Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-024-02476-2

HG-XAI: human-guided tool wear identification approach through augmentation of explainable artificial intelligence with machine vision

Aitha Sudheer Kumar[1] · Ankit Agarwal[2] · Vinita Gangaram Jansari[2] · K. A. Desai[1] · Chiranjoy Chattopadhyay[3] · Laine Mears[2]

Received: 8 November 2023 / Accepted: 29 July 2024

Abstract

Identifying tool wear state is essential for machine operators as it assists in informed decisions for timely tool replacement and subsequent machining operations. As each wear state corresponds to a unique mitigation strategy, timely identification is vital while implementing solutions to minimize tool wear. The paper presents a novel Human Guided-eXplainable Artificial Intelligence (HG-XAI) approach for identifying the tool wear state by integrating human intelligence and eXplainable AI with a pre-trained Convolutional Neural Network (CNN), Efficient-Net-b0 model. The tool wear states were identified based on different wear mechanisms during the machining of IN718. The study considers four distinct tool wear states, i.e., Flank, Flank+BUE, Flank+Face, and Chipping, representing abrasion, adhesion, diffusion, and fracture wear mechanisms. The image-based datasets were created to depict various tool wear states by machining IN718 at varying surface speeds. The effectiveness of the proposed HG-XAI approach was evaluated by comparing its prediction accuracy with a standalone Efficient-Net-b0 model lacking human intelligence and XAI. Further, the scalability of the HG-XAI approach was examined by predicting wear states from images acquired at different cutting parameters. The results from the present study showed that the HG-XAI approach can predict the tool wear state with an accuracy of 93.08% and is scalable to variations in cutting conditions. Also, the proposed approach can be extended while developing vision-based on-machine tool wear monitoring systems.

Keywords Tool wear · Convolutional Neural Networks (CNNs) · Explainability · Grad-CAM · Human intelligence

Introduction

Machining of Nickel-Based Superalloys (NBSAs) is characterized by rapid work hardening, low thermal diffusivity, and extreme abrasive behavior (Pleta et al. 2014). Tool wear progression is stochastically influenced in this case and varies considerably even under identical machining conditions (Potthoff et al. 2023a; Kumar et al. 2023). Such variation of tool wear can be attributed to the simultaneous occurrence of multiple wear mechanisms, including abrasion, adhesion, and diffusion (Thakur et al. 2009). It results in significantly different tool wear states such as flank wear, crater wear, excessive built-up edge formation, or chipping than conventionally reported in the literature (Sarıkaya et al. 2021). The existing numerical or computational models are

✉ K. A. Desai
kadesai@iitj.ac.in

Aitha Sudheer Kumar
kumar.118@iitj.ac.in

Ankit Agarwal
agarwa3@clemson.edu

Vinita Gangaram Jansari
vjansar@clemson.edu

Chiranjoy Chattopadhyay
chiranjoy.chattopadhyay@flame.edu.in

Laine Mears
mears@clemson.edu

[1] Department of Mechanical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan 342030, India

[2] International Center for Automotive Research, Clemson University, Greenville, SC 29607, USA

[3] School of Computing and Data Science, FLAME University, Pune, Maharashtra 412115, India

Published online: 14 August 2024

ineffective for tool wear state identification during machining NBSAs due to the variability in wear progression, presence of multiple wear states, complex wear geometry with curved portions, and intricate tool-workpiece interactions (Liang et al. 2019; Pimenov et al. 2023). The traditional wear monitoring solutions presented in the literature consider flank wear only and cannot be employed readily during the machining of NBSAs. Therefore, a unique monitoring and identification strategy is imperative to identify the tool wear states, guide the timely replacement, and avoid consequences such as poor surface finish, higher cutting forces, catastrophic tool failure, or damage to the machine tool. The periodic inspection of tool wear by humans or machine operators can be an alternative to predictive monitoring systems. It involves human judgments to identify wear patterns, dominating wear mechanisms, wear progression, and anomalies that might not be apparent from computational models.

Over the years, manufacturers relied on experienced human operators to identify and mitigate tool wear during machining operations. The perception and agility of experienced machine operators can achieve robust detection abilities despite variations in tool-work materials or cutting conditions (Yan et al. 2023). However, continuous monitoring involving humans is impractical as decisions are subjected to consistency, fatigue, distraction, and skillset variations (Zhang and Wang 2016). The shortage of experienced operators and retirements due to aging populations further restrict the availability of skilled human resources. In recent years, researchers have been conceptualizing resilient human-centric manufacturing systems that combine the agility of humans with the precision of computational models through Information and Communication Technologies (ICT) (Yao et al. 2022; Zhang and Van Luttervelt 2011). The present work attempts to augment computational models with human intelligence to predict and categorize tool wear states more effectively during the machining of NBSAs than standalone computational models.

Numerous methods have been proposed in the literature to monitor tool wear states. These methods can be broadly classified into indirect and direct approaches. The indirect approaches correlate process signatures such as forces (Shah et al. 2023), vibrations (Shen et al. 2021), acoustic emissions (Shen et al. 2021), temperature (He et al. 2021), and surface roughness (Potthoff et al. 2023b) with tool wear states. These approaches utilize computational models to determine tool wear-related parameters from the data accumulated through single or multiple sensors. However, the sensor data is susceptible to significant noise and outliers, which may yield poorly fitted correlations and lowered prediction accuracy of the models (Zhang and Zhang 2013). The other limitation of online tool wear monitoring solutions is the requirement of manual fine-tuning of threshold values by experts having an understanding of the wear phenomenon with variations in tools, work materials, or process parameters. It is challenging to ensure the onsite availability of such human resources for manual adjustments (Wright and Bourne 1988; Wang et al. 2016). Alternatively, direct approaches are considered more reliable and accurate as they observe tool conditions explicitly by capturing images using an optical sensor, camera, or microscope (Dai and Zhu 2018). The images are analyzed for changes in appearance, geometric shape, and surface properties of the cutting tool.

Recent advancements in machine vision technologies linked to image acquisition hardware, image processing techniques, and deep learning algorithms favor direct monitoring approaches. The image processing algorithms augmented with machine vision hardware can effectively capture cutting-edge images for measuring the width and area of flank wear (Peng et al. 2021; Agarwal et al. 2022) and depth of crater wear (Prasad and Ramamoorthy 2001). The tool wear state is classified as initial, intermediate, and worn-out based on the measured flank and crater wear values (Agarwal et al. 2022). The image enhancement operations minimize reflections, and image binarization detects wear area boundaries. The optimal parameters and techniques for both steps can vary on a case-to-case basis, limiting the generalizability of image processing algorithms (García-Pérez et al. 2023). These challenges are addressed by integrating deep learning algorithms with machine vision systems. Deep learning algorithms utilize labeled image datasets having cutting edges with variations of wear patterns for model training. Bergs et al. (2020) developed a Convolutional Neural Network (CNN) model using diverse cutting tool images captured with a microscopic camera followed by image enhancement, labeling, and training to classify tool types and detect the flank wear area. The cutting inserts during the turning operation are categorized as acceptable (OK) or unacceptable (NOK) using a trained Efficient-Net-b4 algorithm. The model achieved 97.8% accuracy with round inserts but lacked scalability to rhombic inserts (García-Pérez et al. 2023). Martínez-Arellano et al. (2019) encoded the time-series force signals into images using the Gramian Angular Summation Fields (GASF) technique and categorized wear phases of the end mill as break-in, steady, severe, and failure with a test accuracy of 80%.

The 'black box' nature of deep learning algorithms restricts utilities as the logic for reaching a particular decision is not provided, leading to a lack of trust among users. In the case of predictions with lower confidence, a computational tool explaining model predictions can be integrated with humans to enrich the decision-making. Zhang et al. (2018) presented a Function Context Behavior Principle State and Structure (FCBPSS) framework for product design problems to explain the learning mechanism and information abstraction through various layers of deep learning architecture. Further, the researchers developed eXplainable Artificial

Intelligence (XAI) tools to extract layer-specific information from the network architecture and identify the most influential features or factors leading to a particular decision (Lee et al. 2022). The additional information provided by XAI tools can be utilized by humans while making final predictions. Post-hoc explanation techniques such as Gradient Class Activation Maps (Grad-CAM) (Selvaraju et al. 2017), Local Interpretable Model-agnostic Explanation (LIME) (Ribeiro et al. 2016), and SHapely Additive exPlanation (SHAP) (Lundberg and Lee 2017) are developed to explain CNN model predictions. Among these techniques, Grad-CAM has been commonly employed with CNN models to provide visual explanations in the form of a feature/attention map. For example, Papenberg et al. (2023) used Grad-CAM to analyze the ability of CNN while identifying various wear regions from the input image of an end mill. The Grad-CAM results showed that wear on the tool face was significantly relevant for CNN compared to cutting-edge chipping. García-Pérez et al. (2023) generated attention maps using Grad-CAM for visualizing worn regions of turning inserts for evaluating the final classification decision of a CNN model. It was observed that the model focuses on the correct wear region for round inserts but fails to identify the same for rhombic inserts. The feature maps obtained using Grad-CAM can be used to determine inter-class probability scores from the Softmax layer of CNN architecture to show the confidence level in classification decisions (Zhao et al. 2023).

The review of previous literature related to tool wear monitoring using machine vision shows that the studies estimate tool wear parameters (flank wear width and area) or classify the wear state (i.e., initial/intermediate/worn-out) or condition (i.e., OK/NOK; worn/unworn). The literature does not include any study classifying tool wear states (i.e., flank, crater, built-up edge, or chipping) representing wear mechanisms during the machining of NBSAs. The present work utilizes a CNN-based image classification model using Efficient-Net-b0 to classify wear states commonly observed during the machining of NBSAs. The literature also highlighted lower prediction or classification accuracy despite advancements in machine vision technologies, including using XAI to understand model predictions. It is observed that the higher accuracy models also lack scalability and struggle to adapt to newer sets of images with differences in the image acquisition system, tool geometry, and tool-workpiece material combination. As inferences corresponding to the wear state prediction are unclear, correlation with the wear mechanism and deployment of a suitable mitigation strategy is challenging. Therefore, the present work proposes augmenting XAI as a support system for human decisions with a trained deep learning (CNN) model. HG-XAI algorithm enables machine operators or humans to efficiently determine tool wear states during the machining of NBSAs through an XAI-based decision-support system. Figure 1 shows the difference between traditional vision-based approaches presented in previous studies and HG-XAI developed in this work. Further, the study also recognizes the unavailability of a labeled image dataset representing wear states during the machining of NBSAs. Hence, a labeled dataset is also generated to capture wear state evolution from the beginning to the end of tool life during the end milling of Inconel 718 (IN718). Section 2 outlines four distinct tool wear states based on the wear mechanisms observed during the machining of IN718. The elements for developing the HG-XAI approach are outlined in Sect. 3 along with training of the Efficient-Net-b0 model. Section 4 compares the prediction abilities and scalability of the HG-XAI approach with the standalone Efficient-Net-b0 model. The summary of contributions from the present work is enumerated in Sect. 5.

