Proceedings of Engineering and Technology Innovation, vol. 33, 2026, pp. 67-81

Tool Wear Prediction Based on EMD–PSO–BiGRU Hybrid Model

Miaomiao Xin[1], Sze Song Ngu[1,*], Wanzhen Wang[2], Xiaomei Ni[2], Ting Wang[2], Pengpeng Wang[2], Yuxiang Zhang[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

[2]School of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, Jinan, China

Received 15 December 2025; revised 07 March 2026; accepted 12 March 2026

https://doi.org/10.46604/peti.2026.15987

Abstract

To improve milling tool wear prediction accuracy, which is critical for intelligent manufacturing efficiency and cost reduction, a hybrid model based on empirical mode decomposition (EMD), particle swarm optimization (PSO), and bidirectional gated recurrent unit (BiGRU) is proposed. Raw machining signals are decomposed into intrinsic mode functions (IMFs) via EMD; the Pearson correlation coefficient (PCC) is then used to screen wear-related IMFs to eliminate redundancy. Subsequently, PSO is applied to optimize BiGRU parameters, hidden layer neurons, and learning rate, to reduce the risk of local optima. Validated on the PHM2010 dataset, the model increases R^2 by 4.1%, reduces root mean square error (RMSE) by 13.9%, achieves a mean absolute percentage error (MAPE) of 5.915%, and outperforms improved subtraction-average-based optimizer (ISABO)-optimized BiGRU with 2.7% higher R^2 and faster convergence. The contributions lie in the EMD–PCC screening strategy and the integrated hybrid model, providing a practical solution for industrial tool wear prediction.

Keywords: tool wear, empirical mode decomposition, bidirectional gated recurrent unit, particle swarm optimization

1. Introduction

Accurate tool wear prediction enables the transition from “reactive maintenance” to “proactive maintenance,” significantly reducing unexpected downtime, improving processing quality, and lowering production costs [1]. Data-driven deep learning (DL) algorithms have emerged as a primary research focus in tool wear prediction. Leveraging its ability to adaptively decompose nonstationary data, EMD can effectively extract feature information during the machining process. Furthermore, the BiGRU can capture bidirectional temporal dependencies, making it suitable for simulating dynamic wear processes. This section presents the research background, literature review, research gaps, purposes, and the overall study structure, therefore justifying the necessity of the proposed hybrid model and its relevance to Industry 4.0.

Artificial intelligence breakthroughs have driven prognostics and health management (PHM) innovation. Conventional techniques have been increasingly superseded by data-driven solutions, which provide advantages in end-to-end prediction. In milling processes, sensors can capture multidimensional signals such as force, vibration, and audio emission [2]. These signals contain rich information on tool wear and can be analyzed and modeled to enable reliable wear condition prediction. However, machining signals are typically nonlinear, nonstationary, and mixed with strong noise, making feature extraction and model development difficult. This is one of the core industrial challenges in tool condition monitoring (TCM).

Tool wear prediction methods fall into three categories: physics-based, machine learning (ML), and DL approaches [3]. Physics-based methods build models on wear mechanisms, with clear physical interpretability for wear laws. Nevertheless, idealized assumptions cause prediction deviations in complex milling conditions.

* Corresponding author. E-mail address: 22020271@siswa.unimas.my

ML approaches with excellent nonlinear fitting capabilities have been widely applied to tool wear prediction. These methods mine data value through manual feature engineering, and then map wear states using support vector regression (SVR), K-nearest neighbors (KNN), random forest (RF), and back propagation (BP) neural networks, making them suitable for small-to-medium-sized datasets. Their efficacy relies on manual feature design, which is time-consuming, subjective, and limited to shallow features that fail to capture deep wear correlations or temporal dependencies in sequential wear data [4], thereby constraining model generalization in complex machining environments.

DL outperforms conventional approaches via automatic feature extraction from raw signals, cutting domain knowledge reliance and enabling deeper representation learning for complex data patterns.. Long short-term memory (LSTM) addresses recurrent neural networks (RNNs)' gradient vanishing: Marani et al. [5] used it to achieve <5% tool wear prediction error. Ertargin et al. [6] proposed a convolutional neural network (CNN)–LSTM model with strong generalization; Liu et al. [7] introduced optimized deep learning methods that can also be effectively extended to tool wear prediction.

BiGRU, an improved RNN variant, is widely used in tool wear prediction for its simpler structure, faster convergence than LSTM, and bidirectional temporal dependency capture. PSO [8], genetic algorithm (GA) [9], and ISABO [10] optimize DL parameters. However, most studies fail to integrate parameter optimization with signal preprocessing, limiting prediction performance. Milling signals have interference components, requiring pretreatment [11]. EMD, an adaptive method, can decompose nonstationary signals into IMFs to isolate noise-related components. The EMD-ML integration is explored in [12,13] for forecasting and diagnosis tasks.

Sayyad et al. [14] comprehensively reviewed data-driven remaining useful life estimation for the milling process, covering sensors, algorithms, datasets, and future directions, which lays a solid foundation for subsequent data-driven tool wear prediction research. Sayyad et al. [15] further proposed a tool wear prediction method using LSTM variants and hybrid feature selection techniques. Their work provides valuable references for the combination of deep learning variants and feature optimization in wear prediction.

EMD preprocessing has notable merits but has three core issues: lack of quantitative IMF selection criteria, inefficient manual BiGRU tuning, and overreliance on single optimization algorithms. AI-based tool wear prediction emphasizes hybrid models and emerging technologies. Rajakannu et al. [16] proposed a digital twin (DT)-assisted drill bit fault diagnosis system with principal component analysis (PCA)-MLP/CNN, achieving high accuracy via simulated data; however, it is limited to fault classification and lacks parameter optimization, reducing adaptability to complex milling conditions.These limitations highlight the need for a more integrated framework that combines adaptive feature extraction with efficient parameter optimization for robust tool wear prediction.

Another recent and relevant study in AI-based tool wear prediction was conducted by Rajakannu et al. [17], who proposed an airborne acoustic emission-based intelligent wear detection method for CNC drill bits. Their approach combines complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) for signal preprocessing and convolutional long short-term memory (ConvLSTM) for spatiotemporal data modeling, aiming to capture both spatial characteristics and temporal dependencies of acoustic emission signals related to drill bit wear. The study demonstrated high detection accuracy in complex machining environments, highlighting the potential of CEEMDAN in mitigating noise interference and ConvLSTM in handling spatiotemporal wear data.

To more intuitively reflect the research status of existing tool wear prediction methods in model architecture, signal preprocessing, and parameter optimization, and to further highlight the research gaps, a comprehensive comparison and summary of representative studies in this field are presented in Table 1.

