**COMPUTER AIDED ENGINEERING**

# Workpiece image-based tool wear classification in blanking processes using deep convolutional neural networks

Dirk Alexander Molitor[1] · Christian Kubik[1] · Ruben Helmut Hetfleisch[1] · Peter Groche[1]

## Abstract
Blanking processes belong to the most widely used manufacturing techniques due to their economic efficiency. Their economic viability depends to a large extent on the resulting product quality and the associated customer satisfaction as well as on possible downtimes. In particular, the occurrence of increased tool wear reduces the product quality and leads to downtimes, which is why considerable research has been carried out in recent years with regard to wear detection. While processes have widely been monitored based on force and acceleration signals, a new approach is pursued in this paper. Blanked workpieces manufactured by punches with 16 different wear states are photographed and then used as inputs for Deep Convolutional Neural Networks to classify wear states. The results show that wear states can be predicted with surprisingly high accuracy, opening up new possibilities and research opportunities for tool wear monitoring of blanking processes.

## 1 Introduction

Increasing quality requirements, high production rates and progressively more complex product geometries pose manufacturers with the challenges of a systematic automation and an efficient monitoring of blanking processes. Therefore, sensors are increasingly integrated into the processes and attempts are made to identify correlations between process anomalies and features of the recorded time series. Conventional approaches monitor the time series with the help of thresholds [1], linear discriminant functions [2] or envelope curves and can thus distinguish binary process states from each other. In particular, the occurrence of punch wear, which has a negative impact on the resulting product quality [3], is a widely researched application scenario. Knowledge about the current wear can help manufacturers to reduce downtimes and flexibly adapt maintenance intervals to the punch wear state. Modern monitoring approaches predict, for example, the edge radius of the punch by means of multiple linear regressions, whereby features from force signals are used as input variables [4].

Other authors predict that artificial intelligence and machine learning will have enormous effects on blanking processes [5], while the first applications have already been presented. For example, neural networks can be used to approximate the effects of fluctuating process parameters on geometric product properties, whereby training data can be obtained through FEM simulations, as presented by Stanke et al. [6] or Hambli and Guerin [7]. It was also shown that wear conditions can be accurately classified by extracting features from force signals and inputting them into Support Vector Machines [8].

Such time-series-based monitoring approaches have two different key disadvantages. First, the sensors are heavily loaded during the process due to high force and acceleration peaks, which can cause contact losses or slipping. Second, they need to be integrated as close as possible to the tool in order to realistically display the dynamic effects of the process [9]. To circumvent these issues, other data that contain correlations to wear conditions must be recorded and processed in real time.

An efficient approach to overcome this problem is the recording of image data. Image-based sensors enable contactless monitoring of production processes and depict optically perceptible process characteristics in great detail [10].

✉ Dirk Alexander Molitor
  dirk.molitor@ptu.tu-darmstadt.de

1   Institute for Production Engineering and Forming Machines, Technical University of Darmstadt, Otto-Berndt-Straße 2, 64287 Darmstadt, Germany

In combination with the advances made in the field of deep learning over the past decade, images can be processed in real time and correlations with labels recorded in production processes can be explored. However, to the best of the authors' knowledge, no image-based, data-driven monitoring approaches of blanking processes can be found in the literature. Therefore, this paper shows how images of blanked workpieces can be used to predict wear conditions of the punch. For this purpose, 7440 photos of 1860 workpieces, produced by punches with 16 different edge radii, are taken and used as inputs of Convolutional Neural Networks (CNN). Subsequently, it is demonstrated to what extent the CNN can classify the workpiece images according to the 16 wear states.

The paper is organized as follows. Section 2 gives an overview of the blanking process, its wear phenomena, corresponding monitoring approaches and basics about CNNs. Section 3 presents the methodology consisting of the experimental setup, data generation process and CNN based modelling. Section 4 presents the results obtained from a pretrained and a self-created CNN as well as their performance comparison. Finally, the results are summarised in Sect. 5 and new research opportunities are discussed.

## 2 Wear monitoring in blanking processes

### 2.1 Blanking processes

Blanking is one of the most frequently used sheet metal forming process and is part of the value chain in the automotive, information technology, power electronics and consumer goods sector. Blanking is a manufacturing process in which the final workpiece is separated from a sheet by applying a shearing force [11]. In the blanking process, the entire outer geometry of the workpiece is obtained in one working stroke whereby grid-shaped discard is left on the sheet metal strip [12]. Although, blanking is a separation process according to DIN 8588, the literature classifies this manufacturing technique as a sheet metal forming process.

As shown in Fig. 1a, blanking processes can be divided into three phases according to their force displacement curve [14]. In the punch-phase (I), the tool hits the sheet and starts to elastically deform the system consisting of tool, material and press. If the stresses are further increased the material tends to a plastic deformation until the shearing stresses exceed the shear fracture limit, the material breaks and the elastic energy stored in the systems is abruptly released. In the push-phase (II), the workpiece is completely separated from the discard and pushed through the sheet metal strip. At the end of the push-phase the tool passes the bottom dead center and is pulled out of the die during the withdraw-phase (III). While the maximal forces during the punch-phase
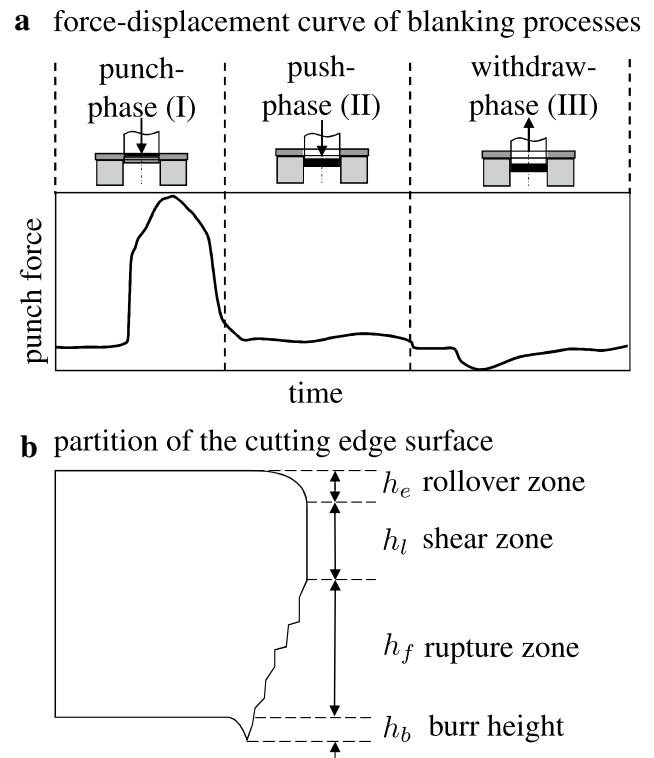
**a** force-displacement curve of blanking processes

punch-phase (I)   push-phase (II)   withdraw-phase (III)

punch force

time

**b** partition of the cutting edge surface

$h_e$ rollover zone

$h_l$ shear zone

$h_f$ rupture zone

$h_b$ burr height

**Fig. 1** Phases of the blanking process indicated by a force-displacement curve [4] (**a**) and forming zones of the cutting edge surface [13] (**b**)

depend on process parameters (clearance, cutting edge radii, cutting line), material properties (tensile strength, sheet thickness) or press settings (stroke speed, stiffness of the press), the pushing as well as the withdrawing forces are mainly due to contact normal forces influenced by ongoing wear on the lateral punch surface [15]. As an indicator for the quality of blanked workpieces, the cutting edge surface is quantified according to Lange [12]. As shown in Fig. 1b, the form error found on the blanked surface is divided into rollover zone $h_e$, shear zone $h_l$ and rupture zone $h_f$. In addition, the burr height $h_b$ is a key indicator of poor product quality and is directly influenced by the wear state [16].

