



A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges

Vahid Nasir¹ · Farrokh Sassani¹

Received: 23 March 2021 / Accepted: 20 May 2021 / Published online: 31 May 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Data-driven methods provided smart manufacturing with unprecedented opportunities to facilitate the transition toward Industry 4.0-based production. Machine learning and deep learning play a critical role in developing intelligent systems for descriptive, diagnostic, and predictive analytics for machine tools and process health monitoring. This paper reviews the opportunities and challenges of deep learning (DL) for intelligent machining and tool monitoring. The components of an intelligent monitoring framework are introduced. The main advantages and disadvantages of machine learning (ML) models are presented and compared with those of deep models. The main DL models, including autoencoders, deep belief networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), were discussed, and their applications in intelligent machining and tool condition monitoring were reviewed. The opportunities of data-driven smart manufacturing approach applied to intelligent machining were discussed to be (1) automated feature engineering, (2) handling big data, (3) handling high-dimensional data, (4) avoiding sensor redundancy, (5) optimal sensor fusion, and (6) offering hybrid intelligent models. Finally, the data-driven challenges in smart manufacturing, including the challenges associated with the data size, data nature, model selection, and process uncertainty, were discussed, and the research gaps were outlined.

Keywords Smart manufacturing · Tool condition monitoring · Data-driven manufacturing · Tool wear · Intelligent machining monitoring · Machine learning · Deep learning · Feature selection · Neural networks, Artificial intelligence

1 Introduction to data-driven smart manufacturing

The manufacturing industry has been dramatically shaped through integrating with the emergence of concepts such as Internet of Things (IoT) [1–3], cloud computing [4], mobile Internet, and artificial intelligence (AI) [5–7]. In this regard, AI has facilitated the transition from automated manufacturing toward smart manufacturing. The rapidly growing size of data in the industry and the necessity to deal with big data acquisition, storage, and processing highlight the need for data-driven manufacturing as an indispensable element of smart

manufacturing. According to Tao et al. [8], the main modules of data-driven smart manufacturing include the manufacturing module, the data driver module, the real-time monitor module, and the problem processing module (Fig. 1). They discussed that the manufacturing module receives the raw material as input and offers the finished product as the output. The data driver module assures implementing the smart manufacturing approach throughout the different stages of the data lifecycle and powers the real-time monitoring module and problem-processing module [8]. Real-time monitoring is a critical step that enforces quality improvement. It can be used for process health and condition monitoring and preventive maintenance. The advancement in intelligent monitoring paves the way for smart and efficient problem processing and decision-making in a system.

Intelligent machining is an industrial application in which the data-driven smart manufacturing approach can be implemented. Machining or subtractive manufacturing has a wide range of applications in different manufacturing sectors. It encompasses various traditional processes, including turning, planning, drilling, sawing, etc., or processes such as electrical

✉ Vahid Nasir
vahid.nasir@ubc.ca

Farrokh Sassani
sassani@mech.ubc.ca

¹ Department of Mechanical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada



Fig. 1 The four modules of data-driven smart manufacturing according to [8]. The data collected from the manufacturing processes should be pre-processed and analyzed and be used for real-time monitoring of the manufacturing processes. The running state analysis may provide insight

into the health and condition of the product, equipment, or the process through descriptive, diagnostic, or predictive analytics, which are used to enhance the functionality of manufacturing processes

discharge or abrasive (grinding, etc.) machining. Online monitoring of the machining processes provides insight through descriptive, diagnostic, or predictive analytics and enforces smart decision-making. Machining monitoring and its consequent decision-making typically deal with the following topics (Fig. 2):

- Workpiece condition monitoring, including surface integrity [13], roughness [14], or waviness monitoring [15–17] during different machining processes.
- Tool condition monitoring [18, 19], including the tool wear detection and prediction.
- Process-machine interaction [20, 21] and machine stability, including its dynamic behavior and chatter detection [22].
- Process sustainability [23], cleaner production [24], and energy-efficient manufacturing [25]. This typically includes reducing the materials footprint, such as the minimum quantity lubrication concept [26], or studying the health and occupational hazard during machining processes such as monitoring and minimizing the dust emission [27] or noise pollution [28].

The data-driven approach has provided unprecedented opportunities for the machining processes. Data is the key element when embracing an intelligent machining approach. While the focus of conventional machining monitoring was on the machining tools and equipment, sensing technology [29, 30], and signal processing [31, 32], the emerging topics

on big data analytics, information fusion, data mining, machine learning (ML), and deep learning (DL) have revolutionized the state-of-the-art research on intelligent machining monitoring by making a dramatic shift toward data-driven approaches. However, much of the advancement in smart manufacturing is associated with the development in data collection, and despite the unprecedented growth in data acquisition (advanced sensory and imaging techniques) and storage, data processing still tends to be low in manufacturing sectors [33].

This study is a critical review of the applications, opportunities, and challenges associated with the data-driven approach applied to intelligent machining and tool monitoring focusing on DL methods for machining monitoring. Following the introduction, the manuscript is organized into five main modules: (1) intelligent monitoring framework, in which the main steps toward developing a data-driven monitoring system are introduced and discussed; (2) machine learning vs. deep learning, in which the most commonly used ML and DL models used for machining monitoring are discussed and compared. Reviews of ML models are covered in other publications [29, 30, 34]; however, due to growing interest in adopting DL models in machining and tool monitoring, the focus of this review will be on DL models; (3) applications of DL in machining and tool monitoring, in which the state-of-the-art progress on employing autoencoders (AE), deep belief networks (DBNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) for intelligent machining and tool monitoring are reviewed and discussed; (4) data-driven opportunities, in which the opportunities provided by

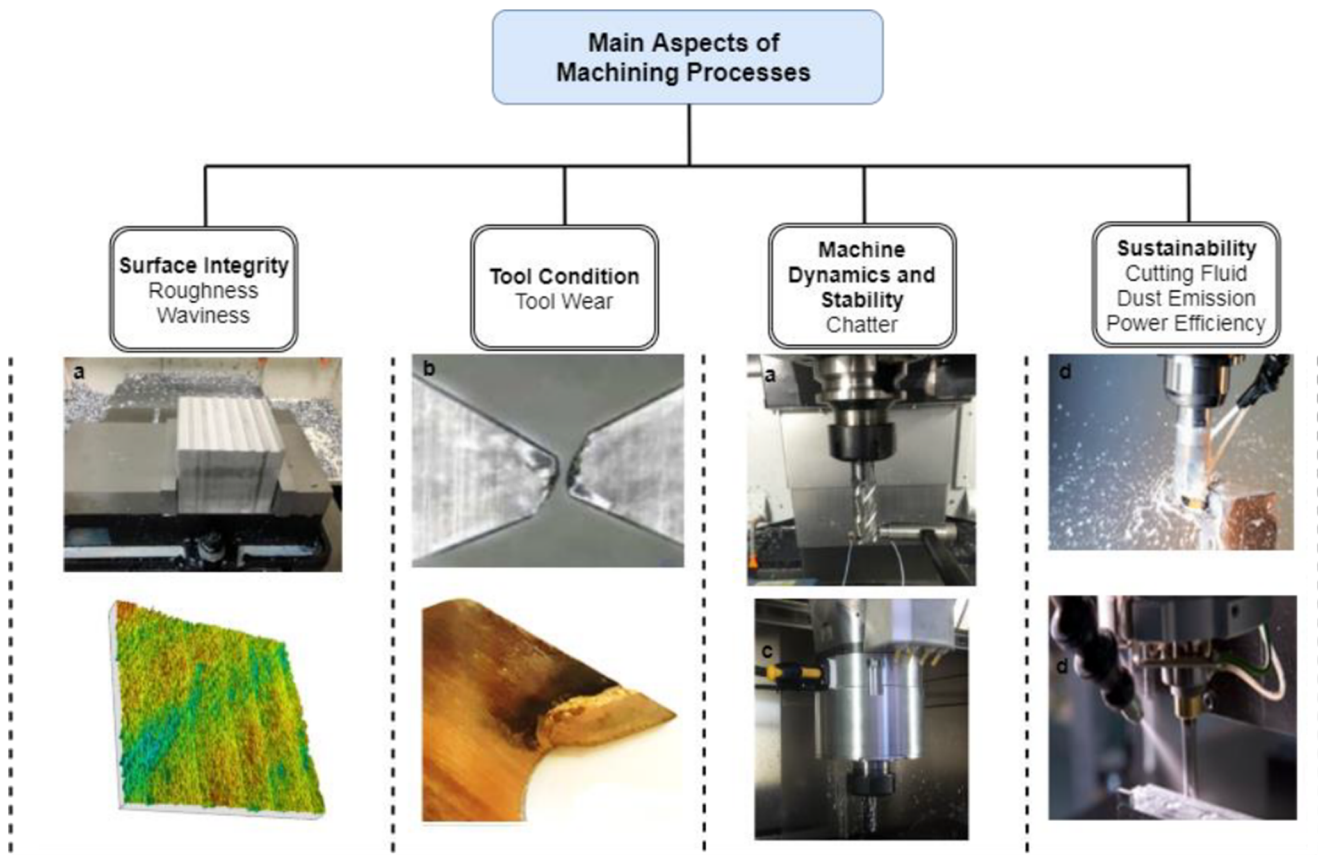


Fig. 2 Main aspects of machining processes for monitoring and optimization (panels a, b, c, and d are from [9–12], respectively)

ML and DL for smart manufacturing and specifically intelligent machining are presented; and (5) data-driven challenges and prospects, in which the cutting-edge challenges in front of smart manufacturing and machining industry to fully embrace the data-driven approach are discussed and the future trends are outlined.

2 Intelligent monitoring framework

The transition toward intelligent machining processes necessitates developing intelligent systems for process health, tool condition, and surface integrity monitoring. Developing an intelligent monitoring framework consists of the following main steps, as shown in Fig. 3.

2.1 Sensor selection

The most widely used sensors for monitoring the machining processes are current, sound, vibration, acoustic emission, and temperature sensors [29]. The machining conditions, the distance between the sensor and the cutting zone, difficulty in sensor placement and positioning, the low- vs. high-frequency nature of the machining process, and the presence of cutting fluids or dust can impact the sensor selection, e.g., choosing

between the contact and non-contact sensors. Typically, data collection occurs from different resources using various sensors and machine vision systems with complex data structures [35], which are challenging to integrate. However, IoT has enabled the manufacturing industry to use the data collected from different resources to be used in real-time monitoring [36, 37].

2.2 Signal processing

Before applying sophisticated feature selection methods and machine learning models, appropriate signal processing is the key to enhancing the performance of monitoring systems. Signal segmentation, de-noising, time and frequency analysis, and wavelet transform for decomposing the signal into its high- and low-frequency components have been widely practiced before extracting statistical features from the processed signals [38].

2.3 Feature selection

Optimal features are advised to be selected before developing an intelligent decision-making model. The growing awareness of the importance of feature engineering on the monitoring systems' performance raised the attention to the feature

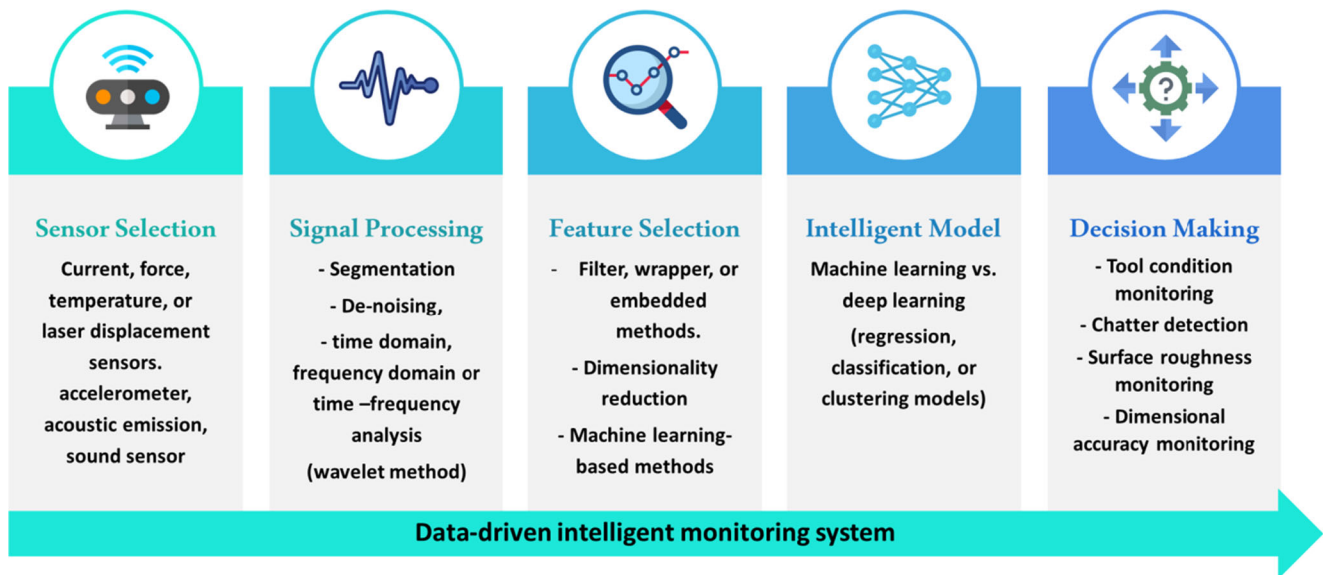


Fig. 3 Steps in developing a data-driven intelligent monitoring system

selection techniques. The most widely used feature selection methods include the filter (correlation, chi-square test, ANOVA, etc.), wrapper (forward selection, backward elimination, stepwise selection, etc.), and embedded methods (e.g., random forest) [39]. Feature selection can be performed using combinatorial search algorithms for feature subset selection [40]. However, despite the effectiveness of optimal feature selection on enhancing the quality of machining monitoring, many studies on machining monitoring include feature extraction by trial and error or by referring to similar literature without conducting in-depth research for optimal feature selection [29, 41].

2.4 Intelligent model

Choosing the type of intelligent model depends on the monitoring purpose, based on which one may select a model for regression, classification, or clustering. The most widely used models for decision-making in the machining processes in ascending order are artificial neural networks (ANNs), fuzzy models, neuro-fuzzy (ANFIS) models, and Bayesian networks [30]. A growing interest has been observed in using DL decision-making models in the past years [18]. Model selection depends on the size and complexity of the acquired data and the level of pre-processing (de-noising, feature extraction, etc.) performed on the raw data.

2.5 Decision-making

Intelligent systems have been primarily used for tool condition monitoring (TCM) [42, 43]; including the tool wear classification, flank wear prediction, chatter detection, and tool temperature monitoring [44]. Following the TCM, intelligent

systems were utilized during the machining processes for monitoring the surface roughness and waviness [45], chip condition [46] (chip formation, size, and breakage, dust emission, etc.), machining environment monitoring (e.g., airborne dust emission), and process condition monitoring (process fault, variation, state, etc.) [47, 48]. The monitoring scope defines the type of intelligent decision-making model (e.g., using classifiers for tool wear detection vs. regression-based models for flank wear prediction).

3 Machine learning vs. deep learning

The data-driven approach should satisfy the current industrial needs in different aspects. Wuest et al. [49] outlined the manufacturing requirements concerning the data-driven methods as follows:

- The ability to handle high-dimensional problems
- The ability to reduce possibly complex nature of results and present transparent and concrete advice
- The ability to adapt to the changing environment with reasonable effort and cost
- The ability to advance the existing knowledge by learning from results
- The ability to work with the available manufacturing data without special requirements toward capturing very specific information at the start
- The ability to identify relevant processes intra- and inter-relations and ideally correlation and/or causality

In this regard, ML and DL are among the foremost important tools in satisfying the abovementioned requirements.

3.1 Machine learning

ML has revolutionized the advanced manufacturing and process monitoring systems and is one of the main pillars of industrial informatics in the transition toward the industry 4.0 concept. By unveiling the hidden patterns in high-dimensional data, ML has raised attention to concepts such as multi-sensor utilization and data-driven methods in manufacturing [5]. ML models can be accurately employed for classification, clustering, and regression with less subjectivity and human error by understanding the data pattern through the learning process. Figure 4 lists the main advantages and disadvantages of most ML models.