Table 1 Comparison of representative tool wear prediction methods in recent studies

Reference	Prediction method/model	Signal preprocessing technique	Parameter optimization algorithm	Research focus	Limitations
[3]	KNN/SVR (ML)	Time/frequency domain feature extraction (manual)	None	Vibration monitoring	Over-reliance on manual feature engineering; low ability to capture temporal dependencies
[5]	LSTM (DL)	Spindle current signal normalization	Manual tuning	High-speed turning tool wear prediction	Complex model structure; slow training efficiency; no intelligent optimization
[8]	PSO-BP (ML+ optimization)	XGBoost feature selection	PSO	CNC milling tool wear prediction	Shallow network; unable to extract deep spatiotemporal features of wear data
[10]	ISABO-IBiLSTM (DL + optimization)	None (raw signal input)	ISABO	Cutting tool wear prediction	No signal preprocessing; unable to eliminate noise interference in machining signals
[16]	CEEMDAN-ConvLSTM (DL + preprocess)	CEEMDAN adaptive decomposition	Manual tuning	CNC drill bit wear detection	No intelligent parameter optimization; easy to fall into a local optimal solution
[17]	PCA-MLP/CNN (ML/DL + preprocess)	PCA feature dimensionality reduction	None	Drill bit fault diagnosis	Only for fault classification; no quantitative wear prediction; no optimization mechanism
[7]	BKA-CNN-BiGRU-attention (DL + optimization)	None (raw signal input)	BKA	Wind power time-series prediction (BiGRU)	No signal preprocessing; for energy prediction, not directly applicable to tool wear

Based on existing research and Table 1, three key gaps exist in tool wear prediction. First, EMD-derived IMF components have redundant information without quantitative screening [18], degrading feature quality. Second, DL models lack effective optimization; manual tuning or single techniques [19] limit their performance. Third, systematic analysis of optimization strategies is insufficient [20], and hybrid models lack integrated preprocessing, screening, and optimization, failing to meet industrial synergy needs.

To fill the research gaps, this study proposes an EMD-PSO-BiGRU tool wear prediction approach, aiming to enhance feature extraction, optimize BiGRU parameters, and verify EMD-PSO synergy. It decomposes signals via EMD, screens IMFs with PCC, optimizes BiGRU via PSO, and compares model performance with different settings. The key innovations include EMD-correlation screening, systematic PSO-ISABO comparison, and integrated preprocessing, feature screening, and parameter optimization for improved industrial applicability.

2. Methodology

This section first introduces the EMD signal preprocessing approach, highlighting its adaptive benefit in accommodating nonstationary milling signals and the quantitative screening logic based on PCC. The structural properties of the BiGRU model are then elaborated to demonstrate how it overcomes the limitations of conventional RNNs and captures the temporal patterns

of tool wear via bidirectional information transfer. The PSO algorithm is then presented, along with an explanation of how it can be used to achieve global optimization of the fundamental BiGRU parameters by modeling bird flock foraging behavior. Finally, the fused EMD–PSO–BiGRU model is developed, specifying the synergistic mechanism and comprehensive process for each module.

2.1 EMD

Signals recorded during milling processes, such as force and vibration, are affected by fluctuations in the cutting parameters, equipment interference, and other factors, resulting in typical nonlinear and nonstationary characteristics. Conventional filtering systems rely on predefined filtering parameters and struggle to react to dynamic signal changes. In contrast, EMD, as an adaptive signal processing method, can perform complete decomposition based on the characteristics of the signal without the need for manual basis function adjustment, providing significant advantages in machining signal processing.

The selection of EMD is theoretically motivated by its unique ability to address the nonlinear and nonstationary nature of milling signals, a core challenge that traditional linear signal processing methods fail to resolve. Unlike wavelet transform or variational mode decomposition (VMD), which require predefined basis functions or hyperparameters, EMD is a data-driven adaptive decomposition method that decomposes raw signals based on the local characteristic time scales of the data itself, ensuring that the extracted features are physically consistent with the wear evolution mechanism.

The basic principle behind EMD is to decompose a complex signal into numerous IMFs and one residual component. An IMF must meet two conditions: first, the number of extreme points of the signal equals or differs by no more than one from the number of zero-crossing points; second, the mean value of the upper and lower envelopes of the signal is zero at all times. The decomposition process proceeds as follows:

The process involves identifying the local maximum and minimum points of the original signal and constructing upper and lower envelopes using cubic spline interpolation. The envelope mean is then calculated and subtracted from the original signal to obtain a preliminary IMF candidate component. The process is repeated until the candidate component satisfies the IMF conditions. Finally, a series of IMFs and residual components is output.

The mathematical expression [21] is given by:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + Res(t) \tag{1}$$

where $x(t)$ denotes the original milling signal; $IMF_i(k)$ denotes the i-th *IMF* for i = 1, 2, ..., n (n represents the total number of *IMF* components); $Res(t)$ denotes the residual component, reflecting the overall signal trend. Low-order *IMF* components correspond to high-frequency signal fluctuations, whereas high-order *IMF* components are primarily noise interference, and the residual component represents the baseline trend of the signal.

To eliminate redundant information, this study quantitatively screens sensitive *IMF* components using the PCC. Taking the tool wear value (*VB*) as the reference index, the correlation coefficient between each *IMF* component and *VB* is calculated as follows:

$$\rho_{IMF_i,VB} = \frac{\sum_{k=1}^{m}\left[(IMF_i(k) - \overline{IMF_i})(VB(k) - \overline{VB})\right]}{\sqrt{\sum_{k=1}^{m}(IMF_i(k) - \overline{IMF_i})^2}\sqrt{\sum_{k=1}^{m}(VB(k) - \overline{VB})^2}} \tag{2}$$

Here, m represents the number of samples; $IMF_i(k)$ denotes the k sample value of the IMF_i component; $\overline{IMF_i}$ denotes the mean value of the IMF_i component; $VB(k)$ denotes the tool wear value corresponding to the k sample; and $\overline{VB}$ denotes the mean tool wear value.

2.2 BiGRU Prediction Model

The tool wear process exhibits strong temporal dependence, with the previous machining conditions directly affecting the current wear level. Conventional RNNs are prone to the gradient vanishing problem, hindering their ability to capture long-term temporal dependencies. Although LSTM models address this issue, they have high complexity (a high number of parameters) and poor training efficiency. In contrast, the BiGRU, an upgraded type of RNN, captures temporal aspects using bidirectional information transmission while simplifying the gating process. BiGRU increases training speed without sacrificing prediction accuracy, making it ideal for tool wear prediction tasks.

BiGRU is selected as the core predictive model based on its theoretical superiority in modeling temporal dependencies of tool wear evolution and its computational efficiency for industrial deployment. Tool wear is a cumulative, time-series process where the current wear state is strongly correlated with historical machining conditions. The key theoretical motivation is BiGRU's bidirectional architecture: It processes the EMD-refined feature sequence both forward (capturing historical wear trends) and backward (incorporating future contextual information), which aligns with the physical reality of tool wear propagation.

The PHM2010 dataset [22] collects full life cycle data on tools from new to failure, resulting in individual long-term time-series sequences. Standard unidirectional RNNs can only detect unidirectional dependencies between the past and the present. By introducing two independent gated recurrent unit (GRU) layers, the BiGRU processes (Fig. 1) the input sequence in both forward and backward directions. At each time step *t*, the final hidden state is a combination of the forward hidden state $\overrightarrow{h_t}$ and the backward hidden state $\overleftarrow{h_t}$.