Generation of labeled datasets

The prediction and generalization abilities of vision-based CNN models in identifying tool wear states depend on the quality and diversity of image datasets used during the training. Vision-based tool wear monitoring has been extensively studied for Stainless Steel (Zhang and Zhang 2013), Titanium (Hu et al. 2019), and IN718 alloys (García-Pérez et al. 2023) with numerous datasets publicly available for benchmarking the classification or regression models (Pan et al. 2023). The publicly available datasets are intended for classifying the wear zones or tool conditions, primarily based on the flank wear width. Also, these datasets are developed by performing the slot-cutting operation using linear toolpaths with a focus on monitoring the bottom cutting edge of an end mill. The present work considers the machining of IN718, which results in wear patterns repeating stochastically at a significantly higher rate than other metals (Potthoff et al. 2023a). Therefore, a trochoidal toolpath is commonly employed to reduce radial engagement, prolong the tool life, and allow higher depths of cut (Pleta et al. 2014). The wear phenomenon is analyzed for side-cutting edges due to their dominance during material removal action. As wear image datasets are unavailable for side cutting edges of an insert with a trochoidal toolpath, a newer case-specific dataset is required to identify tool wear states while machining IN718.

Tool wear states for IN718

ISO 8688-2 (1989) describes various states of tool wear evolution during the end milling operation. As the machining of IN718 is characterized by significant Built-Up Edge (BUE) formation and frequent cutting-edge chipping in addition to rapid flank wear, the definition of wear states requires appropriate modifications. The present work considers four commonly observed tool wear states associated with the

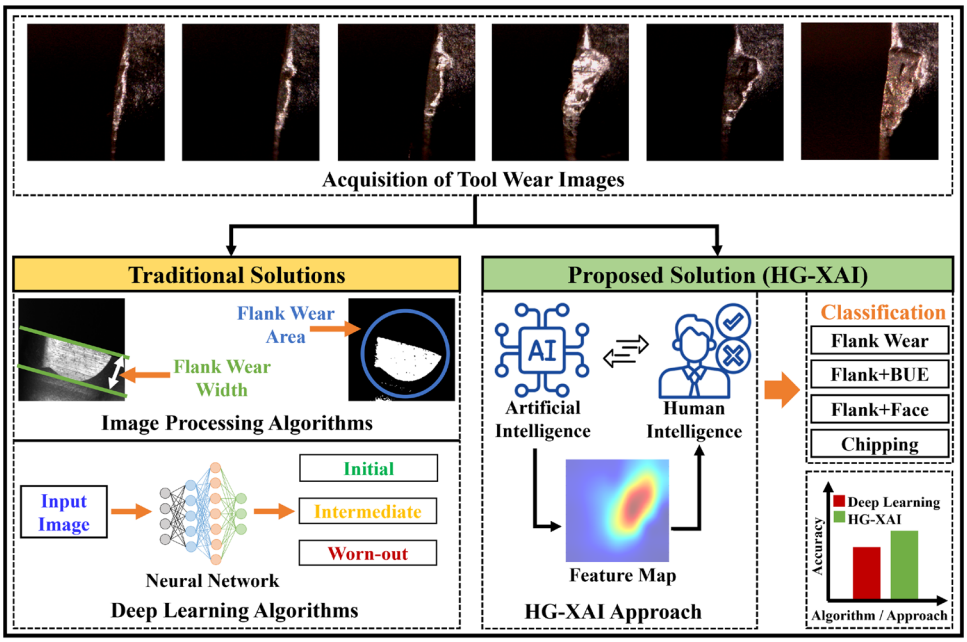


Fig. 1 Traditional and proposed solutions for monitoring tool wear during machining of NBSAs

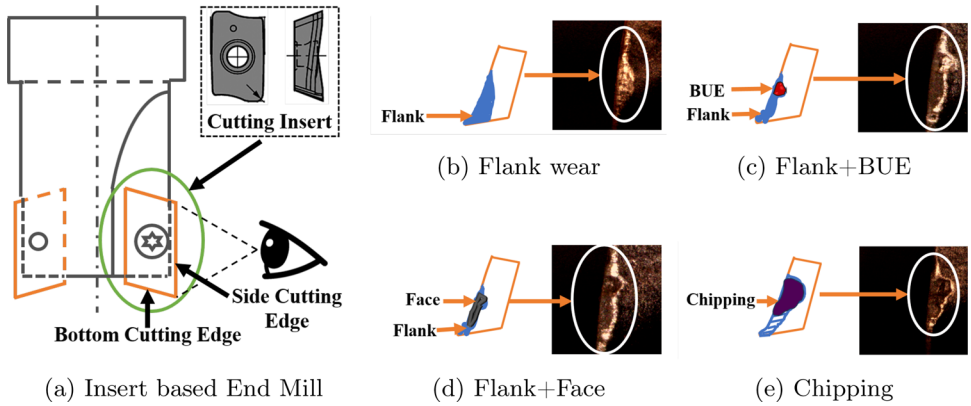
machining of IN718; Flank wear (Flank), Flank wear with the formation of BUE (Flank+BUE), Flank wear accompanied with deterioration of the rake face (Flank+Face) and Cutting edge chipping (Chipping). The flank wear is characterized by progressive wear of the cutting edge due to the rubbing action with the work material. The lower thermal conductivity and intense friction at the chip-tool interface result in elevated temperature and BUE formation on the wear land. The BUE eventually separates from the cutting edge with continued machining, resulting in a gradual material loss from the rake face. The harsh thermal cycles and suddenly increased cutting forces due to BUE formation or tool wear result in the cutting edge being chipped off. Figure 2 shows the schematic representation and actual wear state images along the side cutting edge of an end mill insert.

Machining experiments

The trochoidal toolpath experiments were conducted on a 3-axis vertical milling machine (OKUMA Genos M560-V) using a 2-flute indexable end mill with carbide inserts and IN718 workpiece material. In the present study, the machining experiments were performed at three surface speed values: 25, 45, and 60 m/min, while the other parameters, such as feed per tooth, step depth, and type of tool path, were kept constant. These machining parameters were selected based on the cutting tool manufacturer's recommendations (Coromant 2023). Table 1 outlines the machining conditions, cutting tool specifications, and toolpath parameters used during the experiments for generating different tool wear states during the machining of IN718. The 40 mm width x 1 mm height slots were machined during the test, and the number of passes completed before the tool failure was recorded. The volume of material removed in each pass was 842.4 mm^3. The experimental setup is shown in Fig. 3. The cutting tool was removed from the machine spindle after each pass, and images of side cutting edges depicting tool wear were captured. The subsequent section explains the image acquisition setup and criteria followed for labeling various wear states.

Image acquisition and labeling

The high-resolution microscopic images of the worn area on a side cutting edge for both inserts were captured offline using a Dino-Lite digital microscope arrangement depicted in Fig. 4. The Dino-Lite microscope has a compact and lightweight construction, compatible with most operating systems, and can be operated remotely using wired or wireless connections through the in-built software. The Dino-Lite microscope was preferred as subsequent study plans to integrate the proposed algorithm with a vision-based on-machine wear state detection system. The tool was placed horizontally with the side flank area under the view of a vertically mounted lens arrangement in a controlled lighting environment, as shown in Fig. 4a. A ring-type LED lightning system was used to adjust the illumination intensity by controlling four-quarter sections of a ring. Two images were captured after each pass corresponding to both inserts of a cutting tool. The images were captured at a resolution of 2592 x 1944 and a magnification factor of 200x offers a field of view of 1–2 mm, eliminating the need for a separate Region Of Interest (ROI) extraction. The greater Depth of Field (DOF) combines multiple images acquired at different focus distances to effectively capture depth features like crater and chipping



(a) Insert based End Mill (b) Flank wear (c) Flank+BUE (d) Flank+Face (e) Chipping

Fig. 2 Schematic and actual images of tool wear states

Table 1 Design of experiment

Test No.	Surface speed (m/min)	Passes completed before tool failure (n)	Insert (i)	Images captured (n*i)
1	25	30	2	60
2	45	14		28
3	60	10		20
Trochoidal step size		: 1.5 mm		
Axial depth of cut		: 1 mm		
Feed per tooth		: 0.25 mm/tooth		
Slot width		: 40 mm		
Volume removed in each pass		: 842.4 mm^3		
Cutting tool diameter		: 15.875 mm		
Helix angle		: 30°		

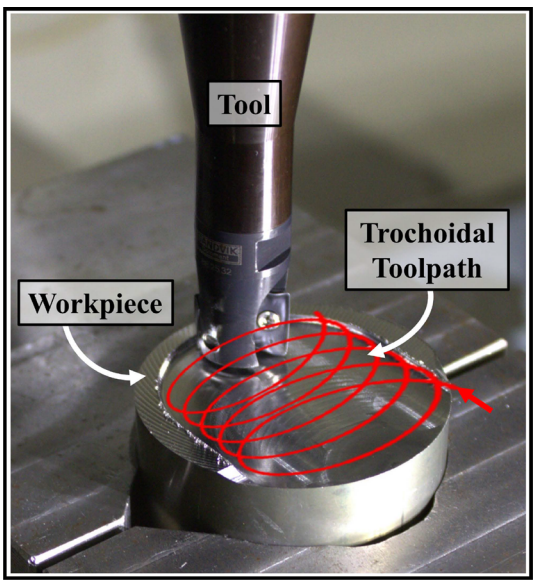


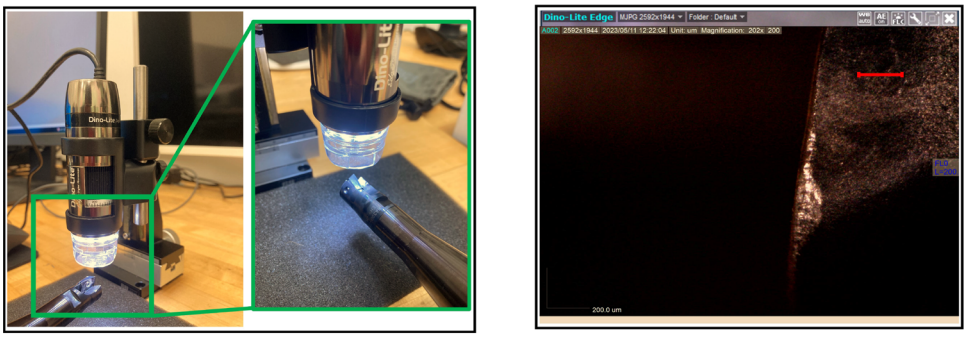
Fig. 3 Machining setup for tool wear experiments

in the final image. Figure 4b shows one of the sample images acquired using the camera setup at these parameters.

