

Article

Multi-Sensor Heterogeneous Signal Fusion Transformer for Tool Wear Prediction

Ju Zhou ^{1,2,†}, Xinyu Liu ^{3,†}, Qianghua Liao ¹, Tao Wang ^{1,*}, Lin Wang ²  and Pin Yang ⁴

¹ Tech X Academy, Shenzhen Polytechnic University, Shenzhen 518055, China; zhouju@szpu.edu.cn (J.Z.); liaoqianghua@szpu.edu.cn (Q.L.)

² Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; lin.wang1@siat.ac.cn

³ College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China; liu1223xy@email.swu.edu.cn

⁴ School of Mechanical Engineering and Automation, University of Science and Technology Liaoning, Anshan 114044, China; yangpin1030@163.com

* Correspondence: charlietree@szpu.edu.cn

† These authors contributed equally to this work.

Abstract

In tool wear monitoring, the efficient fusion of multi-source sensor signals poses significant challenges due to their inherent heterogeneous characteristics. In this paper, we propose a Multi-Sensor Multi-Domain feature fusion Transformer (MSMDT) model that achieves precise tool wear prediction through innovative feature engineering and cross-modal self-attention mechanisms. Specifically, we first develop a physics-aware feature extraction framework, where time-domain statistical features, frequency-domain energy features, and wavelet packet time–frequency features are systematically extracted for each sensor type. This approach constructs a unified feature matrix that effectively integrates the complementary characteristics of heterogeneous signals while preserving discriminative tool wear signatures. Then, a position-embedding-free Transformer architecture is constructed, which enables adaptive cross-domain feature fusion through joint global context modeling and local feature interaction analysis to predict tool wear values. Experimental results on the PHM2010 demonstrate the superior performance of MSMDT, outperforming state-of-the-art methods in prediction accuracy.

Keywords: tool wear prediction; multi-sensor signal; feature fusion; Transformer models; deep learning



Academic Editor: Yuan Yao

Received: 5 June 2025

Revised: 15 July 2025

Accepted: 4 August 2025

Published: 6 August 2025

Citation: Zhou, J.; Liu, X.; Liao, Q.; Wang, T.; Wang, L.; Yang, P.

Multi-Sensor Heterogeneous Signal Fusion Transformer for Tool Wear Prediction. *Sensors* **2025**, *25*, 4847. <https://doi.org/10.3390/s25154847>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In Computer Numerical Control (CNC) machining systems, cutting tools are critical components. Accurate tool wear prediction is important for two reasons: First, inadequate analysis of sensor signals (including acoustic emission, force, and vibration) leads to premature tool replacement, resulting in underutilization of remaining tool life. Second, undetected latent wear can degrade product quality and may even cause tool breakage and machine stoppage in severe cases [1–3]. Consequently, accurate tool wear prediction enables optimal resource utilization, reduces production costs, and prevents machining incidents.

Current monitoring approaches primarily rely on two types of signals to predict tool wear conditions: direct measurement signals (e.g., visual images [4] and laser contours [5]) and indirect inference signals (e.g., cutting force [6], vibration [7], acoustic emission [8],

and spindle motor load [9]). However, monitoring systems based on single-sensor signals exhibit inherent limitations: direct signals remain vulnerable to environmental interference, while indirect signals, being uni-dimensional in nature, cannot fully characterize complex wear mechanisms. The fundamental drawback of single-sensor approaches lies in their ability to capture only localized wear features, rendering them incapable of overcoming environmental sensitivity and establishing robust correlations with wear states. These limitations consequently compromise the reliability and generalizability of monitoring systems.

To overcome the bottlenecks of single-sensor systems, the use of multi-sensors to acquire diverse signals for monitoring tool wear conditions has emerged as a mainstream research trend. By integrating signals such as force, vibration, and acoustic emission, researchers are able to extract more comprehensive information about the tool wear process [10,11]. For example, Silva et al. [12] demonstrated the feasibility of fusing acoustic emission and power signals for tool wear monitoring, revealing that the multi-sensor approach yields significantly superior results compared to single-sensor methods. However, the data acquired from different sensor sources inherently exhibit heterogeneity in terms of sampling rates, physical units, and information granularity. The key challenge lies in developing effective fusion strategies that can fully exploit the complementary information while addressing the dimensional inconsistencies among these heterogeneous data streams.

With the rapid development of computer technology, numerous machine learning and deep learning approaches have been introduced into the field of tool wear monitoring [13]. These studies primarily consist of two steps: firstly, feature engineering, which involves extracting key features from signals, and subsequently, training high-performance prediction models based on these features.