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{3}$$

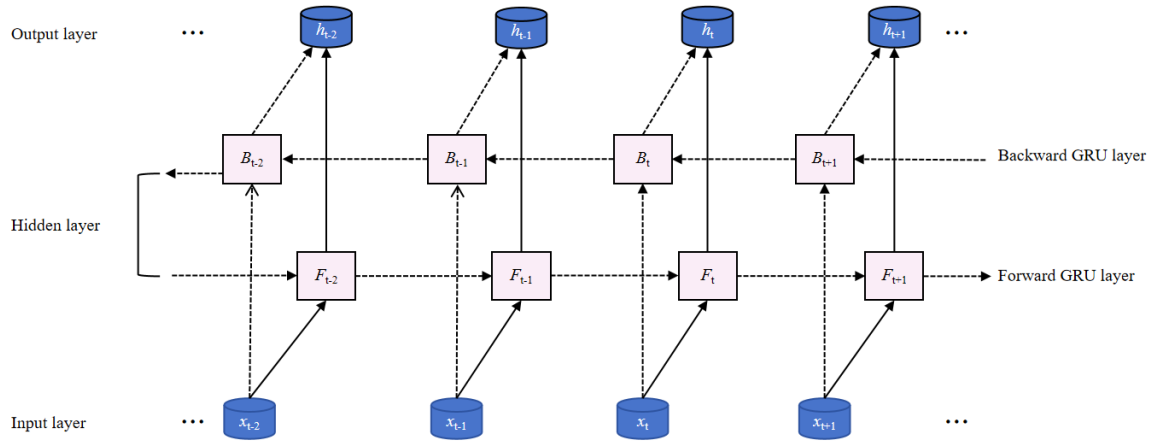

Fig. 1 Structure of the BiGRU cell

2.3 Parameter Optimization Using PSO

The prediction performance of the BiGRU model is heavily dependent on the number of hidden layer neurons (N), learning rate (η), and number of iterations (T). Manual parameter tuning is prone to producing local optimal solutions, and a single optimization method struggles to meet the search needs of parameter combinations. By simulating bird flock foraging behavior, the PSO algorithm achieves global optimization of many parameters, with the advantages of a rapid convergence speed and simple parameter sets, which can significantly increase the parameter configuration efficiency of the BiGRU model.

The integration of PSO is theoretically justified by the need to solve the non-convex optimization problem of BiGRU hyperparameter tuning and maximize the synergistic effect of EMD and BiGRU. The performance of deep learning models is highly sensitive to hyperparameters, and manual tuning or grid search is prone to falling into local optima. PSO, a population-

based stochastic optimization algorithm, is theoretically well-suited for this task due to its balance of global exploration (via particle inertia) and local exploitation (via social/cognitive learning). Compared with recent optimization algorithms such as ISABO, PSO's theoretical advantage lies in its faster convergence speed and lower computational complexity. essential for integrating with EMD's decomposition process and ensuring the entire model pipeline is feasible for edge computing in smart factories.

The fundamental notions of the PSO algorithm include particles, particle locations, particle velocities, personal bests, and global optimal solutions. The technical know-how lies in tailoring PSO to the hybrid EMD–BiGRU framework: This study defines the optimization objective as the minimization of validation set RMSE, directly linking PSO's search process to the model's predictive performance on wear data.

2.4 Tool Wear Prediction Model

The proposed EMD–PSO–BiGRU model accurately predicts tool wear using signal preprocessing, parameter optimization, and temporal prediction. The model combines the feature extraction capability of EMD, the global optimization capability of PSO, and the temporal modeling capability of BiGRU, along with regularization technology to improve generalization. The flow chart is depicted in Fig. 2.

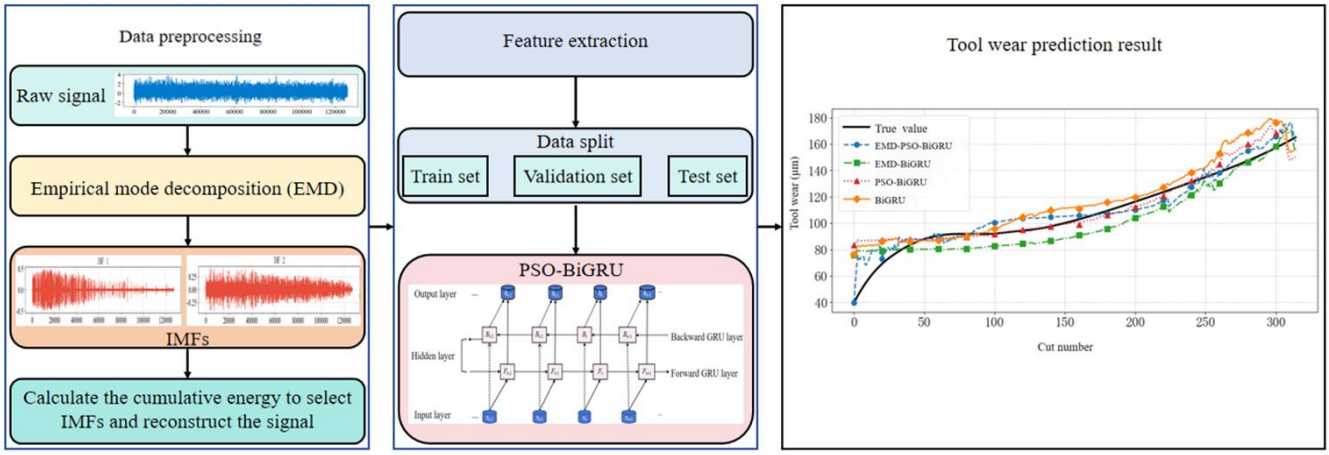

Fig. 2 Tool wear prediction model

3. Experiment Works

This section describes the experimental design and implementation. The machining equipment, tool–workpiece settings, and signal acquisition scheme are specified according to the PHM2010 benchmark dataset. Data preprocessing and feature engineering are then presented, followed by the model training strategy. Model effectiveness is statistically evaluated through ablation experiments and multimodel comparisons using RMSE, mean absolute error (MAE), and other evaluation metrics.

3.1 Data Description

(1) Dataset Collection Process

The dataset used in this study is the publicly available PHM2010 tool wear prediction dataset, a benchmark dataset widely adopted in the field of tool wear prediction, which ensures the reliability, comparability, and reproducibility of the experimental results. The specific machine conditions and experimental parameters of this dataset are detailed in Table 2, including hardware configurations, cutting conditions, and tool wear measurement parameters. The dataset encompasses raw sensor signals, cutting parameters, and tool wear measurements, with tool wear quantified using a LEICA MZ12 Microscope.

