Measurement: Sensors 47 (2026) 102013

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Tool wear prediction using multi-domain features and quantised attention embedded CNN

Aitzaz Ahmed Murtaza [a], Muhammad Hamza Zafar [b], Syed Kumayl Raza Moosavi [b], Muhammad Faisal Aftab [b], Filippo Sanfilippo [b,c,*]

[a] Department of Mechanical Engineering, Capital University of Science and Technology, Islamabad, 44000, Pakistan
[b] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[c] Department of Software Engineering, Kaunas University of Technology, Kaunas, 51368, Lithuania

ARTICLE INFO

Keywords:
Tool wear prediction
Time–frequency domain
Quantisation
Multi-CNN attention
Predictive analytics

ABSTRACT

In precision manufacturing, accurate prediction of tool wear is pivotal for maintaining production quality and efficiency. This paper introduces a novel predictive model that employs a multi-branch convolutional neural network (multi-CNN) enhanced by an attention mechanism and Bayesian optimisation to forecast tool wear with high accuracy. The proposed approach is evaluated using the publicly available dataset, which comprises force and torque measurements acquired from machining processes. By incorporating time-domain (TD), frequency-domain (FD), and discrete wavelet transform (DWT) features, the model comprehensively analyses signals from machining processes, enabling robust predictions. The attention mechanism strategically emphasises significant features, augmenting the model's predictive capability. Demonstrated results show superior performance with a mean absolute error (MAE) of 2.8, root mean squared error (RMSE) of 3.4, and an R^2 score of 0.986, significantly outperforming conventional models. Additionally, post-training quantisation optimises the model for deployment on edge computing devices, maintaining effectiveness while being lightweight, ideal for real-time applications in resource-constrained environments. This study exemplifies how advanced machine learning techniques can improve predictive maintenance in manufacturing which not only enhances operational efficiency but also reduces downtime.

1. Introduction

In today's advanced manufacturing landscape, machining technology plays an essential role in producing highly demanded components for the automotive, aerospace, and precision machinery sectors. The quality of these manufactured parts is pivotal, directly influencing production efficiency and accuracy. This quality largely hinges on the wear condition of the cutting tools [1]. Over time, the tool surfaces deteriorate due to various factors such as the material being machined, cutting forces, vibrations, cutting depth, and spindle speeds etc. [2]. Early prediction of tool wear is crucial for enhancing production quality and optimising the lifespan of the tools. Traditionally, tools are replaced at predetermined intervals, which can lead to either underuse or overuse, given the complex, non-linear nature of tool wear. This traditional method often falls short in terms of cost-effectiveness. Mostly, about 50% to 80% of tool life is actually used, as mentioned in a study [3]. Moreover, approximately 20% of machine tool downtime stems from tool failures. In particular, failure to promptly replace tools when wear exceeds critical thresholds adversely impacts the surface quality of machined workpieces [4]. Therefore, accurately predicting tool wear using advanced methods not only maximises tool utility but also minimises the risk of producing substandard products by detecting tool wear at its onset.

Predicting tool wear presents significant challenges due to its dependence on multiple variables, each tool exhibiting a distinctive wear trajectory influenced by these factors. Measuring tool wear during the machining process is particularly challenging due to obstructions from the closed machining environment and the presence of cutting fluids. With advancements in manufacturing technology, modern systems are increasingly equipped with numerous sensors, leading to the generation of large volumes of data. This data is vital for the real-time evaluation of tool conditions. Consequently, tool condition monitoring (TCM) increasingly utilises data-driven methods to manage the extensive information produced by these advanced systems [5,6].

The data-driven method for predicting tool wear unfolds in three primary phases: data acquisition, feature extraction, and the development and deployment of a predictive model [7,8]. Numerous sensors

* Corresponding author at: Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway.
E-mail address: filippo.sanfilippo@uia.no (F. Sanfilippo).

https://doi.org/10.1016/j.measen.2026.102013
Received 26 November 2025; Received in revised form 4 May 2026; Accepted 14 July 2026
Available online 20 July 2026

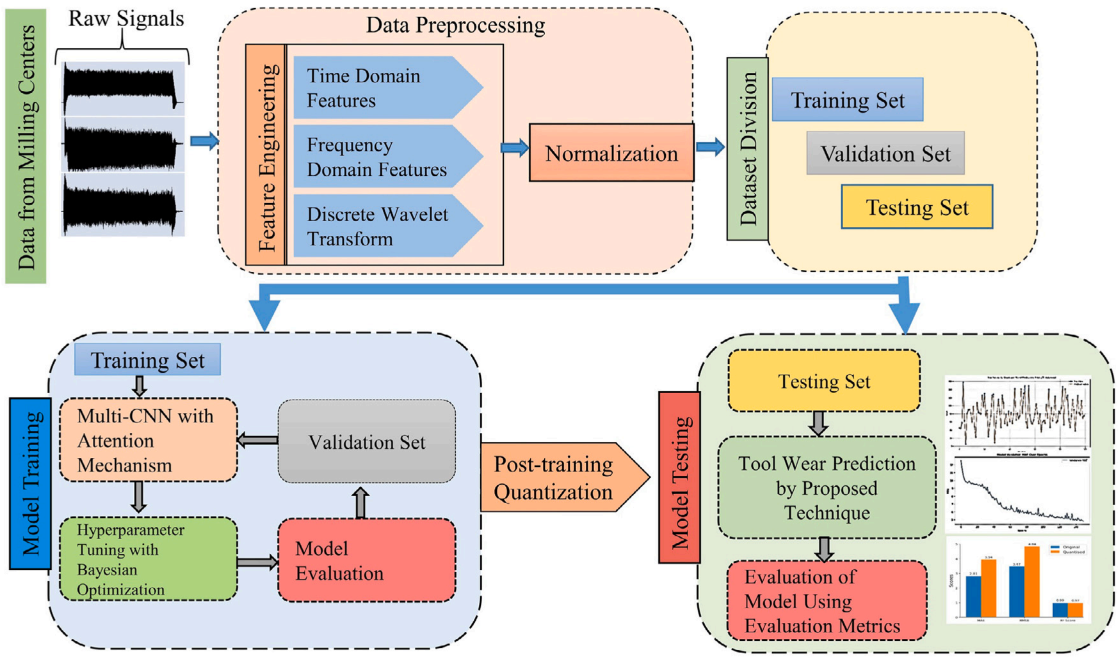


Fig. 1. Flow diagram illustrating the proposed Multi-CNN-based approach for tool wear prediction.

installed on CNC machines facilitate data acquisition for CNC machine tool wear and tool condition monitoring (TCM), including sensors for measuring cutting forces and vibrations. This gathered data serves to monitor the condition of the tool. While many researchers have utilised cutting force data for TCM [9,10], others have emphasised the use of vibration data to achieve similar ends [11–13]. Vibration sensors are straightforward to install, and their vibration amplitude varies with changes in tool wear, making them widely used for monitoring tool condition [14,15].

From the raw sensor data, various useful features are extracted for use in classification or regression models. Key features for predicting tool wear include those from the time domain, frequency domain, and time–frequency domain [16–18]. Time domain features are statistical measures of the raw signals, such as mean, variance, skewness, kurtosis, and root mean square. Frequency domain features encompass aspects of the raw signals such as dominant frequencies, spectral power, harmonic ratios, and spectral amplitudes. In the time–frequency domain, the representation captures both temporal variations and frequency components. Techniques like wavelet transforms are employed to extract features that vary across both time and frequency scales [19]. The extracted features are given as input to a predictive model, which utilises this data to forecast tool wear. The predictive model typically employs advanced machine learning algorithms which estimate the rate and extent of wear by analysing patterns within the features.

Much research is being done to build a predictive model to predict tool wear with great accuracy effectively. Mainly, these predictive models are based on machine learning approaches. In a research work [20], support vector regressor (SVR) is utilised to predict the tool wear and it achieved a mean absolute error of 10 μm. Similarly, another work [12] presents SVM-based prediction of micro-milling tool wear using vibration and sound signals, achieving the predictive accuracy of up to 97.54%. The study highlighted in [21] employed the Relevance Vector Machine (RVM) to predict tool wear, emphasising RVM's advantage over SVM in providing both predicted values and corresponding confidence intervals, unlike SVM which lacks probabilistic significance in its predictions. Researchers have used auto-encoders for feature engineering along with ML-based predictive models to enhance prediction accuracy of tool wear. For example, a research study [22] introduces a novel modelling framework that employs multiple stacked sparse auto-encoders combined with a non-linear regression function. Experimental validation using real manufacturing data has shown that this method achieves superior predictive accuracy when compared to conventional approaches.

Deep Learning (DL) has demonstrated promising outcomes in tool condition monitoring and wear prediction, making it a central focus of research in this field. This success is largely attributed to the vast amounts of data produced by modern manufacturing systems [23]. Recurrent Neural Networks (RNNs) have achieved notable results in processing time series data due to their ability to retain historical information, which is crucial in handling the evolving nature of manufacturing data. RNNs maintain a memory of recent input patterns, enabling them to map prior inputs to target vectors effectively [24,25]. This capability allows the network to preserve a memory of previous inputs within its internal state. However, standard RNNs may struggle with capturing long-term dependencies. To address this limitation, a variant of RNN, known as Long Short-Term Memory (LSTM) networks, has been employed to mitigate issues related to the backpropagation of errors from gradients vanishing or exploding. A study referenced in [26] demonstrated promising results by employing a Long Short-Term Memory (LSTM) network to predict tool wear from features extracted via a Stacked Auto-encoder (SAE). Similarly, another study in [27] utilised a stacked LSTM approach to extract deep features, which were then input into a non-linear regression model for predicting tool wear. Additionally, features extracted through Singular Value Decomposition (SVD) were utilised in a bidirectional Long Short-Term Memory (BiLSTM) network to effectively predict tool wear, as shown in [28].

Due to their computational complexity and the growing volumes of data, RNNs and LSTM networks demand significant computational resources and cannot perform parallel processing due to their sequential dependency on previous data points. In contrast, Convolutional Neural Networks (CNNs) offer a faster alternative for predicting tool wear, while also maintaining high accuracy in the results they produce. A study outlined in [29] demonstrated the use of CNNs for estimating tool wear, employing wavelet packet-based features. The study's performance comparisons revealed that CNNs outperformed various machine learning models across different datasets. CNNs were also proved effective in predicting tool wear values from modest set of images of the tools, as reported in [30].

