urso and Quarto, 2023). However, their effectiveness is tempered by the need for extensive training data and the high computational costs associated with their operation, which can limit their practical application in resource-constrained settings. Thus, the application of neural networks in manufacturing should consider the availability of data and computational resources (Pang et al., 2022).

The need for a solution that is both cost-effective and capable of quick deployment is particularly strong in this context, given that SMEs may not have extensive datasets typically required to train AI models effectively (Bauer et al., 2021). Research indicates that SMEs encounter various operational difficulties. These are exacerbated by external factors such as the COVID-19 pandemic, leading to disruptions in supply chains, workforce availability and market demand, which further strain SMEs' limited resources. The pandemic has underscored the importance of technology adoption for SMEs as a survival strategy, pushing them towards innovative technologies to sustain operations during challenging times (Zamani, 2022).

Given these constraints, developing AI solutions that can function efficiently with smaller datasets and require minimal upfront investment can be particularly valuable for SMEs. This approach would not only address the need for low-cost and rapid deployment, but also cater to the specific operational contexts of SMEs, enabling them to leverage AI for competitive advantage despite their size and resource limitations (Gladysz et al., 2023). Despite expectations that manufacturing would lead in new technology adoption, the penetration and effectiveness of AI in this sector are surprisingly limited, as identified in various scenarios (Villalonga et al., 2021).

The SME project partner has implemented power skiving for manufacturing as new manufacturing technology in its production plant. Gear or power skiving is less common than traditional grinding, honing, shaving and skiving (Nagata et al., 2017; Spur et al., 1999). Until recently, computational limitations are the reason the method had never reached series production grade. The efficiency of power skiving in comparison to traditional methods has led to a revival of this technique, with an increased frequency of its use in production technology today (Antoniadis et al., 2004). However, current challenges of gear production with power skiving are high frequency vibrations caused by cutting power fluctuations during the cutting process, which lead to form errors in cut gear profiles, leads, and chip packing under high-speed cutting conditions, which increase tool wear. Companies have limited experience and empirical values with power skiving (Nagata et al., 2017).

This results in tools wearing out quickly. The cost of a new tool is approximately €5000, in contrast to re-sharpening costs of €500. The SME project partner has noticed an increase in tool expenses due to improvable decisions in sending tools for re-sharpening by the maintenance department. Currently, both shop floor and maintenance workers visually inspect the tool, which typically has around 48 teeth, to determine whether it requires re-sharpening or is still good for production. This task is inherently complex for humans, as wear and damage can be minor, and only detectable from certain angles and under specific lighting conditions (Shen et al., 2021; Xia et al., 2024). Additionally, the assessment is subjective, leading to variability in decisions among different workers. These difficulties can result in a reduction of the RUL of power skiving tools.

The latest developments in AI, especially image classification and segmentation, are promising because these could detect the wear on images of the cutting edges. Consequently, this would result in the formulation of more stable and improved decisions. Furthermore, it would relieve workers of unnecessary tasks and provide support in light of the increasing shortage of skilled workers (Brixiova et al., 2009). Robotics are becoming affordable for SMEs, just as they are becoming easier to program. Unsurprisingly, their use is on the rise: a study by Fernández-Macías et al. (2021) found that, between 1995 and 2015, robot stock per thousand workers rose from 9,410 to 13,417. The combination of AI combined with robots could enhance manufacturing, as shown by Emaminejad and Akhavian (2022) through their adoption in the architecture, engineering and construction (AEC) industries and by Dhouib and Zouari (2023) with non-productive time optimisation of a robotic arm for drilling circular holes. Through precision and efficiency, human errorcan be reduced to meet complex requirements.

There are two types of algorithms:

i generative algorithms, which can learn from normal (or wear-free) tools alone

ii discriminative algorithms, which need normal and anomalous (or worn out) tools.

This distinction is crucial for two reasons. Firstly, anomalies are usually rare, and when production with new tools and processes begins, it may be that no worn-out data for training are available. Secondly, by definition, 'anomalous' encompasses everything that is not normal, making it virtually impossible to capture all possible anomalies with adequate training data (Pang et al., 2022; Xia et al., 2021). This last point is particularly important and deserves a closer look. Our approach is designed to be robust to unseen defects, and makes it much easier for SMEs to adopt to the application of AI in manufacturing.

The objective of our industrial anomaly detection approach involves training a generative model with only defect-free images. A key consideration for manufacturers is how these training datasets are generated. For example, some basic algorithms require extensive datasets for reliable operation (typically around one thousand images or more for each product type and defect), which can be costly and often unfeasible (Yang et al., 2022). The necessity for examples of tools with wear and breakouts to train algorithms marks another distinction since this is a state which is usually undesirable in manufacturing. Therefore, we focus on a generative approach with few-shot training that can operate effectively with significantly fewer examples, usually in the hundreds.

The aim is to develop a method for the SME to classify images taken by a low-cost robot of power skiving tools, to decide whether the tool can continue to be used or is worn out, and to improve the overall RUL of the tools.

The remaining paper is structured as follows. In Section 2, related work and research gap of power skiving as well as the wear anomalies are highlighted. In addition, anomaly detection and the state-of-the-art is discussed in greater detail. It is from this that the research gap is highlighted, which leads to the formulation of the research questions. Section 3 presents the method and the proposed experimental design to answer these questions. Section 4 then presents the results of the experiments. To further enhance the contributions made by this paper, an assistance procedure for companies is discussed in Section 5. Finally, Section 6 summarises the paper and outlines further research possibilities.

## 2 Related literature

### *2.1 Related work*

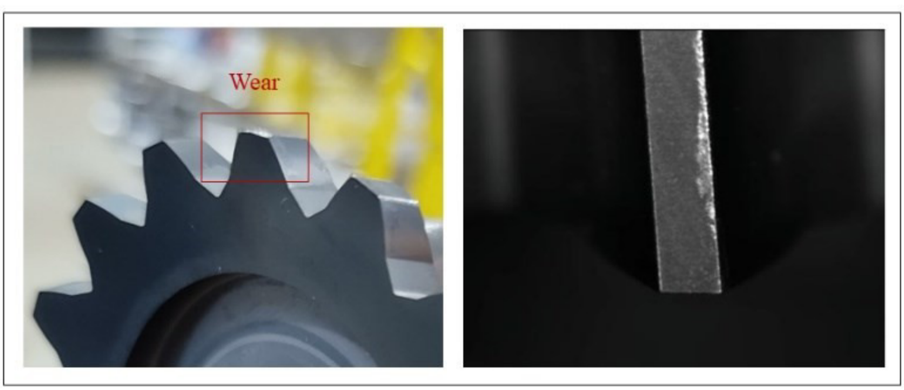
To follow our research, it is essential to understand the current state-of-the-art in power skiving wear, anomaly detection and their approaches as well as methods from AI. For the analysis and preparation of the related work, the guidelines of Levy and Ellis (2006) as well as Webster and Watson (2002) are followed.

### 2.1.1 Power skiving, tool wear and remaining useful life

Power skiving was patented in 1910 by George Adams in London (Adams, 1910) but, at that time, it was not economically feasible. It was only the progressive improvement of electronic gear coupling in recent times that has allowed the technology to become economically viable. The cutting speed is generated during power skiving when the workpiece and tool rotate at a relatively high speed and at a defined angle with an offset axis. The inclined position of the axes generates differently directed vectors for the two circumferential speeds, the difference vector of which dictating the cutting speed (Tsai, 2016). The process of power skiving is not trivial, and to prevent collisions between the workpiece and tool flanks, a constructive clearance angle is necessary on the tool. The design of the skiving tool mirrors that of a gear shaping cutter, which is essentially an external gear, but modified to improve the efficiency and accuracy of the machining process (Olivoni et al., 2022).

Tool wear is a critical parameter requiring monitoring in the power skiving process, see Figure 1. A study by Olivoni et al. (2022) provides an extensive literature review on key factors simulated in power skiving. From 23 analysed papers, only six of them focus on tool wear. Most of these focus on chip thickness and cutting forces. From the six remaining papers focusing on tool wear, they predict the tool wear in accordance with Ren et al. (2021) by estimating the distribution of the stress and temperature on the rake face of the cutting tool. No one has done this based on images, which represents a clear research gap. Generally speaking, the power skiving community is small compared with, for example, milling and turning as a result of its narrow focus on gear production. For a deeper understanding of the kinematic of power skiving, see Ren et al. (2021).

**Figure 1** An example of wear on power skiving tool from the SME (see online version for colours)




The RUL of an asset or system refers to the duration from the present moment until its useful life concludes. This concept is extensively applied in the fields of operational research, reliability and statistics, and holds relevance in other areas including material science, biostatistics and econometrics. Most work concentrates on estimating the RUL through different approaches, mostly with statistical data-driven methods (Si et al., 2011), although machine learning-based methods have been used in the recent past (Ferreira and Gonçalves, 2022). Methods are commonly based on event data, such as past recorded failure data as well as on condition monitoring (CM) data, such as vibration data, oil analysis data, temperature and pressure etc.