The dataset with a total of 108 (60 + 28 + 20) images was generated by performing experiments at three surface speeds of 25, 45, and 60 m/min. The Region Of Interest (ROI) was subsequently extracted by cropping each captured image to a size of 800 x 800 pixels to depict the wear area. The dataset was labeled by involving several human experts with an understanding of ISO 8688-2 tool wear standards and several years of experience in machining, as detailed in Sect. 2.1. Further, *'tool not worn'* category was not considered due to the higher initial tool wear rate during machining. It has been observed that the flank wear increases exponentially during the initial machining period using a new tool (Agarwal et al. 2022). Figure 5 shows labeled image datasets for one of the inserts generated after each pass during the machining of IN718 at the surface speed of 25 m/min. It can be seen that the amount of progressive flank wear increases initially (pass 1–4). A series of flank+BUE and flank+face wear stages are apparent during subsequent passes (pass 5–18), primarily due to the formation of BUE and separation from the insert surface subsequently. Finally, cutting-edge chipping is observed from passes 19–30 due to the material loss from the insert surface.

Data augmentation

The training of CNN-based classification models requires a large amount of datasets. It necessitates removing a



(a) Dino-Lite microscope and LED lighting (b) Captured image of worn area

Fig. 4 Offline image acquisition module

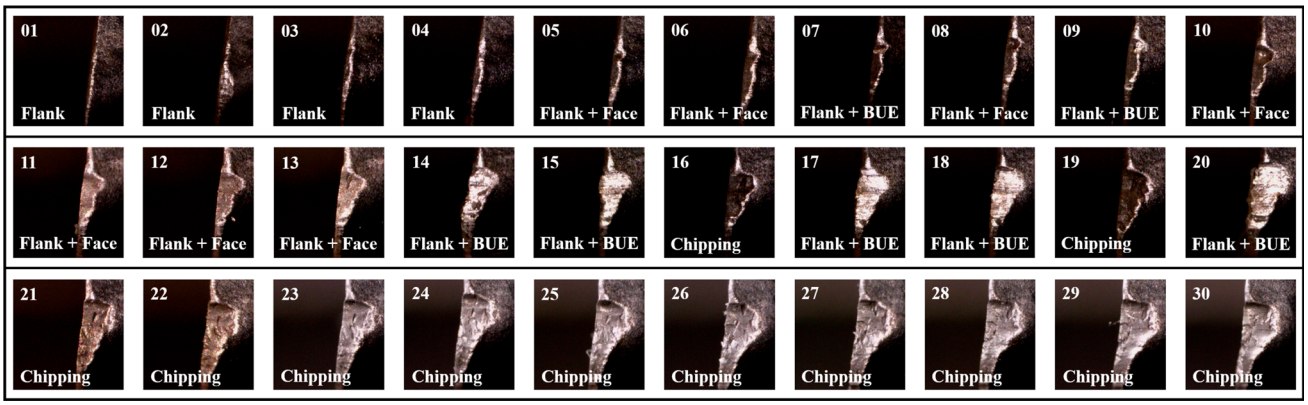


Fig. 5 Labeled image dataset for one sample insert at surface speed of 25 m/min

considerable volume of material by performing numerous experiments and acquiring many images to produce sufficient datasets representing all tool wear states. It can be time-consuming and cost-intensive during IN718 machining due to higher unit material removal costs. Hence, meeting the labeled dataset requirement through end-milling experiments is not practical, and employing data augmentation techniques can be an effective alternative. The present work utilizes various image augmentation techniques, such as horizontal and vertical motion blur (Nath et al. 2023), flipping, rotation, Gaussian Noise (Mandelbrot 1971), Salt & Pepper Noise (Azzeh et al. 2018), Speckle Noise (López-Martínez and Fabregas 2003) that augment and expand the image datasets as shown in Fig. 6. The captured images were blurred horizontally and vertically using a kernel size of 3x3 and 5x5. The images were rotated at the step angle of 10°, and flipped horizontally as well as vertically. Additionally, the images were exposed to artificial noises to resemble the potential variations encountered during an on-machine image acquisition integrated with the milling machine. Gaussian noise was introduced with a mean of zero and a standard deviation of 0.01 (Moghaddam and Jamzad 2007). Similarly, Salt & Pepper and Speckle noise were added to the image with a 2% and 10% noise intensity. Upon augmentation, the originally captured dataset of 60 images with a surface speed of 25 m/min was expanded to 4,400 images, comprising 1,100 images in each tool wear class. The image augmentation operations were not performed on labeled image datasets for the surface speeds of 45 and 60 m/min. These datasets were developed for benchmarking the model performance under conditions similar to the practical implementation and assessing the scalability.

Human-Guided-XAI (HG-XAI) approach

Several CNN-based models have been developed and presented in the literature to perform real-time classification tasks on the manufacturing shop floor. In the previous work, co-authors comprehensively evaluated various deep-learning architectures for surface defect detection of machined components (Singh et al. 2023). The Efficient-Net-b0 model outperformed on evaluation metrics, such as accuracy, precision, and recall during the evaluation. Based on this study, the present work considered Efficient-Net-b0 as a baseline model for developing the HG-XAI approach to identify tool wear states. Efficient-Net-b0 is an effective CNN model demonstrating robust computational abilities with better generalization and prediction accuracy during image-based classification tasks (Tan and Le 2019). Efficient-Net-b0 com-

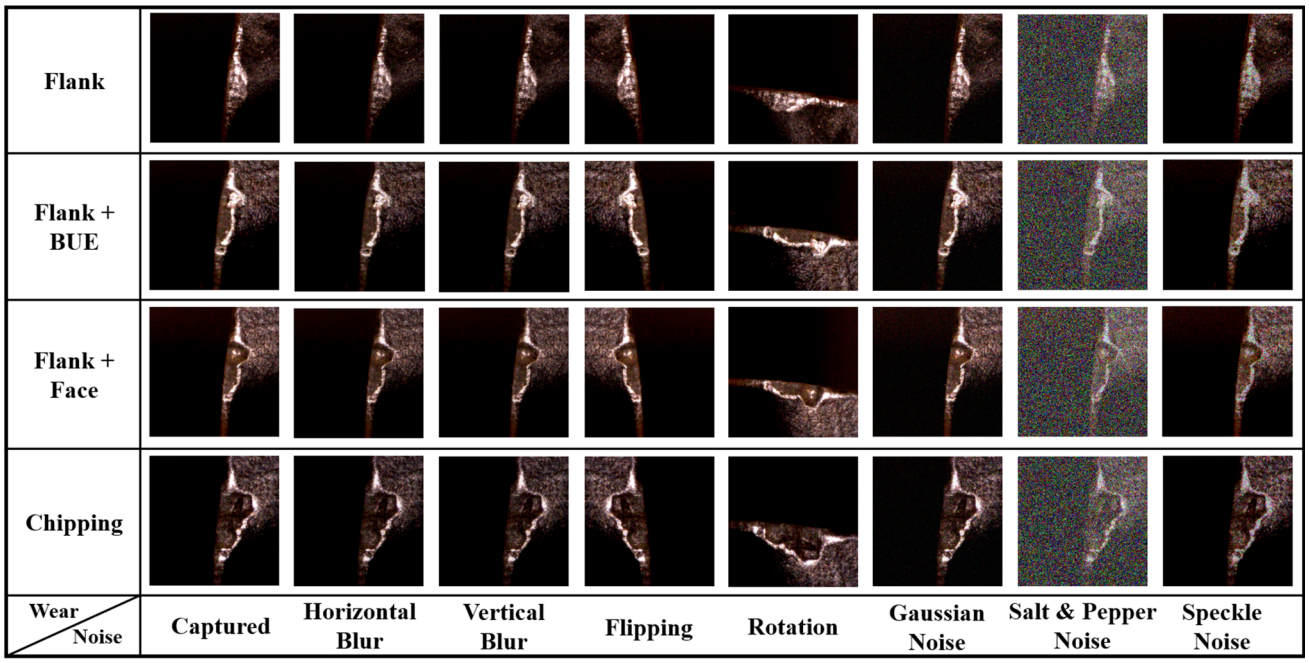


Fig. 6 Data augmentation for different categories of tool wear

prises convolutional and pooling layers culminating in a Fully connected layer as shown in Fig. 7. The convolutional layer detects various local features in an image, while a pooling layer merges semantically similar features. The model processes input and generates output as arrays or feature maps. The feature maps from the last convolutional layer highlight the regions of an input image that contribute the most to the classification results. The fully connected layer computes the weighted sum of array information from the last convolutional layer, which generates the output classification category by analyzing relationships between image features and respective classes. Finally, the Softmax layer normalizes raw scores from the preceding layer and performs elemental operations to estimate probability distribution over all categories. The class with the highest probability score is chosen as the classification result. The feature maps can also be represented using Grad-CAM to understand the classification results further (refer to Fig. 7).