In recent years, the demand for deploying DL models in resource-constrained environments has driven significant progress in model lightweighting technologies [31–33]. Techniques such as model pruning [34], knowledge distillation [35], and quantisation [36] have emerged as effective approaches for reducing model size and computational complexity while maintaining accuracy. Among these, Post-Training Quantisation (PTQ) has gained particular attention due to its simplicity and applicability without requiring additional retraining. PTQ reduces the precision of model weights and activations, often from 32-bit floating-point (FP32) to 8-bit integers (INT8), making it highly suitable for edge devices with limited memory and processing power [37].

Advancements in PTQ have demonstrated its effectiveness across various domains, including image recognition [38,39], object detection [40,41], and natural language processing [42]. These studies highlight how quantisation techniques can deliver lightweight models with minimal performance degradation, enabling deployment on platforms such as embedded systems, microcontrollers, and mobile devices [43]. Despite this progress, the application of PTQ in predictive maintenance tasks, such as tool wear prediction, remains underexplored. This paper contributes to this gap by employing PTQ to develop a quantised model for tool wear prediction, achieving significant reductions in computational cost while maintaining predictive accuracy.

The predominant methods for predicting tool conditions have primarily focused on state-related aspects rather than time-based monitoring. Where time-based monitoring was employed, the complexity of the algorithms required substantial computational resources [44]. Additionally, with the ever-increasing volume of data, these models risk becoming impractical in the future. A significant challenge with these models is the labour-intensive task of manual feature extraction. Moreover, none of these models were specifically designed to operate on edge computing devices, which typically have limited computational capabilities. To address these challenges, this study proposes the following enhancements:

1. Incorporating time-domain, frequency-domain, and time-frequency domain features to capture spatio-temporal patterns in the data, enabling a more comprehensive analysis.
2. Utilising a multi-branched convolutional neural network (multi-CNN) designed for tool condition monitoring (TCM), which efficiently handles sequential data and captures temporal dependencies.
3. Implementing post-training quantisation within the multi-CNN framework to reduce model size and computational demands, thereby increasing inference speed. These energy-efficient models are ideal for battery-powered devices commonly used in edge computing environments.

Fig. 1 provides a visual representation of the overall system flow for the proposed multi-CNN-based tool wear prediction technique. It outlines each step of the process, from data acquisition and preprocessing to feature extraction and the final prediction of tool wear. The remainder of this paper is organised as follows: Sections 2 and 3 provide a detailed theoretical background on the proposed multi-CNN and feature extraction techniques respectively. Section 4 outlines the experimental dataset and the architecture of the predictive model. Section 5 presents results and a comparative analysis, including findings from the implementation of the proposed technique. Finally, the concluding remarks are summarised in Section 6.

2. Proposed technique

Convolutional Neural Networks (CNNs) were originally designed for analysing 2-dimensional data and have been widely used in tasks like image classification and object detection. With some modifications they are used for processing data including time series, audio signals, and textual sequences. These variations offer considerable computational benefits. These modified convolutional neural networks are particularly effective for real-time and cost-effective applications, especially on edge devices where low computational power is required [45]. In this work, the proposed technique utilises a multi-CNN architecture integrated with an attention mechanism. This architecture is further optimised through Bayesian optimisation technique. A detailed model representation is given in Fig. 2 This approach is designed for efficiently processing temporal data from different feature domains. The attention mechanism emphasises the most relevant features and the Bayesian optimisation tunes the model's performance. The combination of these approaches help in building a model that is easy to converge, requires less computational power and outperforms other renowned approaches in terms of efficiency and effectiveness. Moreover, while keeping the focus on deploying the model to the edge devices, the Post Training Quantisation (PTQ) is also applied to the model to further reduce its computational requirements hence making it suitable for the battery powered edge devices.

2.1. Multi-branch Convolutional Neural Network (multi-CNN) architecture

The model consists of three parallel CNN branches processing time-domain, frequency-domain, and time–frequency domain features. Each branch follows an identical structure with two consecutive convolutional and max-pooling layers, followed by flattening. For input X (representing either Xtime, Xfreq, X_{dwt}), the first convolutional layer applies:

$$h_1(X) = \sigma\left(\sum_{j=1}^{f_1} W_{\mathrm{conv1},j} * X + b_{\mathrm{conv1},j}\right) \tag{1}$$

followed by max-pooling:

$$h_1'(X) = \mathrm{MaxPooling1D}(h_1(X), 2) \tag{2}$$

The second convolutional layer processes these features:

$$h_2(X) = \sigma\left(\sum_{j=1}^{f_2} W_{\mathrm{conv2},j} * h_1'(X) + b_{\mathrm{conv2},j}\right) \tag{3}$$

with subsequent max-pooling:

$$h_2'(X) = \mathrm{MaxPooling1D}(h_2(X), 2) \tag{4}$$

The outputs are flattened and concatenated into a single feature vector:

$$h_{\mathrm{concat}} = \left[h_{\mathrm{flatten}}(X_{\mathrm{time}}); h_{\mathrm{flatten}}(X_{\mathrm{freq}}); h_{\mathrm{flatten}}(X_{\mathrm{dwt}})\right] \tag{5}$$

This concatenated vector combines features from all three domains for downstream processing.

2.2. Attention mechanism

An attention mechanism is applied to the concatenated feature vector, computing weights to highlight relevant features. The attention weights α and attention-weighted features are calculated as:

$$\alpha = \mathrm{softmax}(W_{\mathrm{att}} \cdot h_{\mathrm{concat}} + b_{\mathrm{att}}) \tag{6}$$

$$h_{\mathrm{att}} = \alpha \odot h_{\mathrm{concat}} \tag{7}$$

where W_{att} and b_{att} are learnable parameters, and h_{concat} represents the concatenated feature vector.

The attention-weighted features h_{att} are obtained by applying element-wise multiplication between the attention weights and the concatenated vector.

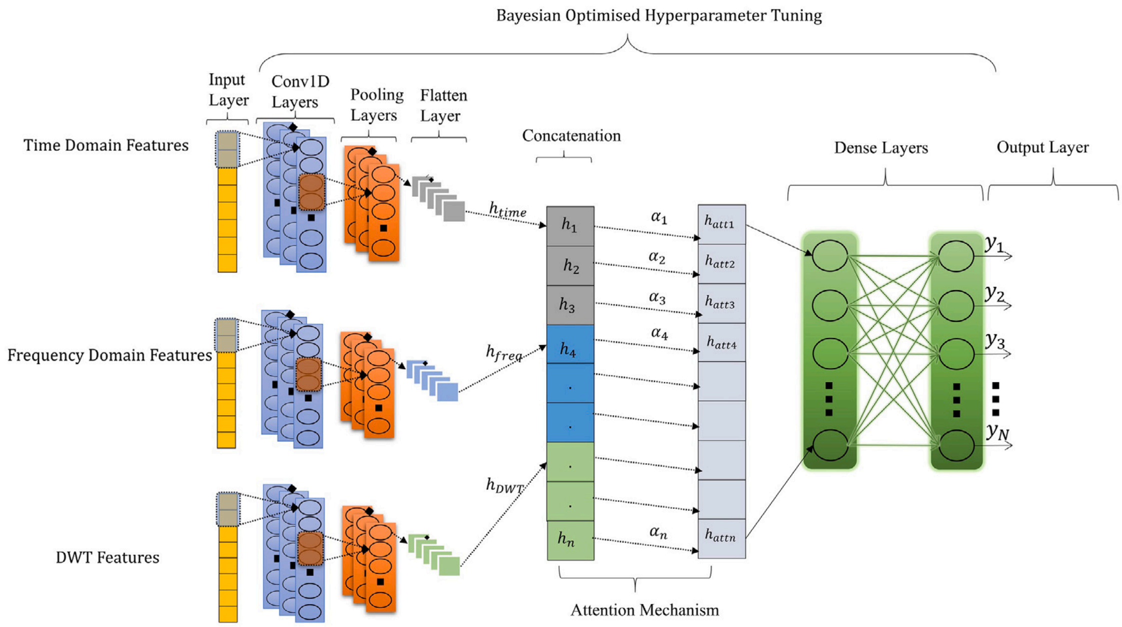


Fig. 2. Multi-CNN with attention mechanism and Bayesian optimisation for hyperparameters tuning.

2.3. Fully connected layers and output

After the attention mechanism, the attention-weighted feature vector h_{att} is passed through fully connected layers. These layers further refine the feature representation and combine them for the final prediction. Dropout is applied during training to prevent overfitting by randomly setting a fraction of the neurons to zero during each training step.

The final output is produced using a linear activation function, appropriate for regression tasks:

$$\hat{y} = W_{\text{out}} \cdot h_{\text{dropout}} + b_{\text{out}} \tag{8}$$

where W_{out} and b_{out} are the weights and biases of the output layer.

2.4. Bayesian optimisation for hyperparameter tuning

Bayesian optimisation is employed to optimise the hyperparameters of the model. The goal is to minimise the validation mean absolute error (MAE) by selecting the best configuration of hyperparameters, such as the number of filters f_1, f_2, kernel sizes k_1, k_2, dropout rate p, and learning rate η. This is formulated as:

$$\lambda^* = \arg\min_{\lambda} \mathbb{E}_{p(\lambda)}[\text{MAE}_{\text{val}}(\lambda)] \tag{9}$$

where λ represents the hyperparameter set, and MAE_{val} is the validation MAE. Bayesian optimisation iteratively selects the hyperparameters expected to improve model performance.

2.5. Training and evaluation

The model is trained using the Mean Squared Error (MSE) loss function, defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{10}$$

where y_i is the true target value, $\hat{y}_i$ is the predicted value, and N is the number of samples. The Mean Absolute Error (MAE) is used as the evaluation metric, computed as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{11}$$

Early stopping is applied to prevent overfitting and stopping training if the validation performance does not improve after a specified number of epochs.

To further enhance the model's applicability on resource-constrained devices, such as edge computing systems, we employ post-training quantisation. This step ensures the model remains computationally efficient while maintaining prediction accuracy, making it suitable for real-time tool wear prediction in industrial settings.