In feature engineering research for tool wear monitoring, researchers have employed various signal processing methods to extract features from sensor data. For vibration signals, Rmili et al. [14] extract average power as a feature to represent tool wear state, while Elangovan et al. [15] focus on statistical feature extraction, and Huang et al. [7] utilize the Short-Time Fourier Transform (STFT) for time–frequency analysis to obtain key features. For vibration and acoustic signals, Lin et al. [16] apply the Fast Fourier Transform (FFT) for feature extraction, while Rafezi et al. [17] not only extract time-domain statistical features but also analyze spectral characteristics and employ wavelet packet decomposition to explore specific frequency band information. However, these feature extraction methods exhibit certain limitations. First, the extracted feature types remain relatively simplistic, with different methods being applied to different signals, lacking comprehensive exploration of the multi-dimensional characteristics. Second, the feature design relies on manual expertise, resulting in limited effectiveness and adaptability when facing different signal types or varying working conditions. More critically, these methods struggle to automatically capture deep correlation features between signals, leading to insufficient information utilization. This feature extraction homogeneity constrains the generalization capability of monitoring systems in complex industrial scenarios.

In tool wear modeling research, researchers have progressively shifted from traditional machine learning methods to deep learning models. The deep Convolutional Neural Network (CNN) model developed by Fu et al. [18] demonstrates superior performance to conventional Support Vector Machine (SVM) in drilling process monitoring. By incorporating the deep Residual Network (ResNet) and Long Short-Term Memory (LSTM) network, Sun et al. [19] propose the hybrid ResNet-LSTM architecture, which achieves more accurate multi-step prediction through multi-sensor signal inputs. Wang et al. [20] construct the CNN-BiLSTM model, which further enhances monitoring performance. Although these deep learning approaches have made significant advancements, current models exhibit notable limitations: CNN-based methods are constrained by local receptive fields, hindering their ability to model

global dependencies in long sequences; LSTM and its variants, while proficient in processing temporal data, suffer from inherently inefficient training due to recursive computation. Furthermore, these conventional network architectures lack sufficient capability for differentiated feature processing of multi-source heterogeneous sensor signals, failing to adaptively capture inter-feature relationships and achieve effective fusion.

To tackle the aforementioned issues, we propose a Multi-Sensor Multi-Domain feature fusion Transformer (MSMDT) approach for tool wear prediction. For heterogeneous signals from multi-sensors, including acoustic emission, cutting force, and vibration, this approach first introduces a unified physical perception-based feature extraction method. This method systematically processes each sensor's signal through complementary time-domain statistical analysis, frequency-domain energy representation, and wavelet packet time–frequency decomposition. These multi-domain features are then integrated into a compact and comprehensive feature matrix, which not only preserves the physical interpretability of specific sensors but also enables cross-modal correlation analysis. Furthermore, to comprehensively model the complex nonlinear interactions within the feature matrix, we develop the MSMDT framework based on the Transformer architecture. By designing a feature fusion encoder and eliminating the position embedding process, the MSMDT enables efficient parallel processing and integration of multi-sensor multi-domain features. This facilitates information interaction among these features and ultimately enhances the accuracy of tool wear prediction. The main contributions of this paper are as follows:

1. Multi-domain feature fusion strategy for multi-sensor heterogeneous signals: Effectively integrates multi-source industrial sensor signals through collaborative modeling of time-domain, frequency-domain, and time–frequency-domain features, providing feature-re-engineering-free extensibility to diverse sensor types while leveraging complementary signal characteristics.
2. Position-embedding-free MSMDT network design: Enables parallel processing and real-time collaborative prediction of cross-sensor information, enhancing effective feature fusion for heterogeneous temporal signals.
3. Breakthrough in single-sensor and single-feature dependency limitations: The proposed method can adaptively extract deep-level features that characterize tool wear and automatically predict its progression, achieving promising results.

The rest of this paper is structured as follows. Section 2 presents the related work, Section 3 describes the proposed MSMDT, Section 4 presents the experimental results and comparative analysis, and Section 5 summarizes and discusses this study.