It is effective to associate humans in the decision-making process as CNN models cannot address inherent process variations and scalability issues. Considering the challenges associated with CNN models in classifying images containing multiple wear state features while machining IN718, augmenting humans in the decision-making can be effective. The HG-XAI involves human intelligence with additional inferences provided by the CNN model in making subtle decisions for images with multiple wear conditions or images categorized with lower probability scores. HG-XAI requires a trained Efficient-Net-b0 model, integrating the Grad-CAM algorithm with the last convolutional layer for generating feature maps, prediction probabilities for each class from the Softmax layer, and threshold level (P_{max}) for bifurcating the decision-making process. HG-XAI uses the P_{max} score or probability of predicted class as a threshold for determining whether a computational model or humans will make predictions from the image. The output of the Efficient-Net-b0 will be accepted as a final decision for images classified with P_{max} score greater than the defined threshold level. The images with a prediction probability smaller than P_{max} will be sent to humans for decision-making. The feature information is extracted from the last convolutional layer using the Grad-CAM algorithm for these images and provided to the human. HG-XAI approach displays feature maps along with captured images for cases with probability less than the threshold value (P_{max}). The comparison of the feature map and captured image enables humans to appreciate algorithm errors and predict the correct wear class. It is assumed that humans have sufficient skill sets or are provided with adequate training to identify tool wear states correctly. The overall flow of information while implementing the HG-XAI algorithm is shown in Fig. 7.

Efficient-Net-b0 model

Tan and Le (2019) presented Efficient-Net-b0 architecture with a rethinking model-based scaling at the International Conference on Machine Learning (ICML) in 2019. It uses the compound coefficient technique for network scaling and showed exceptional performance for image classification tasks in various domains (Papenberg et al. 2023; Singh et

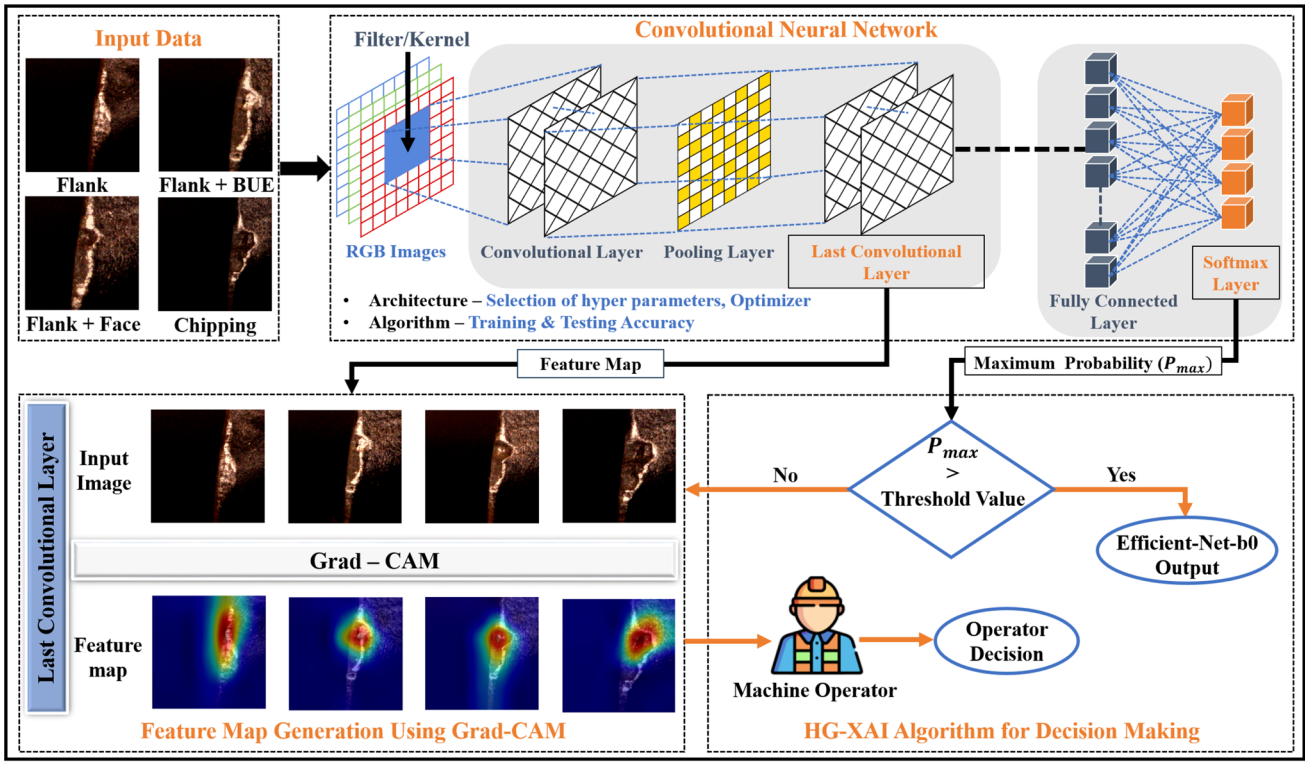


Fig. 7 Elements of HG-XAI tool wear state identification framework

al. 2023). The network architecture consisting of various layers is shown in the top portion of Fig. 7. The training of Efficient-Net-b0 for classifying tool wear states during the machining of IN718 is the first step in developing the HG-XAI framework. The model training is accomplished using an augmented image dataset generated by performing end-milling experiments at the surface speed of 25 m/min. The prediction accuracy and generalization abilities of the Efficient-Net-b0 model are critical in determining the number of images correctly classified with higher prediction probabilities. The model also determines the number of images sent to humans for decision-making.

The optimum performance of Efficient-Net-b0 during the training can be achieved by selecting appropriate hyperparameters such as learning rate, number of epochs, batch size, training/validation split ratio, and validation frequency. The optimizer selection is another critical element in attaining superior prediction abilities for image-based CNN models. The optimizer utilizes adaptive learning rates to adjust the step size dynamically, leading to faster convergence and efficient model training (Cheng et al. 2023). The present work considers three typical optimizers, Root Mean Squared Propagation (RMSProp), Adaptive Moment Estimation (Adam), and Stochastic Gradient Descent Momentum (SGDM) for CNN model training. RMSProp employs an exponentially weighted moving average to adjust gradient accumulation adaptively (Xu et al. 2021) and quick convergence for problems with convex solution space. Adam is a variant of the RMSProp combining momentum and bias correction for achieving robust and efficient prediction abilities. Adam overcomes the slower convergence issue of RMSProp by introducing momentum variable (Zaheer and Shaziya 2019). SGDM is another optimizer preferred for image classification tasks with small image datasets. SGDM is less sensitive to most hyperparameter variations and requires fine-tuning the momentum and learning rate (Ye et al. 2018). The hyperparameters are varied in the following range for achieving desired performance of the Efficient-Net-b0 model (Wilson and Martinez 2003); learning rate values—0.01, 0.005, and 0.001; epochs—30, 40, and 50; batch size - 32, 64, and 128; training/validation split ratio—70/30, 80/20, and 90/10; validation frequency—10, 25, and 50; and momentum—0.9.

Training and optimizer selection

The training of the Efficient-Net-b0 model was conducted using an augmented image dataset generated at a surface speed of 25 m/min. The augmented dataset contains 4400 images with 1100 images in each wear class. From the augmented dataset, 4000 images were utilized during training, and 400 images were reserved for the model testing. The details of computational resources used during the

Table 2 Computational resources for CNN training

Software Platform	:	MATLAB - 2021b
CPU Processor	:	Intel i9 Processor
Operating System	:	Windows 10 Pro - 64-bit
RAM	:	64 GB

model training are summarized in Table 2. The experimental trials were performed to achieve consistent model performance. The satisfactory model performance was obtained at the following hyperparameters: learning rate—0.01; batch size—32; training/validation split ratio—80/20; validation frequency—50 iterations. The hyperparameters were not changed while evaluating the optimizers. The accuracy scores and the corresponding hyperparameters for all three optimizers are detailed in Table 3. It can be seen that the training accuracy achieved using all three optimizers is identical.

The validation dataset consists of 800 images (20% of 4000 for training), with 200 images in each tool wear class. The dataset was utilized for the performance assessment of the model and to select an appropriate optimizer while classifying previously unseen images. The classification results for three optimizers are compared using confusion matrices shown in Fig. 8. The confusion matrix compares predictions of a classification model with the ground truth or the actual class. The entries along the diagonal are correct predictions, while off-diagonal entries show misclassifications. It can be seen that misclassifications were obtained during the validation experiments for all the optimizers with RMSProp showing the highest number of misclassifications. The misclassifications were reduced using the Adam optimizer, and the best performance was obtained using the SGDM optimizer. It is challenging to comment on the prediction abilities of the optimizers as the variations were marginal in results. The classification accuracy of all three optimizers was consistent across the Flank, Flank+BUE, and Chipping wear categories. However, misclassifications in the Flank+Face wear category for each optimizer necessitated further investigations. The inferences about CNN model predictions can be obtained using various tools offering explainability for the results. The present work employed the Grad-CAM approach to understand CNN model predictions (Selvaraju et al. 2017).

The Grad-CAM algorithm was integrated with the Efficient-Net-b0 model to analyze reasons for the consistent misclassifications in the Flank+Face wear class and evaluate a suitable optimizer for the model developed herein. The integration facilitates the identification of dominant features and patterns in the image used for decision-making. A tool wear image was input to the trained Efficient-Net-b0 model, and feature maps in the last convolutional layer were used to generate a gradient feature map representing specific regions in the gradient color varying from red to blue. The red represents the dominant features, while the blue represents the least important features in the decision-making by the model. These feature maps are stored using a transparent image format and combined with the tool wear image to identify the dominant features. The stepwise process of implementing the Grad-CAM algorithm with the CNN model is shown in Fig. 9.