Table 2 Machine conditions of the dataset

Hardware configurations	Models and parameters	Cutting conditions	Parameters values
CNC milling machine	Roders Tech RFM760	Spindle speed	10,400 r/min
Force measuring instrument	Kistler 9265B	Feed rate	1,555 mm/min
Charge amplifier	Kistler 5019A	-	-
Data acquisition card	NI DAQ Data acquisition card	Sampling frequency Axial depth of the mill	50/kHz 0.2 mm
Workpiece material	Inconel 718, Rectangle	Radial depth of the mill	0.125 mm
Tool	Flute ball-end carbide milling cutter	Feed per pass	0.001 mm
Tool wear measurement	LEICA MZ12 microscope	Cooling conditions	Dry Cutting

(2) Experimental Setup

The dataset used in this study is the publicly available PHM2010 tool wear prediction dataset, a benchmark dataset widely adopted in the field of tool wear prediction for vertical milling processes. The details of the experimental setup of this dataset are shown in Fig. 3.

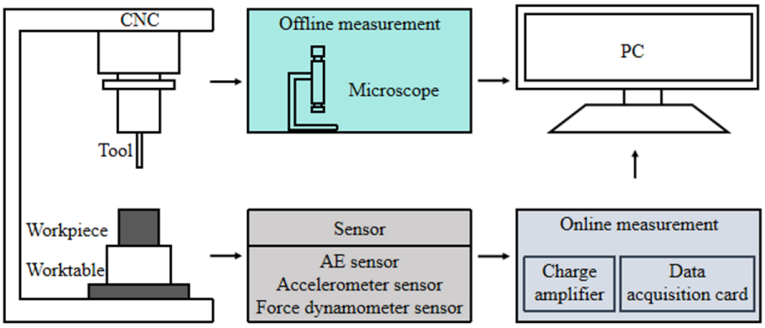

Fig. 3 Details of the experimental setup

The dataset was collected using a Roders Tech RFM760 CNC milling machine, equipped with a three-flute ball-end carbide milling cutter. The workpiece material is rectangular Inconel 718. The cutting conditions include a spindle speed of 10400 r/min, a feed rate of 1555 mm/min, an axial depth of cut of 0.2 mm, a radial depth of cut of 0.125 mm, a feed per pass of 0.001 mm, and dry cutting. The force measuring instrument is Kistler 9265B, matched with a Kistler 5019A charge amplifier; the data acquisition was performed using a NI DAQ data acquisition card with a sampling frequency of 50 kHz. Tool wear was measured using a LEICA MZ12 microscope. In the experiment, each tool was cut along the X direction for 108 mm. Three tools were tested, and each tool was cut 315 times to fully collect the cutting monitoring signals under different conditions. Collecting monitoring signals using signal collecting equipment reveals the evolution law of the machining process, provides sufficient information and features for wear prediction, and improves the model's generalization ability and engineering applicability.

(3) Dataset Size and Justification

The experiment adopted a standardized end-face milling protocol. A square tool path was implemented with a strictly controlled single-pass milling length of 108 mm, maintaining constant feed rate and cutting depth. Post-experiment wear values were recorded, while monitoring equipment continuously tracked milling forces, vibration signals, and acoustic emission root

mean square values across three axes. All experimental data were archived in folders C1, C4, and C6 for subsequent analysis. The compressed dataset is approximately 4.63 GB.

Three-flute ball-end cemented carbide cutting tools (C1, C4, and C6) were used to mill the Inconel 718 workpiece. The signals were recorded using six-channel force and vibration sensors. Charge amplifiers amplified the sensor outputs, which were then delivered to the signal acquisition equipment at a frequency of 50 kHz. To ensure precise machining dynamics, force and vibration sensors were installed on the workpiece or fixture. After observing and collecting the wear value for each tool operation, the average value was calculated and used as the wear signal sample. Simple data processing was employed to facilitate subsequent wear-level prediction.

(4) Suitability of the Proposed Model for the Dataset and Industrial Conditions

The suitability of the proposed EMD–PSO–BiGRU hybrid model for the PHM2010 dataset and real industrial operating conditions is fully justified. The PHM2010 dataset contains nonlinear, nonstationary machining signals mixed with strong noise, which is consistent with real industrial milling signal characteristics. It also has an appropriate sample size, diverse cutting parameters, and full wear stage coverage, making it ideal for model training and verification. The EMD component of the proposed model can adaptively decompose raw nonstationary signals to separate wear-related information from noise, while PCC screening eliminates redundant components. The BiGRU component effectively captures bidirectional temporal dependencies of sequential data with a simple structure and fast convergence, and PSO optimizes its key parameters to avoid local optima, enabling the model to fully adapt to the dataset's characteristics.

3.2 Data Processing

Raw machining signals from the PHM2010 dataset contain high-frequency noise and redundant data, which increases the computational burden of the model and reduces prediction accuracy [11]. To improve the quality of input features, a multi-step data processing strategy was adopted. The rationality of key parameter selection was verified by statistical analysis and experimental comparison based on the characteristics of milling signals and industry-recognized criteria [3, 5, 12, 20]. The specific processing steps are as follows:

(1) The original signals were first processed with a 5th-order Butterworth low-pass filter to eliminate high-frequency interference from machine tool vibration and electromagnetic radiation [11] and then downsampled to 20,000 Hz. The modified data were then subjected to EMD to extract IMF components (Figs. 4-5).

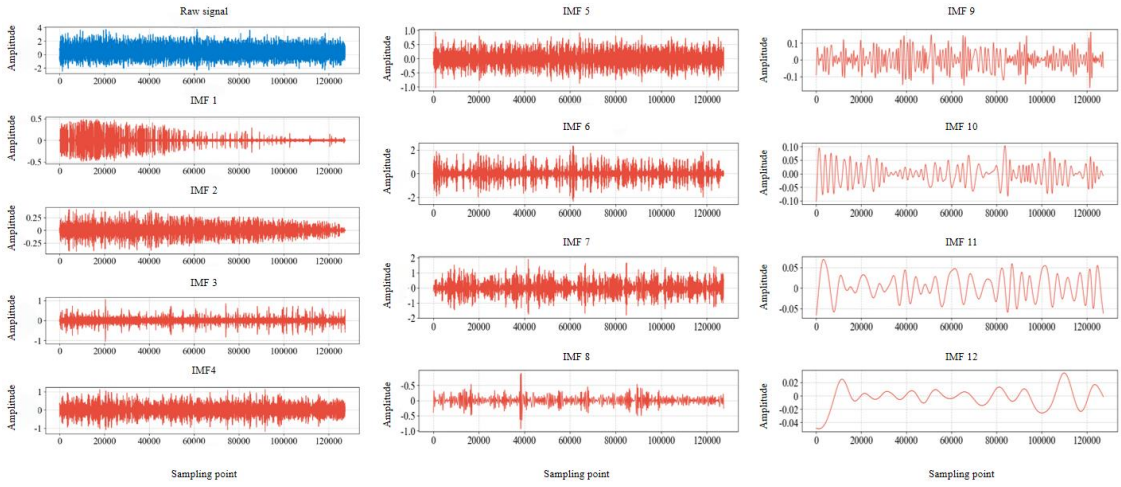

Fig. 4 Original signal and IMF components (IMF1-IMF12)

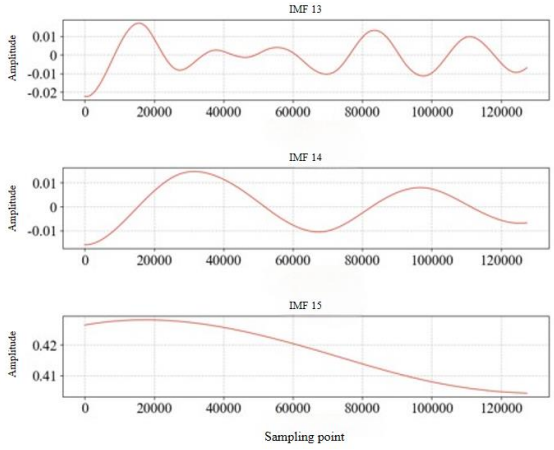

Fig. 5 IMF components (IMF13-IMF15)

(2) The energy of each IMF component was determined, and the modal components that accounted for 90% of the total energy were selected. This threshold is a classic criterion in adaptive signal decomposition and denoising [12, 20], and is verified by the energy characteristics of milling signals. The IMF signals from the first six channels were combined and reconstructed to facilitate the extraction and calculation of key features, as shown in Fig. 6.