While the appearance of numerically controlled machine tools happened in the late 1950s and early 1960s [50], ML methods such as neural networks were used for tool condition monitoring in the late 1980s [51] and early 1990s [52, 53]. Using ML methods became popular during the 1990s to determine CNC machining parameters focusing on fuzzy systems, ANNs, and probabilistic inference approaches [54]. Since then, different ML models such as multilayer perceptron (MLP) networks, radial basis functions (RBF), or Kohonen self-organizing maps (SOM) were utilized for machining monitoring [55]. A survey in 1997 reported that MLP ANN trained with back propagation was used in over 60% of researches [55]. ML has been extensively employed for machining monitoring in the following years. The most widely used methods were ANNs, fuzzy-based and neuro-fuzzy models, support vector machines (SVMs), Bayesian networks, and the hidden Markov model [29, 30]. The advantages and disadvantages of these models were discussed by Abellan-Nebot and Subirón [30] and Ademujimi et al. [56] and are listed in Table 1.

ML has provided unprecedented improvements in data pattern recognition while keeping high accuracy and objectivity with less human error. They have also provided opportunities for handling big and high-dimensional data. However, the performance of ML models directly depends on the quantity and quality of the acquired data, which in many cases are limited and costly to collect. Many ML models are also prone to overfitting issues or false-positive message alerts and may have generalization problems. Thus, even with a high accuracy level, many models still suffer from the acceptable level of precision. Moreover, many ML models may not be easily interpreted and are criticized because of their black-box nature (e.g., ANNs). Despite the opportunities, ML has provided in adopting data-driven models in smart manufacturing (in general) and in machining monitoring (in specific), there are still challenges associated with it as follows:

- 1) The main challenges of ML models specifically associated with ANNs stem from the model overfitting and generalization issues. Generalization is the model's capability to provide reasonable output from input data that the model has not seen during the learning process. Overfitting is the most common challenge associated with a network having backpropagation training algorithms [57]. In this case, there is always a trade-off between focusing on "learning" at the expense of temporarily ignoring "generalization" [58].
- 2) Many ML models, specifically the ANNs, are considered black-box, which are typically hard to interpret [59]. Some models, such as decision trees or random forest, offer descriptive analysis by indicating the relative

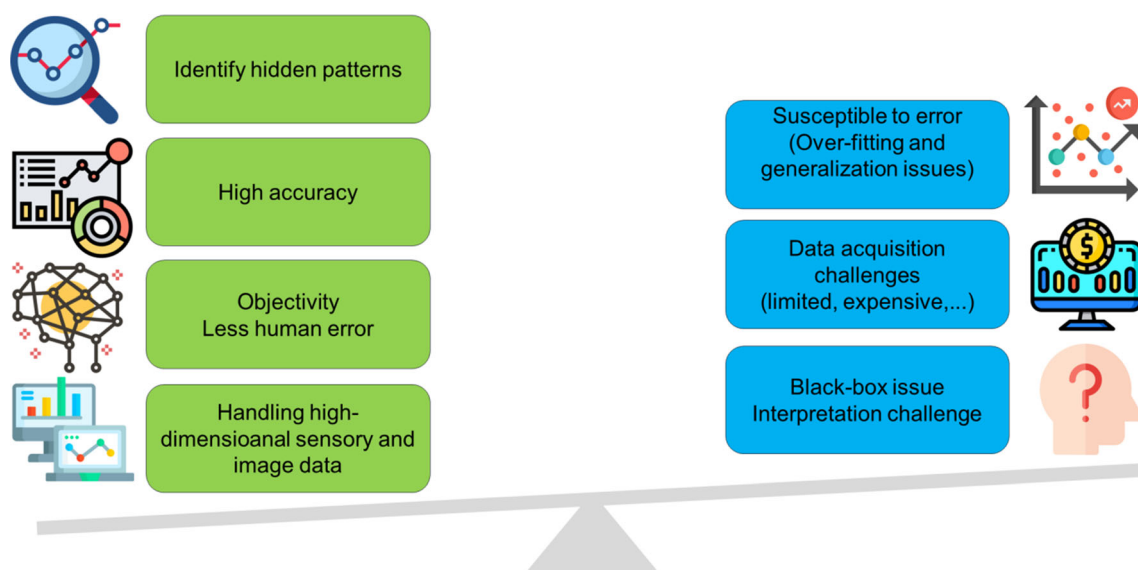


Fig. 4 The main pros and cons of conventional machine learning models

Table 1 The pros and cons of the most widely used machine learning models used in tool condition monitoring [30, 56]

Model	Advantages	Disadvantages
ANN	<ul style="list-style-type: none"> - Modeling nonlinear and complex data behavior - Modeling datasets with incomplete or noisy data - Modeling with minimal prior knowledge of data 	<ul style="list-style-type: none"> - Challenging to interpret - Design the model architecture by trial and error - No guidance for model hyperparameters tuning - Prone to overfitting
Fuzzy	<ul style="list-style-type: none"> - Easy to understand (based on linguistic model) - Handling imprecise data - High precision and extrapolation capability 	<ul style="list-style-type: none"> - Lack of real-time response - Not able to handle a high number of input variables - Lower learning and generalization capability than ANN
Neuro-fuzzy	<ul style="list-style-type: none"> - Benefiting from both the fuzzy system and neural structure - No need for prior knowledge of the process - Good generalization capability - Works well with small and medium-size data 	<ul style="list-style-type: none"> - Difficulty in hyperparameter tuning for a large number of fuzzy sets - Difficulty in handling a high number of input variables - Finding rules and membership functions can be challenging for the user
SVM	<ul style="list-style-type: none"> - Handling nonlinear datasets using kernel functions - Less overfitting issue comparing to ANNs - Good generalization with small training data 	<ul style="list-style-type: none"> - Difficulty in selecting the suitable kernel functions and their parameters - Lower learning speed and more computationally expensive (slow convergence rate for large datasets)
Bayesian	<ul style="list-style-type: none"> - Suitable for modeling stochastic systems and allow probabilistic inference - Presenting the causal relationships between variable - Efficient in avoiding data overfitting - Handling missing data - Utilize prior knowledge 	<ul style="list-style-type: none"> - More arbitrary and subjective than classical probability - Number of possible network architectures can rapidly grow with the number of nodes - High computational cost
Hidden Markov	<ul style="list-style-type: none"> - A probabilistic model with a robust statistical foundation - Handling time-series data - Handling inputs of variable length 	<ul style="list-style-type: none"> - Computationally expensive training process - Size of training data cause a challenge

importance of features and their contribution to the model's performance [60, 61].

- 3) Another main challenge with ML models in many engineering applications is the size and dimension of the acquired dataset. Typically, the machining experiments are conducted at a laboratory scale with limited observations. On the other hand, the data are typical signals acquired from the force, vibration, acoustic emission, or sound sensors with a high sampling rate. Thus, the datasets are usually not very big but are high-dimensional.

The challenge to analyze high-dimensional datasets necessitates employing dimensionality reduction and feature selection techniques to improve the performance of the ML model. These opportunities and challenges will be covered in more detail in Sections 5 and 6.

3.2 Deep learning

The conventional ML models are typically shallow. In ANNs, the conventional neural networks (NNs) usually have a maximum of two hidden layers with limited data processing capability in their raw form [62]. In many applications, analyzing big or high-dimensional data with conventional NNs requires feature engineering as a priori. Data are usually processed through dimensionality reduction methods such as principal component analysis (PCA) or data mapping methods like SOM [16]. Thus, two or more models should be linked together to form a hybrid intelligent system capable of analyzing complex data. While the term deep learning refers to employing numerous hidden layers in the structure of an ANN [63], it is mainly different from the traditional ML models in how representations are learned from the raw data

[64]. A deep model learns representations of data with multiple levels of abstraction [64, 65]. In other words, the learning process yields a high-level meaning in data through employing a high-level data abstraction [66]. DL models have more hidden layers than ML ones; however, what makes DL unique and different from traditional ML is the high-level feature engineering capabilities of deep models, where complex feature construction and abstraction are performed in the model structure during the learning process (Fig. 5).

The high-level abstract representation and feature engineering capabilities make DL models robust to data variation [67]. Also, the deep networks’ hierarchical structure enables them to model the complex nonlinear relationship in big data. On the other hand, ML models typically face difficulty in analyzing very big and high-dimensional datasets. Thus, feature selection serves as a dimensionality reduction approach enabling MLs to process big datasets. The challenge is those big datasets acquired in real industrial applications are typically polluted by noises and include outliers and different types of anomalies, making the feature selection a challenging task. Table 2 compares the typical characteristics of ML and DL models.

The most commonly used DL networks in intelligent manufacturing are autoencoders and their variants, deep belief networks (DBNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) [67].

3.2.1 Autoencoder

Autoencoders (AEs) are unsupervised feed-forward neural networks, where their output tries to return the input data (Fig. 6). It comprises encoder and decoder steps, in which the former transport the input data into a latent representation,

and the latter reconstructs the input from this representation. Gradient-descent-based algorithms are usually employed to tune the model’s hyperparameters by minimizing the reconstruction error. The main variants of AE are de-noising and sparse autoencoders (SAEs) [67].

3.2.2 Deep Belief Network

Deep Belief Network (DBN) comprises a stacking of multiple restricted Boltzmann machines (RBMs) [68]. There is a connection between the layers in DBN but not between the neurons within a layer. The layer-by-layer structure of the network provides a hierarchical feature representation [69], which is used to construct a high-level representation of input data. During the unsupervised training process, the DBN reconstructs its input through learning a probability distribution. RBM is a generative stochastic feed-forward ANN that is an effective tool for feature engineering. Training a DBN includes training multiple RBMs, where the hidden layer of the lower RBM is deemed the model training data, and the RBM output is used as the training data of the upper RBM (Fig. 7). After training all RBMs, fine-tuning process is performed by applying a backpropagation algorithm with the training data as output [70].

3.2.3 Convolutional neural network

The hidden layer of a convolutional neural network (CNN) includes a series of convolution layers, which abstracts the input tensor to a feature map using multiple local kernel filters. Filters are convolved across the width and height of the input data and generate two-dimensional feature maps. The output

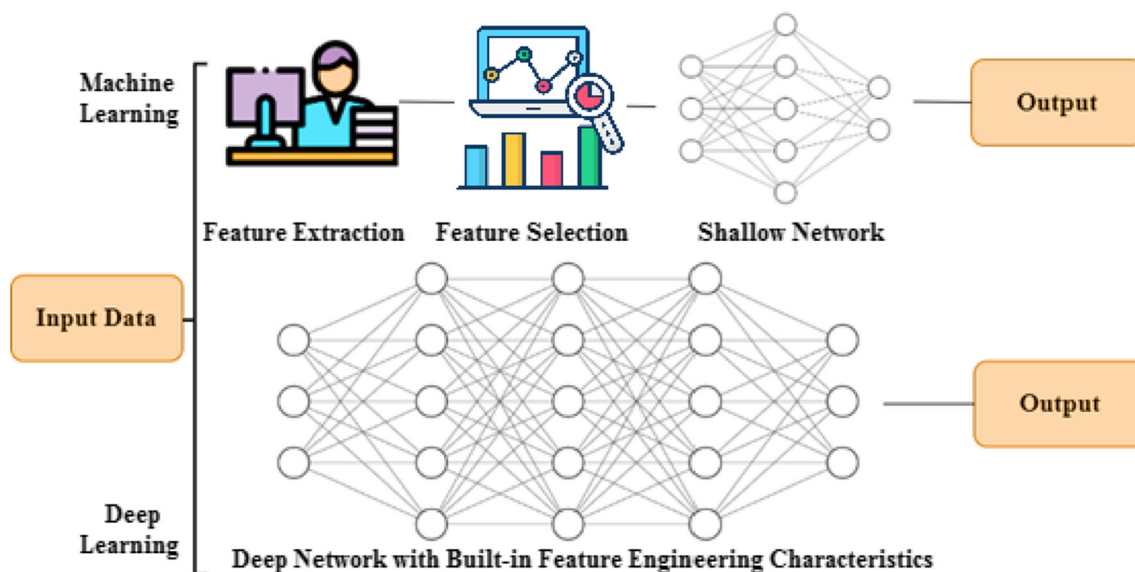


Fig. 5 Machine learning and deep learning difference concerning feature engineering

Table 2 Overall comparison between the deep learning and machine learning models

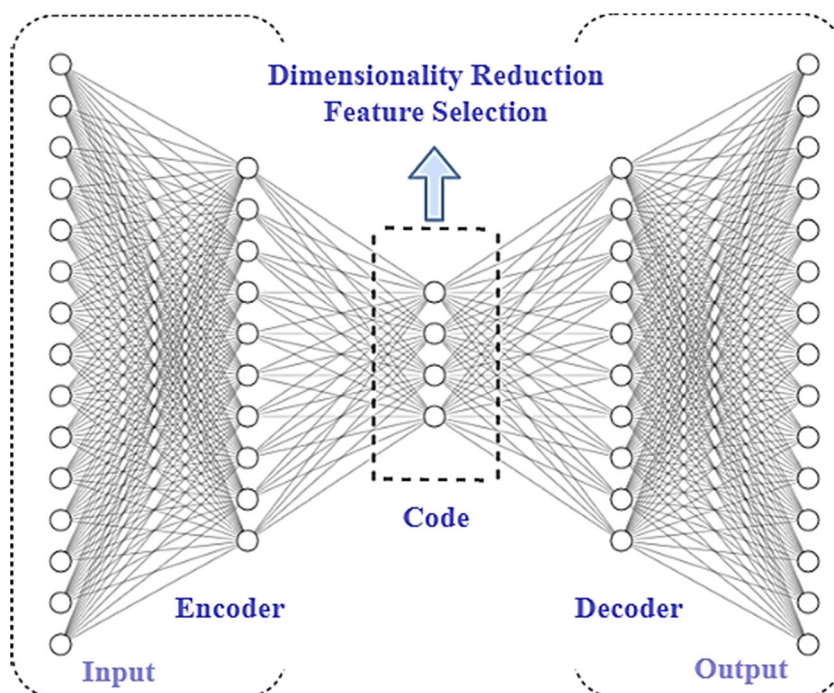
	Deep learning	Machine learning
Feature engineering	Minimal or no need for manual feature extraction and selection as feature engineering is built-in and automatically executed in a deep model.	Feature extraction and selection requires expert user knowledge and employing feature engineering techniques.
Dataset size	High performance when dealing with big data. Deep models perform much better with more data. Large-scale data acquisition may not be easy in many fields and restrict deep models.	Suitable for small- and medium-size data, specifically when big data acquisition may be challenging.
High-dimensional data	High-performance when dealing with high-dimensional data (sensory, image, etc.).	Challenge with high-dimensional datasets. An appropriate dimensionality reduction method or feature selection is needed.
Model tuning	No standard for training and tuning the deep models. The many parameters, such as network topology or activation/transfer functions, are typically tuned by trial and error.	Network tuning is typically done by trial and error. Due to the fewer number of parameters in a shallow network, hyperparameter tuning is relatively easier.
Interpretation	Hard or impossible to interpret the model due to the high number of variables in many different layers.	Hard to interpret but less challenging than deep models. The effect of some parameters in shallow networks or some simple methods such as linear regression may be studied.
Computational cost	Computationally expensive for training very large datasets.	Typically less computationally expensive models compared to deep learning.

of the convolution layer is obtained by stacking the obtained feature maps for all filters, which are then passed to the next layer for down-sampling, which is typically a pooling layer that reduces the spatial dimension of its input (Fig. 8). The dimension of the data can be further reduced by stacking more convolutions and pooling layers to achieve more abstracted features. The obtained feature map may be connected to a fully connected layer for supervised regression or classification.