The RUL can be commonly defined as:

$$f(x_t \mid Y_t) = f(x_t) = \frac{f(t + x_t)}{R(t)} \tag{1}$$

- $t$ time as age or usage
- $X_t$ as the random variable of the RUL at time $t$
- $Y_t$ is the history of operational profiles and CM information up to $t$
- $R(t)$ is the survival function at $t$.

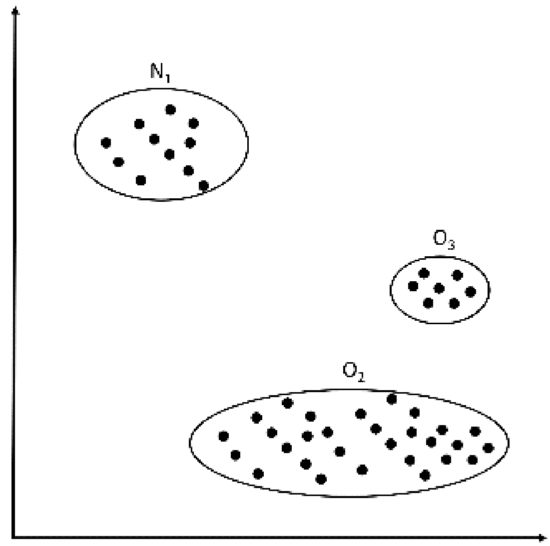
The availability of $Y_t$ will certainly provide more information about the RUL of an asset. However, it is a non-trivial task to incorporate $Y_t$ into the estimation of $X_t$. This has led to a large body of literature in the past decades on the formulation and estimation of the RUL to focus on the property of $f(x_t)$ or $E(x_t)$ without the influence of $Y_t$ (Si et al., 2011). For a derivation, see Banjevic (2009).

The related work approaches require a significant amount of data as well as a wide variety of data sources (Ferreira and Gonçalves, 2022; Si et al., 2011), which makes these very complex. Therefore, it is not suitable for SMEs or our use case. This is why we aim to extend the RUL through better maintenance decisions of the tools based on images and not by estimating it through a probability function. Every time the shop floor or maintenance worker can make a better decision based on our AI model, and the power skiving tool does not break down in the field, but is re-sharpened, the RUL increases.

### *2.1.2 Anomaly detection*

Anomaly detection is the process of identifying data instances that deviate from what is considered normal or expected (Sarafijanovic-Djukic and Davis, 2019). Normal data refers to the majority of data instances. Abnormal data, on the other hand, are the deviants in data instances, which are related to wear in the field of RUL (Wei et al., 2018). The anomaly score describes the degree of abnormality for each data point (Pang et al., 2022). In manufacturing, improvements have been made in anomaly detection. D'Urso and Quarto (2023) introduced an artificial neural network (ANN) that predicts defects in metal-material extrusion, enhancing part selection by avoiding unnecessary post-processing. Concurrently, Dhouib and Zouari (2023) implemented a deep learning approach to detect real-time defects in fused filament fabrication, targeting irregular filament sizes and poor surface quality, thereby improving the technology's industrial applicability. Figure 2 describes anomalies in a two-dimensional dataset. Regions labelled as 'N1' and 'N2' contain the majority of observations and are therefore considered normal data instance regions. Conversely, the 'O3' area, and the 'O1' and 'O2' data points are located far from the bulk of the data points and are therefore considered anomalies. These anomalies may occur due to data errors, but sometimes indicate a new basic process that was not previously known.

**Figure 2** Illustration of anomalies in two-dimensional dataset




### *2.1.3 Discriminative and generative models*

Regarding learning approaches of models for anomaly detection there are discriminative and generative models. These models represent two different approaches to divide the data space into classes. Discriminative models learn the boundaries between classes, while generative models understand how the data is embedded into the space. Each approach is suited to specific tasks (Pang and Li, 2022).

Discriminative models are widely used in fields such as natural language processing, handwriting recognition, time series analysis, etc. by ANNs, convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These are the fields where discriminative models are effective and better used for deep learning as they work well for supervised tasks (Pang et al., 2022; Pang and Li, 2022; Xia et al., 2021). Generative models are used for tasks such as exploratory data analysis of high dimensional datasets, image denoising, image compression anomaly detection, and even generating new images (Pang et al., 2022; Xia et al., 2021) with deep learning and machine learning approaches including autoencoder, Boltzmann machine and self-organising maps.

Figure 3 shows the approach of a discriminative model. This is followed by an explanation of the downsides of this approach.

Discriminative algorithms look at the training data of the two classes (normal and anomalous) and attempt to extract discriminating features when constructing the decision boundary. As such, these algorithms are likely not robust on unseen and novel defect types.

To understand the limitations of a discriminative algorithm, it is important to note that anomalous images, which are defined as those that are not normal, occupy a small volume in the feature space. In contrast, the surrounding space of anomalous images is vast. Therefore, it is difficult to gather enough diverse examples of such images for training purposes (Pang et al., 2022).

**Figure 3** The two subfigures show a simplified two-dimensional feature space of a discriminative model (see online version for colours)

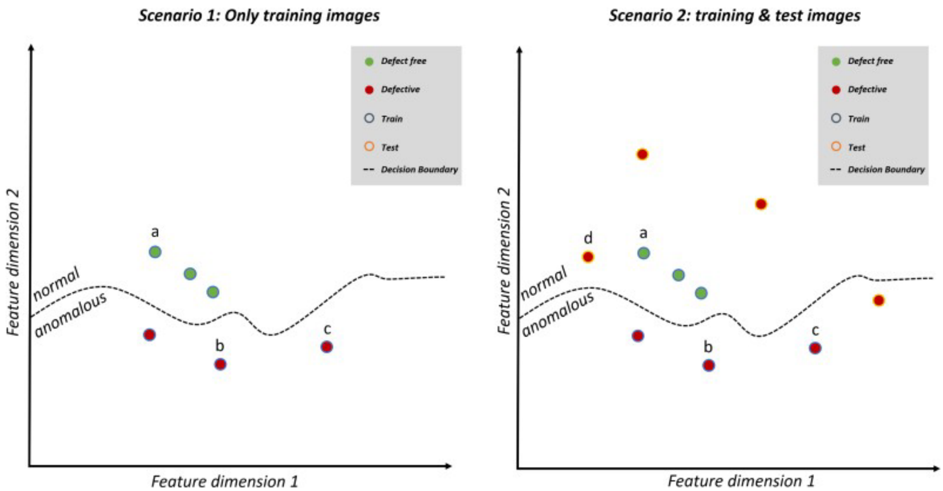


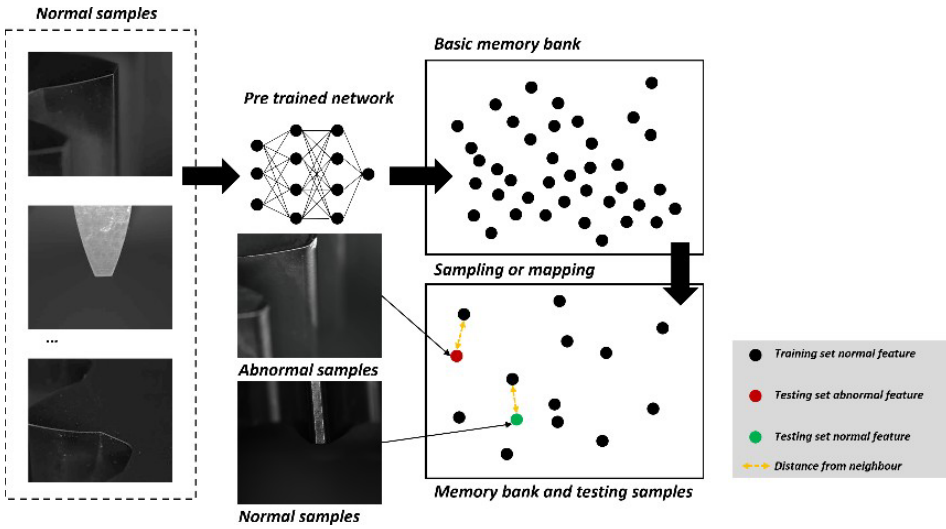

Figure 3 shows that the training images (outlined in blue) only cover the lower part of the space, resulting in a decision boundary that is effective in distinguishing defects in that area. However, it does not account for the possibility of defective products being located further above, to the right, or to the left, as is the case with unseen example (d). This highlights the limitations of discriminative models when defect types are not included in the training set. The decision boundary may perform well on the training data, but it may not accurately classify previously unseen defects that fall within the 'normal' side of the boundary. In this example, image (d) is closer in feature space to the non-defective images than the defective images (b) and (c).