2.6. Post-training quantisation

Given the computational constraints of edge computing devices, the proposed model is optimised using post-training quantisation (PTQ). PTQ reduces the model size and computational complexity by converting the weights and activations from floating-point to lower precision formats, such as 8-bit integers. This optimisation is crucial for deploying models on resource-constrained devices while maintaining accuracy, especially for real-time applications like tool wear prediction. In this study we adopted uniform quantisation technique for post-training quantisation of neural network architecture [46,47]. Uniform quantisation linearly maps full-precision real numbers to low-precision integers [48].

We utilise TensorFlow Lite for Post-Training Quantisation (PTQ). The floating-point model is converted to TensorFlow Lite format and undergoes mixed-precision quantisation, where critical layers maintain floating-point precision while others use 8-bit integers. This optimisation requires a representative dataset of 1000 training samples to calibrate the model parameters. The representative dataset is defined as:

$$\text{Representative Dataset:} \quad \text{Rep} = \{\text{X\_train}_i \mid i = 1, \ldots, 1000\}$$

The quantised model combines 8-bit integer (INT8) and floating-point (FP32) operations to balance efficiency and accuracy. After evaluation on the test dataset, the final model is saved in TensorFlow Lite format, ready for edge deployment with reduced computational requirements while maintaining performance.

Table 1
Time domain features and their equations.

Feature name	Equation
Mean	$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$
Standard deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$
Variance	$\mathrm{Var} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$

3. Feature engineering

Feature engineering plays a critical role in extracting meaningful information from raw sensor signals and improving the predictive capability of data-driven models. In this study, features are extracted from three complementary domains: time domain, frequency domain, and time–frequency domain using discrete wavelet transform (DWT). These domains capture statistical characteristics, spectral content, and non-stationary signal behaviour, respectively.

The raw dataset consists of six time-series signals acquired during the milling process, namely force/torque signals along the X, Y, and Z axes obtained from the machine controller, and force signals along the X, Y, and Z axes measured by the dynamometer. Feature engineering is performed independently for each signal which resulted in a high-dimensional multi-domain feature representation.

3.1. Time-domain features

Time-domain features capture the statistical properties of the signal, such as central tendency and variability. These features are computed directly from the raw signal and provide insights into the overall distribution and behaviour over time. For each of the six signals, the following six time-domain descriptors are computed:

1. Mean
2. Standard deviation (std)
3. Variance (var)
4. Minimum
5. Maximum
6. Median

This results in 36 time-domain features in total (6 features × 6 signals). The mathematical definitions of the primary statistical descriptors are provided in Table 1, while min, max, and median describe the smallest, largest, and middle values of the signal, respectively.

3.2. Frequency-domain features

Frequency-domain features capture the behaviour of the signal in the frequency spectrum, providing insights into how the signal's power and energy are distributed across various frequency bands. These features are especially useful for understanding periodicities and oscillations within the data.

For each signal, the power spectral density (PSD) is computed, from which the following five frequency-domain descriptors are extracted:

1. PSD Sum: The total power of the signal across all frequencies, calculated as the sum of the power spectral density values.
2. PSD Max: The maximum power at any frequency, highlighting the strongest frequency component.
3. Dominant Frequency: The frequency at which the PSD reaches its maximum value.
4. Spectral Centroid: A measure that indicates the "centre of mass" of the spectrum, often related to the perceived pitch of the signal.
5. Spectral Entropy: A measure of the distribution of power across the spectrum, indicating the complexity or randomness of the signal.

Table 2
Frequency domain features and their equations.

Feature name	Equation
PSD sum	$\sum_f \mathrm{PSD}(f)$
PSD max	$\max(\mathrm{PSD}(f))$
Dominant frequency	$f_{\mathrm{dominant}} = \arg\max(\mathrm{PSD}(f))$
Spectral centroid	$\frac{\sum_f f \cdot \mathrm{PSD}(f)}{\sum_f \mathrm{PSD}(f)}$
Spectral entropy	$-\sum_{i=1}^{N} p_i \log(p_i)$

Table 3
Discrete Wavelet Transform (DWT) features and their equations.

Feature name	Equation
Standard deviation (Coeff)	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (c_i - \mu)^2}$
Energy (Coeff)	$\mathrm{Energy} = \sum_{i=1}^{N} c_i^2$
Entropy (Coeff)	$\mathrm{Entropy} = -\sum p(c_i) \log p(c_i)$

Table 4
Summary of extracted features across different domains.

Feature domain	Features per signal	Number of signals	Total features
Time domain	6	6	36
Frequency domain	5	6	30
DWT domain	24	6	144
Total	–	–	**210**

This yields 30 frequency-domain features in total (5 features × 6 signals). The mathematical expressions for these features are presented in Table 2.

3.3. DWT features

The discrete wavelet transform (DWT) captures both time and frequency information from the signal, making it a powerful tool for analysing non-stationary data. DWT decomposes the signal into multiple components, representing different frequency bands with corresponding time resolutions.

For each signal, the DWT is decomposed into six wavelet coefficients corresponding to different frequency bands. These coefficients provide information on how the signal behaves at different frequency scales, while the statistical measures summarise the distribution of energy and variation within each coefficient. For each coefficient, four statistical descriptors such as mean, standard deviation (std), energy and entropy were computed.

This results in 24 DWT-based features per signal (6 coefficients × 4 descriptors), and 144 DWT features across all six signals. These features provide a detailed multi-resolution representation of the signal dynamics relevant to tool wear progression. Equations used for computing std, energy and entropy from each wavelet coefficient are given in Table 3.

3.4. Summary of final features

Across all feature domains, a total of 210 features are extracted per sample prior to deep learning-based representation learning. These features are obtained by aggregating statistical and spectral descriptors computed independently for each of the six time-series signals acquired during the milling process. The extracted feature set includes 36 time-domain features, 30 frequency-domain features, and 144 discrete wavelet transform (DWT)–based features, as summarised in Table 4.

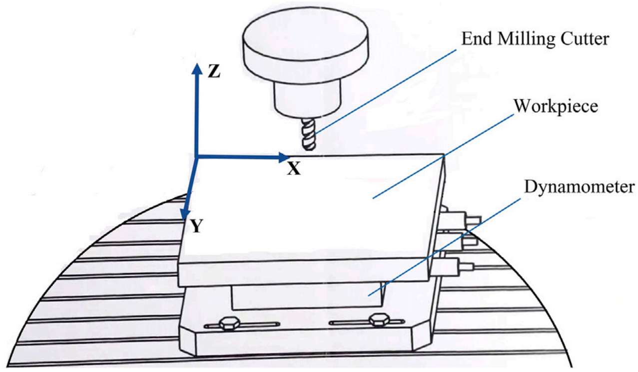


Fig. 3. Schematic illustration of the experimental setup reported in [50].

4. Experimental analysis

To evaluate the validity and effectiveness of the proposed method, this study utilises a publicly available dataset [49] derived from machining experiments conducted at the Institute of Production Engineering and Machine Tools, Leibniz Universität Hannover, Garbsen, Germany. The experimental trials, including tool wear measurements and data acquisition procedures, were performed by the original authors of the dataset. Detailed descriptions of the experimental setup, measurement methodology, and dataset structure are reported by Denk et al. [50]. The dataset is hosted on Mendeley Data (Version 3) and is distributed under a Creative Commons Attribution 4.0 (CC BY 4.0) license.

4.1. Experimental setup

The experiments employed three 5-axis CNC milling machines to wear a total of nine solid carbide end-mill cutters with four edges, coated with TiN-TiAlN, from their unused state to the end of their service life. Each tool was used exclusively on a single workpiece, which was part of a single batch of cast iron, ensuring material consistency. The machining process consisted of 44 shoulder milling runs per workpiece layer, with a consistent depth of cut of 2 mm maintained throughout the experiment. The flank wear land width VB of the cutting edges was measured using a digital microscope with 100x magnification, and a threshold of approximately 150 μm was used as the criterion for tool failure.

Despite all tools sharing the same specifications, they were sourced from different production batches. To ensure consistency, workpieces were from the same cast iron batch, and the cutting process parameters and numerical control (NC) instructions were held constant throughout all experiments. Fig. 3 depicts the experimental setup, including the positioning of the cutting tool, workpiece, and dynamometer.

Tool wear was assessed by calculating the average VB value across all four cutting edges of the tool. The workpieces were mounted on a dynamometer that measured forces across X, Y, and Z axis of the workpiece during machining. From the CNC machine's control system torque/force data across three axis of the tool was also recorded. Together, these multivariate time series data formed the basis for the predictive modelling of tool wear.

4.1.1. Process flow

The process flow of the experimental setup is shown in Fig. 4. The flow begins with the 5-axis CNC milling machine performing shoulder milling on the workpiece, which is fixed on a dynamometer. The dynamometer measures the forces on the workpiece across all three axis during the machining process. Whereas, the torque/force data across X, Y, and Z axis of the tool are simultaneously captured from the machine's control panel. All measured data is fed into a data acquisition system, where it is processed and stored. These datasets are then utilised to train and evaluate the predictive modelling framework for tool wear.

4.2. Dataset description

The dataset used in this study is comprised of 6418 files, each containing time-series data for individual runs and the corresponding VB value. The dataset includes data collected from force sensors and machine controls. Table 5 presents an overview of the total samples alongside detailed label information for each run. The labels serve as identifiers for the different experimental conditions under which the data was collected. The time-series data recorded against these labels provides a robust framework for analysis. For each 4.41-second run, the time-series data was gathered at various frequencies, allowing for a detailed examination of the temporal dynamics and variability in the VB values across different runs.

Force sensor data was recorded at a high sampling rate of 25 kHz to capture precise force dynamics, while machine control signals were sampled at varying frequencies and standardised to a 500 Hz frequency for consistency. The time-series dataset comprises 14 features, two of which correspond to time stamps ensuring synchronisation between machine and sensor data. Out of the 14 features, only 6 are used in this predictive modelling based on their relevance in VB prediction. These features are given in Table 6.

The selected features undergo preprocessing to extract time-domain, frequency-domain, and DWT characteristics. These processed features are critical for developing accurate predictive models. Time-domain features capture signal behaviour over time, while frequency-domain features highlight underlying patterns and trends. DWT features provide both temporal and spectral information, essential for predicting tool wear. These features are input into the predictive model, optimising tool wear predictions and supporting more efficient maintenance scheduling.

According to the dataset documentation, the high-frequency force signals were temporally aligned with the machine control signals by adjusting the time offset of the force measurements to match the controller recordings [50]. In the present study, all time-domain, frequency-domain and discrete wavelet transform features were extracted from normalised signals computed over identical time windows. Thus ensuring feature comparability across sensor modalities and machining runs.