### 2.2 Wear phenomena during blanking

Failures of blanking tools are often caused by major wear mechanisms referred to in the literature as adhesion, abrasion, surface fatigue and tribochemical wear [12]. Especially, abrasive wear related to the rounding of the punch cutting edge due to the trend of processing high-strength materials is a major issue manufactures have to deal with [17]. Hohmann et al. showed in their work that high-strength materials ($R_m > 600$ MPa) tend to abrasive wear on the tip while soft graded steels ($R_m < 350$ MPa) tend to adhesive wear on the lateral tool surface [15]. Many authors have

shown that abrasive wear directly influences product quality in terms of cutting edge surface and burr height of the blanked workpieces [18]. They have carried out experimental and numerical studies on the correlation between cutting edge quality and abrasive wear state on the tool. One of the first studies in this area was conducted by Meade and Matsuno, who investigated the influence of wear on the burr height [19]. Cheung et al. explored the influence of different process variables on the wear state of the blanking tool and correlated it with burr height and the required blanking force [20]. Kubik et al. systematically investigated the influence of semi-finished products and tool parameters on the quality of the cutting edge surface. They showed that the rounding of the cutting edge significantly increases the burr height and enhances the fraction of the sheared surface [14]. In addition to experimental investigations, numerical methods are increasingly used to predict tool wear in blanking processes. Hambli implemented a finite element (FE) wear model to predict tool wear in blanking steel and correlated tool wear with burr formation [21]. A more advanced approach was developed by Cheong and Kim, who took the effect of progressive tool wear into account of a FE model using a modified Archard wear model and a Lemaitre damage model. In addition, the geometry of the tool was adjusted based on the predicted wear volume, the tool replacement time was predicted, and the product quality was quantified based on the decrement of the hole area [22]. Due to the non-linearity of the blanking process and the variety of process variables, a

detailed description of the wear state by analytical, empirical or numerical models is only possible to a limited extent. In this context, data-driven approaches based on acquired process data offer the possibility to identify the cause of wear as well as to quantify the extent of the worn out region.

## 2.3 Data driven methods for process monitoring

Since the 1990s, data-driven monitoring approaches have been researched and used in blanking processes. The approaches, which are mostly based on time series, differ considerably. Conventional monitoring approaches define envelope curves or thresholds that enable time series or extracted features to be monitored, whereby only binary statements about faultless $p_c$ or faulty process states $p_f$ can be made. More advanced approaches use machine learning algorithms that offer the potential to detect different error patterns $p_{f,i}$. Recent publications in other production processes use deep learning algorithms, which can deal with a variety of raw data. Additionally there is no need for hand-crafted feature engineering, extraction and selection. An overview of the abstracted processes of the different approaches is provided in Fig. 2.

### 2.3.1 Conventional monitoring approaches

One of the first qualitative investigations on time series-based monitoring of blanking processes go back to Breitling
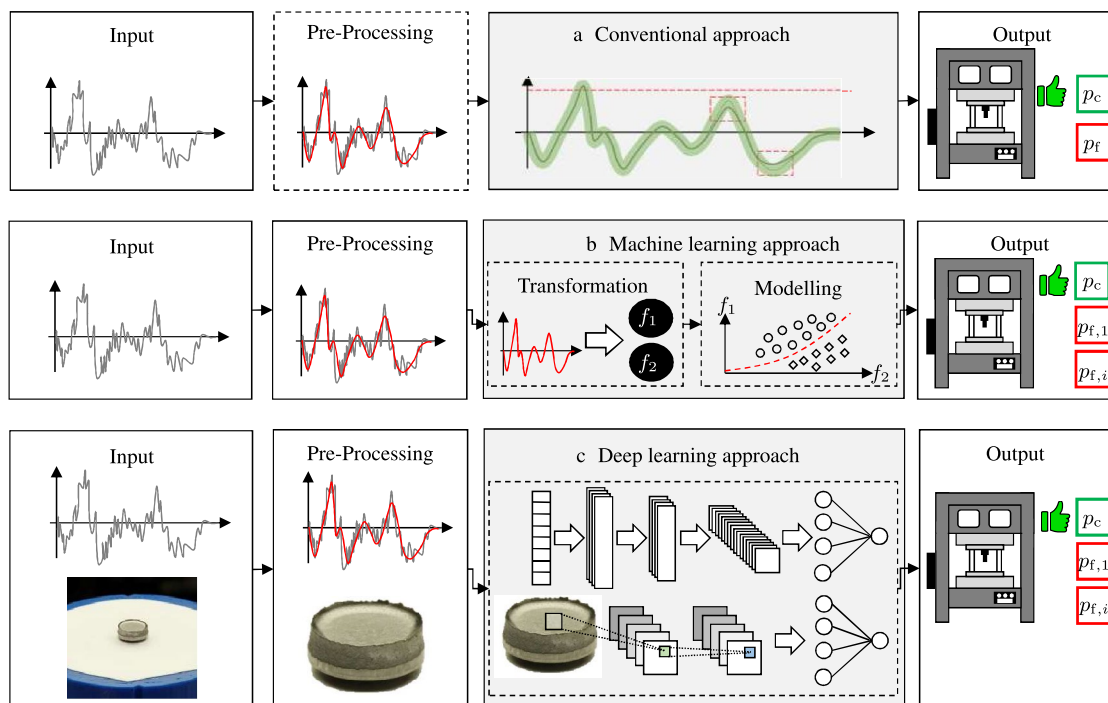


**Fig. 2** Overview over data driven monitoring approaches

et al. [23], who integrated different sensors both in the press frame and in the progressive die. They showed that the optical appearance of the signals differs when using different materials, stroke speeds and punch-die clearances, from which the fundamental suitability of monitoring blanking processes through time series is justified. The first real-time monitoring of tool wear in a blanking process was published by Lee et al. [2] who used force signals to build an autoregressive model in order to estimate blanking force peaks. The coefficients of the autoregressive model were used as inputs for a linear discriminant function that classifies between new and worn tool. Koh et al. showed how wavelet transform-based Haar coefficients from a force signal can be used to detect faulty conditions in a stamping process [24], similar wavelet-based approaches were investigated on force [25] and acceleration signals [26]. Zhang et al. applied bispectral analysis to acceleration signals to identify defective products [27]. Recent publications monitor stamping processes by recording acoustic emissions. Ubhayaratne et al. revealed correlations between wear and lubricant conditions with features of acoustic emission signals in the time [28] and frequency domain [29]. The potentials of applying time-frequency transformations to acoustic signals, such as the short-time Fourier transform, wavelet packet transform and Hilbert-Huang transform, were presented by Shanbhag et al. [30].