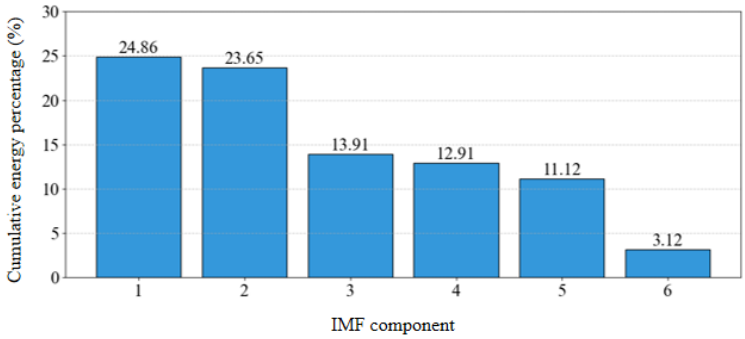

Fig. 6 IMF components with a cumulative proportion of 90%

(3) Time-and frequency-domain characteristics were extracted to reduce data storage and model complexity. The fusion of these features yields the necessary time-frequency-domain information.

(4) A discrete wavelet packet transform was used for multiresolution analysis. Statistical features in the time-frequency domain are derived from eight feature sub-bands (Fig. 7), resulting in a more discriminative time-frequency feature set for future modeling.

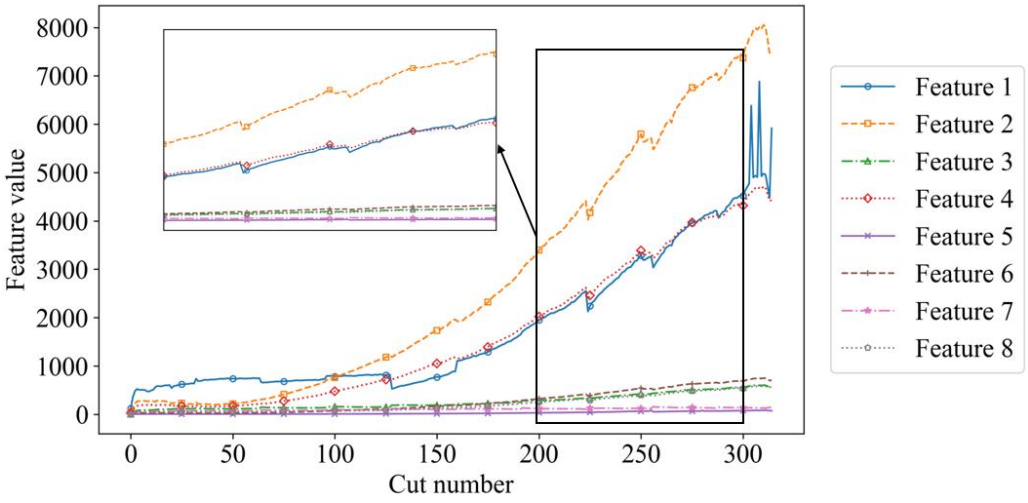

Fig. 7 Time-frequency-domain characteristic map (taking C1 as an example)

(5) The PCC was used to quantify the correlation between the extracted 24-dimensional features and the actual tool wear value (VB) [12], and features with a PCC greater than 0.5 were retained to eliminate redundant features with weak correlation with wear [3,19]. The wear features of the three tools (C1, C4, and C6) are integrated pairwise, and the final wear value label is derived by averaging the wear values across several flutes of each tool. Table 3 details the dataset split.

Table 3 Specific division of the dataset

Name	Training set	Test set
Dataset 1	C4+C6	C1
Dataset 2	C1+C6	C4
Dataset 3	C1+C4	C6

3.3 Model Training

To demonstrate the significant improvement of EMD on prediction outcomes, the data of the first tool were used as an example. The prediction performance obtained with and without EMD processing was compared based on the RMSE. The results are plotted using a bar chart in Fig. 8, showing that the EMD-processed data have a considerably lower RMSE than the raw data [23].

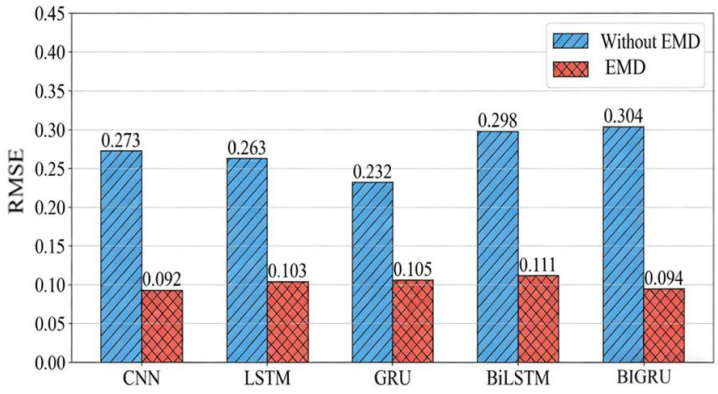

Fig. 8 Comparison of RMSE with and without EMD processing (using C1 as an example)

To validate the performance of the BiGRU-based prediction model, comparative tests were conducted with several typical DL and ML models on identical training and validation sets under consistent experimental parameters and environment. The comprehensive experimental results, together with detailed evaluation metrics, are clearly presented in Table 4.