Therefore, it is recommended that generative algorithms which focus on learning the concept of normality are used. Such algorithms can be trained solely from normal images. Algorithms can also benefit from defective images in their training set to improve robustness to specific types of defects. While it is important to note that this is not a requirement, for the reasons set out above, this approach is very suitable for industrial anomaly detection.

Generative unsupervised anomaly detection can be achieved through feature embedding-based methods, such as teacher-student architecture, one-class classification, distribution map and memory bank approaches. Another area is reconstruction-based methods, including autoencoders, generative adversarial networks, transformer and diffusion models (Liu et al., 2024).

Currently, memory-based methods are the most effective in industrial anomaly tasks (Liu et al., 2024). These methods do not require a loss function for training and models can be constructed quickly. Their performance is ensured by a robust pre-training network and additional memory space (Yang et al., 2022). The main difference between memory bank- based methods and one-class classification-based methods is that memory-based methods, such as support vector data description, require additional memory space to store image features. Figure 4 illustrates that these methods need minimal network training and only require the sampling or mapping of the collected normal image features for inference. During the inference phase, the characteristics of the test image are compared to those stored in the memory bank. The probability of the test image being abnormal is determined by its spatial distance from the normal features stored in the memory bank (Liu et al., 2024).

**Figure 4** Architecture of memory bank-based methods (see online version for colours)




Memory bank-based methods, such as semantic pyramid anomaly detection (SPADE) (Cohen and Hoshen, 2020), the self-organising map for anomaly detection (SOMAD) (Li et al., 2021), PatchCore (Roth et al., 2022) and fast adaptive patch memory (FAPM) (Kim et al., 2023) share the characteristic of being pre-trained, which is a concept of transfer learning (TL) that achieves the best results on the ResNet (Zagoruyko and Komodakis, 2016) backbone (Liu et al., 2024).

TL is a machine learning technique that involves adapting a model trained for one task to a related task. This approach is widely used in various domains, including natural language processing, image classification, and specialised fields such as atmospheric particle classification and disease prediction. TL addresses challenges such as data scarcity and can significantly improve model performance by leveraging pre-existing knowledge from related tasks (Weiss et al., 2016).

PatchCore by Roth et al. (2022) represents a significant advancement in industrial image anomaly detection, significantly improving the performance of MVTec AD. PatchCore has two unique features. Firstly, the memory bank of PatchCore is coreset-subsampled to ensure low inference costs while maximising performance. PatchCore determines whether the test sample is abnormal by measuring the distance between the test sample's nearest neighbour feature in its memory bank and other features. This reweighting process enhances PatchCore's robustness (Liu et al., 2024).

Despite challenges in collecting diverse and sparse abnormal data, supervised learning still offers potential in anomaly detection by utilising a small set of abnormal samples alongside a larger set of normal ones. Chu and Kitani (2020) introduced a semi-supervised approach to address significant data imbalances in anomaly detection. They leverage changes in loss values during training as indicators of anomalies, employing a reinforcement learning-based neural batch sampler to identify the loss curve differences between normal and anomalous data regions, thus enhancing detection capabilities (Liu et al., 2024). Additional resources and literature reviews on anomaly detection can be found in Liu et al. (2024), Alloqmani et al. (2021) and Yang et al. (2022).

### *2.2 Research gap*

There are very few comprehensive studies on the application of deep learning – both discriminative and generative models – for anomaly detection and wear analysis in power skiving tools within an industrial setting. Rather, related work has focused mainly on conditional monitoring data for tool wear prediction.

Although Bergs et al. (2020) presented a CNN for classifying tools and an additional network based on the U-net architecture for segmenting wear on cutting tools, there are few areas where we aim to contribute. Firstly, while they focused on various tools, we concentrate on power skiving tools, an area where research is scarce, and which we aim to enrich. Secondly, they trained the model with 3000 images. We aim to use a generative approach with only normal images to detect wear, as this is a more feasible approach for industry applications. Additionally, we develop an automated system that includes image recording, wear detection and segmentation to assist workers in making better decisions and improving overall RUL.

This gap highlights the need for empirical research on how such AI-driven methodologies can improve efficiency and decision-making in manufacturing automation.

Building on this identified research gap, the following research questions are formulated to directly address these unexplored areas in the context of power skiving tool maintenance using AI-driven methodologies:

> ***RQ1:*** *Is it feasible for a generative deep learning model trained solely on normal data images to detect anomalies in power skiving tools with an accuracy level suitable for industrial applications compared to a discriminative model?*
>
> ***RQ2:*** *Can the generative deep learning model accurately segment and highlight wear on images of power skiving tools, providing insights for decision-makers?*
>
> ***RQ3:*** *How does the learning curve for accurately classifying defect-free images in anomaly detection models vary with the number of training examples provided?*
>
> ***RQ4:*** *Overall, does the implementation of a deep learning-driven, automated imaging system support the decisions regarding tool re-sharpening in gear production?*

## 3 Materials and methods

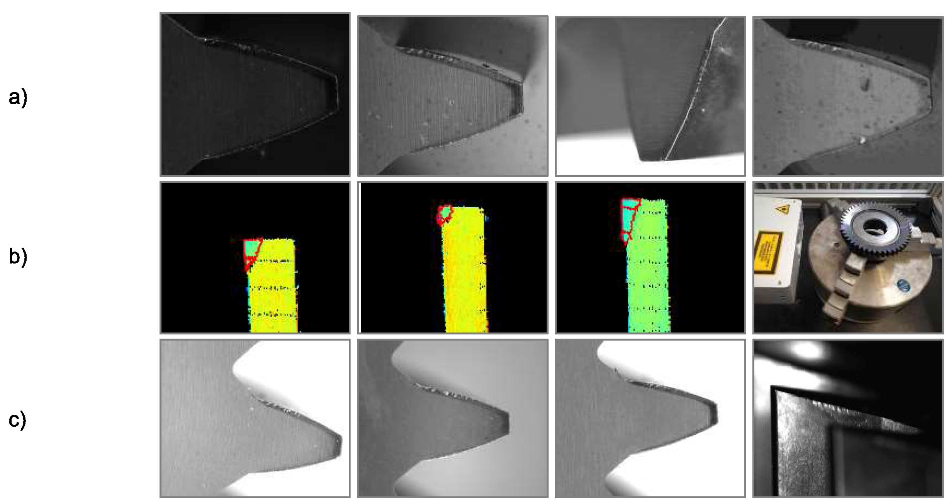
In order to answer the aforementioned research questions and expand upon existing investigations, we propose an experimental design that is adapted to the shortcomings mentioned in the previous sections regarding an appropriate metric as well as the highlighted methods. Our methodology integrates advanced AI techniques with robotic automation to predict and enhance the RUL of power skiving tools.

### *3.1 Robotic image automation setup*

An image generation automation system was required to record all relevant surfaces of each tooth of the power skiving tool. This includes the rake face, as well as the five clearance surfaces. It is necessary to consider each tooth individually, since recording only the rake face from above is not sufficient. The physical resolution must be within the range of <1 μm due to the common wear of [10,100] μm in power skiving. This means that teeth can only be measured individually. A resolution of 2.5 MP (mega pixels) is sufficient witha field of view of 10 × 10 mm.

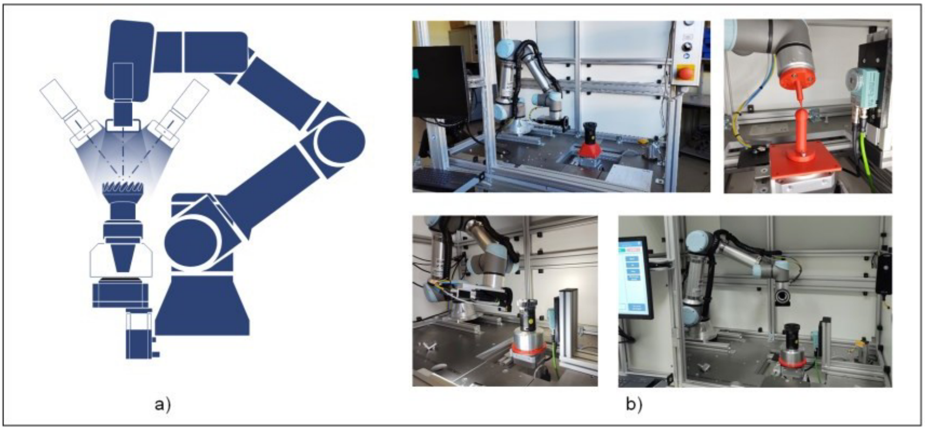
Figure 5 displays some of the results from the different approaches pursued. One approach was to introduce a laser distance measurement. Although the results were encouraging, they were ultimately discarded due to limited resolution and difficulties in handling all relevant surfaces. In imaging approaches, the use of orthogonally aligned illumination (coaxial illumination with strictly parallel light guidance, see Figure 6(a)) provides the most informative images. However, the necessary kinematic guidance of the camera system to reach all relevant surfaces makes the large-volume setup of such lighting a hindrance. Additionally, this variant is very cost intensive. As an alternative, we investigated the use of incident light illumination with a ring light (Figure 6(c)). With properly aligned illumination, we were able to achieve very good images, as demonstrated by the comparison of images (a) and (b) in Figure 6, both of which clearly show the individual grinding grooves on the rake face of the tool.