### 2.3.2 Machine learning approaches

The use of machine learning algorithms to monitor production processes opens up the possibility of classifying conditions in a much more detailed way or predicting quantitatively labels using regression algorithms. For this reason, machine learning algorithms are widely used in academic research as well as in industrial production, e.g., for condition monitoring or predictive maintenance.

The first application of a machine learning algorithm in a stamping process goes back to Jin and Shi, who used principal component analysis (PCA) to extract features from force signals. The PCA features were used as input variables for a decision tree to classify the combined presence of different process states, such as material thickness, lubrication, stroke speed and blank holder pressure [31]. Ge et al. extracted coefficients of an autoregressive model from the force signal of a stamping process and fed them into a hidden Markov model. Their approach can distinguish different fault types such as deviations from the desired material thickness, misfeed or slug with 87.5% classification accuracy [32]. In another publication, the authors applied support vector machines and showed that accuracies of almost 100% can be achieved [33].

Unsupervised learning algorithms are also used in stamping processes. Bassiuny et al. applied empirical mode decomposition to force signals of a stamping process and obtained intrinsic mode functions from which they extracted the signal energy and the Hilbert marginal spectrum as features. Then, they used learning vector quantisation to detect misfeeds and sheet thickness deviations [34]. Ge et al. used a self-organizing map to classify five different fault types [35]. Bergs et al. recorded force signals in a blanking process, extracted 14 statistical features from the time domain and applied PCA to them. The Euclidean and Mahalanobis distance of the first two PCA components were used to detect anomalies in consecutive punches. The authors showed that the process is subject to extensive fluctuations even with constant process settings [36].

### 2.3.3 Deep learning approaches

The application of machine learning algorithms to time signals is accompanied by the application of transformations, hand-crafted feature engineering, extraction and selection. Features must be extracted either model-based, e.g., by autoregressive models or PCA, in the time, frequency or time-frequency domain [37]. Deep learning approaches tackle this problem by unifying the process steps of data transformation and model building, consequently expanding the variety of usable data [38]. For example, the application of filtering and pooling operations in CNNs can simplify the high-dimensional representation of image data so that it can be used efficiently for classification and regression tasks. Publications from stamping, blanking and in general sheet metal forming processes rarely apply deep learning algorithms. Huang and Dzulfikri transformed acceleration signals into the frequency domain and used the entire power density spectrum as input for a one dimensional CNN. They demonstrated that they can distinguish seven different wear states with more than 99% accuracy [39]. Unterberg et al. recorded magnetic barkhausen noise from different material coils, created recurrence plots of the time series and used them as inputs for a two dimensional CNN. They classified whether certain material sections belong to the beginning, middle or end of a coil, as these sections have different semi-finished product properties [40].

Approaches that use deep learning for the purpose of tool condition monitoring are relatively new. This is evidenced by the fact that a comprehensive review paper from 2013 did not present any deep learning applications [41], whereas recent review papers discussed and emphasised the suitability of such algorithms [42]. For example, Gouarir et al. recorded triaxial force signals in a milling process, applied a method called Gramian Angular Summation Field to the time series to obtain a two-dimensional representation and classified the force signals according to three wear states using CNN [43]. Cao et al. also classified tool wear conditions by transforming recorded time series

into two-dimensional representations, which are then used as inputs for a CNN [44]. Recent publications used image data in combination with deep learning algorithms to detect wear states. A particularly interesting approach was taken by Bergs et al. who took 50 microscopic images of each of eight different tools. Then, they generated artificial images using data augmentation techniques and fed the images into a Fully Convolutional Network (FCN). The task of the FCN was to separate worn-out areas of the tool from non-worn-out areas, so that each pixel was labeled according to whether it represents a worn area or not. The authors showed that the deep learning approach is superior to classical computer vision algorithms and that local wear can be detected using FCN [45]. Marei et al. took 327 microscope images of the cutting tool flank of a CNC machine and applied different CNN transfer learning models to them to classify and regress 14 different wear states of the tool. The ResNet-18 model led to both the highest classification accuracy of 84% and the lowest normalized mean absolute error of 0.0773 [46]. Other image-based CNN approaches can be found in the literature, such as predicting the remaining useful life of cutting wheels [47] or classifying different types of wear in a face milling process [48]. These findings prove that image-based methods are highly suitable for the detection of wear states. Since wear conditions are reflected in the geometric properties of the products, especially in blanking processes [3], the authors of the present paper choose the approach of taking images of workpieces and correlating them with the existing tool wear using CNN.

## 2.4 Convolutional neural networks

CNNs are one of the most important deep learning algorithms and are used in different application areas, such as object detection, object recognition, pose estimation and text recognition [49]. Ever since CNN showed unbeatable performance in high-level image classification competitions compared to shallow learning algorithms [50], they have become a standard in image classification applications.

CNNs consist of an input layer and an output layer, between which convolutional, pooling, flatten and, if required, fully connected layers are arranged. The typical structure of a CNN is shown in Fig. 3. The input variable of image-processing CNNs are the original images, which can be represented as a two-dimensional matrix $X_{in}$ in the case of black-and-white images, or as a three-dimensional tensor $\mathbf{X}_{in}$ in the case of RGB images. One of the most important components of CNNs are convolutional layers. In convolutional layers, feature maps are generated by applying filters, also called kernels. If $X_j^l$ is the $j$-th feature map in the $l$-th layer, it is calculated according to

$$X_j^l = f\left( \sum_{i=1}^{N} X_i^{l-1} * W_{ij}^l + b_j^l \right), \tag{1}$$

where $W_{ij}^l$ is a weighting matrix, $*$ is the computational symbol of a two dimensional convolution with predefined filters and $b_j^l$ are bias parameters of the $l$-th layer. Here, $f(\cdot)$ is a non-linear activation function, whereby especially in image processing deep networks Rectified Linear Unit (ReLU)

$$f(x) = \max[0, x] \tag{2}$$

is used. The convolutional layer is usually followed by pooling layers, which reduce the dimensionality of the previous feature map according to predefined rules. In common CNNs, average pooling or max pooling layers are used. After a mostly alternating sequence of convolutional and pooling layers, a flatten layer follows, which transforms the multi-dimensional feature map into a one dimensional representation. At the end of the network topology, several fully connected layers usually form a multilayer perceptron (MLP), which transforms the one dimensional input to output values. Within the MLP, the output of the $l$th hidden layers is calculated by