Table 4 Experimental results of the test set

Model	RMSE			MAE			MAPE			R^2		
	C1	C4	C6	C1	C4	C6	C1	C4	C6	C1	C4	C6
EMD-Lasso	0.117	0.237	0.233	0.080	0.210	0.206	8.610%	25.813%	24.630%	0.815	0.608	0.621
EMD-KNN	0.122	0.231	0.237	0.091	0.209	0.237	8.963%	22.480%	22.480%	0.797	0.609	0.609
EMD-SVR	0.127	0.226	0.209	0.082	0.206	0.177	9.945%	24.385%	22.677%	0.783	0.642	0.695
EMD-CNN	0.092	0.094	0.094	0.072	0.071	0.071	30.928%	31.066%	30.946%	0.885	0.880	0.870
EMD-LSTM	0.103	0.096	0.112	0.073	0.068	0.096	30.144%	30.696%	29.766%	0.856	0.884	0.829
EMD-GRU	0.105	0.108	0.102	0.089	0.069	0.086	30.193%	31.636%	30.214%	0.851	0.843	0.851
EMD-BiLSTM	0.111	0.115	0.125	0.095	0.094	0.110	31.750%	31.829%	32.526%	0.834	0.822	0.788
EMD-BiGRU	0.094	0.097	0.105	0.075	0.070	0.066	30.671%	30.959%	31.773%	0.879	0.873	0.857

The Lasso model (among the ML models) achieves an RMSE of 0.117, whereas the BiGRU model (among the DL models) achieves an RMSE of 0.094, indicating that both perform relatively well in controlling the deviation between predicted and actual values in their respective categories. The MAE values of Lasso (0.080) and BiGRU (0.066) are the lowest in their respective categories, indicating strong predictive accuracy. In terms of MAPE, Lasso achieves the lowest value (8.610%) among the ML models, while BiGRU has a comparatively low value of 30.671% among the DL models. Based on the coefficient of determination (R^2), Lasso (0.815) and BiGRU (0.879) have good fitting effects. Overall, the DL models demonstrate superior capability in minimizing prediction deviation and improving data fitting, whereas Lasso remains competitive among the ML models, particularly in controlling relative error.

4. Results and Discussion

The feature extraction process is the core premise of the proposed model, which directly affects the quality of input features and the final prediction performance. Poorly extracted features will lead to insufficient model learning, low prediction accuracy, and weak generalization ability, while high-quality features can effectively reduce the model's learning burden and improve its prediction stability [24]. Based on the nonlinear and nonstationary characteristics of the machining signals in the PHM2010 dataset, the effectiveness of the EMD–PCC combined feature extraction strategy is discussed from four consecutive steps: signal preprocessing, EMD adaptive decomposition, IMF component screening based on PCC, and feature fusion.

4.1 Feature Extraction Process

Feature extraction and its effective input to the classification algorithm are critical for the proposed model's performance. Raw PHM2010 machining signals undergo 5th-order Butterworth low-pass filtering and downsampling for noise reduction and computational efficiency. EMD adaptively decomposes preprocessed signals into IMFs, from which those with a correlation coefficient >0.1 with the original signal are selected and reconstructed. The get_feature function extracts 24-dimensional features, including 12 time-domain, 4 frequency-domain via FFT, and 8 wavelet packet via db3 wavelet decomposition. PCC (threshold ≈0.8) is then applied to screen wear-related features, with C1/C4/C6 tool features saved separately and merged pairwise for cross-validation datasets.

The screened features are input into the EMD–PSO–BiGRU algorithm for wear prediction. Min-Max normalization scales features to [0,1] using only training set parameters to avoid data leakage. Feature sequences are then constructed for the BiGRU model to fit the time-series characteristics of tool wear data. The EMD–PSO–BiGRU model adopts a three-layer stacked architecture (input layer + BiGRU hidden layer + fully connected output layer) for regression prediction, with a fixed batch_size=32 for model training to balance training stability and computational efficiency, with 100 epochs and an early stopping strategy (patience=15) to prevent overfitting. The BiGRU hidden layer neurons and learning rate are optimized by the PSO algorithm (population size=20, maximum iteration=50, initial inertia weight=0.9, learning factors c1=c2=1.7), with the optimal number of hidden neurons optimized in the range of [64,128] and the optimal learning rate optimized in [0.001,0.1].

The GRU gate unit uses tanh as the activation function to capture the bidirectional temporal dependence of wear time-series data, and sigmoid as the gating activation function to control the weight of inter-temporal information transmission. The fully connected output layer is configured with a neuron and uses a linear activation function (instead of softmax) to match the continuous value prediction task of tool wear. The Adam optimizer is used to optimize the mean squared error (MSE) loss function of the model. Model performance is comprehensively evaluated via RMSE, MAE, MAPE, and R^2, ensuring reliable prediction of tool wear status.

4.2 Experimental Results and Comparative Analysis

Separate ablation experiments were conducted to evaluate the usefulness of the PSO algorithm and the EMD approach in improving prediction accuracy. Control tests were designed to examine the prediction performance of the PSO and enhanced ISABO algorithms, demonstrating the superior prediction performance of the PSO method. The optimization parameters and their ranges are shown in Table 5. The population size is set to 20, the maximum iteration number for optimization is 50, the initial value of inertia weight is 0.9, and the learning factors c_1, c_2 are set to 1.7. Tables 6 and 7 and Figs. 9-14 provide a full account of the experimental outcomes.

Table 5 Optimal parameters and their search ranges

Parameter name	Range of values
Number of neurons in the first layer	[64,128]
Number of second-layer neurons	[64,128]
Learning rate	[0.001,0.1]

Table 6 Experimental results of the PSO algorithm on the test set

Model	C1				C4				C6			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
EMD-PSO-BiGRU	0.079	0.057	5.915 %	0.915	0.095	0.072	7.062 %	0.878	0.094	0.071	13.110 %	0.881
EMD-BiGRU	0.094	0.075	30.671 %	0.879	0.097	0.070	30.959 %	0.873	0.105	0.066	31.773 %	0.857
PSO-BiGRU	0.179	0.129	13.057 %	0.762	0.185	0.167	14.853 %	0.761	0.164	0.136	16.536 %	0.812
BiGRU	0.202	0.175	44.032 %	0.713	0.196	0.161	43.961 %	0.731	0.179	0.160	45.90 %	0.777

Table 7 Experimental results of the ISABO algorithm on the test set

Model	C1				C4				C6			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
EMD-ISABO-BiGRU	0.120	0.091	8.829%	0.891	0.132	0.107	10.418 %	0.880	0.213	0.178	15.059 %	0.871
EMD-BiGRU	0.094	0.075	30.671 %	0.879	0.097	0.070	30.959 %	0.873	0.105	0.066	31.773 %	0.857
ISABO-BiGRU	0.118	0.092	9.558%	0.802	0.117	0.086	8.472%	0.816	0.215	0.179	14.975 %	0.730
BiGRU	0.202	0.175	44.032 %	0.713	0.196	0.161	43.961 %	0.731	0.202	0.175	44.032 %	0.713

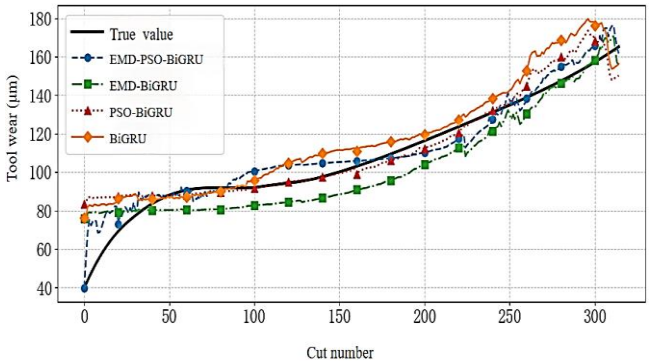

Fig. 9 Training results of the PSO algorithm (C1)