A dataset of 24 distinct images (6 images in each tool wear class) was selected randomly, and their respective feature maps were generated using all three optimizers, as shown in Fig. 10. The first set of columns represents the input image dataset, and the subsequent sets represent feature map information used by RMSProp, Adam, and SGDM optimizers. It can be seen that the features used for arriving at the Efficient-Net-b0 model output are noticeably inconsistent while using RMSProp and Adam optimizers when compared with input images. It can be seen that many predictions are made using random features in an input image, and the algorithm fails to detect appropriate features. Nevertheless, the SGDM optimizer predictions are consistent with input image features for all 24 cases. This further substantiates results obtained in Fig. 8, where the least misclassifications were observed using the SGDM optimizer. It can be concluded that the explainability using the Grad-CAM algorithm can be pivotal for optimizer selection and help better generalize the Efficient-Net-b0 model for tool wear classification tasks. The Efficient-Net-b0 model trained with the SGDM optimizer was used further for integration with the HG-XAI framework.

Implementation of HG-XAI

It is shown in the literature that the wear patterns during the machining of IN718 are stochastically influenced and vary considerably even under identical machining conditions (Potthoff et al. 2023a). CNN models such as Efficient-Net-b0 cannot address these issues effectively due to the shortage of training datasets and the presence of features with different wear types in a single image. It is also shown that these models are not scalable and prediction accuracy is lowered considerably at conditions different than training (García-Pérez et al. 2023). The HG-XAI approach shown in Fig. 7 introduces humans in the decision-making to identify wear states accurately for maximizing machining efficiency and productivity. The test image dataset was provided as input to the trained Efficient-Net-b0 model with the SGDM optimizer discussed in the previous subsection. HG-XAI approach extracts the probability values for the predicted class P_{max} from the Softmax layer of a trained model for an input image. The feature map information is extracted from the last convolutional layer for images with P_{max} values lower than the

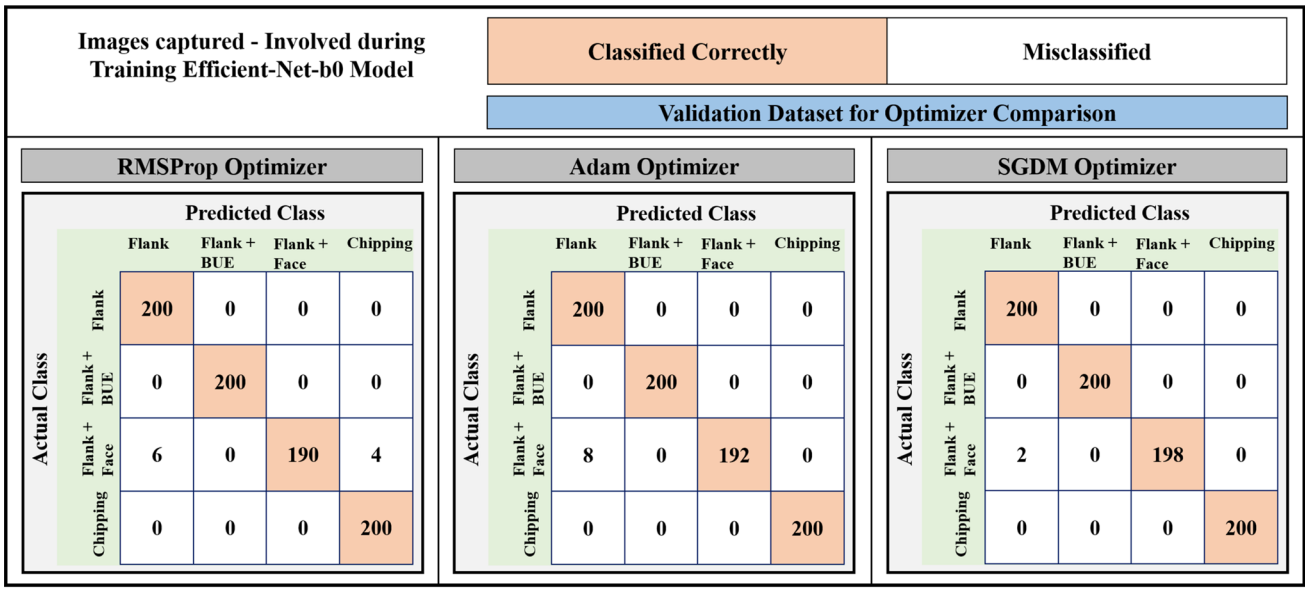


Fig. 8 Confusion matrices for the three optimizers

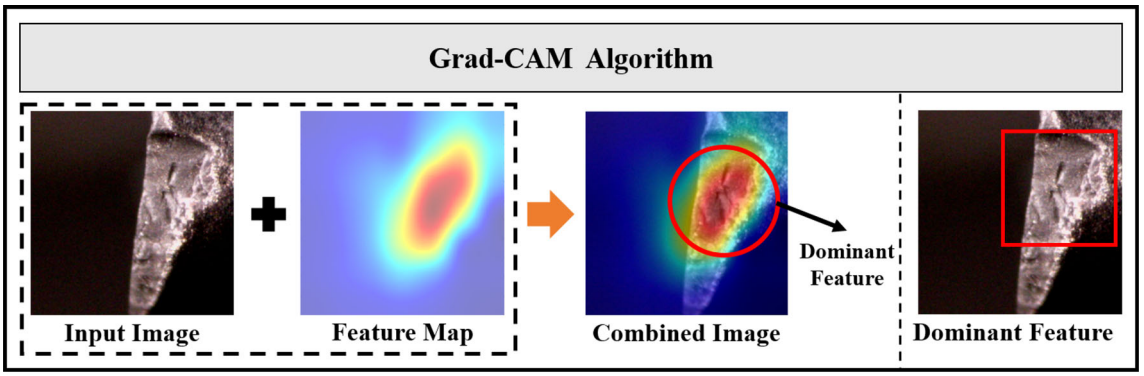


Fig. 9 Generating Gradient Class Activation Map (Grad-CAM) for an input image

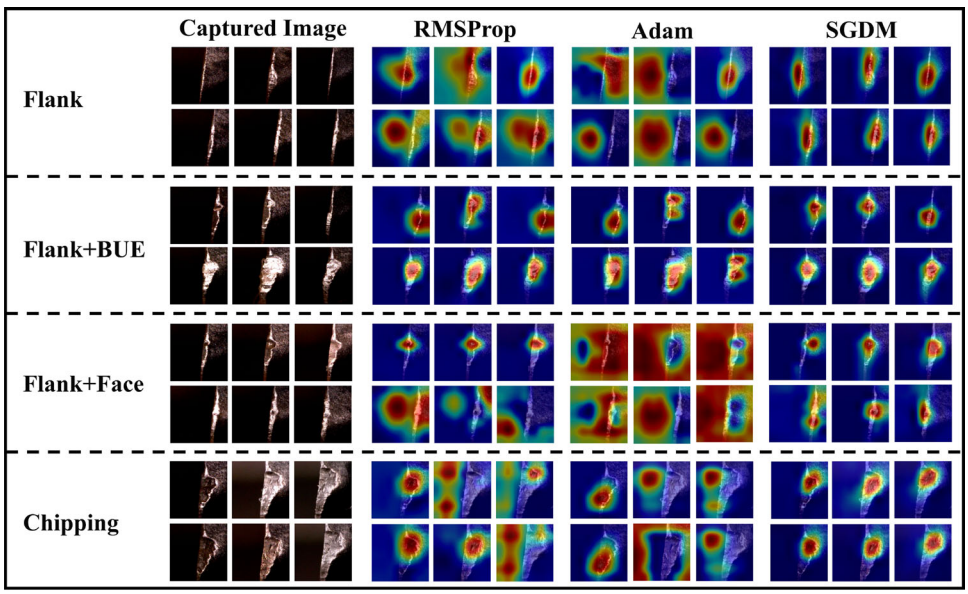


Fig. 10 Comparison of feature maps for optimizers

Table 3 Comparison of training accuracy with hyperparameters and optimizers

Optimizer	Learning rate	Epochs	Batch size	Split ratio	Validation frequency	Training accuracy
RMSProp	0.01	30	32	80/20	50	98.88%
Adam	0.01	30	32	80/20	50	98.59%
SGDM	0.01	30	32	80/20	50	99.30%

threshold value. The extracted information in the form of a feature map and an input image is provided to humans.

The threshold value has to be selected such that the number of images sent to humans for decision-making is limited. If the number of decisions made by humans is large, the cognitive load on the human increases, and the advantages of combining humans with the computational algorithm are offset. After considering the threshold value as P_{max}, the results are bifurcated into two classes; predictions having probability scores greater and smaller than the threshold level. The predictions with a probability score greater than the threshold level are accepted as model predictions while the remaining results are sent to the HG-XAI for subsequent actions. The feature maps for remaining predictions are generated and provided to the humans with captured images for generating the decisions.

Analysis and discussions

The trained Efficient-Net-b0 model with SGDM optimizer was used to demonstrate the effectiveness of the HG-XAI approach using image datasets generated from end milling experiments at 25, 45, and 60 m/min surface speeds. As mentioned in Sect. 3.2, 400 images from the augmented dataset with 100 images in each wear class was reserved as previously unseen test dataset under similar cutting conditions. The feature characteristics of these images are similar to the training dataset as these are generated from image augmentation techniques. The images corresponding to machining experiments at 45 m/min and 60 m/min surface speeds were not used during the training or augmentation as these datasets are planned for evaluating the practical implementation and scalability of the HG-XAI approach. The number of images was fewer in both these cases compared to the 25 m/min experiment, as the cutting tool wears out faster, or cutting edges are chipped off due to higher surface speeds. It can be seen that either no or very few results are recorded in the Flank+BUE and Flank+Face categories. The results are consistent with the process physics as BUE formation in the case of higher cutting speed is short-lived and may not get captured during the experiments. Therefore, no cutting images were recorded in the Flank+BUE category for 60 m/min surface speed experiments.