Several sources of measurement uncertainty are present in the dataset and are described in the original dataset documentation. These include adjustments to the force signal time vectors to align them with machine control signals, which prevents direct interpretation of absolute timing relationships between these signals. The dataset also reports possible aliasing effects in the force and spindle torque signals for tools T4–T6 and a small coordinate system misalignment of approximately 1–2° in the XY-plane for tools T7 and T8. These issues introduce measurement noise and systematic variability into the data. In this study, these uncertainties are not explicitly quantified or modelled. However, they remain present in the released measurements and are therefore encountered during feature extraction, model training, and model evaluation. As a result, the proposed learning-based framework is evaluated using data that reflect practical, non-ideal measurement conditions. An overview of the measurement and data processing chain is shown in Fig. 5.

4.3. Dataset partitioning

The complete dataset was first divided using a run-wise split, where 80% of the machining runs were assigned to the training set and the remaining 20% were reserved as an independent test set. The test set was held out entirely and used exclusively for final model evaluation.

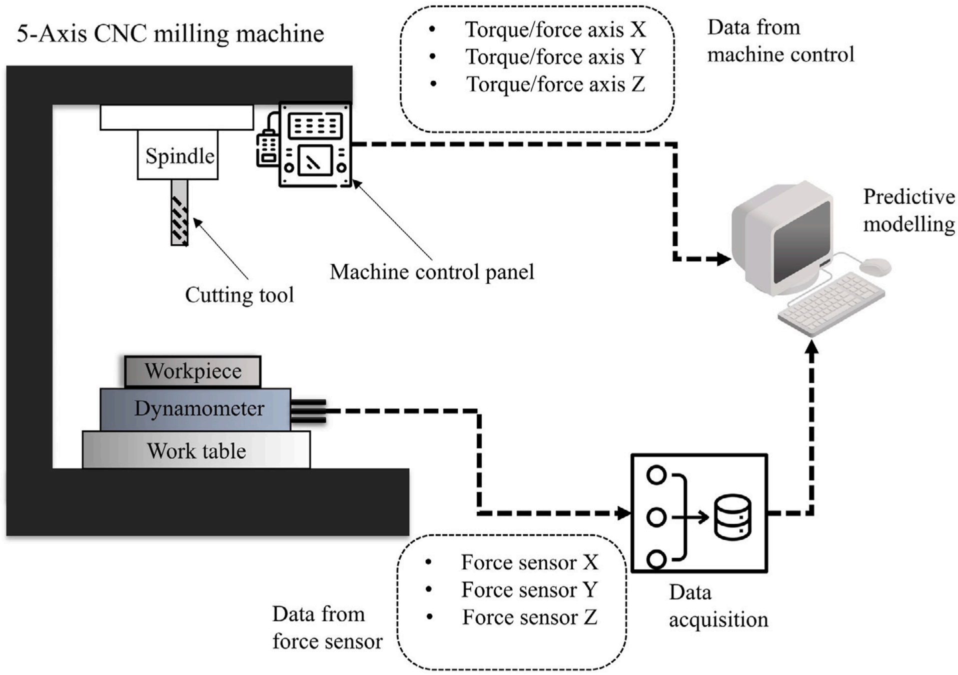


Fig. 4. Process flow diagram illustrating data acquisition and predictive modelling workflow.

Table 5
Label definitions for data collection [50].

ID	Category	Details	Units
M	Machine	Identifies the specific machine tool used.	N/A
T	Tool	Identifies the tool number	N/A
R	Run	Tracks the number of experimental runs for each tool, noting only the successful milling operation.	N/A
C	Cumulated time	Total time the tool was engaged with material prior to the current run.	s
VB	Wear	Measures the width of the tool's flank wear land VB.	μm

Table 6
Time-series features used in predictive modelling.

Name of signal	Unit	Source
Torque axis X/Force axis X	N m	Machine
Torque axis Y/Force axis Y	N m	Machine
Torque axis Z/Force axis Z	N m	Machine
Force sensor X	N	Sensor
Force sensor Y	N	Sensor
Force sensor Z	N	Sensor

During model training, an internal validation subset was created from the training data to support model optimisation and early stopping. Specifically, 10% of the training samples were automatically allocated as a validation set using the `validation_split` mechanism provided by the Keras framework. This validation subset was used solely to monitor the validation loss and guide the early stopping criterion, while the remaining training samples were used for updating the model parameters.

As a result, the effective data distribution corresponds to approximately 72% training, 8% validation, and 20% testing. There is no overlap between the samples used for training, validation, and testing, ensuring that information from the test set is not inadvertently used during model development. The distribution of the flank wear value VB across all subsets spans from early-stage wear to near-failure conditions, enabling a representative and unbiased evaluation of the proposed model's predictive performance.

5. Results and discussion

In this section, a structured evaluation of the proposed tool wear prediction framework is presented. First, the relative importance of input features is analysed to provide insight into their contribution to wear prediction. This is followed by a comparative evaluation against established machine learning and deep learning models to benchmark predictive performance. An ablation study is then conducted to quantify the individual impact of the multi-branch architecture, attention mechanism, and Bayesian optimisation. Subsequently, the Bayesian optimisation process and training behaviour of the proposed model are examined in detail. The results are then compared with state-of-the-art approaches reported in the literature. Finally, the effects of

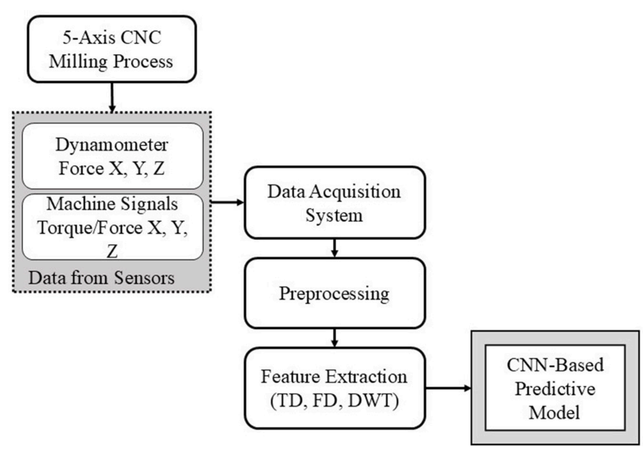


Fig. 5. Data measurement and preprocessing chain diagram illustrating data acquisition and predictive modelling workflow.

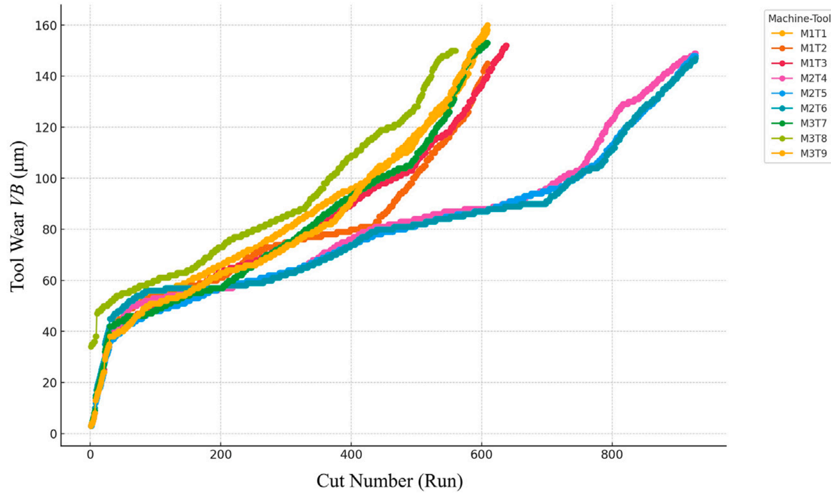


Fig. 6. Tool wear progression across different tools over multiple cuts.

post-training quantisation are analysed in terms of predictive accuracy, model size, inference time and robustness under noisy conditions. As a starting point for this evaluation, the following analysis first examines the observed tool wear progression across different tools and machining runs to establish the underlying wear behaviour present in the dataset.

The graph depicting the relationship between tool wear and cut number across various tools reveals different wear patterns with respect to the cut number as illustrated in Fig. 6. As expected, tool wear generally increases with the number of cuts, indicating progressive degradation as the tool continues to be used. Interestingly, the rate of wear varies between different tools, with some exhibiting a more linear progression while others show accelerated wear after a certain number of cuts. This suggests that certain tools may be more resistant to wear initially but degrade more rapidly once a threshold is reached. Additionally, the variability in wear patterns across tools highlights the potential influence of factors such as tool material, cutting conditions, and specific operational parameters. Another reason for varying tool wear pattern can be the different production batches of the tools although made up with same specifications. Furthermore, the cutting patterns also significantly influence tool wear variability due to their direct impact on how the cutting tool engages with the material. These insights suggest that monitoring wear rates closely and adjusting cutting strategies accordingly could help in optimising tool life and ensuring consistent performance.

5.1. Cumulative importance of features for tool wear prediction

Flank wear VB is fundamentally governed by the physics of cutting forces, particularly those acting in the feed direction and cutting depth direction [51–53]. In this experiment the feed direction is in Y-axis and the Force Sensor Y captures the force acting in this direction. During milling, this force plays a critical role in flank wear progression, as the tool's flank face continuously interacts with the freshly cut workpiece surface. This repeated contact generates frictional forces, leading to abrasive and adhesive wear on the tool flank. Higher feed rates intensify this interaction, increasing contact pressure and accelerating flank wear. Meanwhile, the Z-axis force, which corresponds to the depth of cut, significantly impacts the initial stages of tool wear. A greater cutting depth results in higher cutting forces, subjecting the tool edge to increased stress, which can lead to micro-chipping or edge rounding. However, once the tool edge stabilises, flank wear is predominantly driven by the sustained frictional effects of the Y-axis force. The analysis presented below aligns with these observations, as the cumulative predictive power of the Y-axis force is the highest, while the Z-axis force ranks second due to its involvement in the initial stages of wear. Other features exhibit minimal impact, further reinforcing the dominant role of the Y and Z-axis forces in tool wear prediction.