$$x^l = f\left( W^l \cdot x^{l-1} + b^l \right). \tag{3}$$

In the classification case, the output layer has as many neurons as there are classes, with the softmax function
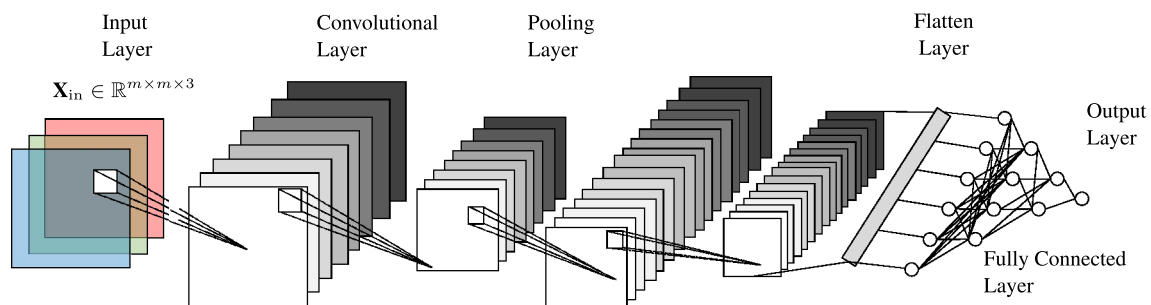


**Fig. 3** Structure of a CNN

$$f_{\text{soft},k}(x_i) = \frac{e^{x_k}}{\sum_{i=1}^{N} e^{x_i}} \tag{4}$$

now serving as the activation function of the $k$-th output neuron.

# 3 Methodology

In order to quantify the state of abrasive wear based on images of the blanked workpieces, a CNN is used to classify sixteen wear states of the cutting edge radii $r_i$. The applied procedure is shown in Fig. 4. In the first step, experiments are carried out and 1860 workpieces are manufactured. Subsequently, the photos are taken, which form a labeled image data set for tool wear classification by assigning the individual cutting edge radii $r_i$. In the model building step, it is investigated which pre-trained image classification model is suitable for wear detection and a self-created model topology is used. The models resulting from the different approaches are hyperparameter optimized and finally their performance is quantified and compared with each other.
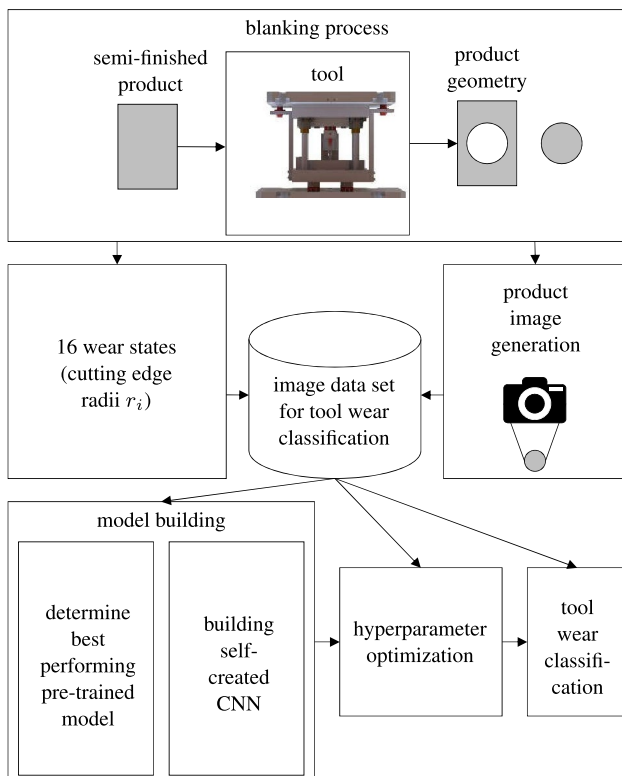
## 3.1 Experimental setup

The experiments to produce the blanked workpieces are performed on a mechanical high-speed press from Bruderer AG (BSTA 810). The machine has a nominal force of 810 kN and stroke rates of up to 1000 spm at a stroke height of 16 mm. The experiments are carried out with a stroke speed of 300 spm and a stroke distance of 35 mm. As an experimental material a micro-alloyed steels with high yield strengths for cold forming approaches (1.0480) is used. Figure 5 shows the experimental set-up, including the blanking tool. Table 1 summarizes the selected parameters of the press and the tool as well as the properties of the experimental material.

## 3.2 Data generation

As input of the CNN, a comprehensive data set is generated with 7440 workpiece images. When generating the data set, two basic principles must be taken into consideration. Product properties that correlate with tool wear must be represented as best as possible and the images should be reproducible and taken from the same perspective. This prevents the model from suffering performance degradation as it has to generalize more than necessary.
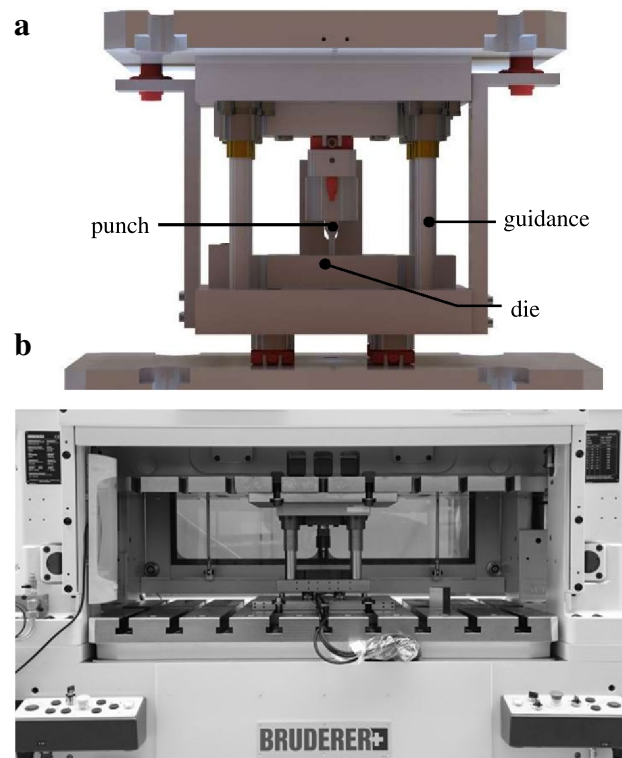