3.2.4 Recurrent neural network

Recurrent neural network (RNN) is a class of feed-forward ANNs, with the capacity to update the current state based on the current input data and past states. Thus, it is ideal for dealing with sequential and time-series data or unsegmented signals through capturing information stored in sequence in the previous elements. RNNs benefit from supervised learning, and model training is performed using Backpropagation

Fig. 6 Schematic of an autoencoder network showing the encoder, decoder, and code layer used for dimensionality reduction and feature selection



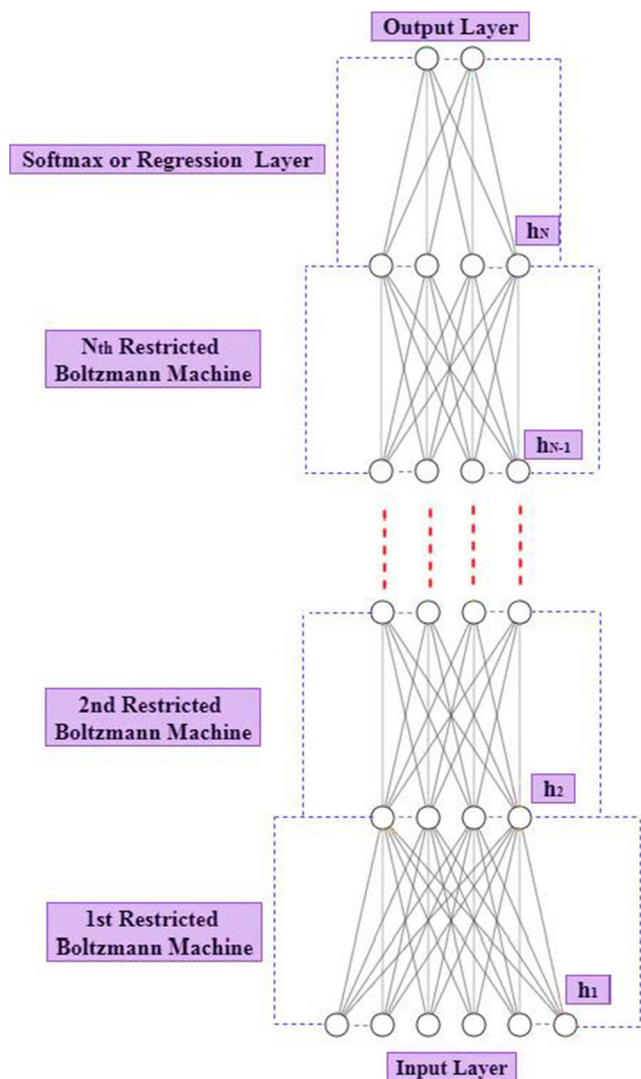


Fig. 7 Schematic of a deep belief network that shows a stack of restricted Boltzmann machines (RBM). The last hidden layer can be linked to a regression model or a softmax layer for classification

Through Time, which suffers from a vanishing gradient. Thus, the standard RNNs cannot effectively handle long-term dependencies in data and are challenging to be employed when the gap in the input data is large [71]. Different variants of

RNNs have been developed, among which, long short-term memory (LSTM) has been widely practiced to address the abovementioned challenge. A cell in the LSTM unit comprises an input, an output gate, and a forget gate, which regulate the flow of information in the cell (Fig. 9). Table 3 summarizes the key features of the abovementioned deep models.

4 Applications of deep learning in machining and tool monitoring

The applications of DL in machine health monitoring are rapidly growing [73]. The applications of DL within the context of intelligent machining are listed in Table 4. It can be seen that the majority of researches focused on monitoring the tool wear condition and prediction of the flank wear and the remaining useful life (RUL) of the tool. Few studies have also been performed on employing DL for chatter detection and surface roughness monitoring. Details on the applications of DL models for machining and tool monitoring are discussed in the following sub-sections:

4.1 Autoencoders

An unsupervised condition monitoring approach using AE is to define and employ an anomaly threshold using the AE reconstruction error. This monitoring technique’s core concept is that the reconstruction error can reveal whether the tool condition is changing or not. In this case, the AE is trained with reference data indicating the base condition (the typically stable situation with no damage and anomaly), and the reconstruction error is calculated. In the monitoring phase, the same AE is fed with new observations, and the reconstruction error is computed again. The basic assumption is that as long as the system condition is not experiencing a major change, the reconstruction error should be stable and small. However, if the reconstruction error goes beyond a defined threshold, then the state of the tool is changing, and it may experience damage such as tool wear. Figure 10 illustrates the schematic of such a

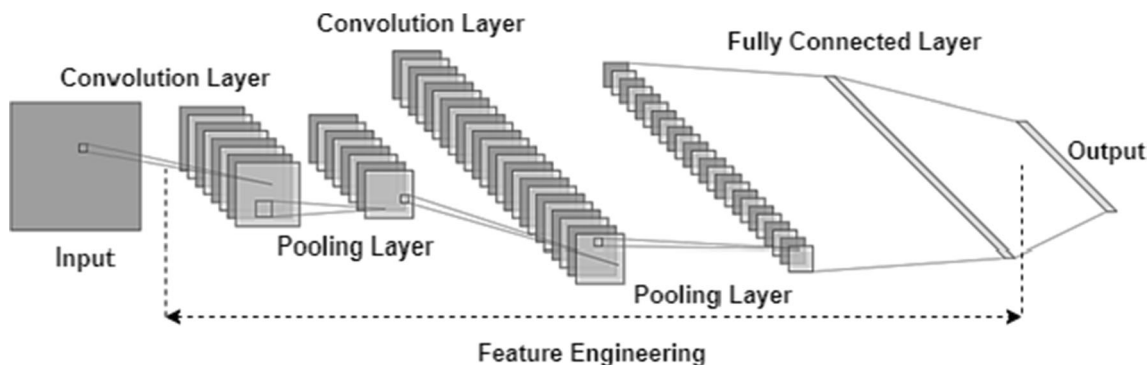


Fig. 8 Schematic of a CNN with two convolution and pooling layers linked to a fully connected layer

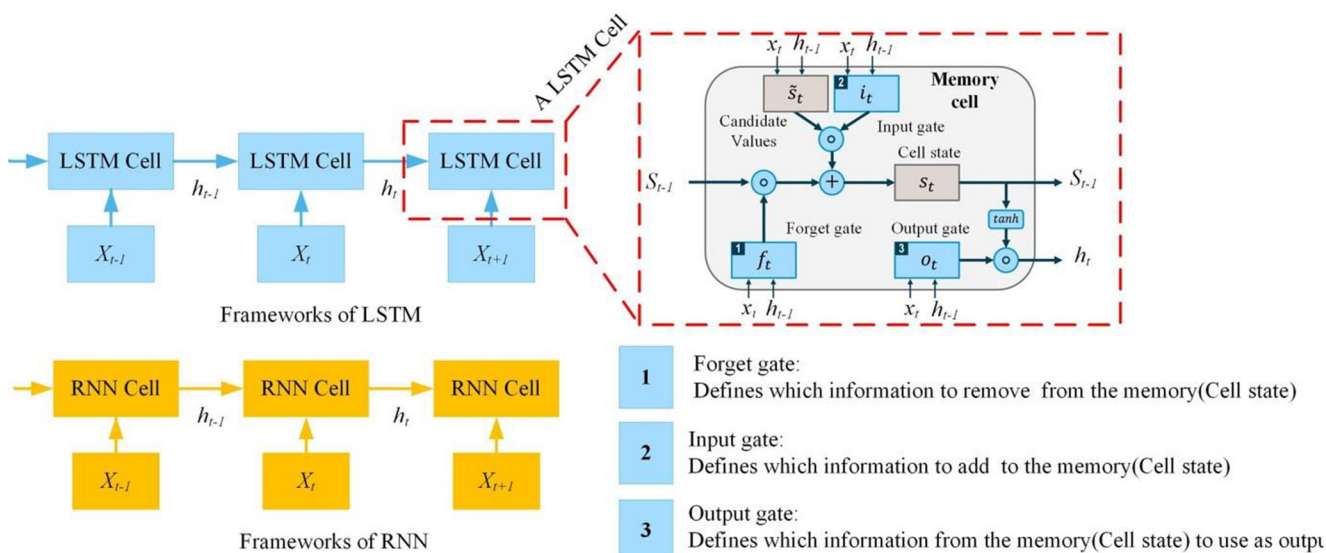


Fig. 9 Schematic of a recurrent neural network and long short-term memory (LSTM) ([72] with permission from Elsevier)

monitoring model. Dou et al. [76] used this approach for tool wear monitoring using the vibration and force signals in the milling process. They directly fed the segments of signals to the SAE model and showed that as tool wear increased, the reconstruction error became more unstable. They could identify four tool wear states using the proposed monitoring model. When more than two states are to be monitored, each state

should be used as a base condition to train an AE and define another threshold to show the next state’s border. For example, three thresholds should be defined to identify the borders between the healthy, initial wear, steady wear, and extreme wear states. Kim et al. [77] also used the AE thresholding-based method to differentiate the new and used tool using the cutting force and current sensors. However, instead of directly

Table 3 Summary of the key features of autoencoder (AE), deep belief network (DBN), convolutional neural network (CNN), and recurrent neural network (RNN)

Model	Description	Opportunities	Key features
AE	Unsupervised networks that transport the input data into a latent representation through encoding and its output reconstructs the input through decoding	Can be used for clustering, dimensionality reduction, data compression, and information retrieval	<ul style="list-style-type: none"> - Capturing the most relevant features is not guaranteed - Can meaningfully handle test data when they are similar to the training data. This is used as a basis for fault detection and state condition monitoring
DBN	Stacking of multiple restricted Boltzmann machines (RBMs) with the connection between the layers but not between the neurons within a layer	<ul style="list-style-type: none"> - More robust to hyperparameter variation comparing to some other models like MLP ANN or SVM [70] - Addressing the limitations of conventional backpropagation - Effective for high-level representation of input data 	<ul style="list-style-type: none"> - Unsupervised layer-by-layer pre-training - Backpropagation applied to the pre-trained model - The hidden layer of one RBM is the input of the next RBM
CNN	Feature abstraction through a series of convolutional and pooling layers	Handling image data (this will allow researchers in machining and tool monitoring to directly use images of the tool or convert the signal to image using techniques like the wavelet transform method)	<ul style="list-style-type: none"> - Designed to process data in the form of multiple arrays [65] - Computational cost and challenging hyperparameter tuning [67]
RNN	A feed-forward network that updates the current state based on the current input data and past states by capturing the temporal dependency in sequential data	<ul style="list-style-type: none"> - Ideal for dealing with time-series data - Can be linked to CNN to make a hybrid model that captures both sequential and spatial correlations in data [72] 	<ul style="list-style-type: none"> - Consider the temporal dependency in sequential data - Suffers from a vanishing gradient and problem in handling long-term dependencies in data [65]

Table 4 Applications of deep learning models in machining and tool monitoring

Model	Application	References
AE	Tool wear state monitoring	[74–82]
	Flank wear and remaining useful life prediction	[83–86]
	Chatter detection	[87]
DBN	Surface roughness prediction	[88]
	Chatter detection	[69]
	Flank wear prediction	[70]
CNN	Tool wear state monitoring	[89–99]
	Flank wear and remaining useful life prediction	[100–107]
	Chatter detection	[108, 109]
	Surface roughness prediction	[110]
RNN	Flank wear and remaining useful life prediction	[111–121]
	Chatter detection	[122]
	Surface roughness prediction	[123]
Combined LSTM-CNN	Tool wear state monitoring	[124]
	Flank wear and remaining useful life prediction	[125–134]

using the signal segments, they manually extracted 36 features from the signals. AE was trained using 80 samples collected from machining with a new tool. The testing data included 20 samples corresponding to the new tool and 218 samples associated with the used tool. Kim et al. [77] showed that the code size and the network architecture impacted the classification performance.

A unique characteristic of an autoencoder is that the neurons in its code layer can be used as a low-dimensional representation of the data. Thus, the feature selection can be achieved through dimensionality reduction in the model. The challenge with employing autoencoder for feature selection is finding the optimal model architecture (number of hidden layers and neurons), especially when dealing with big data. Moldovan et al. [78] determined the state of the tool wear using extracted features from the tool image having an input vector with a dimension of 11,844. The dataset was combined with an autoencoder, and it was shown that the model testing success rate increased by 60% when increasing the number of neurons in the 1st hidden layer to 150. Fine-tuning the autoencoder structures can be challenging and may require trial and error or grid-search techniques for optimizing the model hyperparameters. An approach for feature selection by dimensionality reduction is using stacked AE. In this method, the dimensionality reduction is achieved using stacked autoencoders, in which the code layer of each autoencoder forms the input layer of the subsequent encoder (Fig. 11). Ocha et al. [79] used stacked sparse autoencoders (SSAEs)

for tool wear classification in the milling process of aluminum using force, vibration, and acoustic emission sensors. The study considered four classes of tool conditions, and the input dataset comprised a total of 441 sensory data. In this study, seven sensory features were extracted from the signals as the input of data. Thus, the SSAE did not directly analyze the high-dimensional raw signal. Proteau et al. [80] discussed that AE is effective for dimensionality reduction and 2D visualization of data and showed a better dimension reduction capability compared to PCA for cutting state monitoring using vibration and current signals. AEs have also been used for tool wear monitoring in milling using the current signals [81, 82] and yielded higher monitoring accuracy than methods such as ANN, SVM, or KNN. Ou et al. [82] showed that introducing noise and using a stacked denoising autoencoder improved tool condition monitoring performance. Figure 12 shows their adopted methodology for AE-based tool wear monitoring using.

The AE concept can be integrated with feature fusion to better extract the meaningful features of signals. Shi et al. [83, 85] employed this approach for flank wear prediction using the vibration data acquired during the milling process of aluminum and stainless steel. The sensory data can be augmented using the Fourier and/or wavelet transforms. The time, frequency, and time-frequency data were then separately fed into SAE for feature selection. The selected features in different single domains were then combined and fed into a final AE to yield the input of a nonlinear regression model for tool wear prediction. It was shown that such a model outperformed the conventional machine learning models, however, at the expense of more training and testing time. When a series of AEs are stacked, other than feature transfer learning, weight transfer can also be performed between the AEs to enhance the model performance by improving the weight initialization process. Sun et al. [86] investigated deep transfer learning based on sparse AE for the tool's remaining useful life prediction. AE combined with a hybrid clustering method was used for chatter detection [87].

4.2 Deep belief neural networks

Yu and Liu [88] employed DBN combined with symbol and classification rules for surface roughness prediction and showed that DBNs effectively model complex nonlinear relationships between the machining process variables. DBN was successfully employed to build a feature space for cutting state monitoring (idling, stable cutting, and chatter) using the vibration data collected during the end milling process [69]. The vibration signals were segmented into signals with a dimension of 256. Besides, manual feature extraction was employed in the frequency (12 features) and wavelet (2 features) domains. The performance of DBN was compared with those obtained from ANN, SVM, and k-means clustering. It was