2. Related Work

In tool wear monitoring research, cutting force, vibration, and acoustic emission signals reflect wear states from distinct perspectives. The cutting force signals characterize mechanical load variations at the tool–workpiece interface, while vibration and acoustic emission signals capture structural dynamics and microscopic material damage, respectively. Current studies employ different feature extraction strategies tailored to each signal type: time-domain statistical features such as root mean square and peak-to-peak values are primarily extracted from force signals to quantify overall cutting force intensity [21,22]; frequency-domain features are applied to vibration signals to track energy shifts in specific frequency bands [23–25]; and time–frequency features obtained through wavelet packet transform analyze the non-stationary characteristics of acoustic emission signals [26,27].

Although these domain-specific methods demonstrate effectiveness in individual signal processing, it does not imply other signal domains lack valuable information. For instance, cutting force signals contain meaningful frequency components reflecting tool resonance, and time-domain kurtosis of vibration effectively detects impact events during

wear initiation. Therefore, single-domain features may fail to comprehensively characterize complex wear progression, and cross-sensor feature correlations remain underutilized. This limitation necessitates a unified multi-domain feature extraction framework capable of simultaneous time-domain, frequency-domain, and time–frequency-domain characterization of multi-sensor signals. Such an integrated approach preserves the physical significance of each signal modality while exploiting feature complementarity to achieve more complete wear state representation, thereby providing enhanced information infrastructure for subsequent feature fusion and prediction modeling.

Among various networks, LSTM networks effectively address the vanishing gradient problem of traditional recurrent neural networks (RNNs) through gating mechanisms and memory cell design, making them suitable for capturing long-term temporal dependencies in tool wear processes [28–30]. To better use both forward and backward time-series information, bidirectional LSTM (BiLSTM) has been introduced to capture the dynamic evolution of tool wear more comprehensively and improve the accuracy [31,32]. Currently, hybrid models combining LSTM with other network architectures have become a research focus, such as integrating CNNs for simultaneous local feature extraction and temporal dependency modeling or incorporating attention mechanisms to enhance key feature selection, with these hybrid approaches demonstrating superior robustness and performance in wear prediction [20,33,34].

Although LSTM performs well in tool wear prediction, its inherent sequence dependency leads to low computational efficiency and difficulty in establishing global dependencies across sensors. In contrast, the Transformer architecture demonstrates advantages in multi-sensor data processing [35]. Its core multi-head attention mechanism can process all sequential data in parallel, significantly enhancing computational efficiency. Meanwhile, through the self-attention mechanism, it dynamically models feature correlations at arbitrary positions, making it suitable for capturing complex interactions among multi-source sensor signals. Recently, the Transformer architecture has achieved notable success in the field of fault diagnosis [36–38].

Therefore, this study proposes the MSMDT to predict tool wear states, which addresses the limitations of current research from two aspects. On the one hand, by constructing a multi-domain feature extraction strategy, it effectively tackles the issues of heterogeneous multi-sensor signals and inadequate information utilization. For different temporal signals, the same feature extraction strategy is employed to extract meaningful information, which is then integrated into a compact feature matrix. This approach not only ensures full utilization of information but also achieves a multi-modal consistent representation of tool wear states, providing a unified input format for the model. On the other hand, the MSMDT utilizes the Transformer architecture to effectively fuse multi-domain features from multi-sensors, establishing a reliable wear prediction model. Leveraging the Transformer, long-range dependencies of tool wear states can be established without recursion, and the importance of features extracted from different sensor signals is automatically assigned through the self-attention mechanism, thereby adaptively extracting features highly relevant to the tool wear state. The unified feature extraction strategy and extensible prediction model of MSMDT offer a promising solution for future scenarios involving different sensors and operating conditions.

3. MSMDT

The structure of the proposed MSMDT model is shown in Figure 1. The MSMDT framework primarily consists of four key processes: multi-sensor signal input, multi-domain feature extraction, Transformer encoder processing, and regression decoder output.

To overcome the limitation of poor anti-interference ability in single-sensor systems, a multi-sensor signal input structure is designed. Here, *AE* represents the acoustic emission (AE) signal, while F_x , F_y , and F_z denote the cutting force signals collected in the three orthogonal directions (X, Y, Z), respectively. Similarly, V_x , V_y , and V_z correspond to the vibration signals in the same directions. These seven-channel signals are collectively input into the MSMDT model. In the multi-domain feature extraction phase, 28 features are extracted from each signal to reduce data dimensionality while preserving critical tool wear-related information, forming a high-density feature matrix $X_i \in \mathbb{R}^{7 \times 28}$. These features cover multiple domains, including the time domain, frequency domain, and wavelet packet decomposition time–frequency domain. Unlike traditional tool prediction models, our approach utilizes a Transformer-based backbone network to process the feature matrix X_i . This is achieved through the encoder module of the Transformer, the detailed structure of which is illustrated on the right side of Figure 1. Notably, the MSMDT model omits position embeddings compared to standard Transformer architectures to decouple multi-sensor features and enhance system performance. After extracting deep features in the Transformer encoder, a flattening layer transforms these features into a 196×1 vector to adapt to the MSMDT-specific regression decoder. As tool wear prediction is a regression task, the decoder comprises three fully connected layers with node configurations of $196 \rightarrow 128$, $128 \rightarrow 64$, and $64 \rightarrow 1$, ultimately outputting the predicted tool wear value.