**Fig. 4** Procedure for classifying tool wear conditions in a blanking process using image-based CNN



**Fig. 5** Setup of the blanking tool (**a**) and blanking machine (**b**)

**Table 1** Selected parameters of the process and material

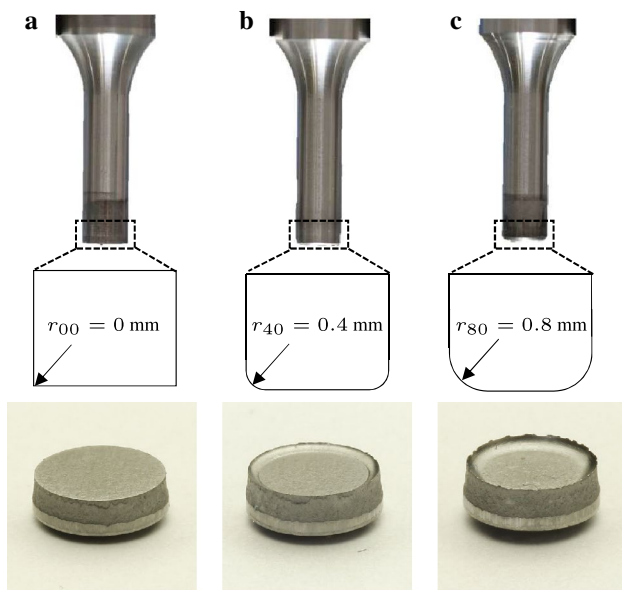| process parameters | |
| --- | --- |
| Stroke speed | 300 spm |
| Stroke distance | 35 mm |
| Clearance | 7.5 % |
| material parameters | |
| Description | HC 260 LA (1.0480) |
| Tensile strength | 365 N mm$^{-2}$ |
| Elongation A$_{80}$ | 27.10 % |
| Sheet thickness | (2.00 ± 0.1) mm |



**Fig. 6** Workpieces produced with cutting edge radii $r_i$ of **a** 0 mm, **b** 0.4 mm and **c** 0.8 mm

To meet these criteria, the images are shot in a illuminated environment using a rigid camera mount from an oblique side view. Each of the 1860 workpieces is photographed four times, with the images taken 90 degrees apart from each other. The camera used is a Canon EOS 6D SLR, which is capable of capturing images at 3648 × 3648 pixels. The images are then cropped to the region of interest consisting of 800 × 800 pixels and finally resized to the target input size of the corresponding CNN. Figure 6 shows three workpieces that have been produced with punches with different levels of cutting edge radii $r_i$.

As the used Deep Learning approach is based on a supervised learning technique, the generated image data set is labeled with the wear states of the punch. It is assumed that abrasive wear causes the rounding of the cutting edge radii of the punch [14, 51]. In order to replicate the desired abrasive wear conditions without the need for time-consuming

long-term experiments, the cutting edges are mechanically rounded by a post machining process. The cutting edge radii $r_i$ are varied in sixteen steps according to

$$r_i \in \{0, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, \\ 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80\}\,\text{mm}. \tag{5}$$

After the images are taken, they are read into Python and then divided into training, validation and test data. Care was taken to ensure that images of the same workpiece are only used in the training or test data set, so that the test data set only consists of images of workpieces that are not known to the model. The RGB images are transformed into NumPy arrays so that the images are available as a tensor $\mathbf{X}_{\text{in}}$. Finally, the tensor elements are normalized to values between zero and one, since CNNs can best process this value range.

### 3.3 Deep learning modeling

In order to find the best possible network topology, two different approaches are chosen. Network topologies can be freely chosen and then hyperparameter optimized. Another approach is transfer learning, which [52] allows the use of pre-trained models whose performance has already been proven in other image classification tasks [53]. The machine and deep learning community was able to demonstrate several years ago that CNNs are capable of learning generic mid-level image representations [54]. As a result, weights and topologies of networks obtained in image classification tasks with millions of labeled images and several thousand classes can also bring advantages to tasks that have comparatively few images and classes [55].

#### 3.3.1 Choice of the transfer learning model

All models presented are developed in Python, popular deep learning libraries such as Keras and Tensorflow [56] are used. Keras provides a variety of pre-trained models

**Fig. 7** Comparison of pretrained models
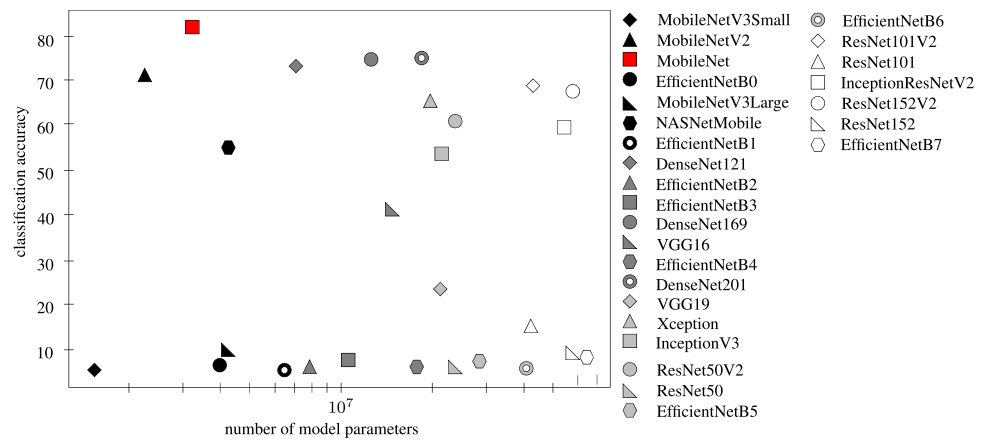


that have been used for the transfer learning approach. The dataset described in chapter 3.2 is divided into 70% test, 15% validation and 15% training data and fed into 27 pretrained models. Sparse categorial cross entropy (SCCE) is used as the loss function and Adam [57] as the optimization algorithm. No additional fully connected layers are added to the models, the inputs of the output layer are taken from the flatten layer and the training of the models is stopped after 5 epochs. The results of the comparison of all pretrained models can be seen in Fig. 7. It is evident that the accuracy of the different models varies greatly. The VGG models [58] deliver classification accuracies between 45 and 50%, whereas different versions of ResNet models achieve accuracies of up to 79%. The MobileNet, which is originally used for object recognition in commercially available smartphones [59], performs best with classification accuracies over 80%. Last but not least due to the comparatively low number of trainable parameters and the associated low model complexity, it was decided to use the MobileNet as transfer learning model in the following.

### 3.3.2 Hyperparameter optimization

In order to determine optimal hyperparameters for the developed models, a comprehensive hyperparameter optimization is carried out for both the self-created CNN and the pre-trained MobileNet. Due to the probability-based search and computational efficiency advantages over conventional optimization methods such as grid search, Bayesian hyperparameter optimization [60] is applied. In the optimization process, the individual runs are trained on 50 epochs and those hyperparameters are selected that lead to the highest accuracies. Table 2 gives an overview of the optimized hyperparameters. Table 3 presents the final network topologies obtained and additional information on the training processes.