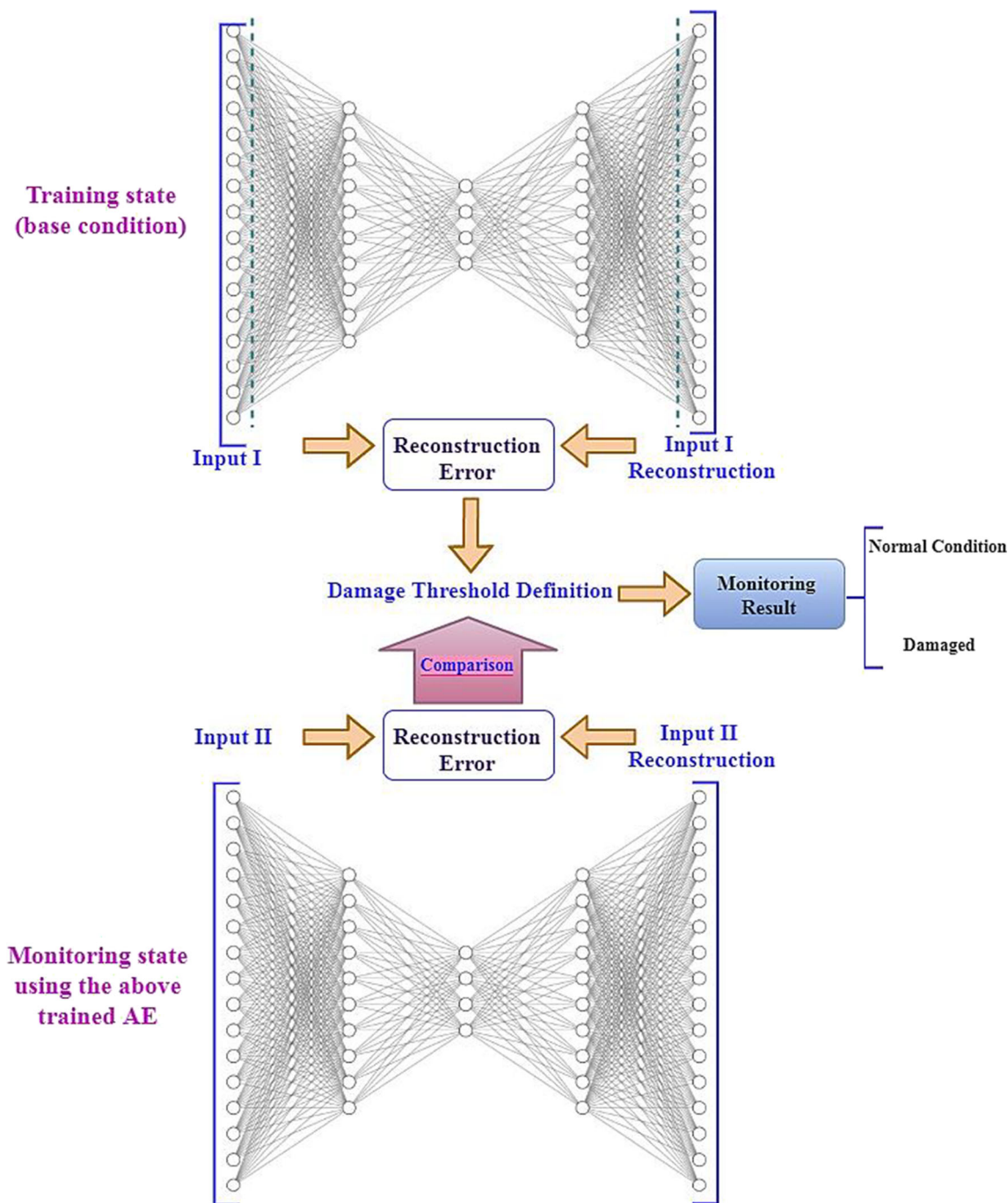


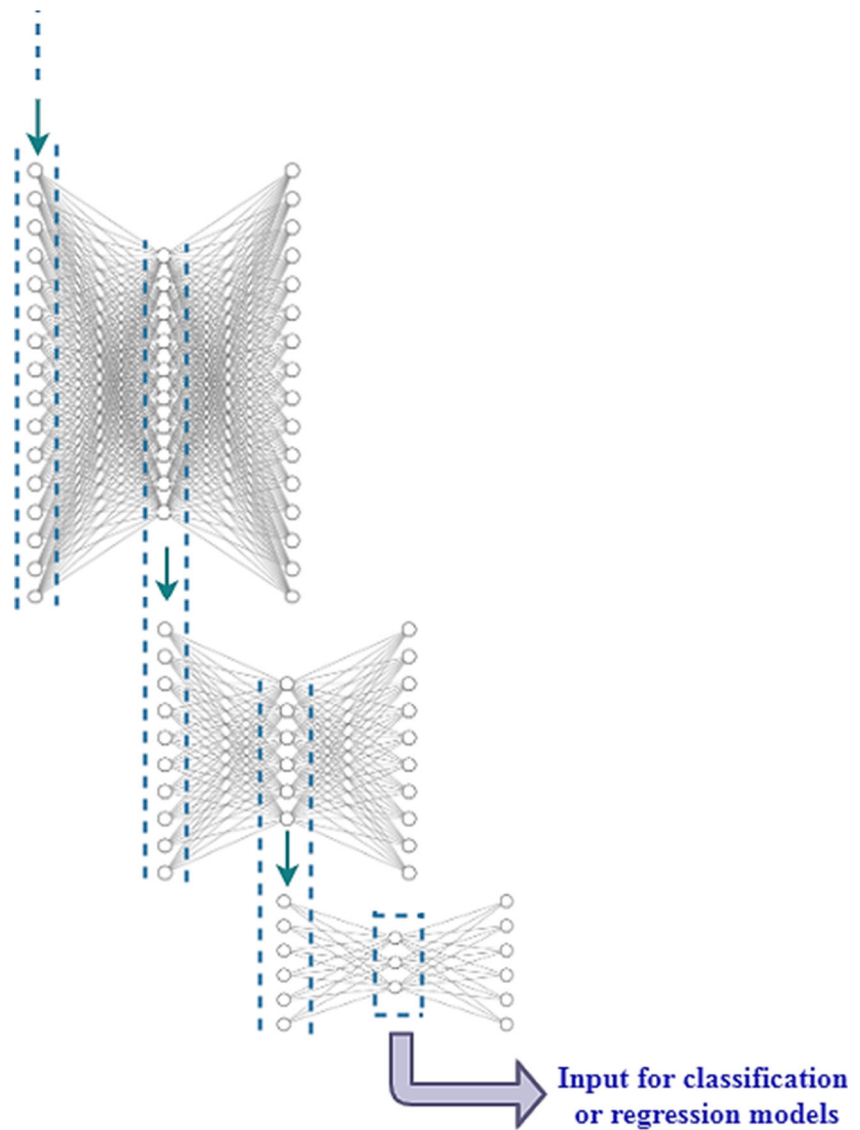
Fig. 10 Using autoencoder for condition monitoring through threshold definition approach

shown that while feature reduction generally improved the performance of ANN and SVM, DBN was more robust to the manual feature extraction and yielded lower error than other models. It was also revealed that comparing to the PCA, the output of DBN could better separate the three monitoring states with a relatively large margin and thus capable than PCA in feature engineering from the sensory data [69].

Chen et al. [70] used DBN for tool wear prediction during the high-speed CNC milling process using the cutting force, accelerometer, and acoustic emission signals. Maximum,

minimum, average, standard deviation, and the time stamp indicating the tool wear evolution and wear rate were chosen as the sensory features and were fed into the DBN. Four DBN layers were used, and for simplicity, every hidden layer was set with the same number of hidden neurons ranging from 10 to 15. The DBN was compared with the support vector regression (SVR) and MLP NN. They showed that while there was no significant difference between the three studied models' performance in terms of coefficient of determination (R^2), ANNs and DBNs required ~60% shorter prediction time than

Fig. 11 A sample architecture of stacked AEs for data reduction and feature selection. In this approach, the neurons in each autoencoder code layer are used as the next encoder's input layer. The last code layer could be linked to a softmax or regression layer for machining and tool condition monitoring



the SVR model. Moreover, while the performance of ANN was fluctuating with changing the number of hidden neurons, epochs, etc., DBN was more robust to hyperparameter variation, which thus outperformed the other models.

4.3 Convolutional neural networks

Compared to the autoencoders and DBNs, more research has been conducted on CNNs for intelligent machining, focusing on tool wear monitoring [89–107]. It has also been used for chatter detection [108, 109] and surface roughness prediction [110]. CNNs are ideal for handling image-based data; however, time-series data or extracted sensory features can be directly fed to 1D CNN. Xu et al. [103] employed a 1D CNN to extract features from the vibration data for tool wear monitoring. Lee et al. [90] also used a 1D CNN for tool condition monitoring in the grinding process using sound signals. The

authors identified the most critical frequency range of the signals (using Fourier transform analysis) and trained the 1D CNN using the audio signals in the time domain, preserving the critical frequency segment. During the machining process, the acquired sensory data can be combined with the experience- and physics-based features from the process to build low-cost models for process monitoring. Li et al. [104] developed a model, which relies on physical analysis to extract useful features to establish a reliable health indicator for tool condition monitoring utilizing the vibration and acoustic signals. Then, they developed a deep CNN model using 20 low-cost processes and cut variables to replace the physics-based model (Fig. 13).

The manually extracted features from different signal channels can be combined to form a multi-domain feature matrix. Huang et al. [105] extracted nine features from force and vibration signals in three directions to form a multi-domain

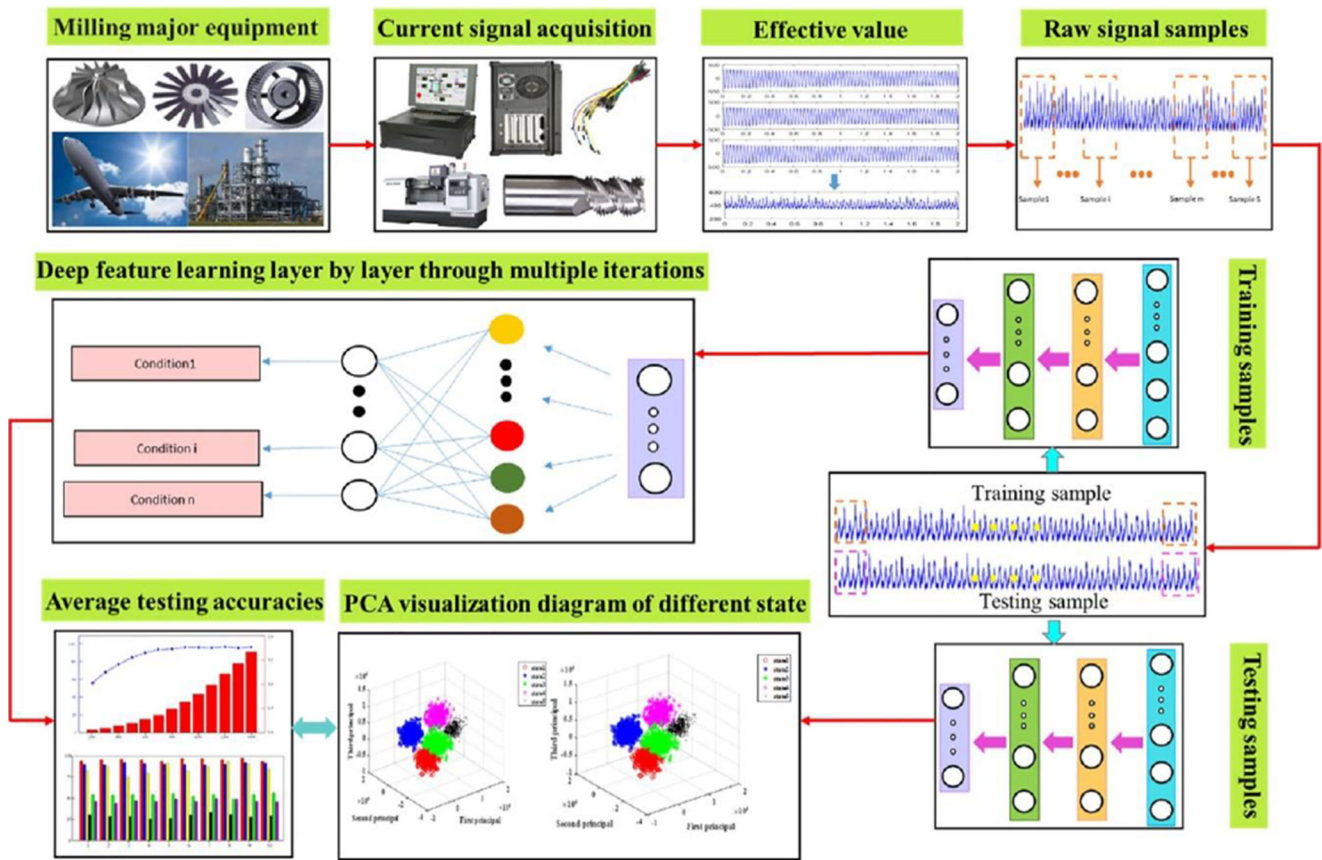


Fig. 12 The AE-based methodology employed at [82] for tool condition monitoring. Current signals were acquired from the CNC machine’s spindle and fed into sparse denoising autoencoders to obtain the low-

dimensional features from the raw current signals. Autoencoders’ outputs were analyzed by an online sequential extreme learning machine for tool condition monitoring ([82] with permission from Elsevier)

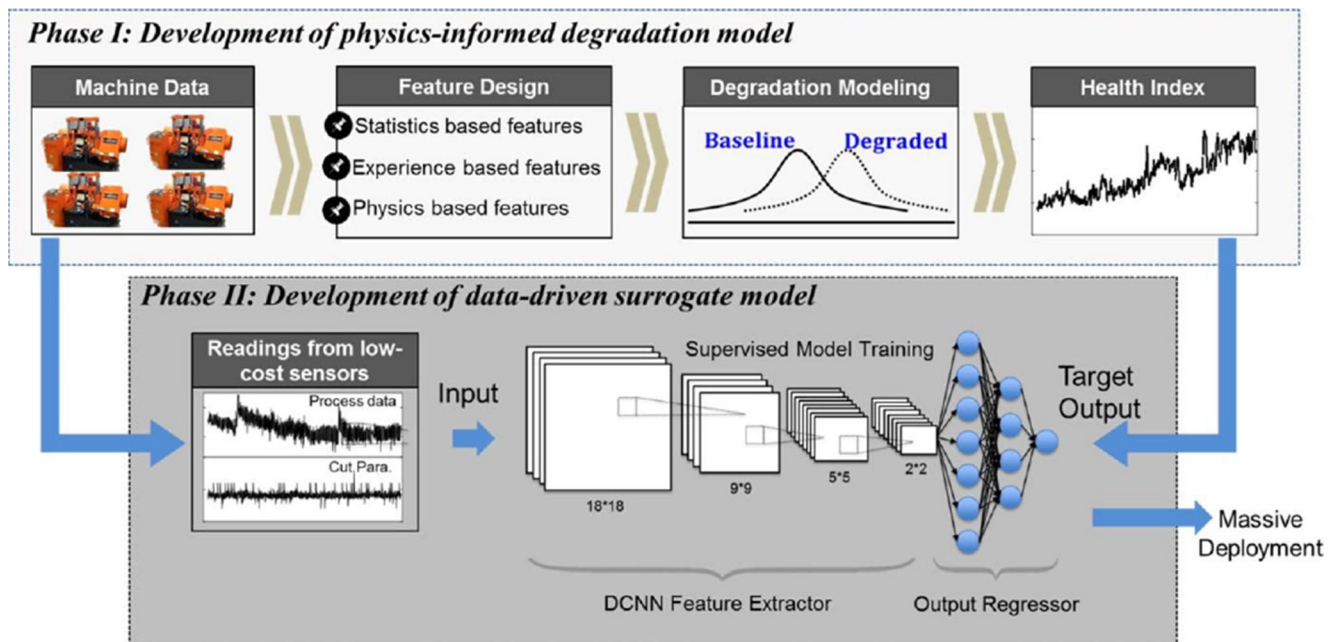


Fig. 13 Developing a low-cost surrogate model for tool degradation monitoring. The model establishes a health index using the statistical-, experience-, and physics-based features. The established index is then

considered the target of the CNN model, which was trained using low-cost sensory data ([104] with permission from Elsevier)

feature matrix for wear prediction in the milling process using CNN. Thus, for each sample data, a total of 54 multi-domain features were extracted to form a column of the original feature matrix. Zhang et al. [101] showed that feature optimization (using recursive feature elimination and cross-validation (RFECV)–based and Isomap-based methods) on the manually extracted features enhanced the performance of CNN for tool wear prediction. Motivated by the similarity between the pixel matrix of high-dimensional image and the raw data matrix of multisensory time-series signal, Huang et al. [106] introduced a reshaped time series stage to represent the multisensory raw signals. Accordingly, the multi-sensory raw signal data were re-shaped and then fed into CNN for tool wear prediction. The method based on reshaped time series convolutional neural network (RTSCNN) was shown to outperform some of the other advanced ML and DL models for tool wear prediction.

In many applications, the time series sensory data were used to construct image-based input data before training the CNN. Reformatting time-series features as images would let the model learn the temporal dependencies on data [91]. Cao et al. [92] discussed that compared to the 1D CNN, the 2-D signal matrix retains more information than a single reconstructed sub-signal. Its associated CNN resulted in higher accuracy than 1D CNN for tool wear state identification. Thus, some research emphasized training CNN using constructed images from the time series sensory data [92–97]. Song et al. [93] used the spindle current clutter signal for tool wear state identification using CNN. They used the Fourier series and the least square method to fit and remove the signal components corresponding to the cutting parameter and extracted the current clutter signal with little dependency on the processing conditions that could best reflect the tool wear condition. Then, they applied image binarization and used the image of the signals as the input of the DL model. Other researchers used mathematical techniques such as Gramian Angular Summation Fields (GASF) to convert the sensory into image data [91, 106]. Martinez-Arellano et al. [95] applied time series imaging on 3-channel force signals using GASF and fed the obtained images to CNN for tool wear classification and achieved classification accuracy above 90% (Fig. 14).