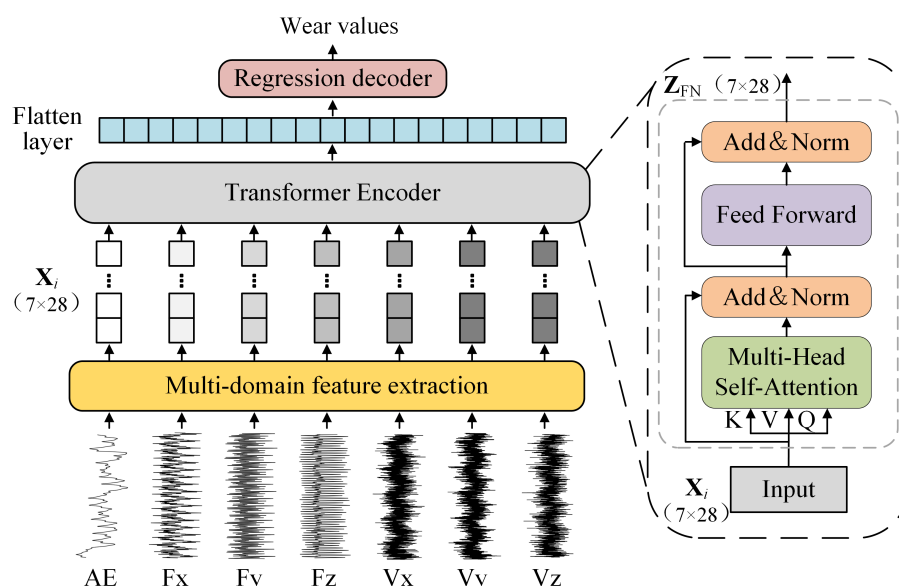


Figure 1. Structure of MSMDT.

3.1. Multi-Sensor Signal Input

During the machining process, relying solely on a single sensor may lead to unreliable tool wear assessment due to inherent sensor limitations, signal transmission issues, and external noise interference. To overcome these limitations, multi-sensor fusion technology has become the predominant approach for comprehensive tool condition monitoring. This study focuses on three complementary sensor modalities: acoustic emission sensor, cutting force sensor, and vibration sensor.

- Acoustic emission sensors are cost-effective and readily available, but they are prone to mechanical noise and require complex signal analysis.
- Cutting force sensors are highly sensitive to tool wear signals and provide accurate monitoring, albeit at a higher cost.
- Vibration sensors offer easy installation, low cost, and abundant information, but they are susceptible to environmental interference.

Currently, these three sensor types are widely adopted in CNC machining systems for tool wear monitoring due to their complementary sensing capabilities. Among them, the acoustic emission signal contains one-dimensional data, and the force and vibration signals contain three directional components. This means that the three sensors have a total of seven signal channels. However, the heterogeneous nature of these signals poses significant challenges for integrated analysis: these signals have different units of measurement and ranges of numerical variation; they contain sensor-specific noise patterns and interference characteristics; and they respond to tool wear with varying sensitivity and latency. This heterogeneity makes it difficult to establish consistent correlations between the raw signals and tool wear states. To address these challenges, a unified multi-domain feature extraction framework is proposed to process diverse temporal signals from multi-source industrial sensors while preserving their essential wear-related information. This framework adaptively handles heterogeneous signals and extracts discriminative features for accurate tool condition assessment.

3.2. Multi-Domain Feature Extraction

Before feature extraction, the raw signals undergo two preprocessing steps: outlier removal via the quartile method and elimination of direct current (DC) components. The outlier removal ensures data integrity by excluding abnormal measurements, while DC component elimination prepares signals for effective frequency-domain feature extraction. These operations enhance signal quality for subsequent multi-domain feature analysis under varying operational conditions, providing optimized inputs for the Transformer model.