**Figure 5** Example images from a comparative test for optical detection of worn tools or cutting edges are shown. The images include: (a) dark field illumination and coaxial illumination using hardware from Keyence; (b) laser triangulation sensor for measuring the front face of the tool tooth from Aku.Automation and (c) incident light illumination proposed by Rauscher (see online version for colours)




A decision was made in favour of a variant with simple incident lighting using a ring light (see Figure 5(c)). The camera system comprises a monochrome 5-megapixel Basler acA2040-35gm camera (with a sensor size of 2/3 inch and a 1/1.8 inch Sony IMX265, manufactured by Basler AG), a macro lens Opto 100-MC100 (with a working distance of 87 mm, aperture of 14, depth of field of 0.4 mm and a field of view of 8.8 × 6.6 mm, manufactured by Opto GmbH), and an LED illumination ring MBJ SRL-07-WT-s (with anouter diameter of 73 mm, manufactured by MBJ Imaging GmbH).

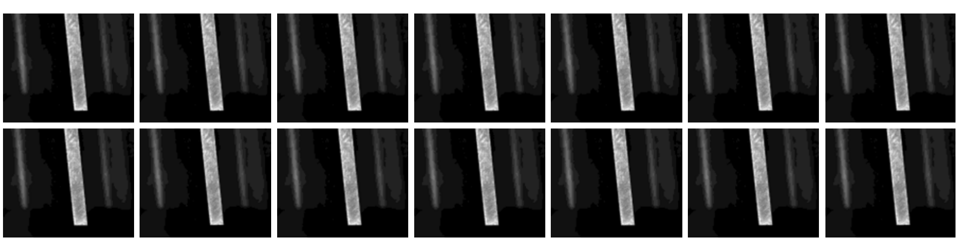
**Figure 6** Sketch of the measuring system with an articulated robot for camera guidance: (a) illustration of the robot and tool and (b) original images of the robot cell (see online version for colours)




Initially, a portal axis system was pursued due to its high accuracy. However, due to ongoing delivery delays, a less accurate and more flexible system was used instead. This system features a lightweight articulated robot (UR 5e, Universal Robots) that is significantly more flexible and will allow for investigations to be extended to other tool geometries in the future.

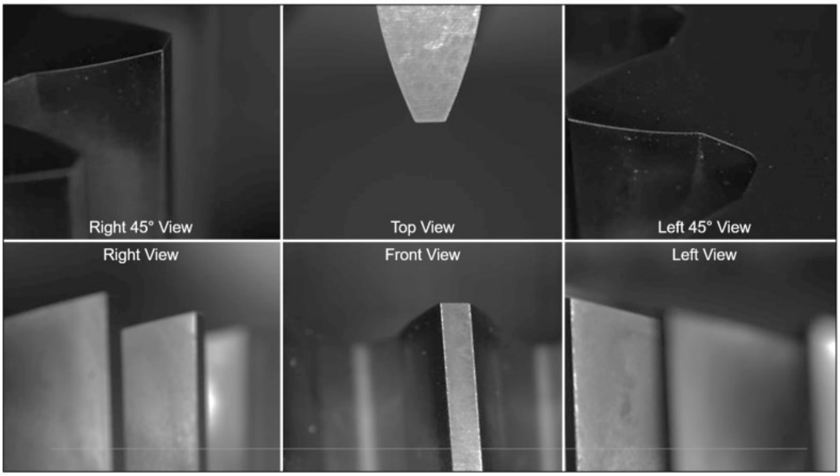
The system requires only one camera despite the articulated arm robot. The accuracy of positioning for a 6-axis articulated arm robot is lower (0.1 mm) compared to a portal axis system (e.g., Festo: 0.01 mm). To verify feasibility, a repeat test was conducted in advance. Two positions were defined, and the robot moved alternately between them. An image was captured by the camera system at one of the positions. The tests conducted on 150 images demonstrated that the robot's positioning and repetition accuracy is adequate for the application. Figure 7 displays a random selection of these images.

**Figure 7** Sample image series of the repeat test



Six surfaces must be considered when analysing the tool geometry, namely, the rake face and the five clearance surfaces (see Figure 8). The positioning of these surfaces depends on the orientation of the tool. The positions are calculated ad-hoc, ensuring that the size of the power skiving tools is limited only by the reach of the robot. The positions are determined using a spherical coordinate system that starts from the centre of the tool.

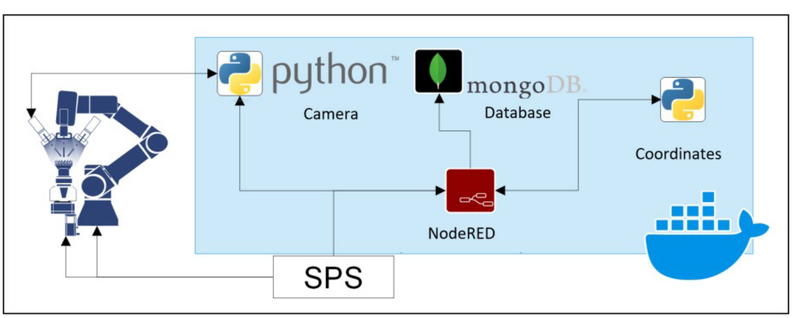
**Figure 8** Images of all six angles from a defective free tooth




The individual elements of the measurement system are programmed using various programming languages and interfaces. To minimise complexity, we aim to modularise according to the microservice architecture. Regardless of the programming language, individual modules are linked together via the REST interface. We encapsulate individual functions and services in virtual containers. These can be developed separately from each other. This type of architecture is standard in scalable server applications from large cloud providers and is also set to be used in edge computing in the future. The reusability and easy maintainability of individual containers are of central importance.

The benefit of this architecture is that it allows for the flexible addition of new functionalities by different developers. The central NodeRED container provides the programmable logic controller (PLC) with functions that would require significant effort to implement using a classic, PLC-based architecture. For instance, during the research project, various models can be made available as containers. The same applies to the development of a user interface to explain the decisions of the models to the operator. The structure is a recent development for many industries and SMEs. Program cycles were created on the PLC side, which also include the selection of different tool types. The PLC also offers an intuitive user interface for the operator. Additionally, it can store data for new tools, which are retrieved from the NodeRED container as required, see Figure 9 for the microservice architecture.

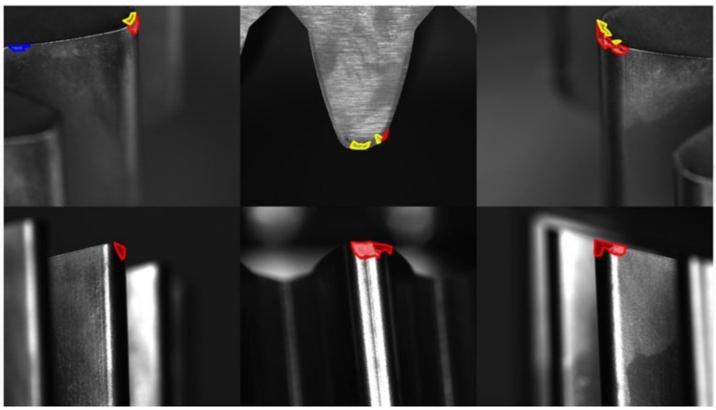
**Figure 9** Structure of the microservice architecture (see online version for colours)



### 3.2 Database and handling

The images captured by the robot were stored in a MongoDB. An open-source software called Label Studio was used to create the image masks with information concerning the wear on the tool. The generated masks are called ground truth because they provide the true answer to the wear localisation. These were uploaded to Label Studio were they got classified and segmented: see Figure 10.

**Figure 10** Tool with wear segmented classified as not OK (see online version for colours)



In total we obtained 600 images. 250 of these are classified as good as they can be used for training for the generative classifier. 350 images are classified with wear in addition, and are segmented into wear, critical wear and outbreaks. The latter are used for the generative supervised segment model. During the label process all images from one tooth are stitched together to provide holistic information for the expert. For training purposes, the images are not stitched together and are fed into the model as a single image.