Another approach to obtain image-based dataset from time series sensory data is through time-frequency analysis and imaging using techniques such as the wavelet transform. Zheng and Lin [96] constructed an image from the 1D force signals using the wavelet and short-time Fourier transform. They designed a CNN using the obtained images and discussed that higher network accuracy was obtained when using wavelet transform for image construction from the force signals. Tran et al. [108] utilized continuous wavelet transform (CWT) on the force signals for chatter detection in the milling process. They applied CWT to the segments of force signals acquired during the stable, transitive, and unstable cutting states. The time-frequency images obtained using CWT

were used to train the CNN, and classification accuracy of 99.67% was achieved. Wavelet packet decomposition [97] and Hilbert envelope analysis [92] have also been used to convert the 1-D signal into a 2-D signal matrix. The general finding is that folding the 1-D spectra into 2-D spectral maps enhances the learning ability of the 2-D CNN [97].

Another approach to train CNN is by directly feeding the machining processes' images [107, 109, 110, 135]. Images of machined surface textures were successfully trained CNNs for chatter detection [109] and surface roughness prediction [110]. Tool wear imaging was also used to train CNN for automatic wear state identification in the face milling process [107]. The network model was pre-trained using an automated convolutional encoder (ACE), and its output was set as the initial value of the CNN parameters (Fig. 15) for tool breakage identification.

4.4 Recurrent neural networks

There has been increasing attention to using RNNs, specifically LSTMs, for TCM during machining processes [111–121]. Recently, LSTM was used for chatter detection [122] and surface roughness prediction [123]. For tool wear classification, the hidden state in the model that is the learned representation of input data can be connected to a softmax layer. In contrast, for tool wear prediction or remaining useful life (RUL), prediction regression layers can be linked to the RNNs. Zhao et al. [113] used a deep LSTM network using three-layer LSTMs with dropout on raw signal and obtained a higher performance for tool wear prediction using deep LSTMs comparing to a basic LSTM. They showed that the prediction accuracy is sensitive to the LSTM architectures, which should be defined by trial and error. They also discussed that when better task-specific LSTMs are desired, the acquired signals can be processed using the wavelet transformation method to obtain better meaningful or noise-free signals to be fed into the model. Aghazadeh et al. [114] employed wavelet transformation on the sensory data and fed the extracted features from the time-frequency domain to LSTM for tool wear prediction. They reported that LSTM outperformed MLP with above 10% in prediction accuracy. Cai et al. [115] developed deep LSTMs for tool wear prediction in milling. They combined the temporal features extracted by LSTMs with the process information to form a new input vector. Examples of process information are the material, feed, depth of cut, etc. [90]. They discussed that having the process information combined with the collected sensory data can significantly improve the prediction accuracy when the machining process runs under various operating conditions. They also reported a higher prediction accuracy using deep LSTMs compared to SVR, MLP, and CNN. Combining the process information (working conditions) with the sensory signals was practiced by Zhou et al. [116] for RUL prediction.

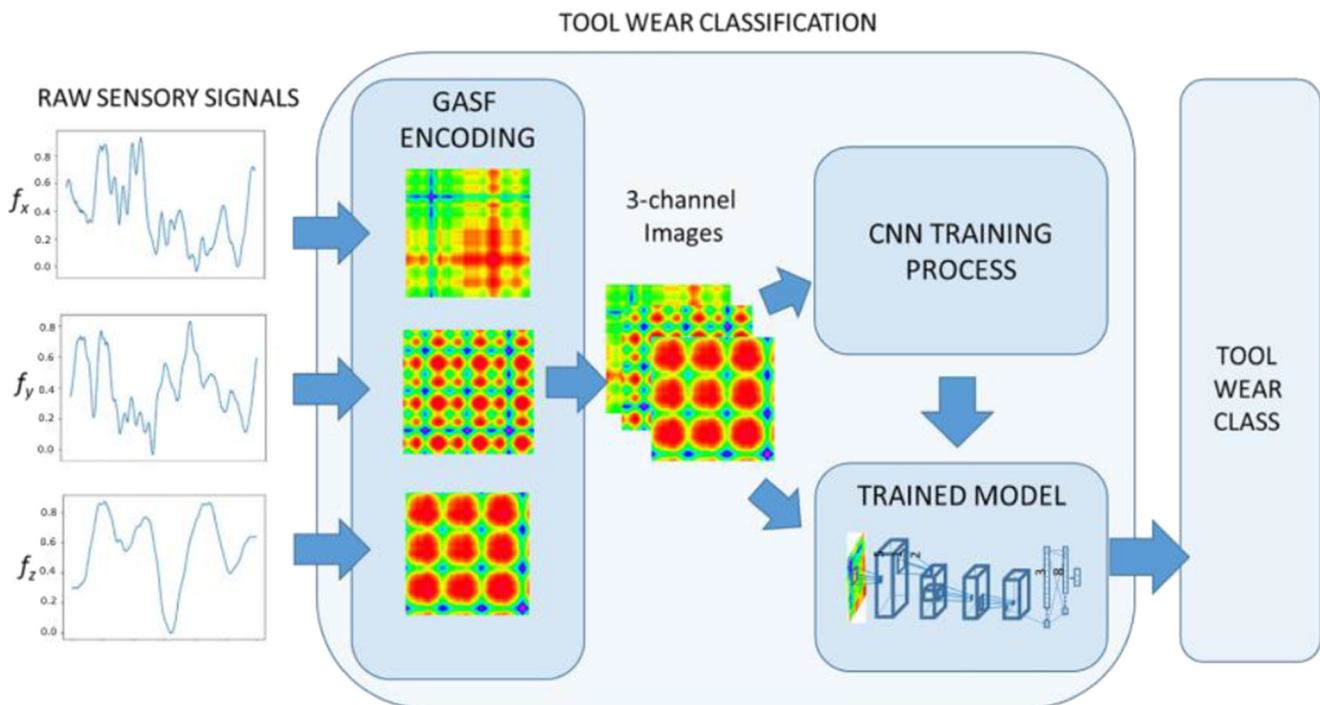


Fig. 14 Schematic of the model for tool wear monitoring through combining time-series imaging and deep learning. Forces in the three dimensions were separately encoded using GASF and formed 3-channel images to train CNN [95]

Hybrid and novel RRN-based networks have been designed to better extract the meaningful feature for process health and condition monitoring. For example, Gugulothu et al. [117] developed an RNN based autoencoder to learn more robust embeddings from the multivariate input time series. Yu et al. [118] applied bidirectional RNNs to the RNN-based autoencoder network for RUL prediction in the milling process and showed the competitiveness of the proposed method. Vashisht and Peng [122] showed that using a low-cost current sensor and LSTM, chatter detection can be achieved with an accuracy of 98%. LSTM was also used to predict the surface roughness in the grinding process using the grinding force, vibration, and acoustic emission signals [123].

4.5 LSTM-CNN

It has been discussed that while LSTM captures the long-term dependency in sequential data, its feature extraction capability is still lower than CNN [127, 128]. This may be an obstacle for LSTM to directly analyze the raw time series data polluted by noise. Xu et al. [72] discussed that, unlike the CNNs, the inherent structure of LSTM does not consider spatial correlation. On the other hand, CNN does not consider the sequential and temporal dependency [72]. Therefore, to overcome the mentioned challenge, combined CNN-LSTM networks were used [124–134]. In these models, the CNN was used for local feature extraction from the original sequential sensory data. The combined CNN-LSTM model was shown to yield superior performance than many other baseline models for tool

wear monitoring [129]. An et al. [127] combined CNN with a stacked bidirectional LSTM (BLSTM) and uni-directional LSTM (ULSTM) for RUL prediction in the milling process. As shown in Fig. 16, CNN was first used for local feature extraction and dimension reduction from raw data. Then two layers of BLSTM and one-layer ULSTM were employed to encode the temporal information. The output of LSTM was connected to regression layers and predicted the RUL with an average prediction accuracy of up to 90%. Such a hybrid model could obtain more in-depth feature engineering with minimal need for expert knowledge for feature selection. A similar approach has been practiced for RUL and tool wear prediction [130, 131]. Niu et al. [130] used a 1D-CNN LSTM network architecture for RUL prediction. The sensory data were decomposed using discrete wavelet transform for denoising, and statistical features were extracted from each sample. It was then fed into the 1D CNN-LSTM network for feature engineering, connected to a fully connected and drop-out layer for RUL prediction. Zhao et al. [131] also showed that the performance of RNNs is improved when combined with CNN for local feature extraction.

Different network architectures can be designed to address the process complexity and extract more meaningful features for process and tool monitoring. Qiao et al. [129] discussed that the features learned by lower layers of a deep learning model are the general features, while the features learned by higher layers are more task-specific and more suitable for tasks such as TCM. Accordingly, they built a BLSTM network on top of a multi-scale convolutional long

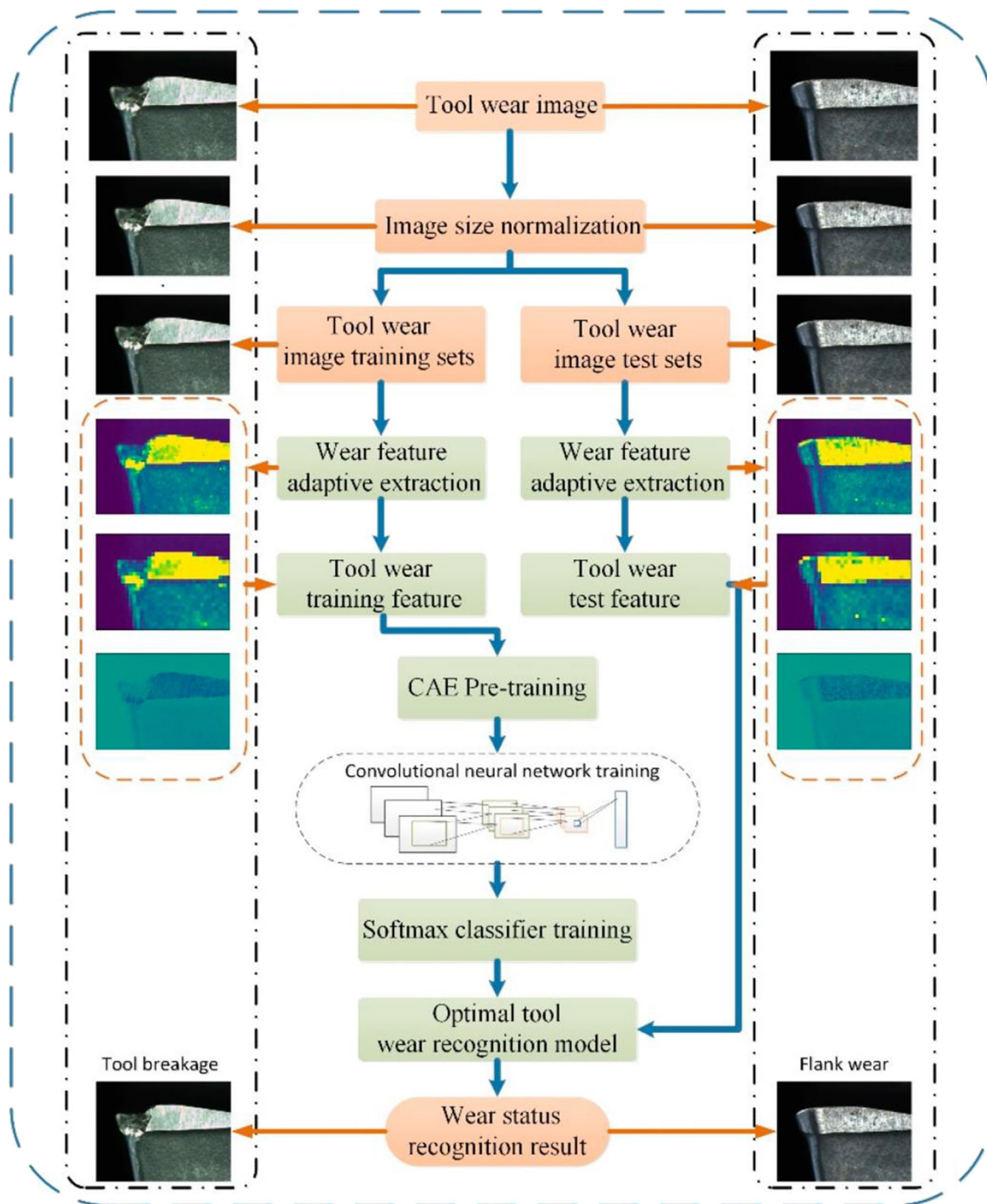


Fig. 15 Tool wear identification process using the tool wear image and CNN [107]

short-term memory model (MCLSTM) to further extract features related to the tool wear prediction tasks. It should be noted that the input data from multi-sensors encompasses multi-scale features, which cannot be captured by the traditional LSTM or traditional CNN due to their lack of multi-scale feature extraction ability [129, 136, 137]. To address that, Qiao et al. [129] employed a multi-scale convolutional long short-term memory model that consisted of different parallel CNN layers. Xu et al. [72] designed a feature-

fusion-based deep model for flank wear prediction. Accordingly, they converted the signals from multiple sensors to images with multi-channels as the model input data. The input data were then fed into different CNNs to extract features in parallel from the multi-source data. The extracted features were then concatenated for multi-sensor information fusion (Fig. 17). The output of this process was then linked into BLSTMs and fully connected layers for tool wear prediction.

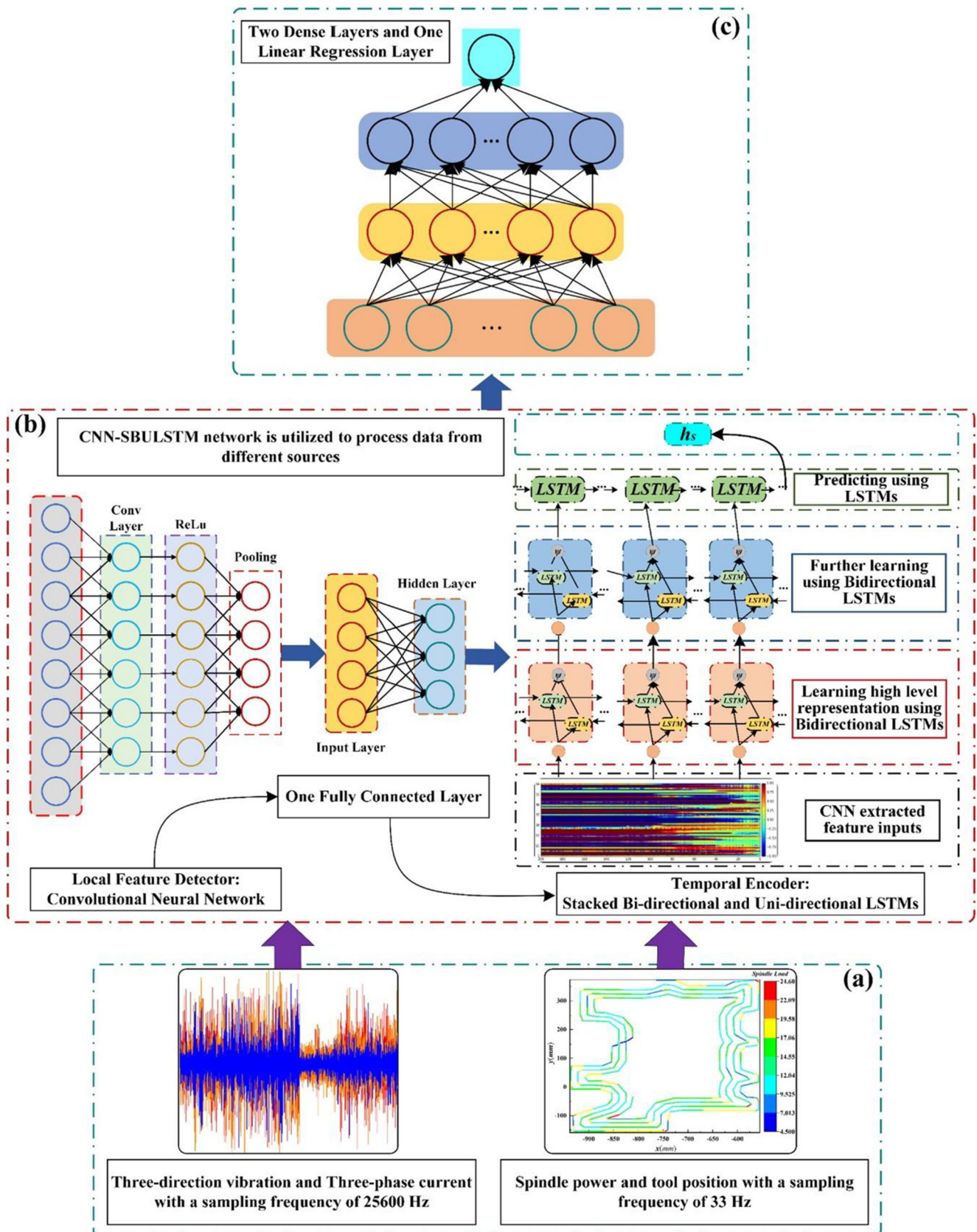


Fig. 16 The CNN-LSTM model for the tool remaining useful life prediction from vibration signals ([127] with permission from Elsevier)