To overcome the limitation of redesigning feature engineering processes when signals or their combinations change, this study develops a unified multi-domain feature extraction framework. This framework further fuses features into a common representation, maximally extracting discriminative features from diverse signals across multiple dimensions for effective tool wear state prediction. The multi-domain feature extraction comprises three distinct domains (time domain, frequency domain, and time–frequency domain), totaling 28 features, with detailed mathematical formulations and physical explanations provided in Table 1. Specifically, the following are considered:

- Time domain (9 features): Statistical descriptors of signal amplitudes, including standard deviation, variance, peak-to-peak, RMS value, skewness coefficient, kurtosis coefficient, crest factor, margin factor, and waveform factor.
- Frequency domain (3 features): Spectral characteristics including center gravity frequency, frequency variance, and mean square frequency.
- Time–frequency domain (16 features): Wavelet packet decomposition energies across 16 sub-bands.

For seven input signals, the use of a multi-domain feature extraction approach yields a compact and informative feature matrix (7×28) that combines physical interpretability with data-driven robustness. Wavelet-derived features compensate for the limitations of a single time- or frequency-domain analysis, and the low dimensionality of the matrix ensures computational efficiency while retaining the ability to discriminate tool wear. More importantly, the inherent adaptability of the framework allows for the free integration of additional sensor inputs or machining conditions through its standardized feature extraction approach. These high-quality fused features are subsequently fed into the Transformer encoder module, where their inherent correlations and global dependencies are further extracted through self-attention mechanisms to establish a more comprehensive wear state representation.

Table 1. Multi-domain feature extraction methods.

Domain	Feature	Expression	Explanation
Time domain (9 features)	Standard deviation	$x_{std} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$	Sample standard deviation
	Variance	$x_{var} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	Sample variance
	Peak-to-peak	$x_{p-p} = x_{max} - x_{min}$	Difference between extrema
	RMS value	$x_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$	Root mean square
	Skewness coefficient	$x_\gamma = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n x_{std}^3}$	3rd standardized moment
	Kurtosis coefficient	$x_g = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n x_{std}^4} - 3$	Excess kurtosis
	Crest factor	$x_c = \frac{ x_{p-p} }{x_{rms}}$	Peak-to-RMS ratio
	Margin factor	$x_m = \frac{x_{p-p}}{\frac{1}{n} \sum_{i=1}^n x_i}$	Peak-to-average ratio
	Waveform factor	$x_w = \frac{x_{rms}}{\frac{1}{n} \sum_{i=1}^n x_i}$	RMS-to-average ratio
Frequency domain (3 features)	Center gravity frequency	$f_{cg} = \frac{\sum_{i=1}^n f_i p_i}{\sum_{i=1}^n p_i}$	Spectral centroid
	Frequency variance	$f_v = \frac{\sum_{i=1}^n (f_i - f_{cg})^2 p_i}{\sum_{i=1}^n p_i}$	Spectral spread
	Mean square frequency	$f_{ms} = \frac{\sum_{i=1}^n f_i^2 p_i}{\sum_{i=1}^n p_i}$	Spectral RMS
Time–frequency domain (16 features)	Wavelet packet energy	$E_{wp} = \sum_{i=1}^n w_{j,k}[i] ^2$	Decomposition energy

Common symbols: x_i —data point; $\bar{x}$ —sample mean; n —data count; f_i —frequency component; p_i —power spectral density; $w_{j,k}[i]$ —wavelet coefficient at level j , node k .

3.3. Backbone Structure

3.3.1. Input Representation and Tokenization

The encoder input is a multi-domain feature matrix $X_i \in \mathbb{R}^{7 \times 28}$, where each sensor channel's feature vector is mapped to an independent token. Specifically, for the k -th sensor's feature vector,

$$\text{Token}_k = X_i[k, :] \in \mathbb{R}^{28}, \quad k = 1, 2, \dots, 7 \quad (1)$$

The input is consequently represented as a token sequence:

$$X_{\text{tokens}} = [\text{Token}_1; \text{Token}_2; \dots; \text{Token}_7] \in \mathbb{R}^{7 \times 28} \quad (2)$$

This representation enables parallel processing of multi-sensor features, where each token encapsulates time-domain statistical features, frequency-domain energy spectrum characteristics, and wavelet packet energy features, physically representing the parallel characterization of multi-sensor features.