Table 7 summarises the individual and cumulative importance of various original features in prediction, or predictive power, including features as Force/Torque Axis X, Y, Z and Force Sensor X, Y, Z. To derive the feature importance scores, a Random Forest Regressor with the Gini Importance method, also known as Mean Decrease in Impurity, is utilised. This technique evaluates each feature's contribution by measuring its role in reducing impurity, variance in regression tasks, across all decision trees in the forest. The dataset includes subfeatures for each primary feature category, such as domain features for time-domain, frequency domain and discrete wavelet transform. The Random Forest model, configured with 50 decision trees and a fixed random state for reproducibility, is trained to predict tool wear VB. Feature importance scores are calculated by aggregating the contributions of all subfeatures within each primary category based on how often and how significantly they contribute to splitting nodes in the decision trees. These scores are then normalised and the results are grouped by primary feature categories to present the importance of primary features for predicting tool wear.

The results indicate that the Force Sensor Y feature holds the most significant predictive power, contributing approximately 90.48% to the overall model performance. This dominance suggests that the Force Sensor Y data is highly influential in predicting the target variable, likely capturing critical variations that correlate strongly with the output.

In contrast, features such as Force/Torque Axis X have minimal predictive power, with contributions close to zero. This indicates that these features may have limited relevance in the current modelling context, contributing little to the accuracy of predictions. The relatively low importance of Force/Torque Axis features compared to Force Sensor data suggests that sensor data may capture more dynamic and relevant information for the model, possibly due to its direct measurement of cutting forces during operation. These findings highlight the need to prioritise more impactful features in model development to enhance predictive accuracy while potentially reducing model complexity by eliminating less relevant variables.

5.2. Comparative analysis of various models

To demonstrate the superior performance of the proposed technique, a comprehensive comparative analysis was conducted against several established models, including Gaussian Naive Bayes (GNB), Deep Neural Networks (DNN), Random Forest (RF), Support Vector

Table 7
Features' individual and cumulative importance in tool wear prediction using Gini Importance.

Original feature	Individual importance	Cumulative importance
Force/Torque axis X	0.0023	0.0023
Force/Torque axis Y	0.0223	0.0246
Force/Torque axis Z	0.0153	0.0399
Force sensor X	0.0083	0.0482
Force sensor Y	0.9048	0.9530
Force sensor Z	0.0439	0.98669

Table 8
Evaluation metrics for various models for VB prediction.

Model	MAE	RMSE	R2-score
Gaussian NB	7.97	12.54	0.82
SVR	5.56	9.70	0.89
RF	8.21	10.68	0.87
DNN	6.53	9.21	0.90
CNN	6.94	8.02	0.92
Proposed technique	2.81	3.46	0.986

Machine (SVM), and CNN. The results, as presented in Table 8, highlight the distinct advantages of the proposed technique across various metrics in predicting tool wear VB. With an outstanding R^2-score of 0.986, the proposed method not only outperforms other models but does so with significant margins in predicting the VB value. Specifically, it achieves a Mean Absolute Error (MAE) of 2.81 and a Root Mean Squared Error (RMSE) of 3.46, markedly lower than those recorded by the benchmark models. This indicates a highly precise predictive capability, setting a new standard in the modelling process.

The table further highlights that the proposed multi-CNN with attention mechanism excels in VB prediction with an R^2-score of 0.986, surpassing the next best, the CNN, which scores 0.92. This is a substantial improvement, reflecting an increase of about 0.07 in the R^2-score. Additionally, the proposed model's MAE and RMSE significantly undercuts that of its closest competitors, demonstrating not only its accuracy but also its consistency in performance across different measures. For instance, compared to the SVR's MAE of 5.56, the proposed model's MAE of 2.81 showcases a great improvement, emphasising its effectiveness and efficiency in predictive modelling. Consequently, the proposed technique emerges as a clearly superior method in the context of predictive analytics, validating its adoption for more accurate and reliable tool wear VB predictions for tool condition monitoring.

5.3. Ablation studies

To evaluate the contributions of the individual components of the proposed model, an ablation study was conducted by systematically removing or modifying key features, such as the attention mechanism, Bayesian optimisation, and multi-branch architecture. The results of these experiments are summarised in Table 9.

The results clearly demonstrate the effectiveness of the proposed model in its full configuration, incorporating all components. The proposed model achieved the best performance, with an MAE of 2.810, an RMSE of 3.460, and an R^2 score of 0.986. When the attention mechanism was excluded, the performance dropped to an MAE of 3.627 and an RMSE of 5.012, highlighting the importance of attention in focusing on relevant features from the time, frequency, and DWT domains. Similarly, removing Bayesian optimisation led to suboptimal hyperparameters, resulting in an MAE of 4.577 and an RMSE of 6.295. The single-branched CNN baseline, which lacks the multi-branch architecture, performed the worst, with an MAE of 6.940 and an RMSE of 8.020, underscoring the significance of the multi-branch design in capturing diverse feature representations.

These results validate the contributions of each component in enhancing the predictive accuracy of the model. The attention mechanism and Bayesian optimisation, in particular, play critical roles in achieving superior performance. The ablation studies demonstrate how the combination of these elements results in a robust and efficient tool wear prediction model.

5.4. Performance of the proposed technique

The proposed technique employed hyperparameter tuning based on Bayesian optimisation, a probabilistic model-based approach that adaptively selects the most promising hyperparameter configurations to evaluate, balancing exploration and exploitation. Specifically, the optimisation process utilised prior uniform probability distributions across pre-defined hyperparameter ranges, reflecting the equal likelihood of each configuration initially. A Gaussian Process served as the surrogate model, iteratively updated using the observed performance (validation MAE) of previously tested configurations.

For each iteration, the acquisition function leveraged the Gaussian Process to select the next hyperparameter set to evaluate, considering both the potential for improvement (exploitation) and the uncertainty in less-explored regions (exploration). The process aimed to minimise the validation MAE over a maximum of 20 trials while maintaining computational efficiency. Early stopping criteria were implemented to prevent excessive computation by halting trials where validation loss increased over five consecutive epochs.

5.4.1. Bayesian optimisation process results

The optimisation process successfully identified an optimal hyperparameter configuration. Table 10 provides a comprehensive summary of all 20 trials, including the validation MAE and corresponding hyperparameter configurations. Whether the best results obtained through this hyperparameters set, is also identified in the table. It can be seen that trial 14 achieved the best performance, with a validation MAE of 2.810, highlighting the efficacy of Bayesian optimisation in refining the model's configuration.

The performance curve in Fig. 7 visually demonstrates the Bayesian optimisation process. The validation MAE varied across trials as the optimisation iteratively refined its surrogate model and acquisition function. Notably, the lowest validation MAE was achieved in trial 14, after which subsequent trials explored configurations with relatively higher MAE, reflecting the balance of exploration and exploitation.

5.4.2. Model training results

The model underwent training over 150 epochs, using the best parameters obtained through Bayesian optimisation process, during which a notable loss reduction curve was observed, as depicted in Fig. 8. Initially, there was a significant improvement in loss reduction during the first 60 epochs, demonstrating rapid progress towards optimal model performance. Beyond this point, the curve began to plateau, transitioning into a near-horizontal line that indicated minimal further reduction in loss. This pattern suggests that the model had effectively reached a point of diminishing returns with respect to learning from additional epochs beyond the sixtieth.

5.5. Attention mechanism analysis and comparison

To improve the interpretability of the proposed model, the learned attention weights were visualised using heatmaps. Since attention weights are computed dynamically for each input instance, the reported values correspond to the average attention across the test dataset, providing a global measure of feature importance. Fig. 9 presents the aggregated attention heatmap, across all sensors comprising of force/torque along the X, Y, and Z axes and the corresponding external force sensor measurements. To enhance interpretability, attention

Table 9
Ablation study results of the proposed model.

Model	MAE	RMSE	R^2 score
Single-Branched CNN CNN	6.940	8.020	0.920
Multi-CNN (Without attention and Bayesian optimisation)	4.577	6.295	0.955
Multi-CNN (Without attention and With Bayesian optimisation)	3.627	5.012	0.972
Proposed Multi-CNN (With attention and Bayesian optimisation)	**2.810**	**3.460**	**0.986**

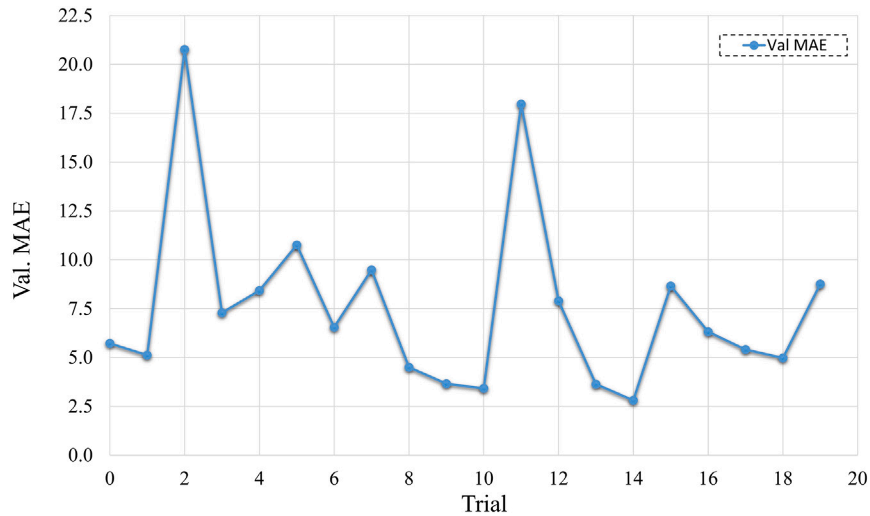


Fig. 7. Validation MAE across Bayesian optimisation trials.

Table 10
Summary of hyperparameters and val. MAE across trials.

Trial	Val MAE	Hyperparameters							Best results
		Filters 1	Kernel size 1	Filters 2	Kernel Size 2	Dense units	Dropout	Learning rate	
0	5.726	104	5	24	3	384	0.2	0.00091	False
1	5.117	112	5	24	3	128	0.5	0.00711	False
2	20.764	112	5	16	3	320	0.1	0.00239	False
3	7.292	16	5	16	5	384	0.4	0.00111	False
4	8.429	120	5	56	3	128	0.4	0.00014	False
5	10.758	32	5	48	3	256	0.4	0.00061	False
6	6.547	104	3	40	5	192	0.3	0.00155	False
7	9.486	16	5	64	5	128	0.1	0.00125	False
8	4.499	24	5	16	5	256	0.1	0.00504	False
9	3.655	96	5	24	3	192	0.1	0.00022	False
10	3.419	104	5	16	5	320	0.3	0.00447	False
11	17.965	48	3	24	3	384	0.3	0.00104	False
12	7.899	40	3	32	3	192	0.5	0.00054	False
13	3.632	120	5	16	5	448	0.2	0.00012	False
14	2.810	72	3	32	3	256	0.1	0.00315	True
15	8.661	56	5	40	3	192	0.3	0.00645	False
16	6.318	88	3	32	3	192	0.4	0.00209	False
17	5.409	112	5	48	3	448	0.1	0.00712	False
18	4.975	88	5	56	5	320	0.1	0.00084	False
19	8.761	48	5	56	3	320	0.3	0.00026	False

weights were aggregated across time, frequency, and wavelet domain features within each group.