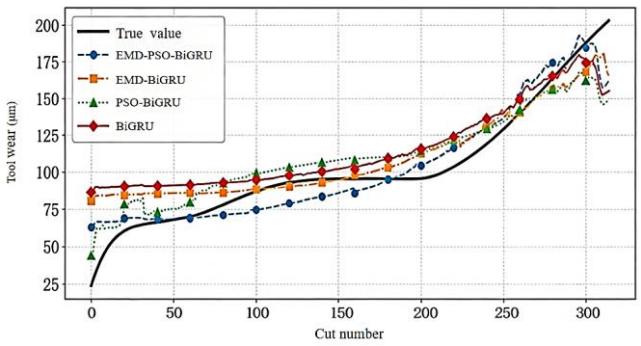

Fig. 10 Training results of the PSO algorithm (C4)

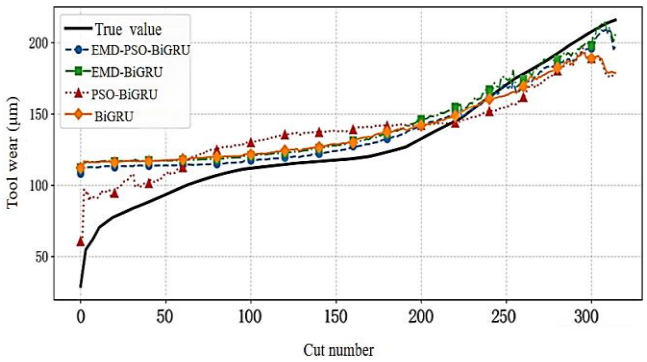

Fig. 11 Training results of the PSO algorithm (C6)

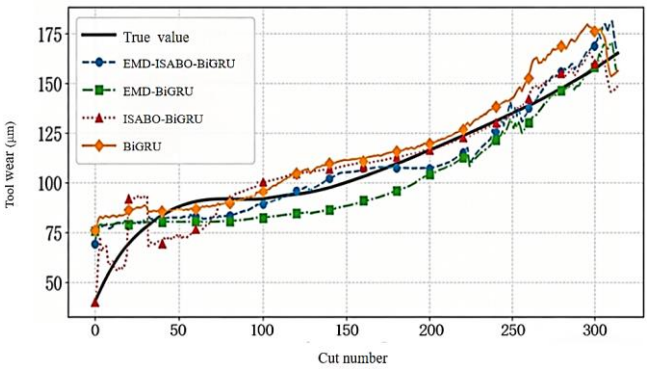

Fig. 12 Training result of the ISABO algorithm (C1)

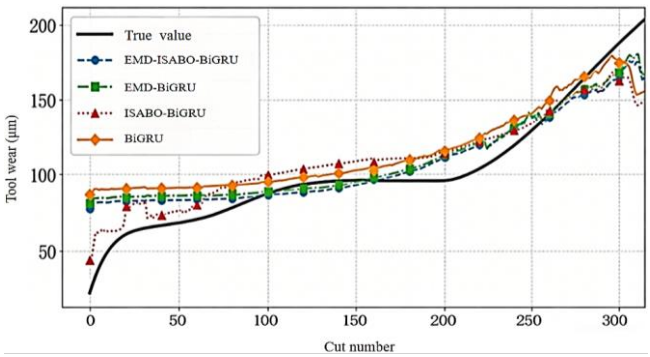

Fig. 13 Training results of the ISABO algorithm (C4)

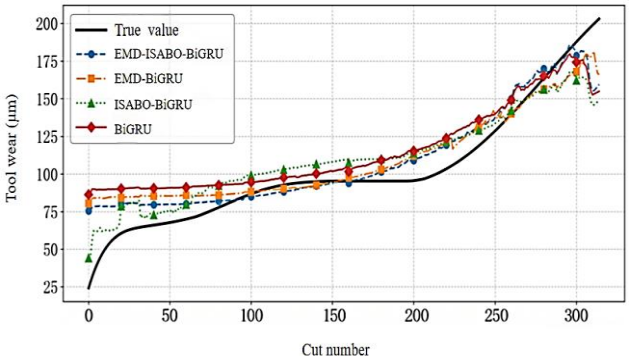

Fig. 14 Training results of the ISABO algorithm (C6)

Quantitative and ablation analyses revealed clear performance hierarchies. The proposed model achieved optimal results across all tools. For C1, it had the lowest RMSE (0.079), MAE (0.057), MAPE (5.915%), and highest R^2 (0.915); C4 and C6 yielded R^2 values of 0.878 and 0.881, respectively. In contrast, the original BiGRU performed worst, with RMSE up to 0.202 and MAPE up to 45.90%, highlighting EMD and PSO's necessity. Ablation tests confirmed their synergistic effect: EMD–PSO–BiGRU reduced RMSE by 15.96–10.48% and increased R^2 by 0.57–4.09% versus EMD–BiGRU (without PSO), and reduced RMSE by 42.68–55.87% versus PSO–BiGRU (without EMD). It also outperformed EMD–ISABO–BiGRU, reducing RMSE by 28.03–62.91% and increasing R^2 by 1.15–2.70%, proving PSO's superior optimization effect.

Qualitative analysis of Figs. 9–14 supported quantitative findings: the proposed model's predicted values closely tracked true wear, with minimal fluctuation, especially in the late wear stage. In contrast, comparison models exhibited obvious deviations. Its superiority stems from three synergistic core components: EMD separates wear information from noise, PCC eliminates redundancy, and PSO optimizes BiGRU to avoid local optima. C1's better performance is attributed to more stable signals, while C6's higher MAPE (13.110%) may result from signal distortion, suggesting potential improvement via advanced noise reduction. These results further demonstrate the robustness and practical applicability of the proposed model under varying signal conditions.

5. Conclusions

This study proposed the EMD–PSO–BiGRU model to tackle nonstationary signal processing and parameter optimization in milling tool wear prediction. EMD with PCC-based screening constructs effective features, and PSO optimizes BiGRU parameters. EMD decomposition and PCC screening effectively extract low-to-medium IMF components carrying over 90% signal energy and strong correlation with tool wear, greatly reducing RMSE and improving R^2. PSO's global optimization further enhances accuracy and convergence speed by avoiding blind parameter tuning. Overall, the synergistic EMD–PSO–BiGRU has achieved superior prediction performance on test sets, outperforming benchmark models.

Limitations include low model interpretability and a lack of quantitative analysis of the EMD–PSO synergy. Future work will integrate attention mechanisms to improve interpretability and establish quantitative relationships between EMD and PSO parameters for adaptive adjustment. Integration with digital twin technology is also planned to reduce reliance on real experimental data, extending industrial applicability in intelligent manufacturing.

Acknowledgments

This work is supervised by professors from Universiti Malaysia Sarawak. This work is also supported by the Engineering Project for Innovation Capability Enhancement of Technology-Based Small and Medium-Sized Enterprises in Shandong Province (No. 202STSGCCZZB0063) and the research program of Qilu Institute of Technology (No. QIT23NN043).