3.3.2. Position-Embedding-Free Inter-Sensor Interaction

Fundamentally distinct from the traditional Transformer architecture, this study removes the position embedding operation. The decision is theoretically grounded in three aspects: Firstly, in tool wear monitoring tasks, the physical arrangement order of sensors generally does not contain semantically meaningful information, and forcibly introducing positional indices may cause feature entanglement, thereby damaging feature independence. Secondly, the self-attention mechanism intrinsically possesses the capability to autonomously learn inter-sensor correlations without requiring predefined spatial dependencies. Thirdly, when integrated with the tokenization strategy described in Section 3.3.1, this architecture exhibits remarkable scalability, where changes in the number of sensors

only necessitate adjustments to the token count while preserving consistent encoding rules, thereby obviating the need for recalibration of position embeddings. The effectiveness of this position embedding removal strategy is further validated through quantitative ablation studies in our experimental section (see Section 4.3.3).

3.3.3. Encoder Computational Architecture

The encoder architecture implements hierarchical feature transformation through stacked self-attention and feed-forward layers, maintaining sensor-specific information flow while enabling cross-modal interaction.

(1) Multi-Head Self-Attention Mechanism

The core operator establishes latent sensor correlations via query–key–value projection:

$$\begin{aligned} Q &= X_{\text{tokens}} W_Q, & W_Q &\in \mathbb{R}^{28 \times d_k} \\ K &= X_{\text{tokens}} W_K, & W_K &\in \mathbb{R}^{28 \times d_k} \\ V &= X_{\text{tokens}} W_V, & W_V &\in \mathbb{R}^{28 \times d_k} \end{aligned} \quad (3)$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \in \mathbb{R}^{7 \times d_k} \quad (4)$$

where d_k controls the latent interaction space dimension. This design complements the position embedding removal in Section 3.3.2, enabling physics-guided feature learning.

(2) Layer-wise Processing

Each encoder layer l processes token sequences through two cascaded sublayers with residual connections and layer normalization:

$$Z^{(l)} = \text{LayerNorm}\left(X_{\text{tokens}}^{(l)} + \text{MHA}\left(X_{\text{tokens}}^{(l)}\right)\right) \quad (5)$$

$$X_{\text{tokens}}^{(l+1)} = \text{LayerNorm}\left(Z^{(l)} + \mathcal{F}(Z^{(l)})\right) \quad (6)$$

where $\text{MHA}(\cdot)$ denotes the multi-head attention mechanism defined in Equation (4) and $\text{LayerNorm}(\cdot)$ represents layer normalization, which standardizes each token's features by subtracting the mean and dividing by the standard deviation.

The feed-forward network $\mathcal{F}$ can be represented as follows:

$$\mathcal{F}(x) = \text{GELU}(xW_1 + b_1)W_2 + b_2, \quad \begin{cases} W_1 \in \mathbb{R}^{28 \times 112} \\ W_2 \in \mathbb{R}^{112 \times 28} \end{cases} \quad (7)$$

Here, the Gaussian Error Linear Unit (GELU) function $\text{GELU}(x) = x\Phi(x)$ serves as the activation function. The final output $X_{\text{tokens}}^{(\text{out})} \in \mathbb{R}^{7 \times 28}$ maintains the dimensional consistency.

3.4. Regression Decoder

The decoder module maps deep features from the Transformer encoder to scalar predictions of tool wear. The encoder output $X_{\text{tokens}}^{(\text{out})} \in \mathbb{R}^{7 \times 28}$ is first flattened into a 1D vector using

$$z = \text{Flatten}\left(X_{\text{tokens}}^{(\text{out})}\right) \in \mathbb{R}^{196} \quad (8)$$

The regression decoder then processes the features through three fully connected layers:

$$h_1 = \text{ReLU}(W_3 z + b_3) \quad (9)$$

$$h_2 = \text{ReLU}(W_4 h_1 + b_4) \quad (10)$$

$$\hat{y} = W_5 h_2 + b_5 \quad (11)$$

In Equation (8), the flatten operation reshapes the 7×28 matrix into a 196-dimensional vector in row-major order. Equations (9)–(11) define the computational workflow of the fully connected layers, where $W_3 \in \mathbb{R}^{128 \times 196}$ and $b_3 \in \mathbb{R}^{128}$ represent the weight matrix and bias vector for the first layer, $W_4 \in \mathbb{R}^{64 \times 128}$ and $b_4 \in \mathbb{R}^{64}$ those for the second layer, and $W_5 \in \mathbb{R}^{1 \times 64}$ and $b_5 \in \mathbb{R}^1$ those for the output layer. The Rectified Linear Unit (ReLU) activation function $\text{ReLU}(x) = \max(0, x)$ provides nonlinear transformations between layers.