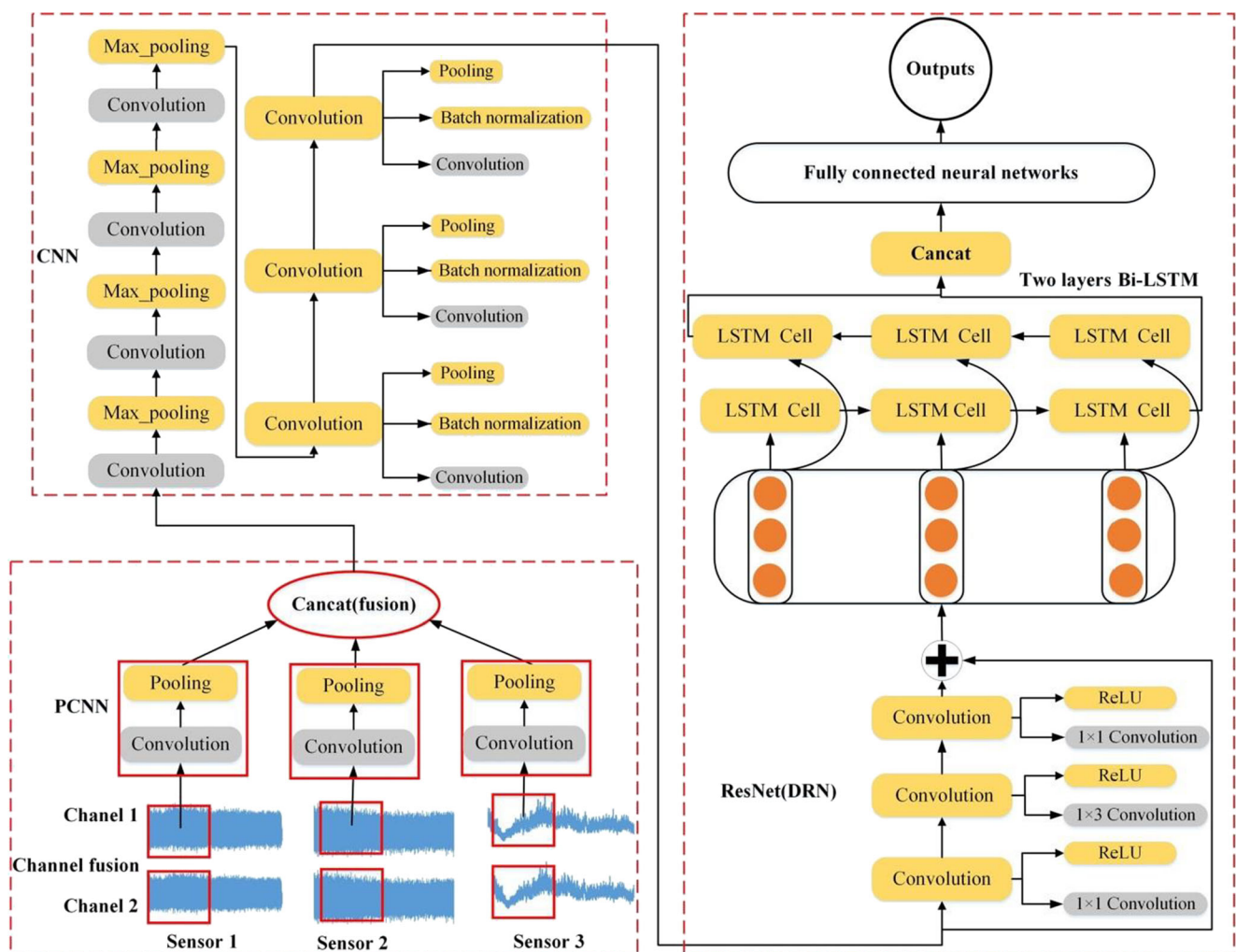


Fig. 17 A feature-fusion-based CNN-LSTM model for flank wear prediction ([72] with permission from Elsevier)

5 Data-driven opportunities

ML and DL have become the core of the data-driven approach for developing intelligent frameworks to monitor the machining processes. The two main steps to developing intelligent monitoring systems that were significantly affected by ML and DL’s advancement were feature selection and intelligent decision-making models. ML and DL have provided smart manufacturing, in general, and intelligent machining, in specific, with unprecedented tools for in-depth and efficient feature engineering. The main opportunities that machine learning may provide for intelligent monitoring are shown in Fig. 18 and summarized as follows.

5.1 Automated feature engineering

Using AI-based methods such as heuristic optimization models for optimal feature selection has been practiced in machining monitoring. Methods based on computational intelligence, such as genetic algorithm and particle swarm

optimization, can find the optimal subset of features that maximizes the monitoring performance. A crucial novelty that ML provided smart manufacturing is in-depth and automated feature engineering. Feature engineering can be different from feature selection as the latter aims to limit the amount of the extracted features by selecting a subset of features that enhance the performance of decision-making models. However, feature engineering aims to build complex models from the raw data to better extract meaningful patterns and information, which cannot be easily achieved using conventional data processing and feature extraction methods. For example, DL yields a high-level meaning in data through high-level data abstraction [66]. Feature engineering may include automatic feature selection. DL-based methods such as the autoencoders or deep belief networks can be used for high-level data abstraction. In the meantime, these models can perform dimensionality reduction and automated feature selection, which has many applications in smart manufacturing [67]. It is also shown in Sections 4.3 and 4.4 that CNNs and RNNs could perform complex feature engineering on image-

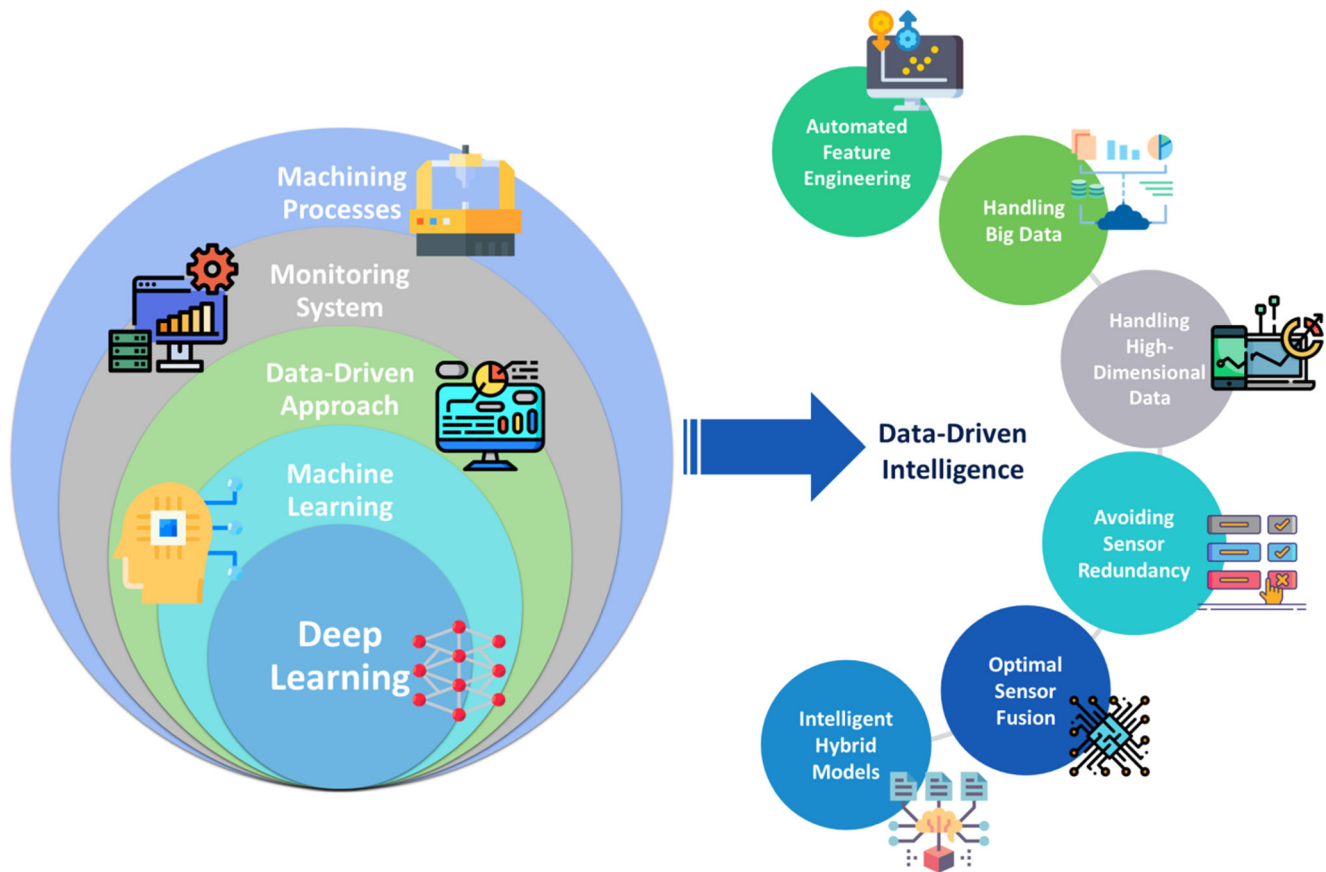


Fig. 18 The elements of monitoring systems based on data-driven and machine learning approaches. The opportunities machine learning provides aimed at achieving such a system are shown

based data and time series. Before the ML advancement, monitoring a model's performance highly depended on the type of signal processing and applying methods such as Fourier or wavelet transforms. While data pre-processing is still a critical aspect in achieving an accurate monitoring model, ML and DL have given more weight to feature engineering in developing a reliable monitoring model.

5.2 Handling big data

Traditional ML models have resulted in efficient performance when dealing with small- and medium-size data. The advancements in deep learning enabled applying data-driven methods to big datasets that can address real industrial challenges.

5.3 Handling high-dimensional data

More important than the impact of sophisticated ML models for handling big-size datasets, ML, in general, and DL, in specific, enabled analyzing the high-dimensional image and sensory data. Tool condition or surface roughness monitoring using image-based techniques can be handled by models such as CNNs. Apart from models such as principal component or linear discriminant analysis, feature selection and

dimensionality reduction can be directly achieved as part of the model, such as what is done in CNNs through pooling layers [65]. Autoencoders, deep belief networks, or self-organizing maps may pre-process the high-dimensional datasets and prepare them to be linked to a more simple and conventional classifier or regression-based decision-making models.

5.4 Avoiding sensor redundancy

Difficulties in optimal feature engineering may restrict a sensor's performance for process monitoring, which thus necessitates having more sensors and performing sensor fusion for better monitoring. However, ML and DL may offer robust tools for better extracting meaningful information from the sensory or image-based data. Combining with techniques such as feature fusion may reduce the need for using multiple source data and avoid sensor redundancy. For example, Ou et al. [82] could develop a framework for tool condition monitoring using low-cost current signals processed with sparse denoising autoencoders. Another example is the CNN model developed by Li et al. [104] for tool degradation monitoring using low-cost sensory data. While sensor fusion can be a powerful means for better understanding a process, it should

not cause sensor redundancy. In other words, sensor fusion should not be practiced due to poor feature engineering.

5.5 Optimal sensor/feature/decision fusion

When multi-source data are needed, ML may facilitate sensor or feature fusion. In-depth feature engineering may let using more simple decision-making models. Decision fusion also can be done using some ML models such as random forest, where the final decision is based on the voting system among a series of weak learners (e.g., decision tree models) [44, 61]. Feature fusion has also been practiced when using DL for machining monitoring. An example includes applying feature fusion to TCM using autoencoders [83, 85] or CNN-LSTM [72].

5.6 Developing intelligent hybrid models

The advancement in ML and DL allowed for designing deeper models for better feature engineering and pattern recognition. Besides, different models could be combined to make intelligent hybrid models. For example, self-organizing maps could be used for unsupervised learning and data clustering, through which tool condition or machine state can be identified. SOM can also be used for feature processing and data reduction and feed another ML model, e.g., for supervised learning. The hybrid SOM-ANFIS model was successfully used for machining monitoring [16]. The most recent trend in developing hybrid models includes combining CNNs and RNNs to enable the model to understand both temporal and spatial features in the dataset [124–134].

6 Data-driven challenges and prospects

There are some challenges (Fig. 19) associated with the data-driven approach, which can be categorized into four main groups as follows:

6.1 Data size challenges

Choosing the suitable model and designing its structure depends directly on the size of the data available. Much of the researches in the literature comprise datasets of small- and medium-size for machining monitoring. The transition toward big data acquisition requires emphasizing choosing the right model and accurate tuning of its hyperparameters. As a rule of thumb, the number of samples should be at least ten times bigger than the number of parameters in a deep learning model [64]. Apart from the size of data, high-dimensional data may need special consideration in terms of feature engineering. Specifically, high-dimensional small datasets are challenging to deal with in terms of generalization and overfitting issues;

thus, data pre-processing or manual feature engineering is needed before using sophisticated ML or DL models for high-dimensional small datasets. This is why many research using DL models for machining and tool monitoring still perform manual feature extraction in their analysis. Another solution is to apply data fusion (augmentation) methods to the small acquired data at the early state of machining experiments and estimate the range of data and generate synthetic data [138], which may not be a suitable approach in many cases (depending on the nature of the process and its variability). Not having access to big data may require performing manual feature engineering and pre-selection before applying DL models for tool condition monitoring [138]. Thus, to employ a fully automated feature engineering approach, intelligent machining needs to embrace data of sufficient size [139].

6.2 Data nature challenges

Data acquisition is a challenging task in manufacturing. In many applications, sensor placement is not readily possible or machine vibration and background noises highly pollute the collected data. Also, many monitoring processes require multi-sensor data acquisition that is not a trivial task as the collected data from different resources may not be synchronized and are challenging to integrate that further complicates utilizing big multi-source data [35]. Another main challenge is that the acquired data are not clean. Kusiak [139] discussed that many manufacturers mistakenly believe that they pose big data for analysis. However, the quality of data is more important than its quantity, since in many cases noisy or irregularly sampled measurements are of little use [139]. Besides, datasets used for anomaly and damage detection are imbalanced in many cases, and the majority of observations belong to machine/process normal conditions. This may create bias in favor of the dominant data group for classification. Up-sampling may not be readily available, and down-sampling may create other data size issues. Another practice is to consider the initial cost and modify the loss function before performing the classification. The classification of imbalanced data can also be turned into the anomaly detection problem using unsupervised learning methods and defining anomaly/damage threshold. In this regard, unsupervised models such as autoencoders have shown promising results for tool condition monitoring [76].