Evaluation of HG-XAI framework

All test images were input to the trained Efficient-Net-b0 model without integrating the HG-XAI approach and predicted classes were obtained. The trained Efficient-Net-b0 was considered a baseline model and results for predicted classes are represented using confusion matrices shown in Fig. 11. It can be seen that the model performance is excellent while classifying images within the Flank, Flank+BUE, and Chipping wear for surface speed of 25 m/min. However, 39 misclassifications were observed within the Flank+Face wear class. These results are consistent with the outcomes obtained during the model training and validation. The model could predict the Flank wear class accurately for images acquired from experiments even at surface speeds of 45 and 60 m/min. However, the performance substantially degraded within the other three classes resulting in lower overall prediction accuracy. The lower prediction abilities in these classes are an indicator of the poor scalability of the Efficient-Net-b0 model.

Figure 12 shows a few misclassification results obtained by the Efficient-Net-b0 model for various tool wear classes. It can be seen that the false classifications are due to the presence of features representing multiple wear classes in the same image. The trained model could not identify multiple features from images due to a lack of training. As images within the Flank+BUE and Flank+Face categories exhibit such characteristics commonly, the prediction accuracy within these two classes is quite poor. For instance, features in Fig. 12a–c resemble a combination of Flank and Flank+Face, with different feature level information. Similarly, Fig. 12d, h contain features combining Flank+BUE and Flank wear, Fig. 12e, i contain features of both Flank+Face and Flank+BUE wear, Fig. 12f exhibit combined features of Chipping and Flank+Face wear and, Fig. 12g contain Chipping and Flank wear features. Further investigations revealed that some of these images are categorized with lower probability values for the predicted class. Although the probability value is less for the predicted class, CNNs will categorize the image based on the highest value. The other wear classes are not considered due to lower probabilities. It is challenging to categorize such tool wear cases with features belonging to multiple classes in an image using the Efficient-Net-b0 model due to limitations such as the non-availability of diverse training datasets. Alternatively, skilled humans can perceive the

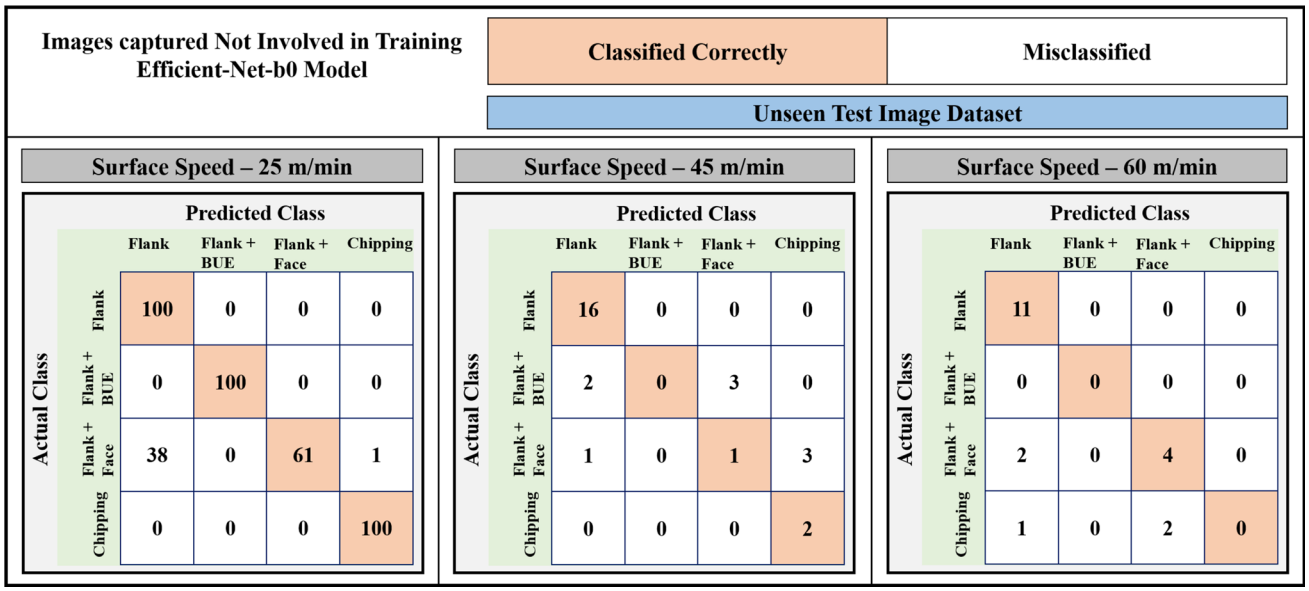


Fig. 11 Confusion matrices for test datasets obtained using Efficient-Net-b0 model (without integrating HG-XAI)

wear state more effectively by observing a cutting tool or an image due to inherent cognition, intuition, prior experience of variability, and adaptability to change based on the understanding of physics.

The trained Efficient-Net-b0 model was integrated with HG-XAI subsequently, and the same test datasets were input for obtaining predicted classes. The threshold value for the P_{max} was set as 0.85 by conducting trial experiments at different values in the range of 0.75−0.95. After considering the threshold value of 0.85, the predicted classes were obtained by combining Efficient-Net-b0 model outputs and human decisions. The predictions with a probability score higher than 0.85 were accepted as the final predicted class. The remaining results with a probability score less than 0.85 were input to the HG-XAI. The feature maps for these predictions were generated and provided to the humans along with captured images for the final prediction. The results from the Efficient-Net-b0 model and humans were combined and compared with the ground truths or actual classes to derive confusion matrices for all three surface speeds as presented in Fig. 13.

The comparison of results in Fig. 11 with Fig. 13 shows that the implementation of HG-XAI significantly improved the overall prediction accuracy of the model. The overall prediction accuracy (combining all three surface speeds) improved from 88.16% (with baseline Efficient-Net-b0) to 93.08% after integrating the HG-XAI approach in the decision-making. Based on these results, it can be concluded that better prediction abilities can be achieved by combining humans and CNN models in the decision-making. The prediction accuracy for images corresponding to 25 m/min surface speed (representing conditions identical to the model training) improved from 90.25% (with baseline Efficient-Net-b0) to 93.5% after integrating the HG-XAI approach. However, the improvements in the case of 45 and 60 m/min surface speed were substantial, from 67.85% and 75% to 85.71% and 95% respectively after integrating the HG-XAI approach. The improved classification accuracy of the Flank+Face category with HG-XAI approach can be attributed primarily to the following factors: In certain images, the overlapping characteristics of Flank+Face and Flank led to a situation where the correct class (Flank+Face) was assigned a lower probability score, while the incorrect classification (Flank) received a marginally higher score. These images, along with feature map information, were provided to the machine operator for decision-making. The conditions corresponding to 45 and 60 m/min were chosen to assess the practical implementation and scalability of the approach. The improved prediction accuracy indicates the robust prediction abilities of the HG-XAI compared to the baseline Efficient-Net-b0 model.

Discussions and subsequent work

The Efficient-Net-b0 model predictions with the integration of the HG-XAI were further analyzed to validate the robustness of the proposed approach in tool wear state monitoring during end milling of IN718. Figure 14 shows sample images with threshold level P_{max} less than 0.85 and sent to humans for decision-making along with feature maps. It can be seen that the images classified with probability values less than 0.85 contain features corresponding to multiple wear classes. The comparison of captured images and feature maps clearly shows that the Efficient-Net-b0 model cannot extract features

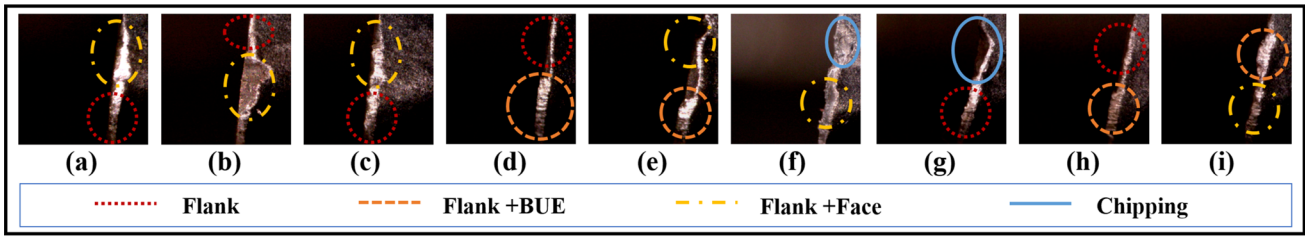


Fig. 12 Misclassified images showing multiple tool wear features

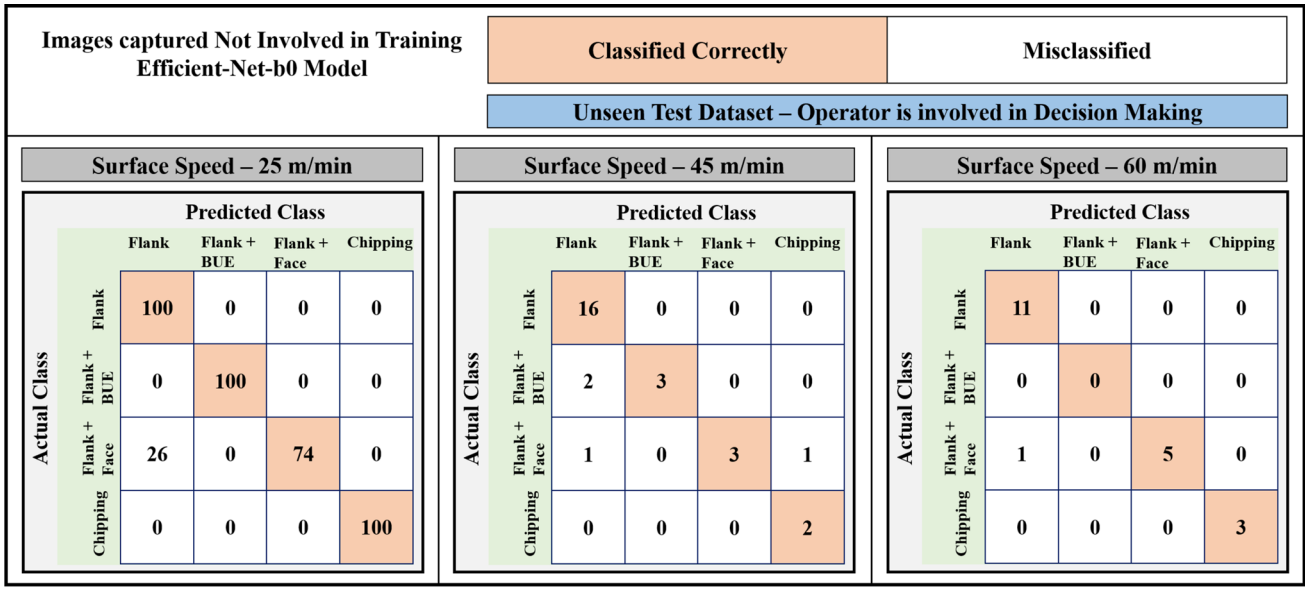


Fig. 13 Confusion matrices for test datasets obtained using Efficient-Net-b0 model (integrating HG-XAI)