**Table 2** Overview over optimized hyperparameters

| Hyperparameter | MobileNet | Self-created CNN |
|---|---|---|
| Amount of FCLs | Not used (0) | 3 |
| Neurons in FCLs | Not used | 512; 512; 512 |
| Batch normalization | Not used | After every FCL |
| Dropout | Not used | Not used |
| L2-regularization | Not used | 0.01 |
| Amount of CL | Not optimized | 7 |
| Learning rate | 0.0001 | 0.001 |
| Trained layers | Last 24 layers | All layers |
| Filters in CL | Not optimized | $2^5; 2^6; 2^6; 2^7; 2^8; 2^8; 2^8$ |

**Table 3** Overview over network topologies and informations on the training process

| Network parameter | MobileNet | Self-created CNN |
|---|---|---|
| Activation function | ReLU | ReLU |
| Amount of layers | 88 | 31 |
| Trainable parameters | 3,223,376 | 2,273,680 |
| Output activation function | Softmax | Softmax |
| Pooling operation used | Average | Max |
| Amount of CL | 27 | 7 |
| Amount of PL | 1 | 6 |
| Input dimension of $\mathbf{X}_{in}$ | $\mathbb{R}^{224 \times 224 \times 3}$ | $\mathbb{R}^{128 \times 128 \times 3}$ |
| Amount of output neurons | 16 | 16 |
| Amount of FCL | 0 | 3 |
| Optimization algorithm | Adam | Adam |
| Loss function | SCCE | SCCE |

## 4 Results

In the present section, the performances of the two models obtained from hyperparameter optimization are compared. To be able to compare the model performances in

a statistically robust way, the two models are both trained 100 times over 200 epochs, whereby the training process is stopped if no significant performance improvements are recorded over 30 consecutive epochs. 64 images of each wear class are used as validation data, another 64 images of each class as test data. The remaining 5392 images are used to train the models.

Figure 8 shows the results of 100 independent model runs as a boxplot, descriptive statistics of the results can be found in Table 4. It can be seen that both models deliver extremely high accuracy values and the wear condition of the tool can be determined by both models with almost 99 % accuracy. The accuracies of the models differ only in the decimal range, whereby the performance of the MobileNet is slightly better. The robustness of the MobileNet also appears to be higher, which is indicated by lower standard deviations and closer confidence intervals of the model performance. This is either due to the initialised weights of the MobileNet and the associated ability to provide generic mid-level image representations, or to the higher number of trainable model parameters.

In order to make statistically reliable statements about the superiority of one machine learning algorithm over another, methods from statistics can be applied. One approach to detect significant differences in performances of classifiers is the Welch's unequal variances $t$ test for differences in mean accuracy [61]. Since it is assumed that the pre-trained MobileNet can generate generic mid-level image representations, a one-tailed hypothesis test is conducted and the significance level is chosen to be $p = 0.01$. If the mean accuracies of MobileNet and self-created CNN are described by $\mu_{\text{Mob}}$ and $\mu_{\text{CNN}}$, respectively, then the null hypothesis is:

$H_0$: The accuracy of the MobileNet is equal to or less than that of the self-created CNN ($\mu_{\text{Mob}} \leq \mu_{\text{CNN}}$).
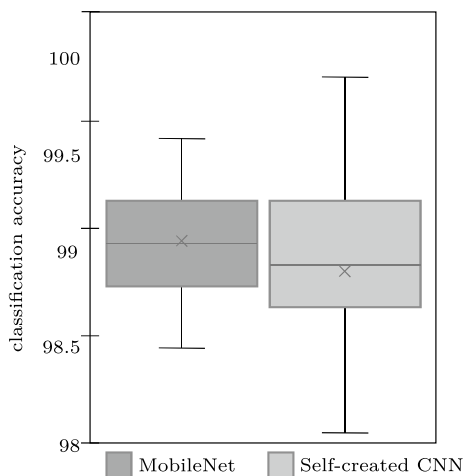
The $t$-statistic value can be calculated by

**Table 4** Descriptive statistics of the model performances

| Statistical parameter | MobileNet | Self-created CNN |
|---|---|---|
| Mean accuracy $\mu$ | 98.94 % | 98.80 % |
| Median accuracy | 98.93 % | 98.83 % |
| Standard deviation $\sigma$ | 0.24 % | 0.37 % |
| 95% confidence intervals | $(98.94 \pm 0.06)\%$ | $(98.80 \pm 0.10)\%$ |
| Maximum accuracy | 99.41 % | 99.71 % |
| Minimum accuracy | 98.44 % | 97.47 % |

$$t_{\text{stat}} = \frac{\mu_{\text{Mob}} - \mu_{\text{CNN}}}{\sqrt{\frac{\sigma_{\text{Mob}}^2}{N} + \frac{\sigma_{\text{CNN}}^2}{N}}}, \tag{6}$$

where $\sigma_{\text{Mob}}$ and $\sigma_{\text{CNN}}$ are the standard deviations of the accuracies of Mobilenet and self-generated CNN, respectively.

The results of the $t$ test can be seen in Table 5. It is shown that the test is significant due to

$$p(t_{\text{stat}} \leq t_{\text{crit}}) < p = 0.01, \tag{7}$$

which is why the null hypothesis must be rejected. This leads to the statistically robust statement that the MobileNet is better suited for the classification of tool wear conditions and leads to higher accuracies.

Table 6 gives an overview over the percentage misclassifications of the 100 trained MobileNets in the form of a confusion matrix, from which the wear states that are difficult to classify for the models can be derived. While there are low numbers of misclassifications at low cutting edge radii and especially with unworn punches, the models seem to become less accurate with increasing wear. Figure 9 shows example workpiece images for each wear class, which can be used to make assumptions about the increasing number of misclassifications with increasing wear. While differences in the geometric properties of the workpieces in the lower wear classes are particularly noticeable in the burr height and the impressions of the punch on the punching surface, which can thus potentially be discriminatory parameters for class classification, a differentiation of the high wear classes on the basis of optical properties is only possible to a limited extent. This is consistent with results from experimental studies in which the burr height grows asymptotically with



**Fig. 8** Performance comparison of MobileNet and self-created CNN

**Table 5** Results of the one-tailed $t$ test for differences in mean accuracies between MobileNet and self-created CNN

| | |
|---|---|
| Observations $N$ per model | 100 |
| Degrees of freedom | 169 |
| $t_{\text{stat}}$ | 3.13 |
| $t_{\text{crit}}$ one-tailed | 2.35 |
| $p(t_{\text{stat}} \leq t_{\text{crit}})$ | 0.001 |

**Table 6** Cumulative confusion matrix for tool wear classification of all 100 trained MobileNets in %

| $r_i$ [mm] | 0 | .1 | .15 | .2 | .25 | .3 | .35 | .4 | .45 | .5 | .55 | .6 | .65 | .7 | .75 | .8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 100 | – | - | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 0.10 | – | 98.47 | .39 | – | 1.11 | – | .03 | – | – | – | – | – | – | – | – | – |
| 0.15 | – | – | 100 | – | – | – | – | – | – | – | – | – | – | – | – | – |
| 0.20 | – | – | .02 | 99.98 | – | – | – | – | – | – | – | – | – | – | – | – |
| 0.25 | – | – | .05 | – | 99.95 | – | – | – | – | – | – | – | – | – | – | – |
| 0.30 | – | – | – | – | – | 100 | – | – | – | – | – | – | – | – | – | – |
| 0.35 | – | – | – | – | – | – | 100 | – | – | – | – | – | – | – | – | – |
| 0.40 | – | – | – | – | – | – | – | 99.97 | – | .03 | – | – | – | – | – | – |
| 0.45 | – | – | – | – | – | – | – | – | 99.73 | – | .27 | – | – | – | – | – |
| 0.50 | – | – | – | – | – | – | – | – | – | 100 | – | – | – | – | – | – |
| 0.55 | – | – | – | – | – | – | – | – | .02 | – | 98 | 1.98 | – | – | – | – |
| 0.60 | – | – | – | – | – | – | – | – | – | – | .52 | 96.86 | .22 | 2.41 | – | – |
| 0.65 | – | – | – | – | – | – | – | – | – | – | .86 | .81 | 98.23 | .08 | .02 | – |
| 0.70 | – | – | – | – | – | – | – | – | – | – | – | 1.03 | 2.05 | 94.67 | 1.89 | .36 |
| 0.75 | – | – | – | – | – | – | – | – | – | – | – | – | – | .03 | 99.97 | – |
| 0.80 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 2.78 | 97.22 |

increasing wear, reaches a plateau and therefore no longer serves as a discriminating feature above a certain degree of wear [62]. From wear class $r_{55}$ upgoing, the punch leaves hardly any marks on the blanking surface. In addition, it can be seen that the burr height over the rotationally symmetrical edge shows strong variances, which make a local feature extraction for the CNN more difficult.