The model is trained using a combined loss function of L1 loss and Mean Squared Error (MSE). They are calculated as follows:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (12)$$

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (13)$$

The composite loss function combines these two components:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{L1}} + (1 - \alpha) \mathcal{L}_{\text{MSE}} \quad (14)$$

In Equations (12) and (13), y_i denotes the actual tool wear value of the i -th sample, $\hat{y}_i$ denotes the predicted value, and N represents the batch size. $\alpha \in [0, 1]$ controls the relative contribution of L1 and MSE losses in Equation (14). Parameter updates are performed via backpropagation:

$$\Theta^{(t+1)} = \Theta^{(t)} - \eta \nabla_{\Theta} \mathcal{L}_{\text{total}} \quad (15)$$

where $\Theta = \{W_3, b_3, W_4, b_4, W_5, b_5\}$ represents the trainable parameter set, η is the learning rate, and t indexes the training iteration. This hierarchical dimensionality reduction architecture enables robust mapping from the high-dimensional feature space to tool wear measurements in physical units.

4. Experimentation and Validation

4.1. Experimental Setup

4.1.1. Datasets and Preprocessing

To validate the effectiveness of the proposed MSMDT in addressing the fundamental challenges of sensor data fusion for tool wear prediction, all experiments in this study are conducted on the publicly available PHM2010 dataset [39]. This choice is particularly valuable for demonstrating the robustness of our approach, as benchmarking on standardized public data provides a rigorous and reproducible foundation to evaluate generalization capabilities across diverse sensor scenarios. In this dataset, each tool executes a 108 mm stroke in the X direction, involving a total of 315 cuts across six tools. The machining process employs the following cutting parameters: spindle speed of 10,400 rpm, feed rate of 1555 mm/min, Y-axis depth of cut of 0.125 mm, and Z-axis depth of cut of 0.2 mm. The signals are initially amplified and subsequently captured by a data acquisition card operating at a sampling frequency of 50 kHz. The sensor-acquired signals comprise seven-dimensional data, including acoustic emission, cutting forces (in the X, Y, and Z directions), and vibrations (in the X, Y, and Z directions), with corresponding tool wear values recorded

simultaneously. In this study, we use three available tools (C1, C4, and C6) from the public PHM2010 dataset [39] for analysis and validation.

The quantification of tool wear follows a standardized protocol in which a wear label for each cycle is generated by averaging three flank (three-flute ball end mill) wear measurements. Using C1 as an example, Figure 2a shows the wear curves on the three flanks as well as the final wear pattern. Figure 2b shows the wear progression curves for tool groups C1, C2, and C3, demonstrating a characteristic three-phase wear pattern across all cases: (1) Initial wear phase: Characterized by a sharp slope increase in wear values. (2) Steady wear phase: Marked by stable tool wear values with smaller gradient amplitudes. (3) Severe wear phase: Exhibiting an abrupt slope surge until functional failure. All curves maintain strict monotonic non-decreasing behavior throughout the cutting process.

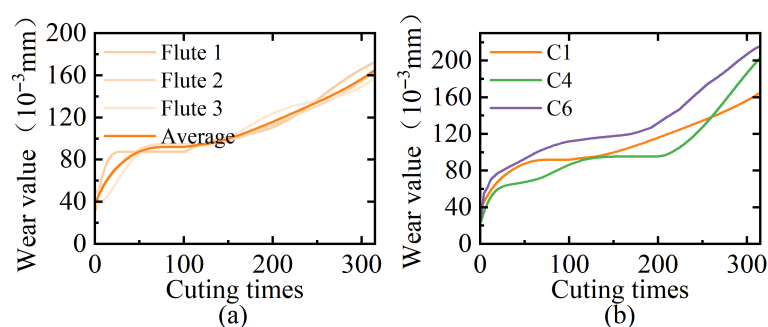


Figure 2. (a) Three flank wear measurements of C1; (b) wear progression curves for C1, C4, and C6.

The preprocessing workflow comprises three critical stages: **label construction** computes cycle labels through averaging flank wear measurements; **feature engineering** extracts time-domain, frequency-domain, and time–frequency-domain features following Table 1 to build a feature matrix; and **cross-validation** adopts a hold-out strategy, as shown in Table 2, where two tool groups (630 samples) train the model and the remaining group (315 samples) independently tests model generalizability.

Table 2. Experimental groups.