The results reveal a non-uniform attention distribution, with the highest importance assigned to Force Sensor Y, followed by Force/Torque Axis Y and Force Sensor X, while Z-axis components exhibit comparatively lower importance. Notably, this pattern is consistent with the feature importance ranking reported in Table 7, where Force Sensor Y is also identified as the dominant contributor. This agreement between attention-based and Gini-based importance reinforces the reliability of the feature selection of the proposed model.

5.5.1. Comparison with existing attention mechanisms

Recent studies in tool wear prediction have incorporated different forms of attention mechanisms to improve feature representation and prediction accuracy. As shown in Table 11, existing attention-based approaches primarily focus on channel-wise or hybrid attention mechanisms applied to convolutional feature maps or sequential representations. In contrast, the proposed model applies attention directly to a concatenated multi-domain feature vector which enables simultaneous weighting of TD, FD and DWT features.

Furthermore, the effectiveness of the proposed attention mechanism is supported by the ablation results given in Table 9, where the inclusion of attention significantly improves predictive performance. In addition, the consistency between the learned attention distribution and the feature importance analysis as given in Table 7, demonstrates that the proposed approach captures physically meaningful relationships in the data while maintaining a relatively simple and computationally efficient architecture.

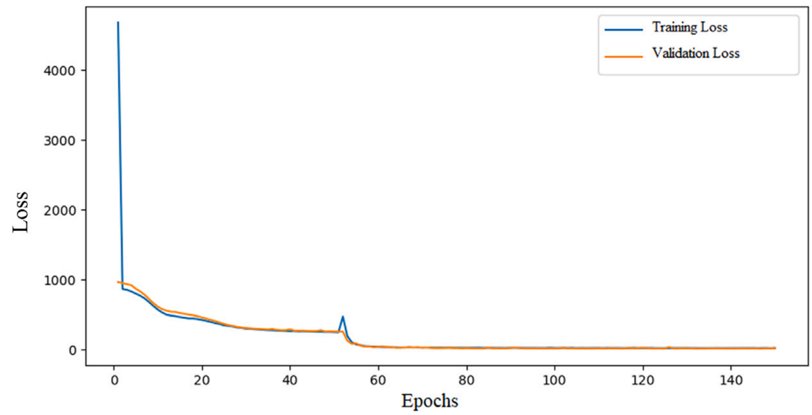


Fig. 8. Training and validation loss curve.

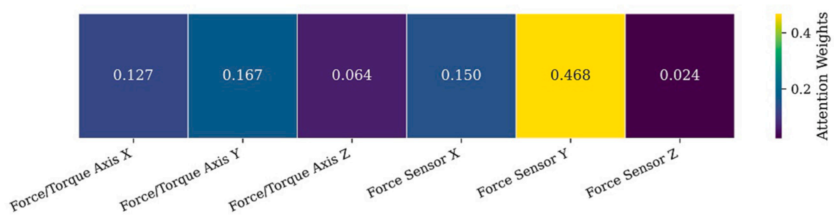


Fig. 9. Attention weight distribution across sensor channels, representing the relative importance learned by the model.

Table 11
Comparison of attention mechanisms in recent tool wear prediction studies.

Model	Attention type	Level	Key mechanism/focus
Parallel CNN [54]	Channel + Residual	Feature maps	Feature map weighting for multi-sensor fusion
HABPDL [55]	CBAM + GAU	Feature + Temporal	CNN + BiLSTM with hybrid attention
MPCAN [56]	ECA + BiLSTM	Feature + Temporal	Lightweight channel + temporal modelling
Proposed model	Feature-wise soft attention	Multi-domain	Cross-domain weighting (TD, FD, DWT)

5.6. Evaluation of the model against existing approaches

Several recent studies have proposed deep learning models, specifically CNNs, for tool wear prediction. In this section, we compare the performance of our proposed model with existing methods in the literature.

1. **Multi-Frequency-Band Deep Convolutional Neural Network (MFB-DCNN)** [57]: This model uses wavelet packet decomposition to extract multi-frequency-band features, which are processed by a deep CNN for sensitive feature extraction. These features are then passed through fully connected layers to predict tool wear. The model's performance is optimised through milling experiments and hyperparameter tuning.
2. **Parallel CNN with Attention** [54]: This model uses parallel convolutional neural networks for multi-scale feature fusion and incorporates a channel attention mechanism with residual connections to adjust the importance of different feature maps.
3. **Hybrid Attention-Based Parallel Deep Learning (HABPDL) Model** [55]: This model combines ResNet and BiLSTM blocks with attention layers (CBAM and general attention unit) to enhance feature extraction. It also uses global average pooling (GAP) to reduce redundant features for better interpretability.
4. **Multi-Input Parallel Convolutional Attention Network (MPCAN)** [56]: This framework employs a multi-input parallel convolutional (MPC) network to extract multi-scale features from sensor data. Efficient channel attention (ECA) assigns varying attention weights to tool wear information, while a Bi-LSTM captures time-series features. The predicted tool wear value is then output through a fully connected network, achieving strong performance.
5. **1D CNN-LSTM Model** [58]: This study combines a multi-channel 1D CNN with long short-term memory (LSTM) for tool wear prediction.

Table 12
Comparison of the proposed model with state-of-the-art methods for tool wear prediction.

Model	MAE	RMSE	R^2 score
MFB-DCNN [57]	8.807	10.7904	0.87764
Parallel CNN with attention [54]	3.503	4.353	–
HABPDL [55]	6.072	7.955	0.933
MPCAN [56]	5.06	6.52	0.9737
1D CNN-LSTM [58]	2.938	3.81	0.9506
Proposed model	**2.81**	**3.46**	**0.986**

The performance of these models, including our proposed model, is summarised in Table 12. The proposed model, which utilises a multi-branched CNN with an attention mechanism, outperforms the existing models in terms of MAE, RMSE, and R^2. The proposed model is optimised using Bayesian optimisation and incorporates features from the time domain, frequency domain, and discrete wavelet transform (DWT), leading to improved tool wear prediction results.

5.7. Performance comparison between original and quantised model

The optimal model achieved in our study was subsequently subjected to post-training quantisation, which yielded satisfactory outcomes. This process significantly reduced the model's computational

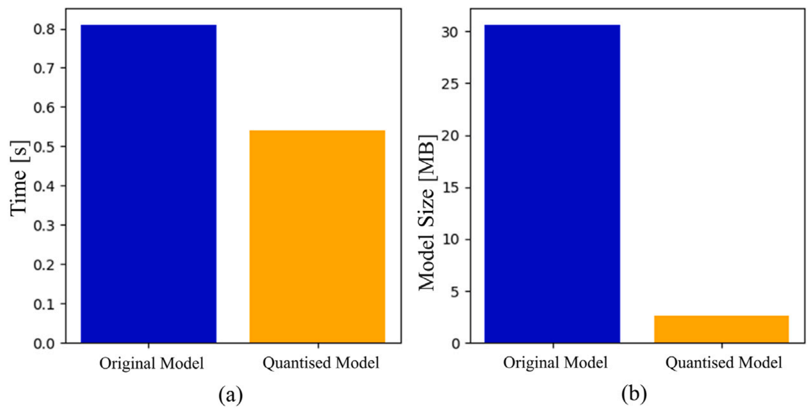


Fig. 10. Comparison of models' inference time (a) and size (b).

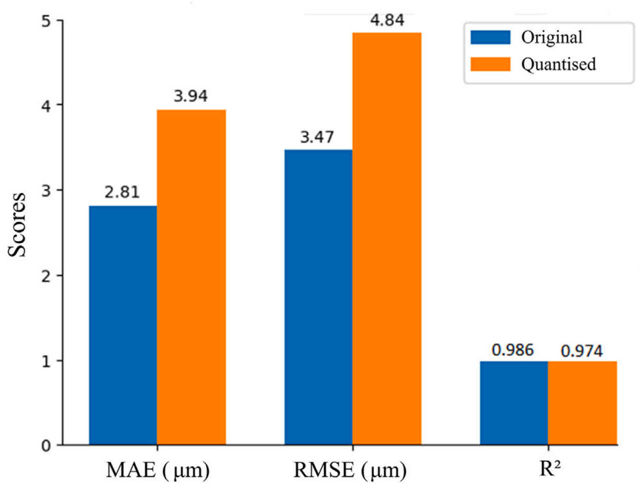


Fig. 11. Comparison of scores for original and quantised models.

needs, reducing the size by a factor of 12 and the inference time by 1.4 times, as illustrated in Fig. 10(a), and Fig. 10(b), demonstrating the effectiveness of quantisation in optimising computational resources. Despite these substantial reductions, the R^2-score was only marginally affected, decreasing from 0.986 to 0.974, a slight reduction of less than 1.5% as depicted in Fig. 11. This minor decrement in performance underscores the quantised model's capability to retain its predictive accuracy while benefiting from considerable improvements in operational efficiency.

This substantial reduction in both size and computational demand makes the quantised model particularly well-suited for deployment on edge computing devices and battery-powered devices, where limited space and power availability are critical constraints. The model's compactness and comparable performance with the original model ensure that it can operate effectively in resource-constrained environments, thus extending its applicability and utility in real-world scenarios where quick, efficient processing is essential.

The quantised model demonstrated a remarkable consistency in its predictive accuracy, closely mirroring the performance of the original model. Both models tended to predict slightly lower than the actual values, exhibiting a negative offset. While a significant negative offset could be disadvantageous, in this instance, the consistent slight under-prediction is manageable and can be accounted for during the calibration of the model on actual devices. The original and quantised models not only aligned closely in their predictive outcomes but also displayed a high degree of consistency across the range of predictions. As illustrated in Fig. 12, the graph of actual versus predicted values for both models confirms this trend. Notably, between the tool wear VB values of 0 and 100, the predictions from both models show remarkable overlap, indicating a strong agreement in their assessments. Beyond the value of 100, however, the predictions begin to diverge slightly, suggesting areas where further refinement of the quantisation process could enhance the intended performance.