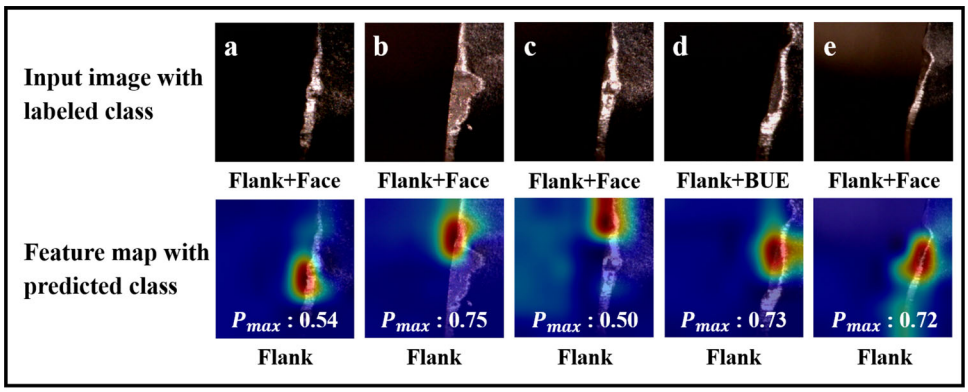


Fig. 14 Images with P_{max} less than 0.85 input to the HG-XAI

easily discernible to humans. Therefore, human involvement in decision-making resulted in better prediction abilities. However, performance improvement with human involvement was less than the expected level, which warranted further investigations.

The individual misclassification results for each surface speed and tool wear class were also analyzed to investigate reasons for the inability of the HG-XAI approach to predict correctly for these cases. For example, 26 Flank+Face wear class images at 25 m/min surface speed (refer Fig. 11) were misclassified as Flank wear. Similarly, the images within the Flank+Face and Flank+BUE classes were wrongly classified as Flank wear. All these images were categorized with the predicted class probability of more than 0.85 threshold value and not filtered by the HG-XAI for decision-making by the humans. Figure 15a–e show some of these images and feature maps having a prediction probability of more than 0.85 but are misclassified by the Efficient-Net-b0 model. It can be

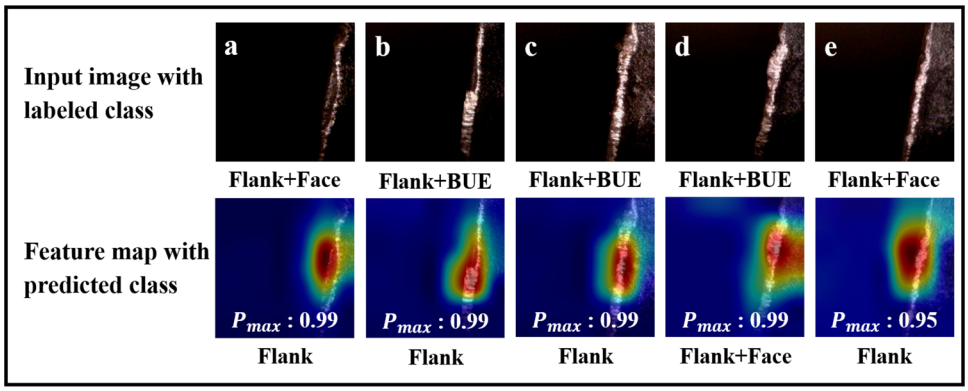


Fig. 15 Misclassified images with high probability score

seen that these images contain multiple features in an image representing Flank, Face, BUE, and a combination thereof in varied proportions. The feature map shows that the Efficient-Net-b0 model identified the most dominant feature based on training and predicted the class having higher probability. As seen in Fig. 15, non-dominating features were identified with lower prediction probabilities and were not referred to humans by the HG-XAI. Also, tool wear is a progressive phenomenon with continuous BUE formation and detachment from the edge, resulting in the wearing of the rake face. The classification approach needs to capture continuously varying wear regions of Flank, Face, and BUE in an image depending on the surface speed and other conditions during the machining of IN718.

The prediction abilities of HG-XAI could have been better if the images with the characteristics described above were transferred to humans for decisions. Humans can effectively categorize such images as the history of wear progression is remembered as patterns rather than a single image at a time being used by CNN models. The presence of multiple features with varying information levels in an image is another reason for Efficient-Net-b0 assigning higher weights to the dominant feature (refer to feature maps in Fig. 15). The classification scheme presented in this work is based on a dominating feature only. Developing a ranking-based algorithm for deciding the classes or transferring decisions to humans can be a subsequent work. The images for scalability experiments were captured offline by removing the cutting tool from the machine spindle. The subsequent studies focus on integrating the image acquisition setup with the CNC milling machine for image-based online monitoring of tool wear states. The FCBPSS framework (Zhang et al. 2018) is presented in the literature as an alternative to the XAI for understanding the working and predictions of a deep learning model. The subsequent work can develop the FCBPSS-based model to identify the tool wear state and compare the performance with the HG-XAI approach.

Conclusions

The paper introduced a Human-guided classification approach (HG-XAI) based on the Efficient-Net-b0 model to categorize tool wear states during the end milling of IN718. The proposed model considers augmenting human intelligence with the Efficient-Net-b0 model to achieve better classification accuracy. The results of the proposed approach are compared with the standalone Efficient-Net-b0 model that does not consider human guidance while deciding the prediction class. The major conclusions from the study are summarized as follows;

- HG-XAI approach presented in this paper can effectively classify the tool wear state as Flank, Flank+BUE, Flank+Face, and Chipping during end milling of IN718 with a prediction accuracy of 93.08%. The accuracy of the proposed approach is substantiated by predicting wear classes for previously unseen test datasets. The good agreement between the predicted and actual class substantiates the utility of the proposed approach.
- The comparison of the HG-XAI approach and standalone Efficient-Net-b0 model without human guidance was also carried out to assess scalability by considering image datasets having surface speeds significantly different than model training conditions. The studies showed improvement in the prediction accuracy from 67.85% and 75% to 85.71% and 95% for representative conditions of 45 and 60 m/min surface speeds. The detailed analysis of misclassified cases shows that the HG-XAI can be improved further by effectively segregating images with multiple wear states and transferring them for human judgment.
- The study performed iterative experiments to generate an image-based dataset depicting the evolution of the wear state from the initial to the failure stage of the tool during the milling of IN718 using a trochoidal toolpath. The datasets have been created for different surface speeds varying in the range of 25–60 m/min.

Acknowledgements The authors would also like to thank the Ministry of Education (MoE), India, for providing financial support as a student fellowship to Mr. Aitha Sudheer Kumar for carrying out this research work. The authors would also like to thank Mr. Karan Bhuta for his help in data collection.

Author Contributions **Aitha Sudheer Kumar**: Conceptualization, Data curation, Formal analysis and investigation, Methodology, Writing—original draft preparation; **Ankit Agarwal**: Conceptualization, Data curation, Validation, Writing—original draft preparation, Project administration; **Vinita Gangaram Jansari**: Conceptualization, Methodology, Formal analysis and investigation, Writing—review and editing; **K A Desai**: Validation, Writing—review and editing, Resources, Supervision; **Chiranjoy Chattopadhyay**: Supervision; **Laine Mears**: Supervision, Writing—review and editing, Fund acquisition, Project administration.

Funding This material is based upon work partly supported by the National Science Foundation under Grant No. 1760809. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Availability of data The dataset used in this study were generated by the authors and is not publicly available. However, dataset will be made available on reasonable request.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Ethics approval The ethics approval is not applicable in this work.

Consent to participate: This work did not apply for the consent to participate.

Consent for publication The authors give consent to the journal for the publication of this work.

References

Agarwal, A., Potthoff, N., Shah, A. M., Mears, L., & Wiederkehr, P. (2022). Analyzing the evolution of tool wear area in trochoidal milling of Inconel 718 using image processing methodology. *Manufacturing Letters, 33*, 373–379. https://doi.org/10.1016/j.mfglet.2022.08.002

Azzeh, J., Zahran, B., & Alqadi, Z. (2018). Salt and pepper noise: Effects and removal. *JOIV International Journal on Informatics Visualization, 2*(4), 252–256. https://doi.org/10.30630/joiv.2.4.151

Bergs, T., Holst, C., Gupta, P., & Augspurger, T. (2020). Digital image processing with deep learning for automated cutting tool wear detection. *Procedia Manufacturing, 48*, 947–958. https://doi.org/10.1016/j.promfg.2020.05.134

Cheng, H., Kong, X., Wang, Q., Ma, H., Yang, S., & Chen, G. (2023). Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions. *Journal of Intelligent Manufacturing, 34*(2), 587–613. https://doi.org/10.1007/s10845-021-01814-y

Coromant S. (2023). Cutting Data Calculation. Sandvik Coromant. Retrieved from June 14, 2024 from www.sandvik.coromant.com/en-us/tools/coroplus-toolguide/cutting-data-calculation?MaterialID=5853467;6416911.