Furthermore, the cutting surface also contains potentially discriminating features that cause a more accurate classification in lower wear classes. While especially in wear classes $\leq r_{20}$ the share of the shear zone in the total height of the



**Fig. 9** Example images of workpieces produced with 16 different punch edge radii $r_i$

workpiece correlates positively with the wear, such tendencies are not recognizable in higher wear classes.

## 5 Conclusion and outlook

While monitoring approaches of blanking processes are nowadays largely based on time series, this paper presents an approach for image-based tool wear monitoring. For this purpose, abrasive wear is simulated by mechanically rounding the edges of punches, whereby 1860 workpieces are manufactured with 16 differently worn punches. By capturing 7440 workpiece images, a labeled data set is generated, which is then processed using Convolutional Neural Networks. A pre-trained CNN called MobileNet as well as a self-generated CNN are hyperparameter optimized and used for the image classification task. The results show that both models have excellent classification accuracies of up to 99% and are thus highly suitable for tool wear classification. Using statistical testing methods, it can be shown that the pre-trained MobileNet is slightly superior to the self-created CNN. In addition, images of the workpieces show that optical properties correlate with the punch wear, but their ambiguity prevents an even more precise classification, especially in the presence of high wear.

These promising results open up new possibilities for future research. Future research efforts should investigate the extent to which models react to workpiece images of different semi-finished products and punch geometries. A highly interesting approach is to train the models on unlabeled data by using domain adaptation techniques in order to generate models that are as generalisable as possible and adaptable
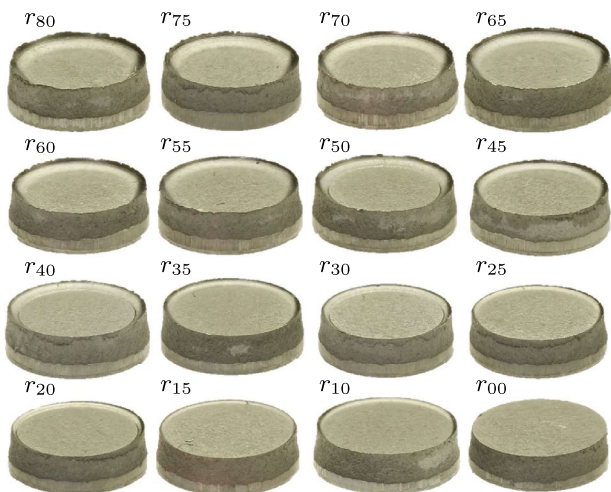
to different applications. Another interesting object of investigation is the performance behaviour of the models with varying data set sizes and how data augmentation techniques can improve model performances. In addition, the workpiece images offer the possibility to apply computer vision algorithms for hand-crafted feature extraction in order to capture product properties as time-efficiently as possible. A process-integrated acquisition of the workpiece images and their fusion with sensor data enables the identification of correlations between time series signals and product properties, which in turn can be used for predictive quality applications.

## Declarations

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

## References

1. Sari DY, Wu TL, Lin BT (2017) Preliminary study for online monitoring during the punching process. Int J Adv Manuf Technol 88(5–8):2275–2285
2. Lee WB, Cheung CF, Chiu WM, Chan LK (1997) Automatic supervision of blanking tool wear using pattern recognition analysis. Int J Mach Tools Manuf 37(8):1079–1095
3. Hambli R, Kobi S, Guerin F, Dumon B (2002) Relationships between blanking force and part geometry vs. clearance, tool wear, and sheet thickness. Qual Eng 15(2):197–207
4. Hoppe F, Hohmann J, Knoll M, Kubik C, Groche P (2019) Feature-based supervision of shear cutting processes on the basis of force measurements: evaluation of feature engineering and feature extraction. Proc Manuf 34:847–856
5. Zheng Q, Zhuang X, Zhao Z (2019) State-of-the-art and future challenge in fine-blanking technology. Prod Eng 13(1):61–70
6. Stanke J, Trauth D, Feuerhack A, Klocke F (2017) Setup of a parameterized fe model for the die roll prediction in fine blanking using artificial neural networks. In: *Journal of Physics: Conference Series*, volume 896, page 012096. IOP Publishing
7. Hambli R, Guerin F (2003) Application of a neural network for optimum clearance prediction in sheet metal blanking processes. Finite Elements Anal Design 39(11):1039–1052
8. Kubik C, Knauer SM, Groche P (2021) Smart sheet metal forming: importance of data acquisition, preprocessing and transformation on the performance of a multiclass support vector machine for predicting wear states during blanking. *Journal of Intelligent Manufacturing* 32
9. Groche P, Hohmann J, Übelacker D (2019) Overview and comparison of different sensor positions and measuring methods for the process force measurement in stamping operations. Measurement 135:122–130
10. Cao J, Brinksmeier E, Fu M, Gao RX, Liang B, Merklein M, Schmidt M, Yanagimoto J (2019) Manufacturing of advanced smart tooling for metal forming. CIRP Ann 68(2):605–628
11. DIN (2013) Din 8588:2013-08—manufacturing processes severing—classification, subdivision, terms and definitions. Technical report, Deutsches Institut für Normung
12. Lange K (1986) Handbook of metal forming. McGraw-Hill, New York
13. VDI (1994) Vdi 2906 blatt 2—quality of cut faces of (sheet) metal parts after cutting, blanking, trimming or piercing; shearing, form of sheared edge und characteristic values. Technical report, Verein Deutscher Ingenieure e.V
14. Kubik C, Hohmann J, Groche P (2021) Exploitation of force displacement curves in blanking—feature engineering beyond defect detection. Int J Adv Manuf Technol 113:261–278
15. Hohmann J, Schatz T, Groche P (2017) Intelligent wear identification based on sensory inline information for a stamping process. In: Majstorovic V, Jakovljevic Z (eds) *Proceedings of 5th International Conference on Advanced Manufacturing Engineering and Technologies*, Lecture Notes in Mechanical Engineering, pp 285–295. Springer International Publishing, Cham
16. Feistle M, Koslow I, Krinninger M, Golle R, Volk W (2017) Reduction of burr formation for conventional shear cutting of boron-alloyed sheets through focused heat treatment. Proc CIRP 63:493–498
17. Xing Z, Chen L, Lei C, Cai T, Yub H (2018) Simulated analysis and experimental investigation on edge qualities of high strength steels hot blanking parts. Proc Manuf 15:619–626
18. Klingenberg W, Singh UP (2004) Principles for on-line monitoring of tool wear during sheet metal punching. In: *Proceedings of the 34th International MATADOR Conference*, pp 169–174. Springer
19. Maeda T, Matsuno K (1967) Wear on shearing tools: (wear on cutting edges of blanking tool for square parts). Bull JSME 10(37):197–205
20. Cheung CF, Lee WB, Chiu WM (2000) An investigation of tool wear in the dam-bar cutting of integrated circuit packages. Wear 237(2):274–282
21. Hambli R (2001) Blanking tool wear modeling using the finite element method. Int J Mach Tools Manuf 41(12):1815–1829
22. Cheon S, Kim N (2016) Prediction of tool wear in the blanking process using updated geometry. Wear 352:160–170
23. Breitling J, Pfeiffer B, Altan T, Siegert K (1997) Process control in blanking. J Mater Process Technol 71(1):187–192
24. Koh CKH, Shi J, Williams WJ, Ni J (1999) Multiple fault detection and isolation using the haar transform, part 2: application to the stamping process. J Manuf Sci Eng 121(2):295–299
25. Jin J, Shi J (2001) Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. J Intell Manuf 12(3):257–268