The analysis of the residuals for the original and quantised models illustrates some difference in prediction consistency between the two. The original model, with a standard deviation of 2.3, demonstrates a higher precision, consistently making predictions closer to the actual values. In contrast, the quantised model's standard deviation of 3.2 indicates a broader spread of residuals, reflecting less consistency and a tendency to under-predict. This variability is visually represented in their respective histograms as given in Fig. 13(a), and Fig. 13(b), where the quantised model shows a shift towards positive values, affirming its under-prediction bias.

The positive shift in the residuals observed in both the original and quantised models arises from several factors. One key reason is the dataset's characteristics, such as the underrepresentation of extreme tool wear conditions or phases of rapid wear progression, which can lead to systematic underestimation. The features derived from time, frequency, and DWT domains, while comprehensive, may not fully capture the non-linear and dynamic nature of tool wear progression, contributing to this bias. In the original model, while the attention mechanism enhances feature weighting and improves overall performance, it does not entirely mitigate these biases and may still struggle with edge cases. For the quantised model, the shift becomes more pronounced due to precision loss during quantisation, which reduces the model's ability to represent subtle patterns accurately. These combined factors result in the observed residual shift.

To address this challenge, future work could explore reinforcement learning approaches to dynamically adjust the tool wear prediction model during machining, based on real-time feedback from sensors. Alternatively, metaheuristic optimisation methods could enhance the model's ability to adapt to varying machining conditions by fine-tuning hyperparameters beyond Bayesian optimisation. Additionally, conducting further machining experiments under diverse operational conditions would provide richer data to improve model robustness and mitigate underestimation bias. These approaches could significantly enhance the predictive accuracy and reliability of the tool wear models, particularly in complex, real-world scenarios.

5.8. Robustness of the model under noise

To evaluate the robustness of the proposed model under noisy and uncertain conditions, Gaussian noise was introduced into the tool wear data at five different noise levels: 1%, 3%, 5%, 7%, and 10%. Both the original and quantised models were tested on the noisy data to assess

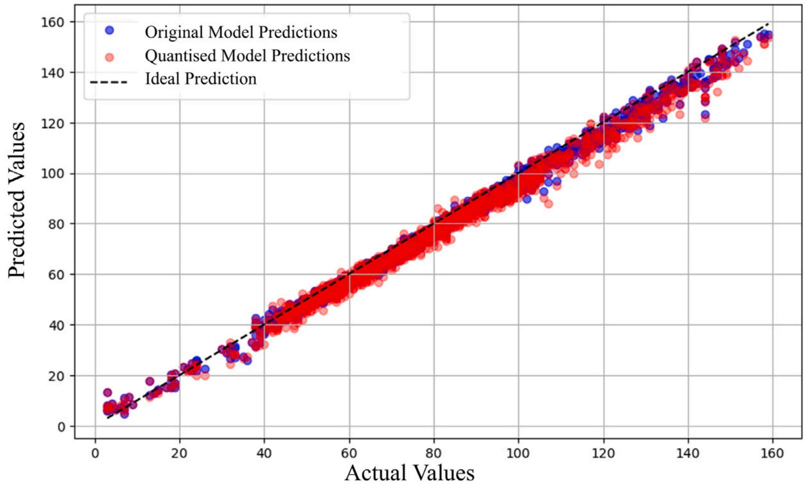
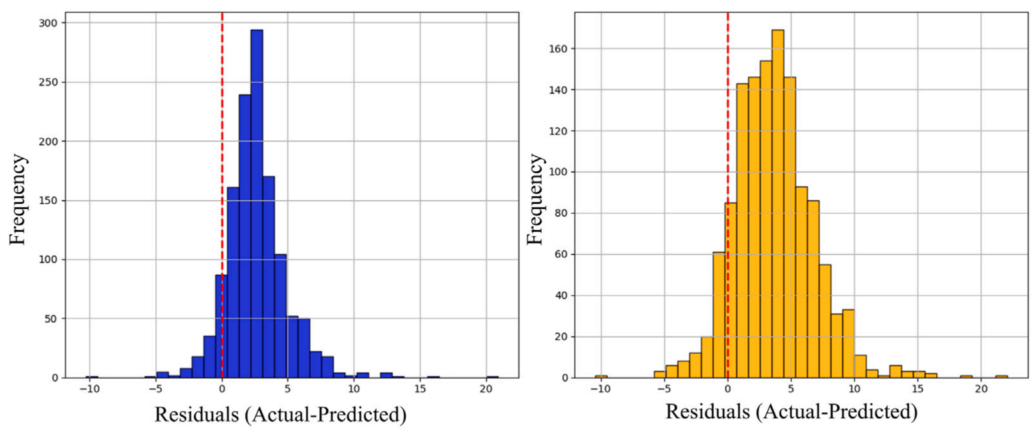


Fig. 12. Comparison of quantised vs. original model predictions.



(a) (b)

Fig. 13. Error comparison of original (a) and quantised (b) models.

Table 13
Evaluation metrics for the original and quantised models under Gaussian noise.

Noise level	Noise %	Original model			Quantised model		
		MAE	RMSE	R^2 Score	MAE	RMSE	R^2 Score
Low	1	2.849	3.543	0.985	4.021	4.989	0.972
	3	3.366	4.226	0.979	4.443	5.542	0.965
Moderate	5	4.247	5.516	0.966	5.205	6.430	0.953
High	7	5.434	7.171	0.942	5.881	7.361	0.939
	10	6.149	8.078	0.926	6.916	8.807	0.912

their performance in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. The results are summarised in Table 13.

The noisy data $\tilde{X}$ was generated using the following equation:

$$\tilde{X} = X + \mathcal{N}(\mu, \sigma^2)$$

where X represents the original data, $\mathcal{N}(\mu, \sigma^2)$ denotes Gaussian noise with mean μ and variance σ^2, σ is the standard deviation of the noise, calculated as $\sigma = \alpha \cdot \mathrm{std}(X)$, where α represents the noise level percentage (e.g., 0.01 for 1% noise).

At low noise levels (1% and 3%), both models demonstrated strong predictive performance with minimal degradation. The original model achieved an R^2 score of 0.985 at 1% noise and 0.979 at 3% noise, while the quantised model achieved R^2 scores of 0.972 and 0.965, respectively. These results indicate robust predictions in the presence of small perturbations. At the moderate noise level (5%), there was a slight performance decline, with the original model achieving an R^2 score of 0.966 and the quantised model achieving 0.953. Whereas, at higher noise levels (7% and 10%), the performance degradation became more noticeable. The R^2 scores for the original model dropped to 0.942 and 0.926 at 7% and 10% noise, respectively, while the quantised model achieved R^2 scores of 0.939 and 0.912.

Although the decline was more pronounced at these noise levels, both models maintained R^2 scores above 0.9, showcasing their ability to handle significant noise. Across all noise levels, the quantised model exhibited slightly higher MAE and RMSE values than the original model, reflecting its susceptibility to noise compared to the original model. However, it achieved comparable R^2 scores, reinforcing its reliability and suitability for resource-constrained environments. These results underline the model's suitability for deployment in environments where noisy or uncertain data is inevitable, showcasing its potential for real-world tool wear monitoring applications.

6. Conclusions

This study presents a comprehensive evaluation of a multi-domain, learning-based framework for accurate and efficient tool wear prediction in precision manufacturing.

- A multi-branch convolutional neural network (multi-CNN) integrating time-domain, frequency-domain, and discrete wavelet transform (DWT) features was developed to capture detailed and dynamic machining characteristics.
- Dedicated feature-processing branches and an attention mechanism enabled effective prioritisation of salient features, leading to improved predictive accuracy.
- Bayesian optimisation was employed for hyperparameter tuning, resulting in strong predictive performance with a mean absolute error (MAE) of 2.8 μm, root mean squared error (RMSE) of 3.4 μm and an R^2 score of 0.986.
- Post-training quantisation substantially reduced model size and inference time, enabling deployment on resource-constrained edge devices with only a marginal reduction in performance metrics (R^2 score from 0.986 to 0.974).
- The proposed framework demonstrated robustness under noise which highlights its suitability for practical predictive maintenance applications.

Future work will focus on enhancing feature selection strategies, integrating reinforcement learning for adaptive maintenance decision-making and validating the framework through real-world machining experiments.

CRediT authorship contribution statement

Aitzaz Ahmed Murtaza: Conceptualisation, Methodology, Software, Writing – original draft. **Muhammad Hamza Zafar:** Conceptualisation, Methodology, Formal analysis. **Syed Kumayl Raza Moosavi:** Conceptualisation, Writing – review & editing, Formal analysis, Investigation. **Muhammad Faisal Aftab:** Conceptualisation, Writing – review & editing, Formal analysis, Investigation. **Filippo Sanfilippo:** Writing – review & editing, Investigation, Formal analysis.

Declaration of competing interest

NONE

All authors claim that there is not any conflict of interest regarding the above submission. The work of this submission has not been published previously. It is not under consideration for publication elsewhere. Its publication is approved by all authors and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Acknowledgements

This research is supported by the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group at the Top Research Center Mechatronics (TRCM), University of Agder (UiA), Norway. All authors were involved in the preparation and review of the original manuscript.

Data availability

The data used in this work is publically available.

References

[1] T. Mohanraj, J. Yerchuru, H. Krishnan, R.N. Aravind, R. Yameni, Development of tool condition monitoring system in end milling process using wavelet features and Hoelder's exponent with machine learning algorithms, Measurement 173 (2021) 108671.

[2] D. Knittel, M. Nouari, Milling diagnosis using machine learning approaches, in: Surveillance, Vishno and AVE Conferences, 2019.

[3] Y. Li, C. Liu, J. Hua, J. Gao, P. Maropoulos, A novel method for accurately monitoring and predicting tool wear under varying cutting conditions based on meta-learning, CIRP Ann 68 (1) (2019) 487–490.

[4] M. Kuntoğlu, H. Sağlam, Investigation of progressive tool wear for determining of optimized machining parameters in turning, Measurement 140 (2019) 427–436.