6.3 Model challenges

Model selection depends on many factors, including the monitoring objective, the size, and the dimension of the dataset, the feature engineering work performed on the data, the computation time, and hardware limitations. Flank wear prediction and surface roughness or waviness estimation are made using a regression-based model, for which traditional

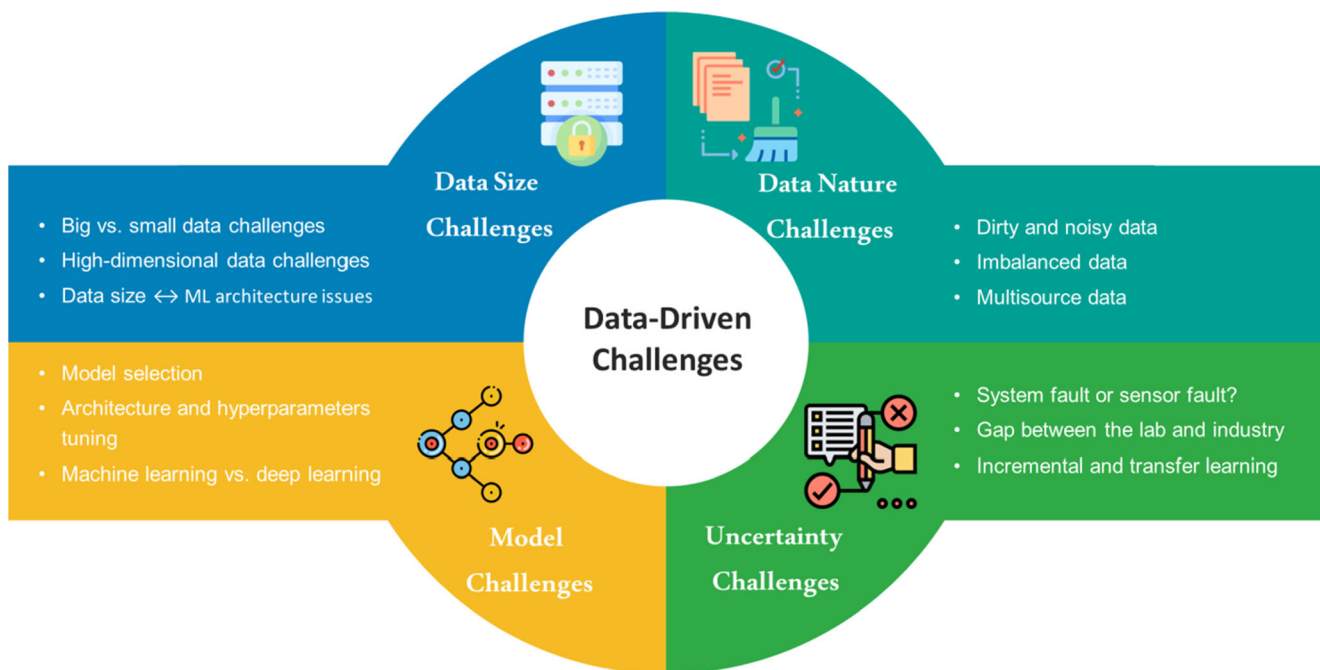


Fig. 19 The challenges associated with the data-driven approach

supervised ML (logistic regression, multilayer perceptron neural networks, neuro-fuzzy, etc.) or deep models such as RNNs or CNNs models can be used. For tool condition monitoring and fault or chatter detection, both supervised and unsupervised models may be used, depending on labeled data availability [67]. While ML-based models such as support vector machines and DL-based models such as CNNs have been widely used for tool wear classification, k-means, fuzzy C-means, self-organizing maps, or hierarchical clustering can be employed for pattern recognition when labeled data is not available. DL-based models such as autoencoders may be utilized in these applications depending on the size and complexity of the acquired data.

Choosing a traditional ML model may necessitate performing more in-depth data pre-processing and feature extraction as it may be challenging to apply them to big-size or high-dimensional data. Depending on the data complexity, DL models may be applied to the data for feature engineering. Models such as autoencoders may be directly used for fault detection and tool condition monitoring [76], or they can be used for data reduction [80] and feature engineering in a hybrid model combined with a supervised regression or classification model.

When dealing with time-series data, where data sequence is critical, RNNs are shown to be a powerful tool [65]. CNNs have also gained popularity for classification and regression on image-based data [65]. There is a growing interest in using DL for intelligent tools and process condition monitoring in recent years. However, many investigations have focused on developing intelligent monitoring systems using ML models. Thus, more focus should be placed on a comparative study

between different ML and DL models in smart manufacturing and intelligent machining. It should be noted that the potential superiority of DL over ML models directly depends on the size and quality of the data. Thus, it is very challenging to generalize their performance and conclude if one approach outperforms the other one. It should also be mentioned that deep models are more difficult to be tuned [67], and the literature lacks a systematic approach for designing the architecture and hyperparameter tuning of deep models.

Finally, it is worth emphasizing that feature selection should not be considered a separate step from the intelligent decision-making model development. An integrated feature engineering combined decision-making approach, which is the case for deep models, should be adopted, and data-driven opportunities applied to automatic feature engineering should be embraced. It should be considered that thorough and in-depth feature engineering may eliminate the need for selecting complex sensory systems or decision-making models and better unveil the hidden pattern in the acquired data.

6.4 Uncertainty challenges

The main concern under this category is whether the research conducted on a laboratory scale to assess a sensing technology or the effectiveness of a data-driven model can reflect the process uncertainty widely seen in real-life situations. Developing a data-driven model using a single source data may not consider sensor failure. The developed model should also consider sensor failure and investigate the difference between the sensor fault and the system fault [140]. Machining

under harsh conditions may experience sensor failure, which could be confused by the process fault. Contact sensors such as a thermometer or vibration-based transducers may be damaged by the cooling fluid, lubricant, dust, machine motor, or blade vibration, etc. [141]. Thus, it is critical to consider the sensor failure and its impact on the performance of the monitoring model. Also, it necessitates investigating the multi-source data acquisition for monitoring even though it may cause data redundancy.

Despite the uncertainty associated with the models generated in laboratory-scale investigations, these models could still be used as a baseline to develop and train models from the real-life collected data. Since the data collection is dramatically growing in size and considering the challenges about uncertainty in the machine and process health and, consequently, the veracity of the data used for model generation, incremental learning should be practiced to extend the existing intelligent models. Transfer learning can also effectively deal with the scarcity of labeled prognosis data in industrial setups [142]. Transfer learning has been practiced for defect detection and fault diagnosis [143–145] and in few studies concerning TCM [86, 98]. Another factor impacting the veracity of the developed models is the machine-to-machine interaction in manufacturing units. The machine-to-machine interaction may cause data variation, which can affect the veracity of the developed expert models. Lee et al. [5] argued that expert models should be assessed so that a developed model does not interfere with other components in a manufacturing system.

A practical approach to transfer advanced monitoring and prediction algorithms from the research laboratory to the industry is through investing in cloud computing due to its enhanced computing efficiency and data storage capability in the cloud center [129]. Caggiano [146] proposed a cloud-based manufacturing process monitoring framework for tool condition monitoring during machining. To reduce the network delay, increase reliability, and defend against security attacks in cloud computing-based systems, a fog computing layer can be used in process monitoring systems [129]. Qiao et al. [129] used a fog computing architecture consisting of an edge computing layer, a fog computing layer, and a cloud computing layer for TCM and showed that introducing a fog computing layer reduced the response latency of the monitoring system. Future studies in cloud and fog computing can further tackle the uncertainty challenges concerning developing online monitoring systems. IoT is paving the way for online monitoring using multi-source data [35–37] that should be further highlighted in intelligent machining monitoring.

7 Conclusions

Data-driven methods have transitioned machining monitoring into embracing machine learning and deep learning

techniques for developing intelligent systems for process health and condition monitoring. Machine learning, in general, and deep learning, in particular, have significantly impacted feature engineering and expert decision-making by allowing for automated feature selection, handling big and high-dimensional data, and avoiding sensor redundancy. It also facilitates optimal data fusion and the development of intelligent hybrid models that can be used for descriptive analytics for product quality inspection, diagnostic analytics for fault assessment, and predictive analytics for defect prognosis.

Despite its huge opportunities, there are still challenges facing a data-driven industrial approach, especially concerning the size and quality of the acquired data. The deep learning concept and its opportunities and limitation should be further investigated and compared with the traditional machine learning models. Comparative studies among different basic deep models and more complex hybrid models should be performed. Small data challenges should be studied by practicing data fusion methods and comparative studies among machine learning versus deep learning. The fusion concept at different levels of sensor, feature, and decision should be assessed and compared. The gap between the laboratory scale results and real-life conditions should be emphasized by investigating the process uncertainty and applying cloud computing. Incremental and transfer learning can play a crucial role in bridging the gap from the laboratory to the industry. To fully comprehend the power of data-driven methods, intelligent machining should focus on big data acquisition. Also, the crucial role of feature engineering should be acknowledged by developing an attitude that integrates feature selection and expert decision-making to better unveil the hidden patterns in data for intelligent monitoring.

Author contribution Conceptualization: Vahid Nasir (V. N.) and Farrokh Sassani (F. S.); literature review (V. N.); manuscript writing (V. N. and F. S.); editing and final review (F. S.).

Funding Not applicable

Availability of data and materials Not applicable. There is no data associated with this manuscript.

Declarations

Ethical approval Not applicable

Consent to participate Not applicable

Consent to publish Not applicable

Competing interests The authors declare no competing interests.

References

- Zhong RY, Ge W (2018) Internet of things enabled manufacturing: a review. *Int J Agile Syst Manag* 11(2):126–154
- Yang C, Shen W, Wang X (2018) The internet of things in manufacturing: key issues and potential applications. *IEEE Syst Man Cybern Mag* 4(1):6–15
- Yang C, Shen W, Wang X (2016, May) Applications of Internet of Things in manufacturing. In: 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp 670–675
- Siderska J, Jadaan KS (2018) Cloud manufacturing: a service-oriented manufacturing paradigm. A review paper. *Eng Manag Produc Serv* 10(1):22–31
- Lee J, Davari H, Singh J, Pandhare V (2018) Industrial artificial intelligence for Industry 4.0-based manufacturing systems. *Manuf Lett* 18:20–23
- Li BH, Hou BC, Yu WT, Lu XB, Yang CW (2017) Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers Inf Technol Electron Eng* 18(1):86–96
- Kumar SL (2017) State of the art-intense review on artificial intelligence systems application in process planning and manufacturing. *Eng Appl Artif Intell* 65:294–329
- Tao F, Qi Q, Liu A, Kusiak A (2018) Data-driven smart manufacturing. *J Manuf Syst* 48:157–169
- Lin YC, Wu KD, Shih WC, Hsu PK, Hung JP (2020) Prediction of surface roughness based on cutting parameters and machining vibration in end milling using regression method and artificial neural network. *Appl Sci* 10(11):3941
- Bhogal SS, Sindhu C, Dhami SS, Pabla BS (2015) Minimization of surface roughness and tool vibration in CNC milling operation. *J Opt* 2015:1–13. <https://doi.org/10.1155/2015/192030>
- Silge, M., & Sattel, T. (2018). Design of contactlessly powered and piezoelectrically actuated tools for non-resonant vibration assisted milling. In *Actuators* (Vol. 7, 2, p. 19). Multidisciplinary Digital Publishing Institute.
- Omair M, Sarkar B, Cárdenas-Barrón LE (2017) Minimum quantity lubrication and carbon footprint: a step towards sustainability. *Sustainability* 9(5):714
- Wang B, Liu Z (2018) Influences of tool structure, tool material and tool wear on machined surface integrity during turning and milling of titanium and nickel alloys: a review. *Int J Adv Manuf Technol* 98(5-8):1925–1975
- Yeganefar A, Niknam SA, Asadi R (2019) The use of support vector machine, neural network, and regression analysis to predict and optimize surface roughness and cutting forces in milling. *Int J Adv Manuf Technol* 105(1):951–965
- Nasir V, Mohammadpanah A, Cool J (2018) The effect of rotation speed on the power consumption and cutting accuracy of guided circular saw: experimental measurement and analysis of saw critical and flutter speeds. *Wood Mater Sci Eng* 15(3):1–7
- Nasir V, Cool J (2020) Intelligent wood machining monitoring using vibration signals combined with self-organizing maps for automatic feature selection. *Int J Adv Manuf Technol* 108:1811–1825. <https://doi.org/10.1007/s00170-020-05505-5>
- Nasir V, Cool J (2019) Optimal power consumption and surface quality in the circular sawing process of Douglas-fir wood. *Eur J Wood Wood Produc* 77(4):609–617
- Serin G, Sener B, Ozbayoglu AM, Unver HO (2020) Review of tool condition monitoring in machining and opportunities for deep learning. *Int J Adv Manuf Technol*:1–22
- Wang M, Wang J (2012) CHMM for tool condition monitoring and remaining useful life prediction. *Int J Adv Manuf Technol* 59(5-8):463–471
- Brecher C, Esser M, Witt S (2009) Interaction of manufacturing process and machine tool. *CIRP Ann* 58(2):588–607
- Chen W, Liu H, Sun Y, Yang K, Zhang J (2017) A novel simulation method for interaction of machining process and machine tool structure. *Int J Adv Manuf Technol* 88(9-12):3467–3474
- Quintana G, Ciurana J (2011) Chatter in machining processes: a review. *Int J Mach Tools Manuf* 51(5):363–376
- Hegab HA, Darras B, Kishawy HA (2018) Towards sustainability assessment of machining processes. *J Clean Prod* 170:694–703
- Mia M, Gupta MK, Singh G, Królczyk G, Pimenov DY (2018) An approach to cleaner production for machining hardened steel using different cooling-lubrication conditions. *J Clean Prod* 187:1069–1081
- Zhou Z, Yao B, Xu W, Wang L (2017) Condition monitoring towards energy-efficient manufacturing: a review. *Int J Adv Manuf Technol* 91(9-12):3395–3415
- Said Z, Gupta M, Hegab H, Arora N, Khan AM, Jamil M, Bellos E (2019) A comprehensive review on minimum quantity lubrication (MQL) in machining processes using nano-cutting fluids. *Int J Adv Manuf Technol* 105(5-6):2057–2086
- Nasir V, Cool J (2020) Characterization, optimization, and acoustic emission monitoring of airborne dust emission during wood sawing. *Int J Adv Manuf Technol* 109(9):2365–2375. <https://doi.org/10.1007/s00170-020-05842-5>
- Licow R, Chuchala D, Deja M, Orłowski KA, Taube P (2020) Effect of pine impregnation and feed speed on sound level and cutting power in wood sawing. *J Clean Prod* 272:122833
- Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Ann* 59(2):717–739
- Abellan-Nebot JV, Subirón FR (2010) A review of machining monitoring systems based on artificial intelligence process models. *Int J Adv Manuf Technol* 47(1-4):237–257
- Zhu K, San Wong Y, Hong GS (2009) Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results. *Int J Mach Tools Manuf* 49(7-8):537–553
- Lauro CH, Brandão LC, Baldo D, Reis RA, Davim JP (2014) Monitoring and processing signal applied in machining processes—a review. *Measurement* 58:73–86
- Kusiak A (2019) Fundamentals of smart manufacturing: a multi-thread perspective. *Annu Rev Control* 47:214–220
- Kim DH, Kim TJ, Wang X, Kim M, Quan YJ, Oh JW et al (2018) Smart machining process using machine learning: a review and perspective on machining industry. *Int J Precis Eng Manuf Green Technol* 5(4):555–568
- Ayvaz S, Alpay K (2021) Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time. *Expert Syst Appl* 173:114598
- Morariu C, Morariu O, Răileanu S, Borangiu T (2020) Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Comput Ind* 120:103244
- Adi E, Anwar A, Baig Z, Zeadally S (2020) Machine learning and data analytics for the IoT. *Neural Comput & Applic* 32:16205–16233
- Peng ZK, Chu FL (2004) Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mech Syst Signal Process* 18(2):199–221
- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Comput Electr Eng* 40(1):16–28
- Nasir V, Cool J, Sassani F (2019) Acoustic emission monitoring of sawing process: artificial intelligence approach for optimal sensory feature selection. *Int J Adv Manuf Technol* 102(9-12):4179–4197. <https://doi.org/10.1007/s00170-019-03526-3>
- Sick B (2002) On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mech Syst Signal Process* 16(4):487–546