### 3.3 Training and evaluation approach

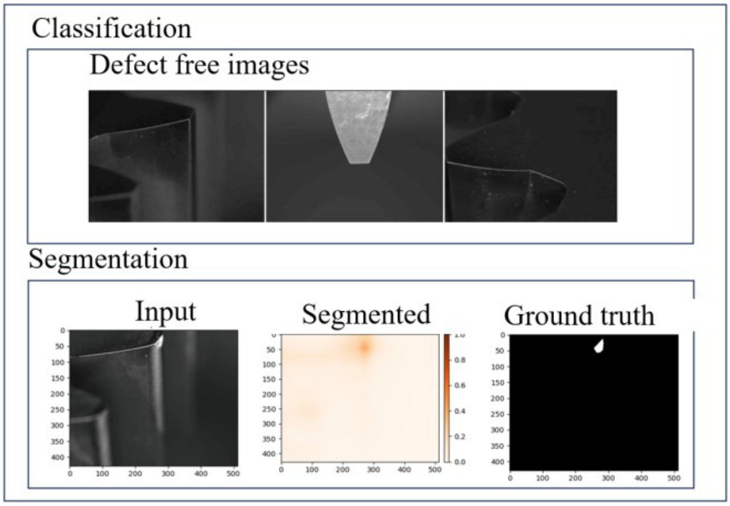
In deep learning, it is standard practice to divide the dataset into separate subsets. The most important subset for the training phase of a neural network (NN) is the training data. Apart from this, a subset of the data, known as the test data, is reserved for evaluating the network's performance on images it has not encountered before. This means that, although the test data come from the same pool as the training data, they are withheld from the network throughout the training phase, ensuring that the network is tested on entirely new data.

The unsupervised generative classifier was trained with 150 images of good nominal examples of the power skiving tool and evaluated on 30 different wear defective images. The supervised generative approach was trained with 150 images of good nominal examples as well as on 50 defective examples, and evaluated on 30 different wear defective images.

For training purposes, the images are scaled down to 1.200 pixels as well as grey scale images.

A comparison of the training procedure of the unsupervised classifier and the supervised segment version with ground truth is shown in Figure 11.

**Figure 11** Unsupervised training examples are at the top (defect free images only) for classification. Segmentation training examples are at the bottom. Original image (left), image during label process (middle), ground truth mask (right) (see online version for colours)




For the first experiment regarding the classifier, we used Accuracy and F1, which are common metrics for evaluating classifiers (Wardhani et al., 2019). In addition, we investigated the image-level anomaly detection performance via the area under the receiver operator curve (AUROC), which is a common metric for this purpose (see Kim et al., 2023; Li et al., 2021; Roth et al., 2022). To measure the segmentation performance, we used pixel-wise AUROC. The AUROC is calculated as the area under the ROC curve. A ROC curve shows the trade-off between true positive rate (TPR) and false positive rate (FPR) across different decision thresholds (Baratloo et al., 2015).

- **True positive (TP) =** the number of cases correctly identified with defect.
- **False positive (FP) =** the number of cases incorrectly identified with defect.
- **True negative (TN) =** the number of cases correctly identified as defect-free.
- **False negative (FN) =** the number of cases incorrectly identified as defect-free.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

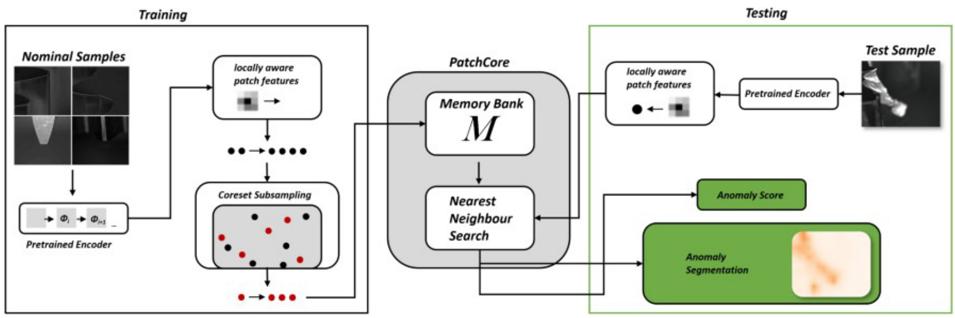
$$F1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

## 3.4 AI model

Since the PatchCore (see Figure 12) algorithm achieved such good results in the MVTec Challenge dataset, we chose this as our main AI model (Bergmann et al., 2021). It meets the requirements of a generative deep learning model, which can only be trained on error-free images. At the same time, it is memory-based, which is the current state-of-the-art.

In addition, comparisons show good few-shot results (Roth et al., 2022) and very fast inference times, which are relevant within the SME application.

**Figure 12** PatchCore algorithm (see online version for colours)




PatchCore by Roth et al. (2022) consists of two relevant parts, local patch features aggregated into a memory bank as well as a coreset-reduction method to increase efficiency.

1 PatchCore uses a network trained on ImageNet and emphasises the importance of features from different levels within this network for image analysis. It assigns features from different hierarchical levels of a pre-trained network to images, focusing on feature maps common in ResNet-style architectures. These features, particularly from the final outputs of the spatial resolution blocks, are critical to the performance of the model to preserve localised information and avoid bias towards unrelated tasks. This method aims to improve anomaly detection by focusing on patch-level features and local neighbourhood aggregation to increase the model's sensitivity to spatial variation, without compromising the resolution of the feature map. For all nominal training samples $x_i \in X_N$, the PatchCore memory bank $M$ is simply defined as in equation (4).

$$\mathcal{M} = \bigcup_{x_i \in \mathcal{X}_N} \mathcal{P}_{s,p}(\phi j(x_i)) \tag{4}$$

2 The coreset reduction method aims to improve the efficiency of anomaly detection by significantly reducing the size of the memory bank and the inference time without compromising performance. This approach uses a coreset subsampling mechanism to ensure that the reduced memory bank retains a similar coverage of nominal features as the original, allowing for faster and nearly equivalent problem solving. This technique is particularly beneficial for handling large datasets and is implemented through an iterative greedy approximation, supported by dimensionality reduction to speed up the selection process.

The PatchCore anomaly detection method uses a memory bank of nominal patch features to score test images by comparing test patch features to their nearest neighbours in the memory bank. This comparison produces an anomaly score that is adjusted based on the rarity of the nearest nominal features. This approach enables robust image-level anomaly scoring and segmentation, where the segmentation map is generated by spatially aligning the patch anomaly scores and refining the resolution as required. For more details, see the original paper on PatchCore (Roth et al., 2022).

As a benchmark we used classic AlexNet architecture with and without pretraining on ImageNet.

We implemented our models in Python 3.7 (Van Rossum and Drake, 2010) and PyTorch (Paszke et al., 2019). Experiments were run on a GPU Nvidia RTX 2080 TI with a CPU Intel Core i9-10900X, 10x 3.70GHz and RAM of 256GB, DDR4. We used torchvision ImageNet-pretrained models. PatchCore uses a WideResNet50-backbone (Zagoruyko and Komodakis, 2016) for direct comparability. Patch-level features were taken from feature map aggregation of the final outputs in blocks 2 and 3. For all nearest neighbour retrieval and distance computations, we used faiss (Johnson et al., 2021).

## 4 Results

This section presents the findings of the benchmark model and the PatchCore experiments. We only report out of sample (i.e., test set) results, because superior results on the in sample (i.e., training set) can be misleading as regards to overfitting.

In the experiment 1 (see Table 1 for results), the models were trained exclusively on a generative basis using 84 defect free images. PatchCore emerged as the top performer, with an AUROC of 0.85, an accuracy of 0.88 and an F1 score of 0.93, indicating a commendable balance between precision and recall. The pre-trained AlexNet model achieved moderate results with an AUROC of 0.66 and an accuracy of 0.70, along with an F1 score of 0.83, indicating a respectable performance. Conversely, AlexNet without pre-training faced significant challenges, as evidenced by its AUROC of 0.48, a low accuracy of 0.07, and an F1 score of 0. This highlights the critical role of pre-training in deep learning frameworks. The SVM model also showed remarkable effectiveness with an AUROC of 0.59, an accuracy of 0.86 and an F1 score of 0.92, demonstrating its robustness even with a limited training dataset. The experiment 2 (see Table 2 for results) introduced a discriminative training regime where, in addition to the 84 defect-free images, 268 images with defects were also used to train all models except PatchCore, which continued to be trained generatively using only the defect-free images. Notably, the performance metrics for PatchCore remained consistent with Experiment 1, highlighting its robustness and effectiveness in generative training. The pre-trained AlexNet showed a significant improvement in this setting, with an AUROC of 0.83, an accuracy of 0.71, and an F1 score of 0.83, suggesting that discriminative training with a balanced dataset of defect and defect-free images improves its performance. AlexNet without pre-training also showed improvement, with an AUROC of 0.81, an accuracy of 0.58 and an F1 score of 0.61, although it still lagged behind the pre-trained version. The SVM showed a decrease in performance compared to Experiment 1, with an AUROC of 0.64, an accuracy of 0.69, and an F1 score of 0.78, suggesting that the addition of defect images may have introduced complexity that slightly reduced its effectiveness.