26. Ge M, Zhang GC, Du R, Xu Y (2002) Feature extraction from energy distribution of stamping processes using wavelet transform. J Vib Control 8(7):1023–1032

27. Zhang GC, Ge M, Tong H, Xu Y, Du R (2002) Bispectral analysis for on-line monitoring of stamping operation. Eng Appl Artif Intell 15(1):97–104

28. Ubhayaratne I, Xiang Y, Pereira M, Rolfe B (2015) An audio signal based model for condition monitoring of sheet metal stamping process. In: *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, pp 1267–1272. IEEE

29. Ubhayaratne I, Pereira Michael P, Xiang Y, Rolfe BF (2017) Audio signal analysis for tool wear monitoring in sheet metal stamping. Mech Syst Signal Process 85:809–826

30. Shanbhag Vignesh V, Rolfe Bernard F, Arunachalam Narayanan, Pereira Michael P (2020) Investigation of galling wear using acoustic emission frequency characteristics. Lubricants 8(3):25

31. Jin J, Shi J (2000) Diagnostic feature extraction from stamping tonnage signals based on design of experiments. J Manuf Sci Eng 122(2):360–369

32. Ge M, Du R, Xu Y (2004) Hidden Markov model based fault diagnosis for stamping processes. Mech Syst Signal Process 18(2):391–408

33. Ge M, Du R, Zhang G, Xu Y (2004) Fault diagnosis using support vector machine with an application in sheet metal stamping operations. Mech Syst Signal Process 18(1):143–159

34. Bassiuny AM, Li X, Du R (2007) Fault diagnosis of stamping process based on empirical mode decomposition and learning vector quantization. Int J Mach Tools Manuf 47(15):2298–2306

35. Ge M, Du R, Xu Y (2003) Fault detection using hierarchical self-organizing map. In: *IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003*, volume 1, pp 565–570. IEEE

36. Bergs T, Niemietz P, Kaufman T, Trauth D (2020) Punch-to-punch variations in stamping processes. In: *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp 000213–000218. IEEE

37. Du R (2006) Monitoring and diagnosis of sheet metal stamping processes. In: Wang L, Gao RX (eds) Condition monitoring and control for intelligent manufacturing, chapter 8. Springer, pp 193–218

38. LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time-series. In: Arbib MA (ed) The Handbook of Brain Theory and Neural Networks. MIT Press

39. Huang CY, Dzulfikri Z (2021) Stamping monitoring by using an adaptive 1d convolutional neural network. Sensors 21(1):262

40. Unterberg M, Niemietz P, Trauth D, Wehrle K, Bergs T (2019) In-situ material classification in sheet-metal blanking using deep convolutional neural networks. Prod Eng 13(6):743–749

41. Dutta S, Pal SK, Mukhopadhyay S, Sen R (2013) Application of digital image processing in tool condition monitoring: a review. CIRP J Manuf Sci Technol 6(3):212–232

42. Serin G, Sener B, Ozbayoglu AM, Unver HO (2020) Review of tool condition monitoring in machining and opportunities for deep learning. *The International Journal of Advanced Manufacturing Technology*, pp 1–22

43. Gouarir A, Martínez-Arellano G, Terrazas G, Benardos P, Ratchev S (2018) In-process tool wear prediction system based on machine learning techniques and force analysis. Proc CIRP 77:501–504

44. Cao X, Chen B, Yao B, Zhuang S (2019) An intelligent milling tool wear monitoring methodology based on convolutional neural network with derived wavelet frames coefficient. Appl Sci 9(18):3912

45. Bergs T, Holst C, Gupta P, Augspurger T (2020) Digital image processing with deep learning for automated cutting tool wear detection. Proc Manuf 48:947–958

46. Marei M, El Zaatari S, Li W (2021) Transfer learning enabled convolutional neural networks for estimating health state of cutting tools. Robot Comput Integr Manuf 71:102145

47. Li X, Jia X, Wang Y, Yang S, Zhao H, Lee J (2020) Industrial remaining useful life prediction by partial observation using deep learning with supervised attention. IEEE/ASME Trans Mech 25(5):2241–2251

48. Wu X, Liu Y, Zhou X, Mou A (2019) Automatic identification of tool wear based on convolutional neural network in face milling process. Sensors 19(18):3817

49. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26

50. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90

51. Hambli R (2002) Prediction of burr height formation in blanking processes using neural network. Int J Mech Sci 44(10):2089–2102

52. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2020) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76

53. Shao L, Zhu F, Li X (2014) Transfer learning for visual categorization: a survey. IEEE Trans Neural Netw Learning Syst 26(5):1019–1034

54. Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1717–1724

55. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y (2013) Overfeat: integrated recognition, localization and detection using convolutional networks. *preprint*. arXiv:1312.6229

56. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M et al. (november 2-4 2016) Tensorflow: a system for large-scale machine learning. In: *12th USENIX symposium on operating systems design and implementation (OSDI '16)*, pp 265–283

57. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *preprint* arXiv:1412.6980

58. Tammina S (2019) Transfer learning using vgg-16 with deep convolutional neural network for classifying images. Int J Sci Res Publ 9(10):143–150

59. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. *preprint* arXiv:1704.04861

60. Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH (2019) Hyperparameter optimization for machine learning models based on bayesian optimization. J Electro Sci Technol 17(1):26–40

61. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

62. Straffelini G, Bizzotto G, Zanon V (2010) Improving the wear resistance of tools for stamping. Wear 269(9–10):693–697