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] J. Kaiser, D. McFarlane, G. Hawkridge, P. André, and P. Leitão, "A Review of Reference Architectures for Digital Manufacturing: Classification, Applicability and Open Issues," Computers in Industry, vol. 149, article no. 103923, 2023.

[2] S. Zhou and H. Shen, "Study on Deformations of Gold Film Induced by Ultrafast Laser at GHz Burst Mode," Journal of Manufacturing Processes, vol. 84, pp. 469-480, 2022.

[3] K. S. Keerthika, M. Asritha, R. N. Kumari, K. R. A. Hegde, and G. Talla, "Tool Life Prediction through Vibration Monitoring During Drilling of Nimonic 90," Sādhanā, vol. 50, no. 4, article no. 273, 2025.

[4] Y. Zhao, H. Hu, C. Song, and Z. Wang, "Predicting Compressive Strength of Manufactured-Sand Concrete Using Conventional and Metaheuristic-Tuned Artificial Neural Network," Measurement, vol. 194, article no. 110993, 2022.

[5] M. Marani, M. Zeinali, V. Songmene, and C. K. Mechefske, "Tool Wear Prediction in High-Speed Turning of a Steel Alloy Using Long Short-Term Memory Modelling," Measurement, vol. 177, article no. 109329, 2021.

[6] M. Ertargin, O. Yildirim, and A. Orhan, "Mechanical and Electrical Faults Detection in Induction Motor Across Multiple Sensors with CNN-LSTM Deep Learning Model," Electrical Engineering, vol. 106, no. 6, pp. 6941–6951, 2024.

[7] H. Liu, P. Li, and Q. Wang, "Short-Term Prediction of Wind Power Based on BKA-CNN-BiGRU-Attention," Electric Power Systems Research, vol. 253, article no. 112543, 2026.

[8] Z. Lin, Y. Fan, J. Tan, Z. Li, P. Yang, H. Wang, and W. Duan, "Tool Wear Prediction Based on XGBoost Feature Selection Combined with PSO-BP Network," Scientific Reports, vol. 15, article no. 3096, pp.1-16, 2025.

[9] T. E. Dribin, A. Muraro, C. A. Camargo, P. J. Turner, J. Wang, G. Roberts, et al., "Anaphylaxis Definition, Overview, and Clinical Support Tool: 2024 Consensus Report—A GA^2 LEN Project," Journal of Allergy and Clinical Immunology, vol. 156, no. 2, pp. 406-417.e6, 2025.

[10] H. Zeng, H. J. Cao, and J. X. Dong, "Tool Wear Prediction Method Based on ISABO-IBiLSTM Model," China Mechanical Engineering, vol. 35, no. 11, pp. 1995-2006, 2024. (in Chinese)

[11] V. Singh, K. Dey, and S. Padmanabhan, "Signal Processing Techniques for Noise Reduction in Industrial Machinery," Proceedings of 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), Indore, India, IEEE, pp. 1-6, 2024.

[12] A. Redekar, H. S. Dhiman, D. Deb, and S. M. Muyeen, "On Reliability Enhancement of Solar PV Arrays Using Hybrid SVR for Soiling Forecasting Based on WT and EMD Decomposition Methods," Ain Shams Engineering Journal, vol. 15, no. 6, article no. 102716, 2024.

[13] X. Chen, Y. Zhang, G. Ai, L. Wang, H. Zhang, X. Li, et al., "Distance Optimization KNN and EMD Based Lightweight Hardware IP Core Design for EEG Epilepsy Detection," Microelectronics Journal, vol. 151, article no. 106335, 2024.

[14] S. Sayyad, S. Kumar, A. Bongale, A. Bongale, P. Kamat, S. Patil, and K. Kotecha, "Data-Driven Remaining Useful Life Estimation for Milling Process: Sensors, Algorithms, Datasets, and Future Directions," IEEE Access, vol. 9, pp. 110255-110286, 2021.

[15] S. Sayyad, S. Kumar, A. Bongale, K. Kotecha, G. Selvachandran, and P. N. Suganthan, " Tool Wear Prediction Using Long Short-Term Memory Variants and Hybrid Feature Selection Techniques," The International Journal of Advanced Manufacturing Technology, vol. 121, no. 9-10, pp. 6611-6633, 2022.

[16] A. Rajakannu, K. Vijayalakshmi, S. R. AV, and J. Wekalao, "Airborne Acoustic Emission-Based Intelligent Wear Detection for Spatiotemporal Data in CNC Drill Bits Using CEEMDAN and ConvLSTM," Sensors and Actuators A: Physical, vol. 396, article no. 117170, 2025.

[17] A. Rajakannu, R. K. P, V. K, and M. Kamarudeen, "Intelligent Fault Diagnosis of Drill Bit Using Digital Twin Assisted Deep Learning Classifiers of MLP, CNN, Hybrid MLP-CNN, and Hybrid CNN-MLP with PCA Algorithm," 2024 5th International Conference on Computers and Artificial Intelligence Technology (CAIT), IEEE, pp. 552-559, 2024.

[18] S. P. Shuvo, S. P. Shibazee, C. Das, G. Paul, K. Malakar, and J. A. Mahmud, "Twofold Integration Viability of EMD–Hilbert Transform for Optimizing Short-Term Precipitation Modeling," IET Signal Processing, vol. 2025, no. 1, article no. 3385508, 2025.

[19] K. A. Shastry and A. Shastry, "Artificial Intelligence Enabled Gait Monitoring for Parkinson's Disease: An Attention-Augmented BiGRU Approach," SN Computer Science, vol. 6, no. 8, article no. 961, 2025.

[20] Y. E. Erdoğan and A. Narin, "Time-Frequency-Based Separation of Earthquake and Noise Signals on Real Seismic Data: EMD, DWT and Ensemble Classifier Approaches," Sensors, vol. 25, no. 21, article no. 6671, 2025.

[21] K. J. Moore, M. Kurt, M. Eriten, D. M. McFarland, L. A. Bergman, and A. F. Vakakis, "Wavelet-Bounded Empirical Mode Decomposition for Measured Time Series analysis," Mechanical Systems and Signal Processing, vol. 99, pp. 14-29, 2018.

[22] X. Li, B. S. Lim, J. H. Zhou, S. Huang, S. J. Phua, K. C. Shaw, et al., "Fuzzy Neural Network Modelling for Tool Wear Estimation in Dry Milling Operation," Annual Conference of the Prognostics and Health Management Society, vol. 1, no. 1, pp. 1-11, 2009.

[23] Y. Duan, H. Li, M. He, and D. Zhao, "A BiGRU Autoencoder Remaining Useful Life Prediction Scheme with Attention Mechanism and Skip Connection," IEEE Sensors Journal, vol. 21, no. 9, pp. 10905-10914, 2021.

[24] W. Wang, S. S. Ngu, M. Xin, R. Liu, Q. Wang, M. Qiu, et al., "Tool Wear Prediction Based on Adaptive Feature and Temporal Attention with Long Short-Term Memory Model," International Journal of Engineering and Technology Innovation, vol. 14, no. 3, pp. 271-284, 2024.