**Table 1** Experiment 1, classification – generative training with 84 defect free images

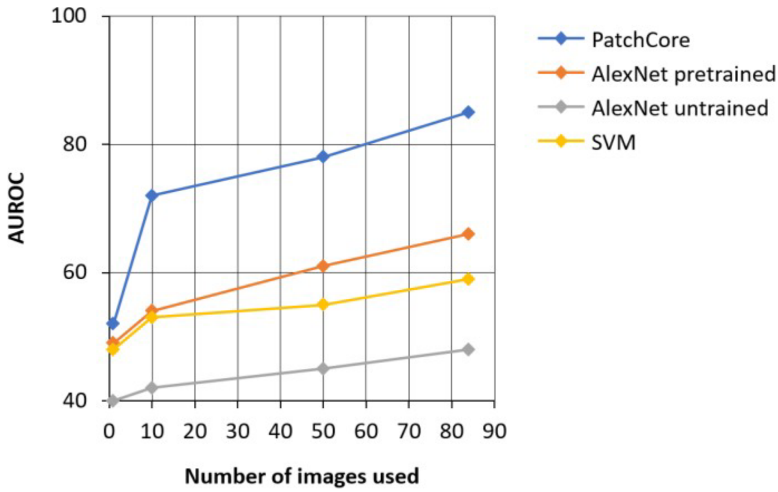
| | *PatchCore* | *AlexNet –pretrained* | *AlexNet – nopretraining* | *SVM* |
|---|---|---|---|---|
| AUROC | ***0.85*** | 0.66 | 0.48 | 0.59 |
| Accuracy | ***0.88*** | 0.70 | 0.07 | 0.86 |
| F1 | ***0.93*** | 0.83 | 0.00 | 0.92 |

**Table 2** Experiment 2, classification – discriminative training with 84 defect free images and 268 defect images

| | *PatchCore* | *AlexNet – pretrained* | *AlexNet – nopretraining* | *SVM* |
|---|---|---|---|---|
| AUROC | ***0.85*** | 0.83 | 0.81 | 0.64 |
| Accuracy | ***0.88*** | 0.71 | 0.58 | 0.69 |
| F1 | ***0.93*** | 0.83 | 0.61 | 0.78 |

PatchCore still trained generative with only 84 defect free images.

**Figure 13** PatchCore shows notably higher sample efficiency in generative setting than other algorithms and achieves higher AUROC with a fraction of nominal training data (see online version for colours)




In experiment 3 (see Table 3 for results) calculated the anomaly pixel AURCO with the threshold values from PatchCore and the ground truth masks to assess the anomaly segmentation skill. A AUROC from 0.94 is very good, to show the capabilities see Figures 18 and 19.

**Table 3** Experiment 3, segmentation – generative training with 84 defect free images, since the other algorithms have no segmentation capability there is no comparison

| | *PatchCore* |
|---|---|
| Anomaly pixel AUROC | 0.94 |

In summary, these experiments demonstrate the strengths and limitations of generative vs. discriminative training approaches in the context of defect detection and classification. PatchCore's consistent performance across both experiments underscores the potential of generative models in scenarios with limited training data. The findings from these experiments contribute valuable insights to the field, particularly for applications requiring high precision in defect detection with constrained training datasets. In Figures 14–19, we demonstrate the input image, the anomaly and the threshold.

**Figure 14** Example from the industry manufacturing dataset with classification. Front view. Image anomaly score 0.13. True negative since there is no wear or breakouts at the power skiving tool (see online version for colours)

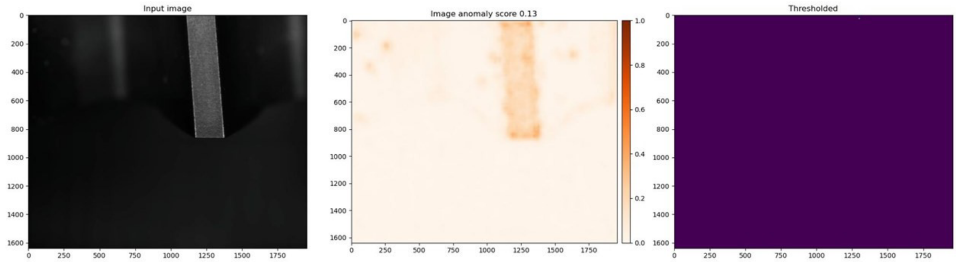



**Figure 15** Example from the industry manufacturing dataset with classification. Front view. Image anomaly score 0.97. True positive since it is a breakout at the power skiving tool (see online version for colours)

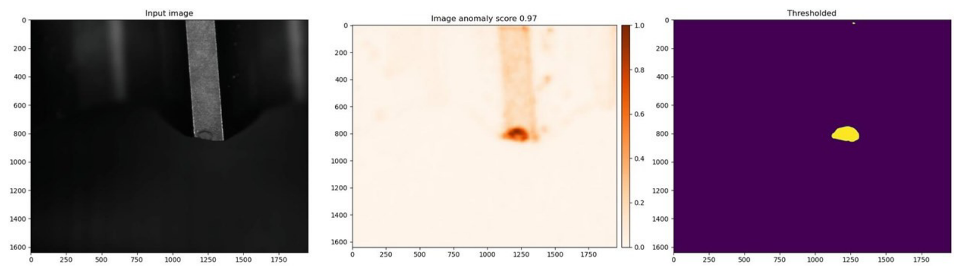



**Figure 16** Example from the industry manufacturing dataset with classification. Top view. Image anomaly score 0.61. False positives since it is just a marking on the power skiving tool (see online version for colours)

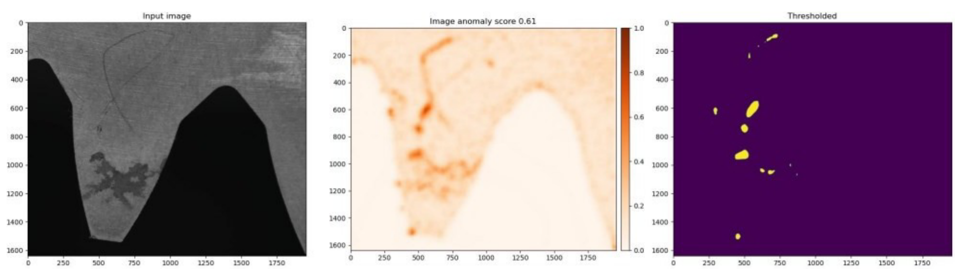



**Figure 17** Example from the industry manufacturing dataset with classification. Front view. Image anomaly score 0.73. True positive since there is wear on the power skiving tool from industry machine usage over time (see online version for colours)

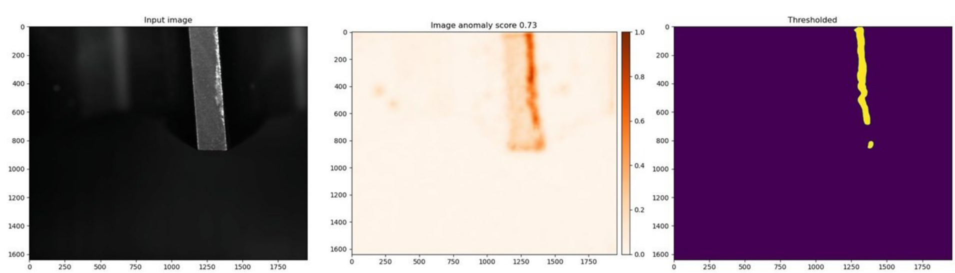


**Figure 18** Example from the industry manufacturing dataset with segmentation and ground truth. Side view. Image anomaly score 0.35. True positive since there is a breakout (see online version for colours)