42. Roth JT, Djurdjanovic D, Yang X, Mears L, Kurfess T (2010) Quality and inspection of machining operations: tool condition monitoring. *J Manuf Sci Eng* 132(4)
43. Stavropoulos P, Papacharalampopoulos A, Vasiliadis E, Chryssolouris G (2016) Tool wear predictability estimation in milling based on multi-sensorial data. *Int J Adv Manuf Technol* 82(1-4):509–521
44. Nasir V, Kooshkbaghi M, Cool J, Sassani F (2020) Cutting tool temperature monitoring in circular sawing: measurement and multi-sensor feature fusion-based prediction. *Int J Adv Manuf Technol* 112:2413–2424. <https://doi.org/10.1007/s00170-020-06473-6>
45. Nasir V, Cool J, Sassani F (2019) Intelligent machining monitoring using sound signal processed with the wavelet method and a self-organizing neural network. *IEEE Robot Autom Lett* 4(4): 3449–3456
46. Bhuiyan MSH, Choudhury IA, Dahari M (2014) Monitoring the tool wear, surface roughness and chip formation occurrences using multiple sensors in turning. *J Manuf Syst* 33(4):476–487
47. Ahmadi H, Dumont G, Sassani F, Tafreshi R (2003) Performance of informative wavelets for classification and diagnosis of machine faults. *Int J Wavelets Multiresolution Inf Process* 1(03): 275–289
48. Tafreshi R, Sassani F, Ahmadi H, Dumont G (2009) An approach for the construction of entropy measure and energy map in machine fault diagnosis. *J Vib Acoust* 131(2)
49. Wuest T, Weimer D, Irgens C, Thoben KD (2016) Machine learning in manufacturing: advantages, challenges, and applications. *Produc Manuf Res* 4(1):23–45
50. Hermann G (1990) Artificial intelligence in monitoring and the mechanics of machining. *Comput Ind* 14(1-3):131–135
51. Rangwala SS (1987) Integration of sensors via neural networks for detection of tool wear states. *Proc Winter Annu Meet ASME* 25: 109–120
52. Dornfeld DA, DeVries MF (1990) Neural network sensor fusion for tool condition monitoring. *CIRP Ann* 39(1):101–105
53. Rangwala, S., & Dornfeld, D. (1990). Sensor integration using neural networks for intelligent tool condition monitoring, 219–228.
54. Park KS, Kim SH (1998) Artificial intelligence approaches to determination of CNC machining parameters in manufacturing: a review. *Artif Intell Eng* 12(1-2):127–134
55. Dimla DE Jr, Lister PM, Leighton NJ (1997) Neural network solutions to the tool condition monitoring problem in metal cutting—a critical review of methods. *Int J Mach Tools Manuf* 37(9):1219–1241
56. Ademujimi TT, Brundage MP, Prabhu VV (2017, September) A review of current machine learning techniques used in manufacturing diagnosis. In: IFIP International Conference on Advances in Production Management Systems. Springer, Cham, pp 407–415
57. Panchal G, Ganatra A, Shah P, Panchal D (2011) Determination of over-learning and over-fitting problem in backpropagation neural network. *Int J Soft Comput* 2(2):40–51
58. Montavon, G., Orr, G., & Müller, K. R. (Eds.). (2012). *Neural networks: tricks of the trade* (Vol. 7700). Springer.
59. Lopez C (1999) Looking inside the ANN “black box”: classifying individual neurons as outlier detectors. In: IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339, vol 2. IEEE, pp 1185–1188
60. Palczewska A, Palczewski J, Robinson RM, Neagu D (2014) Interpreting random forest classification models using a feature contribution method. In: Integration of reusable systems. Springer, Cham, pp 193–218
61. Nasir V, Kooshkbaghi M, Cool J (2020) Sensor fusion and random forest modeling for identifying frozen and green wood during lumber manufacturing. *Manuf Lett* 26:53–58
62. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828
63. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR (2018) Deep learning for healthcare applications based on physiological signals: a review. *Comput Methods Prog Biomed* 161:1–13
64. Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2018) Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 19(6):1236–1246
65. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
66. Khan S, Yairi T (2018) A review on the application of deep learning in system health management. *Mech Syst Signal Process* 107: 241–265
67. Wang J, Ma Y, Zhang L, Gao RX, Wu D (2018) Deep learning for smart manufacturing: methods and applications. *J Manuf Syst* 48: 144–156
68. Zhang N, Ding S, Zhang J, Xue Y (2018) An overview on restricted Boltzmann machines. *Neurocomputing* 275:1186–1199
69. Fu Y, Zhang Y, Qiao H, Li D, Zhou H, Leopold J (2015) Analysis of feature extracting ability for cutting state monitoring using deep belief networks. *Procedia Cirp* 31(Suppl. C):29–34
70. Chen Y, Jin Y, Jiri G (2018) Predicting tool wear with multi-sensor data using deep belief networks. *Int J Adv Manuf Technol* 99(5-8):1917–1926
71. Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 31(7):1235–1270
72. Xu X, Tao Z, Ming W, An Q, Chen M (2020) Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion. *Measurement* 165: 108086
73. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 115:213–237
74. Hahn TV, Mechefske CK (2021) Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder. *Int J Hydromechatron* 4(1):69–98
75. Xiangyu Z, Lilan L, Xiang W, Bowen F (2021) Tool wear online monitoring method based on DT and SSAE-PHMM. *J Comput Inf Sci Eng* 21(3):034501
76. Dou J, Xu C, Jiao S, Li B, Zhang J, Xu X (2020) An unsupervised online monitoring method for tool wear using a sparse auto-encoder. *Int J Adv Manuf Technol* 106(5):2493–2507
77. Kim J, Lee H, Jeon JW, Kim JM, Lee HU, Kim S (2020) Stacked auto-encoder based CNC tool diagnosis using discrete wavelet transform feature extraction. *Processes* 8(4):456
78. Moldovan OG, Dzitac S, Moga I, Vesselenyi T, Dzitac I (2017) Tool-wear analysis using image processing of the tool flank. *Symmetry* 9(12):296
79. Ochoa LEE, Quinde IBR, Sumba JPC, Guevara AV Jr, Morales-Menendez R (2019) New approach based on autoencoders to monitor the tool wear condition in HSM. *IFAC-PapersOnLine* 52(11):206–211
80. Proteau A, Zemouri R, Tahan A, Thomas M (2020) Dimension reduction and 2D-visualization for early change of state detection in a machining process with a variational autoencoder approach. *Int J Adv Manuf Technol* 111(11):3597–3611
81. Ou J, Li H, Huang G, Zhou Q (2020) A novel order analysis and stacked sparse auto-encoder feature learning method for milling tool wear condition monitoring. *Sensors* 20(10):2878

82. Ou J, Li H, Huang G, Yang G (2021) Intelligent analysis of tool wear state using stacked denoising autoencoder with online sequential-extreme learning machine. *Measurement* 167:108153
83. Shi C, Panoutsos G, Luo B, Liu H, Li B, Lin X (2018) Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing. *IEEE Trans Ind Electron* 66(5):3794–3803
84. He Z, Shi T, Xuan J, Li T (2021) Research on tool wear prediction based on temperature signals and deep learning. *Wear* 478:203902
85. Shi C, Luo B, He S, Li K, Liu H, Li B (2019) Tool wear prediction via multidimensional stacked sparse autoencoders with feature fusion. *IEEE Trans Ind Informatics* 16(8):5150–5159
86. Sun C, Ma M, Zhao Z, Tian S, Yan R, Chen X (2018) Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. *IEEE Trans Ind Informatics* 15(4):2416–2425
87. Dun Y, Zhus L, Yan B, Wang S (2021) A chatter detection method in milling of thin-walled TC4 alloy workpiece based on auto-encoding and hybrid clustering. *Mech Syst Signal Process* 158:107755
88. Yu J, Liu G (2020) Knowledge-based deep belief network for machining roughness prediction and knowledge discovery. *Comput Ind* 121:103262
89. Brili N, Ficko M, Klančnik S (2021) Automatic identification of tool wear based on thermography and a convolutional neural network during the turning process. *Sensors* 21(5):1917
90. Lee CH, Jwo JS, Hsieh HY, Lin CS (2020) An intelligent system for grinding wheel condition monitoring based on machining sound and deep learning. *IEEE Access* 8:58279–58289
91. Gouarir A, Martínez-Arellano G, Terrazas G, Benardos P, Ratchev SJPC (2018) In-process tool wear prediction system based on machine learning techniques and force analysis. *Procedia CIRP* 77:501–504
92. Cao XC, Chen BQ, Yao B, He WP (2019) Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Comput Ind* 106:71–84
93. Song K, Wang M, Liu L, Wang C, Zan T, Yang B (2020) Intelligent recognition of milling cutter wear state with cutting parameter independence based on deep learning of spindle current clutter signal. *Int J Adv Manuf Technol* 109(3):929–942
94. Terrazas G, Martínez-Arellano G, Benardos P, Ratchev S (2018) Online tool wear classification during dry machining using real time cutting force measurements and a CNN approach. *J Manuf Mater Process* 2(4):72
95. Martínez-Arellano G, Terrazas G, Ratchev S (2019) Tool wear classification using time series imaging and deep learning. *Int J Adv Manuf Technol* 104(9):3647–3662
96. Zheng, H., & Lin, J. (2019). A deep learning approach for high speed machining tool wear monitoring. In 2019 3rd International Conference on Robotics and Automation Sciences (ICRAS) (pp. 63–68). IEEE.
97. 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
98. Mamledesai H, Soriano MA, Ahmad R (2020) A qualitative tool condition monitoring framework using convolution neural network and transfer learning. *Appl Sci* 10(20):7298
99. Zhi G, He D, Sun W, Yuqing Z, Pan X, Gao C (2021) An edge-labeling graph neural network method for tool wear condition monitoring using wear image with small samples. *Meas Sci Technol* 32:064006
100. Xu X, Wang J, Zhong B, Ming W, Chen M (2021) Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism. *Measurement* 177:109254
101. Zhang X, Wang S, Li W, Lu X (2021) Heterogeneous sensors-based feature optimisation and deep learning for tool wear prediction. *Int J Adv Manuf Technol*:1–25
102. Ambadekar PK, Choudhari CM (2020) CNN based tool monitoring system to predict life of cutting tool. *SN Appl Sci* 2(5):1–11
103. Xu X, Wang J, Ming W, Chen M, An Q (2021) In-process tap tool wear monitoring and prediction using a novel model based on deep learning. *Int J Adv Manuf Technol* 112:453–466
104. Li P, Jia X, Feng J, Zhu F, Miller M, Chen LY, Lee J (2020) A novel scalable method for machine degradation assessment using deep convolutional neural network. *Measurement* 151:107106
105. Huang Z, Zhu J, Lei J, Li X, Tian F (2019) Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. *J Intell Manuf*:1–14
106. Huang Z, Zhu J, Lei J, Li X, Tian F (2019) Tool wear predicting based on multisensory raw signals fusion by reshaped time series convolutional neural network in manufacturing. *IEEE Access* 7:178640–178651
107. 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
108. Tran MQ, Liu MK, Tran QV (2020) Milling chatter detection using scalogram and deep convolutional neural network. *Int J Adv Manuf Technol* 107(3):1505–1516
109. Zhu W, Zhuang J, Guo B, Teng W, Wu F (2020) An optimized convolutional neural network for chatter detection in the milling of thin-walled parts. *Int J Adv Manuf Technol* 106(9):3881–3895
110. Rifai AP, Aoyama H, Tho NH, Dawal SZM, Masruroh NA (2020) Evaluation of turned and milled surfaces roughness using convolutional neural network. *Measurement* 161:107860
111. Liu Y, Hu X, Jin J (2019) Remaining useful life prediction of cutting tools based on deep adversarial transfer learning. In: *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*, pp 434–439
112. Liu H, Liu Z, Jia W, Lin X, Zhang S (2020) A novel transformer-based neural network model for tool wear estimation. *Meas Sci Technol* 31(6):065106
113. Zhao R, Wang J, Yan R, Mao K (2016) Machine health monitoring with LSTM networks. In: 2016 10th international conference on sensing technology (ICST), IEEE, pp 1–6
114. Aghazadeh F, Tahan AS, Thomas M (2019, July) Tool condition monitoring method in milling process using wavelet transform and long short-term memory. In *Surveillance, Vishno and AVE conferences*
115. Cai W, Zhang W, Hu X, Liu Y (2020) A hybrid information model based on long short-term memory network for tool condition monitoring. *J Intell Manuf* 31(6):1497–1510
116. Zhou JT, Zhao X, Gao J (2019) Tool remaining useful life prediction method based on LSTM under variable working conditions. *Int J Adv Manuf Technol* 104(9):4715–4726
117. Gugulothu N, Tv V, Malhotra P, Vig L, Agarwal P, Shroff G (2017) Predicting remaining useful life using time series embeddings based on recurrent neural networks. *arXiv preprint arXiv 1709:01073*
118. Yu W, Kim IY, Mechefske C (2019) Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mech Syst Signal Process* 129:764–780
119. Wu X, Li J, Jin Y, Zheng S (2020) Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *Int J Adv Manuf Technol* 106(9):4391–4399
120. Wang, J., Yan, J., Li, C., Gao, R. X., & Zhao, R. (2019). Deep heterogeneous GRU model for predictive analytics in smart manufacturing: application to tool wear prediction. *Comput Ind*, 111, 1–14, 1.

121. Marani M, Zeinali M, Songmene V, Mechefske CK (2021) Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling. *Measurement* 177:109329
122. Vashisht RK, Peng Q (2021) Online chatter detection for milling operations using LSTM neural networks assisted by motor current signals of ball screw drives. *J Manuf Sci Eng* 143(1)
123. Guo W, Wu C, Ding Z, Zhou Q (2021) Prediction of surface roughness based on a hybrid feature selection method and long short-term memory network in grinding. *Int J Adv Manuf Technol* 112(9):2853–2871
124. Chen Q, Xie Q, Yuan Q, Huang H, Li Y (2019) Research on a real-time monitoring method for the wear state of a tool based on a convolutional bidirectional LSTM model. *Symmetry* 11(10):1233
125. Ma J, Luo D, Liao X, Zhang Z, Huang Y, Lu J (2021) Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning. *Measurement* 173:108554
126. Zhang X, Lu X, Li W, Wang S (2021) Prediction of the remaining useful life of cutting tool using the Hurst exponent and CNN-LSTM. *Int J Adv Manuf Technol* 112(7):2277–2299
127. An Q, Tao Z, Xu X, El Mansori M, Chen M (2020) A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. *Measurement* 154:107461
128. Babu GS, Zhao P, Li XL (2016) Deep convolutional neural network based regression approach for estimation of remaining useful life. In: *International conference on database systems for advanced applications*. Springer, Cham, pp 214–228
129. Qiao H, Wang T, Wang P (2020) A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing. *Int J Adv Manuf Technol* 108:2367–2384
130. Niu, J., Liu, C., Zhang, L., & Liao, Y. (2019). Remaining useful life prediction of machining tools by 1D-CNN LSTM network. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1056-1063). IEEE.
131. Zhao R, Yan R, Wang J, Mao K (2017) Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* 17(2):273
132. Qiao H, Wang T, Wang P, Qiao S, Zhang L (2018) A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series. *Sensors* 18(9):2932
133. Wang B, Lei Y, Yan T, Li N, Guo L (2020) Recurrent convolutional neural network: a new framework for remaining useful life prediction of machinery. *Neurocomputing* 379:117–129
134. Zhang X, Lu X, Li W, Wang S (2021) Prediction of the remaining useful life of cutting tool using the Hurst exponent and CNN-LSTM. *Int J Adv Manuf Technol*:1–23
135. Misaka T, Herwan J, Kano S, Sawada H, Furukawa Y (2020) Deep neural network-based cost function for metal cutting data assimilation. *Int J Adv Manuf Technol* 107(1):385–398
136. Qiao H, Wang T, Wang P, Zhang L, Xu M (2019) An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions. *IEEE Access* 7:118954–118964
137. Jiang G, He H, Yan J, Xie P (2018) Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans Ind Electron* 66(4):3196–3207
138. Li DC, Wen IH, Chen WC (2016) A novel data transformation model for small data-set learning. *Int J Prod Res* 54(24):7453–7463
139. Kusiak A (2017 Apr) Smart manufacturing must embrace big data. *Nature*. 544(7648):23–25
140. Taiebat M, Sassani F (2017 Sep) Distinguishing sensor faults from system faults by utilizing minimum sensor redundancy. *Trans Can Soc Mech Eng* 41(3):469–487
141. Nasir V, Cool J (2020) A review on wood machining: characterization, optimization, and monitoring of the sawing process. *Wood Material Sci Eng* 15(1):1–16
142. Diez-Olivan A, Del Ser J, Galar D, Sierra B (2019 Oct 1) Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0. *Inf Fusion* 50:92–111
143. Ferguson MK, Ronay AK, Lee YTT, Law KH (2018) Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. *Smart Sustain Manuf Systems* 2:20180033. <https://doi.org/10.1520/SSMS20180033>
144. Imoto K, Nakai T, Ike T, Haruki K, Sato Y (2018) A CNN-based transfer learning method for defect classification in semiconductor manufacturing. In: *2018 International Symposium on Semiconductor Manufacturing (ISSM)*. IEEE, pp 1–3
145. Wang P, Gao RX (2020) Transfer learning for enhanced machine fault diagnosis in manufacturing. *CIRP Ann* 69(1):413–416
146. Caggiano A (2018) Cloud-based manufacturing process monitoring for smart diagnosis services. *Int J Comput Integr Manuf* 31(7):612–623

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.