[5] K. Javed, R. Gouriveau, X. Li, N. Zerhouni, Tool wear monitoring and prognostics challenges: A comparison of connectionist methods toward an adaptive ensemble model, J. Intell. Manuf. 29 (2018) 1873–1890.

[6] Y. Yang, Y. Guo, Z. Huang, N. Chen, L. Li, Y. Jiang, N. He, Research on the milling tool wear and life prediction by establishing an integrated predictive model, Measurement 145 (2019) 178–189.

[7] A. Fertig, M. Weigold, Y. Chen, Machine learning based quality prediction for milling processes using internal machine tool data, Adv. Ind. Manuf. Eng. 4 (2022) 100074.

[8] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237.

[9] G. Wang, Y. Yang, Q. Xie, Y. Zhang, Force based tool wear monitoring system for milling process based on relevance vector machine, Adv. Eng. Softw. 71 (2014) 46–51.

[10] T. SK, S. Shankar, D. K, Tool wear prediction in hard turning of EN8 steel using cutting force and surface roughness with artificial neural network, Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci. 234 (1) (2020) 329–342.

[11] F. Alonso, D. Salgado, Analysis of the structure of vibration signals for tool wear detection, Mech. Syst. Signal Process. 22 (3) (2008) 735–748.

[12] M.C. Gomes, L.C. Brito, M.B. da Silva, M.A.V. Duarte, Tool wear monitoring in micromilling using support vector machine with vibration and sound sensors, Precis. Eng. 67 (2021) 137–151.

[13] V. Guleria, V. Kumar, P.K. Singh, Prediction of surface roughness in turning using vibration features selected by largest Lyapunov exponent based ICEEMDAN decomposition, Measurement 202 (2022) 111812.

[14] J. Ratava, M. Lohtander, J. Varis, Tool condition monitoring in interrupted cutting with acceleration sensors, Robot. Comput.-Integr. Manuf. 47 (2017) 70–75.

[15] Z. Huang, J. Zhu, J. Lei, X. Li, F. Tian, Tool wear monitoring with vibration signals based on short-time Fourier transform and deep convolutional neural network in milling, Math. Probl. Eng. 2021 (1) (2021) 9976939.

[16] H. Rafezi, M. Behzad, J. Akbari, Time domain and frequency spectrum analysis of soundsignal for drill wear detection, Int. J. Comput. Electr. Eng. 4 (5) (2012) 722.

[17] Z. Huang, J. Zhu, J. Lei, X. Li, F. Tian, Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations, J. Intell. Manuf. 31 (2020) 953–966.

[18] V. Guleria, V. Kumar, P.K. Singh, A novel approach for prediction of surface roughness in turning of EN353 steel by RVR-PSO using selected features of VMD along with cutting parameters, J. Mech. Sci. Technol. 36 (6) (2022) 2775–2785.

[19] S. Laddada, M.O. Si-Chaib, T. Benkedjouh, R. Drai, Tool wear condition monitoring based on wavelet transform and improved extreme learning machine, Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci. 234 (5) (2020) 1057–1068.

[20] M. Cheng, L. Jiao, X. Shi, X. Wang, P. Yan, Y. Li, An intelligent prediction model of the tool wear based on machine learning in turning high strength steel, Proc. Inst. Mech. Eng. B: J. Eng. Manuf. 234 (13) (2020) 1580–1597.

[21] D. Kong, Y. Chen, N. Li, C. Duan, L. Lu, D. Chen, Relevance vector machine for tool wear prediction, Mech. Syst. Signal Process. 127 (2019) 573–594.

[22] C. Shi, B. Luo, S. He, K. Li, H. Liu, B. Li, Tool wear prediction via multidimensional stacked sparse autoencoders with feature fusion, IEEE Trans. Ind. Inform. 16 (8) (2019) 5150–5159.

[23] V. Nasir, F. Sassani, A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges, Int. J. Adv. Manuf. Technol. 115 (9) (2021) 2683–2709.

[24] X. Wang, Y. Huang, Optimized recurrent neural network-based tool wear modeling in hard turning, Trans. NAMRI/SME 37 (2009) 213–220.

[25] J. Schmidhuber, A local learning algorithm for dynamic feedforward and recurrent networks, Connect. Sci. 1 (4) (1989) 403–412.

[26] M. Wang, J. Zhou, J. Gao, Z. Li, E. Li, Milling tool wear prediction method based on deep learning under variable working conditions, IEEE Access 8 (2020) 140726–140735.

[27] W. Cai, W. Zhang, X. Hu, Y. Liu, A hybrid information model based on long short-term memory network for tool condition monitoring, J. Intell. Manuf. 31 (2020) 1497–1510.

[28] X. Wu, J. Li, Y. Jin, S. Zheng, Modeling and analysis of tool wear prediction based on SVD and BiLSTM, Int. J. Adv. Manuf. Technol. 106 (2020) 4391–4399.

[29] F. Aghazadeh, A. Tahan, M. Thomas, Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process, Int. J. Adv. Manuf. Technol. 98 (2018) 3217–3227.

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

[31] C.-H. Wang, K.-Y. Huang, Y. Yao, J.-C. Chen, H.-H. Shuai, W.-H. Cheng, Lightweight deep learning: An overview, IEEE Consum. Electron. Mag. (2022).

[32] G.A. Mukhaini, M. Anbar, S. Manickam, T.A. Al-Amiedy, A. Al Momani, A systematic literature review of recent lightweight detection approaches leveraging machine and deep learning mechanisms in internet of things networks, J. King Saud Univ.-Comput. Inf. Sci. 36 (1) (2024) 101866.

[33] F. Chen, S. Li, J. Han, F. Ren, Z. Yang, Review of lightweight deep convolutional neural networks, Arch. Comput. Methods Eng. 31 (4) (2024) 1915–1937.

[34] Y. He, L. Xiao, Structured pruning for deep convolutional neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2023).

[35] G. Habib, S.M. Kaleem, T. Rouf, B. Lall, et al., A comprehensive review of knowledge distillation in computer vision, 2024, arXiv preprint arXiv:2404.00936.

[36] C. YANG, R. ZHANG, L. HUANG, S. TI, J. LIN, Z. DONG, S. CHEN, Y. LIU, X. YIN, A survey of quantization methods for deep neural networks, Chin. J. Eng. 45 (10) (2023) 1613–1629.

[37] Z. Liu, Y. Wang, K. Han, W. Zhang, S. Ma, W. Gao, Post-training quantization for vision transformer, Adv. Neural Inf. Process. Syst. 34 (2021) 28092–28103.

[38] Z. Tu, J. Hu, H. Chen, Y. Wang, Toward accurate post-training quantization for image super resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5856–5865.

[39] Y. Ding, H. Qin, Q. Yan, Z. Chai, J. Liu, X. Wei, X. Liu, Towards accurate post-training quantization for vision transformer, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 5380–5388.

[40] P. Kluska, M. Zięba, Post-training quantization methods for deep learning models, in: Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings, Part I 12, Springer, 2020, pp. 467–479.

[41] S. Zhou, L. Li, X. Zhang, B. Zhang, S. Bai, M. Sun, Z. Zhao, X. Lu, X. Chu, Lidar-ptq: Post-training quantization for point cloud 3d object detection, 2024, arXiv preprint arXiv:2401.15865.

[42] H. Bai, L. Hou, L. Shang, X. Jiang, I. King, M.R. Lyu, Towards efficient post-training quantization of pre-trained language models, Adv. Neural Inf. Process. Syst. 35 (2022) 1405–1418.

[43] M.W.U. Rahman, M.M. Abrar, H.G. Copening, S. Hariri, S. Shao, P. Satam, S. Salehi, Quantized transformer language model implementations on edge devices, in: 2023 International Conference on Machine Learning and Applications, ICMLA, IEEE, 2023, pp. 709–716.

[44] H. Sun, J. Zhang, R. Mo, X. Zhang, In-process tool condition forecasting based on a deep learning method, Robot. Comput.-Integr. Manuf. 64 (2020) 101924.

[45] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: A survey, Mech. Syst. Signal Process. 151 (2021) 107398.

[46] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, D. Kalenichenko, Quantization and training of neural networks for efficient integer-arithmetic-only inference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2704–2713.

[47] Y. Choukroun, E. Kravchik, F. Yang, P. Kisilev, Low-bit quantization of neural networks for efficient inference, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop, ICCVW, IEEE, 2019, pp. 3009–3018.

[48] J. Fang, A. Shafiee, H. Abdel-Aziz, D. Thorsley, G. Georgiadis, J.H. Hassoun, Post-training piecewise linear quantization for deep neural networks, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16, Springer, 2020, pp. 69–86.

[49] T. Stiehl, Multivariate time series data of milling processes with varying tool wear and machine tools, 2023, http://dx.doi.org/10.17632/zpxs87bjt8.3, Mendeley Data, Version 3.

[50] B. Denkena, H. Klemme, T.H. Stiehl, Multivariate time series data of milling processes with varying tool wear and machine tools, Data Brief 50 (2023) 109574.

[51] B. Denkena, D. Biermann, Cutting edge geometries, CIRP Ann 63 (2) (2014) 631–653.

[52] V.P. Astakhov, Tribology of metal cutting, 2006.

[53] A. Aslan, Optimization and analysis of process parameters for flank wear, cutting forces and vibration in turning of AISI 5140: A comprehensive study, Measurement 163 (2020) 107959.

[54] X. Xu, J. Wang, B. Zhong, W. Ming, M. Chen, Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism, Measurement 177 (2021) 109254.

[55] J. Duan, X. Zhang, T. Shi, A hybrid attention-based paralleled deep learning model for tool wear prediction, Expert Syst. Appl. 211 (2023) 118548.

[56] Q. Liu, D. Li, J. Ma, X. Wei, Z. Bai, A multi-input parallel convolutional attention network for tool wear monitoring, Int. J. Comput. Integr. Manuf. (2023) 1–17.

[57] J. Duan, J. Duan, H. Zhou, X. Zhan, T. Li, T. Shi, Multi-frequency-band deep CNN model for tool wear prediction, Meas. Sci. Technol. 32 (6) (2021) 065009.

[58] B. Li, Z. Lu, X. Jin, L. Zhao, Tool wear prediction in milling CFRP with different fiber orientations based on multi-channel 1DCNN-LSTM, J. Intell. Manuf. 35 (6) (2024) 2547–2566.