Dai, Y., & Zhu, K. (2018). A machine vision system for micro-milling tool condition monitoring. *Precision Engineering, 52*, 183–191. https://doi.org/10.1016/j.precisioneng.2017.12.006

García-Pérez, A., Ziegenbein, A., Schmidt, E., Shamsafar, F., Fernández-Valdivielso, A., Llorente-Rodríguez, R., & Weigold, M. (2023). CNN-based in situ tool wear detection: A study on model training and data augmentation in turning inserts. *Journal of Manufacturing Systems, 68*, 85–98. https://doi.org/10.1016/j.jmsy.2023.03.005

He, Z., Shi, T., Xuan, J., & Li, T. (2021). Research on tool wear prediction based on temperature signals and deep learning. *Wear, 478*, 203902. https://doi.org/10.1016/j.wear.2021.203902

Hu, M., Ming, W., An, Q., & Chen, M. (2019). Tool wear monitoring in milling of titanium alloy Ti-6Al-4V under MQL conditions based on a new tool wear categorization method. *The International Journal of Advanced Manufacturing Technology, 104*, 4117–4128. https://doi.org/10.1007/s00170-019-04125-y

ISO 8688-2 (1989). Tool Life Testing in Milling. Part 2: End Milling. International Organization for Standardization (ISO). www.iso.org/standard/16092.html

Kumar A. S., Agarwal A., Jansari V. G., Desai K. A., Chattopadhyay C., & Mears L. (2023) Vision-Based Tool Wear Classification During End-Milling of Inconel 718 Using a Pre-Trained Convolutional Neural Network. In: ASME International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineers, p V003T03A016, https://doi.org/10.1115/IMECE2023-113344

Lee, M., Jeon, J., & Lee, H. (2022). Explainable AI for domain experts: A post Hoc analysis of deep learning for defect classification of TFT-LCD panels. *Journal of Intelligent Manufacturing, 33*, 1747–1759. https://doi.org/10.1007/s10845-021-01758-3

Liang, X., Liu, Z., & Wang, B. (2019). State-of-the-art of surface integrity induced by tool wear effects in machining process of titanium and nickel alloys: A review. *Measurement, 132*, 150–181. https://doi.org/10.1016/j.measurement.2018.09.045

López-Martínez, C., & Fabregas, X. (2003). Polarimetric SAR speckle noise model. *IEEE Transactions on Geoscience and Remote Sensing, 41*(10), 2232–2242. https://doi.org/10.1109/TGRS.2003.815240

Lundberg S. M., & Lee S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Procedings of the 31st International Confrence on Neural Information Processing Systems, 30, 4768-4777. doi: https://doi.org/10.48550/arXiv.1705.07874

Mandelbrot, B. B. (1971). A fast fractional Gaussian noise generator. *Water Resources Research, 7*(3), 543–553. https://doi.org/10.1029/WR007i003p00543

Martínez-Arellano, G., Terrazas, G., & Ratchev, S. (2019). Tool wear classification using time series imaging and deep learning. *The International Journal of Advanced Manufacturing Technology, 104*, 3647–3662. https://doi.org/10.1007/s00170-019-04090-6

Moghaddam, M. E., & Jamzad, M. (2007). Motion blur identification in noisy images using mathematical models and statistical measures. *Pattern Recognition, 40*(7), 1946–1957. https://doi.org/10.1016/j.patcog.2006.11.022

Nath, V., Chattopadhyay, C., & Desai, K. A. (2023). On enhancing prediction abilities of vision-based metallic surface defect classification through adversarial training. *Engineering Applications of Artificial Intelligence, 117*, 105553. https://doi.org/10.1016/j.engappai.2022.105553

Pan, Y., Xu, G., Xiong, Z., Li, F., Hu, B., Sun, Y., Pan, C., & Wang, Y. (2023). NJUST-CCTD: An image database for milling tool wear classification with deep learning. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-023-11418-w

Papenberg, B., Hogreve, S., & Tracht, K. (2023). Visualization of relevant areas of milling tools for the classification of tool wear by

machine learning methods. *Procedia CIRP, 118*, 525–530. https://doi.org/10.1016/j.procir.2023.06.090

Peng, R., Liu, J., Fu, X., Liu, C., & Zhao, L. (2021). Application of machine vision method in tool wear monitoring. *The International Journal of Advanced Manufacturing Technology, 116*(3–4), 1357–1372. https://doi.org/10.1007/s00170-021-07522-4

Pimenov, D. Y., Bustillo, A., Wojciechowski, S., Sharma, V. S., Gupta, M. K., & Kuntoğlu, M. (2023). Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review. *Journal of Intelligent Manufacturing, 34*(5), 2079–2121. https://doi.org/10.1007/s10845-022-01923-2

Pleta, A., Ulutan, D., & Mears, L. (2014). Investigation of trochoidal milling in nickel-based superalloy Inconel 738 and comparison with end milling. *International Manufacturing Science and Engineering Conference*. https://doi.org/10.1115/MSEC2014-4151

Potthoff, N., Agarwal, A., Wöste, F., Wiederkehr, P., & Mears, L. (2023). Evaluation of contrived wear methodology in end milling of inconel 718. *Journal of Manufacturing Science and Engineering, 145*(10), 101002. https://doi.org/10.1115/1.4062603

Potthoff, N., Liß, J., & Wiederkehr, P. (2023). Experimental setup for in-process measurements and analysis of wear-dependent surface topographies. *Journal of Manufacturing Science and Engineering, 145*, 101013. https://doi.org/10.1115/1.4063133

Prasad, K. N., & Ramamoorthy, B. (2001). Tool wear evaluation by stereo vision and prediction by artificial neural network. *Journal of Materials Processing Technology, 112*(1), 43–52. https://doi.org/10.1016/S0924-0136(00)00896-7

Ribeiro M. T., Singh S., & Guestrin C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1135–1144. https://doi.org/10.1145/2939672.2939778

Sarıkaya, M., Gupta, M. K., Tomaz, I., Pimenov, D. Y., Kuntoğlu, M., Khanna, N., Yıldırım, Ç. V., & Krolczyk, G. M. (2021). A state-of-the-art review on tool wear and surface integrity characteristics in machining of superalloys. *CIRP Journal of Manufacturing Science and Technology, 35*, 624–658. https://doi.org/10.1016/j.cirpj.2021.08.005

Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., & Batra D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision, pp 618–626. https://doi.org/10.1109/ICCV.2017.74

Shah, A. M., Agarwal, A., & Mears, L. (2023). Tool wear area estimation through in-process edge force coefficient in trochoidal milling of Inconel 718. *Manufacturing Letters, 35*, 391–398. https://doi.org/10.1016/j.mfglet.2023.08.072

Shen, Y., Yang, F., Habibullah, M. S., Ahmed, J., Das, A. K., Zhou, Y., & Ho, C. L. (2021). Predicting tool wear size across multi-cutting conditions using advanced machine learning techniques. *Journal of Intelligent Manufacturing, 32*, 1753–1766. https://doi.org/10.1007/s10845-020-01625-7

Singh, S. A., Kumar, A. S., & Desai, K. A. (2023). Comparative assessment of common pre-trained CNNs for vision-based surface defect detection of machined components. *Expert Systems with Applications, 218*, 119623. https://doi.org/10.1016/j.eswa.2023.119623

Tan M., & Le Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International Conference on Machine Learning, pp 6105–6114. https://proceedings.mlr.press/v97/tan19a.html

Thakur, D., Ramamoorthy, B., & Vijayaraghavan, L. (2009). An experimental analysis of effective high speed turning of superalloy Inconel 718. *Journal of Materials Science, 44*, 3296–3304. https://doi.org/10.1007/s10853-009-3445-4

Wang, J., Wang, H., Ding, J., Furuta, K., Kanno, T., Ip, W., & Zhang, W. (2016). On domain modelling of the service system with its application to enterprise information systems. *Enterprise Information Systems, 10*(1), 1–16. https://doi.org/10.1080/17517575.2013.810784

Wilson, D. R., & Martinez, T. R. (2003). The general inefficiency of batch training for gradient descent learning. *Neural Networks, 16*(10), 1429–1451. https://doi.org/10.1016/S0893-6080(03)00138-2

Wright P. K., & Bourne D. A. (1988). Manufacturing intelligence. Addison-Wesley Longman Publishing Co, Inc. https://dl.acm.org/doi/abs/10.5555/31839

Xu, D., Zhang, S., Zhang, H., & Mandic, D. P. (2021). Convergence of the RMSProp deep learning method with penalty for nonconvex optimization. *Neural Networks, 139*, 17–23. https://doi.org/10.1016/j.neunet.2021.02.011

Yan, X., Melkote, S., Mishra, A. K., & Rajagopalan, S. (2023). A digital apprentice for chatter detection in machining via human-machine interaction. *Journal of Intelligent Manufacturing, 34*(7), 3039–3052. https://doi.org/10.1007/s10845-022-01992-3

Yao, X., Ma, N., Zhang, J., Wang, K., Yang, E., & Faccio, M. (2022). Enhancing wisdom manufacturing as industrial metaverse for industry and society 5.0. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-022-02027-7

Ye, J., Ito, S., & Toyama, N. (2018). Computerized ultrasonic imaging inspection: From shallow to deep learning. *Sensors, 18*(11), 3820. https://doi.org/10.3390/s18113820

Zaheer R., & Shaziya H. (2019). A Study of the Optimization Algorithms in Deep Learning. 2019 Third International Conference on Inventive Systems and Control (ICISC), pp 536–539. https://doi.org/10.1109/ICISC44355.2019.9036442

Zhang, C., & Zhang, J. (2013). On-line tool wear measurement for ball-end milling cutter based on machine vision. *Computers in Industry, 64*(6), 708–719. https://doi.org/10.1016/j.compind.2013.03.010

Zhang, W., & Van Luttervelt, C. (2011). Toward a resilient manufacturing system. *CIRP Annals, 60*(1), 469–472. https://doi.org/10.1016/j.cirp.2011.03.041

Zhang, W., & Wang, J. (2016). Design theory and methodology for enterprise systems. *Enterprise Information Systems, 10*(3), 245–248. https://doi.org/10.1080/17517575.2015.1080860

Zhang W. C., Yang G., Lin Y., Ji C., & Gupta M. M. (2018). On definition of deep learning. World Automation Congress (WAC), pp 1–5. https://api.semanticscholar.org/CorpusID:51971897

Zhao, J., Qian, X., Zhang, Y., Shan, D., Liu, X., Coleman, S., & Kerr, D. (2023). A knowledge distillation-based multi-scale relation-prototypical network for cross-domain few-shot defect classification. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-023-02080-w