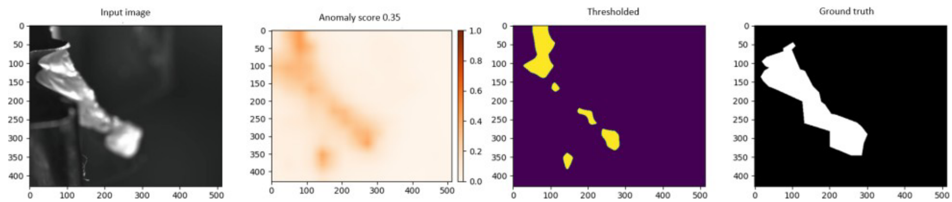
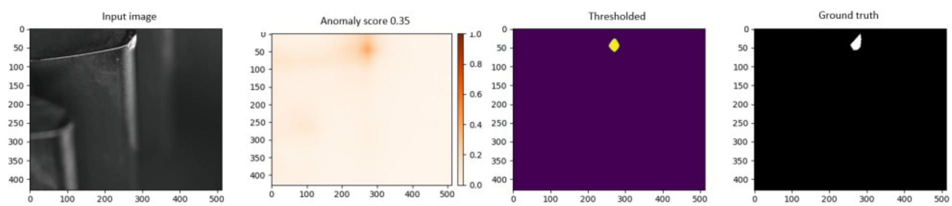


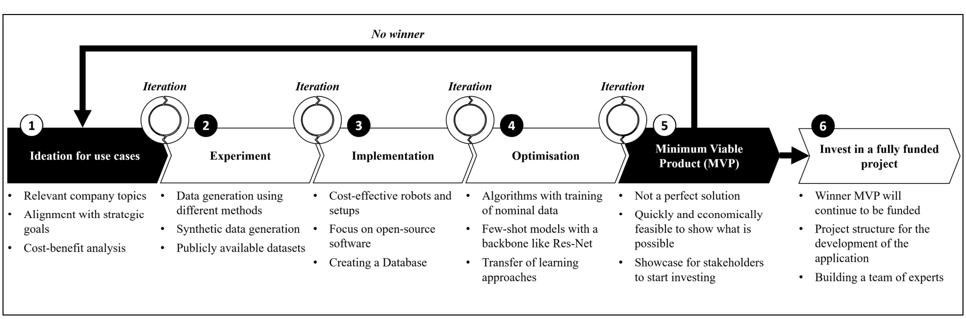

**Figure 19** Example from the industry manufacturing dataset with segmentation and ground truth. Side view. Image anomaly score 0.35. True positive since there is a breakout (see online version for colours)




## 5 Implications for practitioners

Based on our experience in this industrial anomaly detection use case, we have identified a series of steps that we believe will be beneficial for industrial users and summarised them in Figure 20. We see that AI already has its place in industrial applications and can offer improvements and support. Even with just a few error-free images, it was possible to train a generative model that works well. This is very important for industry. It also shows that a low-cost approach is already possible. A basic articulated robot paired with an exposure and camera system effectively trained the model. The few-shot distribution in Figure 13 shows that acceptable results can be obtained quickly. Based on the holistic framework for AI systems in industrial applications developed by Kaymakci et al. (2021), we have developed our own framework for AI applications in industrial settings, focusing on a fast MVP and small datasets, as well as a generative approach.

**Figure 20** Generic step model for AI applications in industrial settings



## 6 Discussion and further research

This paper proposed both an approach to tool type classification and an approach to tool wear detection for cutting tools. Our findings demonstrate that generative deep learning models trained solely on normal (defect free) data outperform traditional discriminative models in classifying anomalies and segmenting wear on power skiving tools. This confirms their industrial-grade accuracy and effectiveness in enhancing decision-making processes for gear tool re-sharpening. The approaches performed on unknown defects well, so a generalisation that allows inferring could be reached. In comparison, Bergs et al. (2020) could not achieve the generalisation on unknown data. Further improvements could be made by augmentation of the training data such as flip, multiply, rotate, blur or contrast normalisation. This could further reduce the needed amount for defect-free training images and lower the hurdle for application.

i Addressing **RQ1**, our research confirmed the feasibility of using generative deep learning models trained only on normal data to detect anomalies in power skiving tools with industrial-grade accuracy, outperforming traditional discriminative models.

ii With respect to **RQ2**, these generative models perform well at segmenting and highlighting tool wear, providing actionable insights for decision-makers.

iii In response to **RQ3**, we observed a direct relationship between the number of training examples and the accuracy of defect-free image classification, highlighting the importance of a robust training dataset. More important we also saw, that the generative approached performed better than the discriminative models with less data.

iv Finally, in response to **RQ4**, the use of an automated deep learning-based image processing system can help to improve the decision-making process in gear tool re-sharpening through the automated classification and segmentation of wear, demonstrating the transformative potential of deep learning in industrial settings.

This paper contributes to both academic research and practical applications by demonstrating the effectiveness of generative deep learning models in industrial anomaly detection, specifically in gear sharpening tools. Research wise, it extends the understanding of the capabilities of generative models in anomaly detection with limited data. Practically, it provides a proven framework for implementing advanced AI in manufacturing, improving decision-making processes, and increasing the efficiency of tool maintenance, thereby bridging the gap between theoretical research and real-world industrial applications. The method may transfer well to other cutting or manufacturing processes where tool degradation occurs. For generalisation, in Section 5 we proposed a step-by-step model for industrial AI use cases to achieve a rapid minimum viable product (MVP) status, on the basis of which a decision can be made whether to build on it or not.

Potential shortcomings of the proposed systems include their dependency on training data with a defect free dataset which may hinder their effectiveness in practical, diverse industrial environments.

Future research could integrate machine data (axis torque, axis speed, vibrations, axis feed speed, etc.) to improve decision-making quality. Another important step would be to develop an explainable artificial intelligence (XAI) framework that the workers acceptance improves. The combination of our approach and the suggested future research could achieve superior results. The image-based anomaly detection highlights a level of wear. With machine data it could be proved or even correlated as to which setting was the cause for too much wear. An XAI framework could explain the relationships to workers.

## References

Adams, G. (1910) *Method of Cutting Van Gears Using a Gear-Like Cutting Tool with Cutting Edges on the Face Surfaces of the Teeth, Germany*, DE243514C.

Alloqmani, A.B.Y., Irshad, A. and Alsolami, F. (2021) 'Deep learning based anomaly detection in images: insights, challenges and recommendations', *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 4, pp.205–215.

Antoniadis, A., Vidakis, N. and Bilalis, N. (2004) 'A simulation model of power skiving', *Journal ofMaterials Processing Technology*, Vol. 146, No. 2, pp.213–220.

Banjevic, D. (2009) 'Remaining useful life in theory and practice', *Metrika*, Vol. 69, Nos. 2–3, pp.337–349.

Baratloo, A., Hosseini, M., Negida, A. and El Ashal, G. (2015) 'Part 1: simple definition and calculation of accuracy, sensitivity and specificity', *Emergency*, Vol. 3, No. 2, pp.48–49.

Bauer, M., van Dinther, C. and Kiefer, D. (2021) 'Machine learaning in SME: an empirical study on enablers and success factors', *A Vision for the Future*, Vol. 1, pp.1–10.

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D. and Steger, C. (2021) 'The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection', *International Journal of Computer Vision*, Vol. 129, No. 4, pp.1038–1059.

Bergs, T., Holst, C., Gupta, P. and Augspurger, T. (2020) 'Digital image processing with deep learning for automated cutting tool wear detection', *Procedia Manufacturing*, Vol. 48, pp.947–958.

Brixiova, Z., Li, W. and Yousef, T. (2009) 'Skill shortages and labor market outcomes in central Europe', *Economic Systems*, Vol. 33, No. 1, pp.45–59.

Chu, W-H. and Kitani, K.M. (2020) 'Neural batch sampling with reinforcement learning for semi-supervised anomaly detection', in Vedaldi, A., Bischof, H., Brox, T. and Frahm, J-M. (Eds.): *Computer Vision – ECCV 2020: 16th European* Conference, 23–28 August, Glasgow, UK, *Proceedings, Part XXVI*, Springer, Cham, pp. 751–766

Cohen, N. and Hoshen, Y. (2020) *Sub-Image Anomaly Detection with Deep Pyramid Correspondences* [Online], Available at http://arxiv.org/pdf/2005.02357.pdf

D'Urso, G. and Quarto, M. (2023) 'Neural network as approach for detection of non-compliant semi-finished additive manufactured parts', *International Journal of Mechatronics and Manufacturing Systems*, Vol. 16, Nos. 2–3, pp.261–279.

Dey, A. and Yodo, N. (2022) 'Performance improvement techniques for neural networks in tool condition monitoring', *International Journal of Mechatronics and Manufacturing Systems*, Vol. 15, Nos. 2–3, p.107.

Dhouib, S. and Zouari, A. (2023) 'Optimising the non-productive time of robotic arm for drilling circular holes network patterns via the dhouib-matrix-3 metaheuristic', *International Journal of Mechatronics and Manufacturing Systems*, Vol. 16, Nos. 2–3, pp.320–338.

Emaminejad, N. and Akhavian, R. (2022) 'Trustworthy AI and robotics: implications for the AEC industry', *Automation in Construction*, Vol. 139, p.104298.

Fernández-Macías, E., Klenert, D. and Antón, J-I. (2021) 'Not so disruptive yet? Characteristics, distribution and determinants of robots in Europe', *Structural Change and Economic Dynamics*, Vol. 58, pp.76–89.

Ferreira, C. and Gonçalves, G. (2022) 'Remaining useful life prediction and challenges: a literature review on the use of machine learning methods', *Journal of Manufacturing Systems*, Vol. 63, pp.550–562.

Gladysz, B., Matteri, D., Ejsmont, K., Corti, D., Bettoni, A. and Haber Guerra, R. (2023) 'Platform-based support for AI uptake by SMEs: Guidelines to design service bundles', *Central European Management Journal*, Vol. 31, No. 4, pp.463–478.

Johnson, J., Douze, M. and Jegou, H. (2021) 'Billion-scale similarity search with GPUs', *IEEE Transactions on Big Data*, Vol. 7, No. 3, pp.535–547.

Kaymakci, C., Wenninger, S. and Sauer, A. (2021) 'A holistic framework for AI systems in industrial applications', in Ahlemann, F., Schütte, R. and Stieglitz, S (Eds.): *Innovation Through Information Systems*, Springer International Publishing, Cham, pp.78–93.

Kiefer, D., van Dinther, C., and Straub, T. (2022) *The Time Has Come – Application of Artificial Intelligence in Small-and Medium-Sized Enterprises*, pp.1–5.

Kim, D., Park, C., Cho, S. and Lee, S. (2023) 'FAPM: fast adaptive patch memory for real-time industrial anomaly detection', *ICASSP 2023*, pp.1–5.

Levy, Y. and Ellis, J.T. (2006) 'A systems approach to conduct an effective literature review in support of information systems research', *Informing Science: The International Journal of an Emerging Transdiscipline*, Vol. 9, pp.181–212.

Li, N., Jiang, K., Ma, Z., Wei, X., Hong, X. and Gong, Y. (2021) 'Anomaly detection via self-organizing map', *2021 IEEE International Conference,* pp.974–978.

Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F. and Jin, Y. (2024) 'Deep industrial image anomaly detection: a survey', *Machine Intelligence Research*, Vol. 21, No. 1, pp.104–135.

Nagata, E., Tachikawa, T., Nakahara, Y., Kurita, N., Nakamura, M., Iba, D. and Moriwaki, I. (2017) 'Gear skiving for mass production', *Proceedings of MPT2017-Kyoto*, Kyoto, 28.02.2017-03.03.2017, pp.1–6.

Olivoni, E., Vertechy, R. and Parenti-Castelli, V. (2022) 'Power skiving manufacturing process: a review', *Mechanism and Machine Theory*, Vol. 175, p.104955.

Pang, G., Shen, C., Cao, L. and van Hengel, A., den (2022) 'Deep learning for anomaly detection', *ACM Computing Surveys*, Vol. 54, No. 2, pp.1–38.

Pang, J. and Li, C. (2022) *On the Connection of Generative Models and Discriminative Models for Anomaly Detection* [Online], Available at http://arxiv.org/pdf/2211.08910.pdf

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein,N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S. (2019) 'PyTorch: An imperative style, high-performance deep learning library', *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA.

Ren, Z., Fang, Z., Arakane, T., Kizaki, T., Feng, Y., Nishikawa, T., Kugo, J., Nabata, E. and Sugita, N. (2021) 'Predictions of cutting force and tool wear in gear power skiving', *ASME 2021 16th International Manufacturing Science and Engineering Conference: Volume 2: Manufacturing Processes; Manufacturing Systems; Nano/Micro/Meso Manufacturing; Quality and Reliability*, Virtual, Online, 21–25 June, American Society of Mechanical Engineers (ASME).

Ren, Z., Fang, Z., Arakane, T., Kizaki, T., Nishikawa, T., Feng, Y., Kugo, J., Nabata, E. and Sugita, N. (2021) 'Parametric modeling of uncut chip geometry for predicting crater wear in power skiving', *Journal of Materials Processing Technology*, Vol. 290, p.116973.

Roth, K., Pemula, L., Zepeda, J., Scholkopf, B., Brox, T. and Gehler, P. (2022) 'Towards total recall in industrial anomaly detection', *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18–24 June. New Orleans, LA, USA, IEEE Computer Society, pp.14298–14308.

Sarafijanovic-Djukic, N. and Davis, J. (2019) 'Fast distance-based anomaly detection in images using an inception-like autoencoder', in Birukou and Novak, K. (Eds.): *Discovery Science*, Springer International Publishing, Cham, pp.493–508.

Shen, Y., Yang, F., Habibullah, M.S., Ahmed, J., Das, A.K., Zhou, Y. and Ho, C.L. (2021) 'Predicting tool wear size across multi-cutting conditions using advanced machine learning techniques', *Journal of Intelligent Manufacturing*, Vol. 32, No. 6, pp.1753–1766.

Si, X-S., Wang, W., Hu, C-H. and Zhou, D-H. (2011) 'Remaining useful life estimation – a review on the statistical data driven approaches', *European Journal of Operational Research*, Vol. 213, No. 1, pp.1–14.

Spur, G., Eversheim, W., Sahm, P.R., Michaeli, W., Maier, H.R., Doege, E., Siegert, K., Schmoeckel, D., König, W., Klocke, F., Dorn, L., Steffens, H-D., Macherauch, E., Müller, H., Pollack, A., Merz, P., Philipp, M., Grefenstein, A., Polley, W., Stein, B., Fries, E., Krieg, G. and Sparrer, M. (1999) 'Produktionstechnologie', in Eversheim, W. and Schuh, G. (Eds.): *Produktion und Management 3: Gestaltung von Produktionssystemen*, Heidelberg, Springer Berlin, pp.247–479.

Tsai, C-Y. (2016) 'Mathematical model for design and analysis of power skiving tool for involute gear cutting', *Mechanism and Machine Theory*, Vol. 101, pp.195–208.

Van Rossum, G. and Drake, F.L. (2010) *The Python Language Reference*, 3rd ed., [Hampton, NH], [Redwood City, Calif.], Python Software Foundation; SoHo Books.

Villalonga, A., Negri, E., Biscardo, G., Castano, F., Haber, R.E., Fumagalli, L. and Macchi, M. (2021) 'A decision-making framework for dynamic scheduling of cyber-physical production systems based on digital twins', *Annual Reviews in Control*, Vol. 51, pp.357–373.

Wardhani, N.W.S., Rochayani, M.Y., Iriany, A., Sulistyono, A.D. and Lestantyo, P. (2019) 'Cross-validation metrics for evaluating classification performance on imbalanced data', *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 23–24 October, 2019, IEEE, Tangerang, Indonesia, pp.14–18.

Webster, J. and Watson, R.T. (2002) 'Analyzing the past to prepare for the future: writing a literature review', *MIS Quarterly*, Vol. 26, No. 2, pp.xiii–xxiii [Online], Available at http://www.jstor.org/stable/4132319

Wei, Q., Shi, B., Lo, J.Y., Carin, L., Ren, Y. and Hou, R. (2018) 'Anomaly detection for medical images based on a one-class classification', *Medical Imaging 2018: Computer-Aided Diagnosis*, 10–15 February, Houston, USA, SPIE, p.57.

Weiss, K., Khoshgoftaar, T.M. and Wang, D. (2016) 'A survey of transfer learning', *Journal of Big Data*, Vol. 3, No. 1, pp.1–40.

Xia, L., Shi, Y., Lin, H., Zheng, H., Cao, X., Chen, B., Zhou, Y. and Sun, W. (2024) 'Segmentation and quantitative evaluation for tool wear condition via an improved SE-U-Net', *The International Journal of Advanced Manufacturing Technology*, Vol. 132, Nos. 9–10, p.5173.

Xia, X., Pan, X., He, X., Zhang, J., Ding, N. and Ma, L. (2021) *Discriminative-Generative Representation Learning for One-Class Anomaly Detection* [Online], Available at http://arxiv.org/pdf/2107.12753.pdf

Yang, J., Xu, R., Qi, Z. and Shi, Y. (2022) 'Visual anomaly detection for images: a systematic survey', *Procedia Computer Science*, Vol. 199, pp.471–478.

Zagoruyko, S. and Komodakis, N. (2016) 'Wide residual networks', *Proceedings of the British Machine Vision Conference*, British Machine Vision Association, York, UK, pp.87.1–87.12.

Zamani, S.Z. (2022) 'Small and medium enterprises (SMEs) facing an evolving technological era: a systematic literature review on the adoption of technologies in SMEs', *European Journal of Innovation Management* [Online], DOI: 10.1108/EJIM-07-2021-0360.