Group	Train Set	Test Set
1	C4, C6	C1
2	C1, C6	C4
3	C1, C4	C6

4.1.2. Model Configuration and Hyperparameters

The experimental platform utilizes an Intel I7-13650H processor and an NVIDIA RTX 4060 GPU, with the model implemented using the PyTorch framework (Version: 2.3.1). For model configuration, we utilize the Adam optimizer with a learning rate of 0.001. The batch size is set to 32 to optimally balance GPU memory constraints and training efficiency. Through empirical analysis of the loss curve, we confirm that 500 training epochs ensure model convergence. A dropout rate of 0.1 is applied to prevent overfitting in the MSMDT architecture while maintaining critical features. The loss function weight α is 0.5, which is determined through grid search experiments in the range of 0.1 to 0.9, reflecting equal importance between L1 and MSE loss components. To guarantee reproducibility, we fix the random seed at 3407. All experimental results in this study can be reproduced using these key parameters.

4.1.3. Evaluation Metrics

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are adopted, and they are calculated as follows:

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i| \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2} \quad (17)$$

where y_i denotes the ground truth wear value, $\hat{y}_i$ represents the predicted value, and M indicates the number of test samples. Lower metric values correspond to higher prediction accuracy. In addition, the absolute error between the predicted and actual values is calculated and visualized to show the model prediction performance.

4.2. Visualization and Analysis of Multi-Sensor Features

To explore the correlations between sensor signals and tool wear, this study first validates the effectiveness of time-domain features. Taking tool C1 as an example, Figure 3 visualizes the temporal characteristics of cutting force, vibration, and AE signals at three critical wear stages (100th, 200th, and 300th cutting cycles). It should be noted that in Figure 3c, the AE signal exhibits a different sampling number compared to the cutting force and vibration signals. This difference arises because the AE signal demonstrates minimal local variations within short sampling intervals. However, as the number of sampling points increases, distinct changes emerge in the AE signal characteristics, reflecting its unique property of being primarily sensitive to transient events rather than continuous processes. The results show that the waveform amplitudes progressively increase with the number of cutting cycles, leading to a direct rise in both peak and RMS values. These time-domain variations exhibit state-sensitive discriminative patterns. This phenomenon may correlate with enhanced cutting resistance caused by material deformation. Consequently, multi-dimensional feature extraction is employed to enhance the distinguishability of patterns between different wear stages, facilitating robust tool wear prediction modeling.

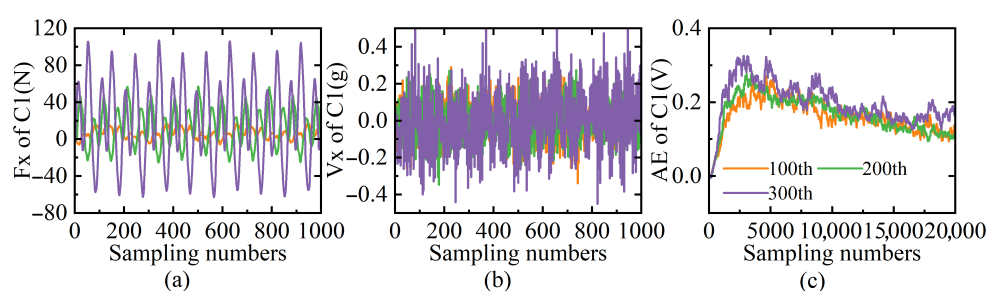


Figure 3. Curves of C1 in (a) cutting force signal; (b) vibration signal; (c) AE signal.

Building upon the established time-domain correlations, this study evaluates the feasibility of multi-sensor feature extraction using the multi-domain methodology in Table 1, which systematically extracts 9 time-domain features, 3 frequency-domain components, and 16 wavelet packet energy ratios. A representative subset of C1 tool features is selected for validation, with normalized distributions visualized in Figure 4. It should be noted that the aforementioned characteristics of the AE signal lead to the distinctive patterns observed in Figure 4d,e: the frequency-domain and partial time–frequency features extracted from the AE signal predominantly maintain zero values or remain constant. The analysis reveals two critical observations: (1) all features exhibit progressive variations

aligned with tool wear progression, confirming their sensitivity to machining state changes; (2) individual feature trajectories show insufficient wear state specificity due to overlapping value ranges between adjacent stages. To address this limitation, the MSMDT prediction model implements coordinated parallel processing of heterogeneous features, enabling simultaneous analysis of time-domain, frequency-domain, and time–frequency-domain characteristics. This integrated approach demonstrates the capability to establish robust correlations between multi-sensor features and tool wear evolution.

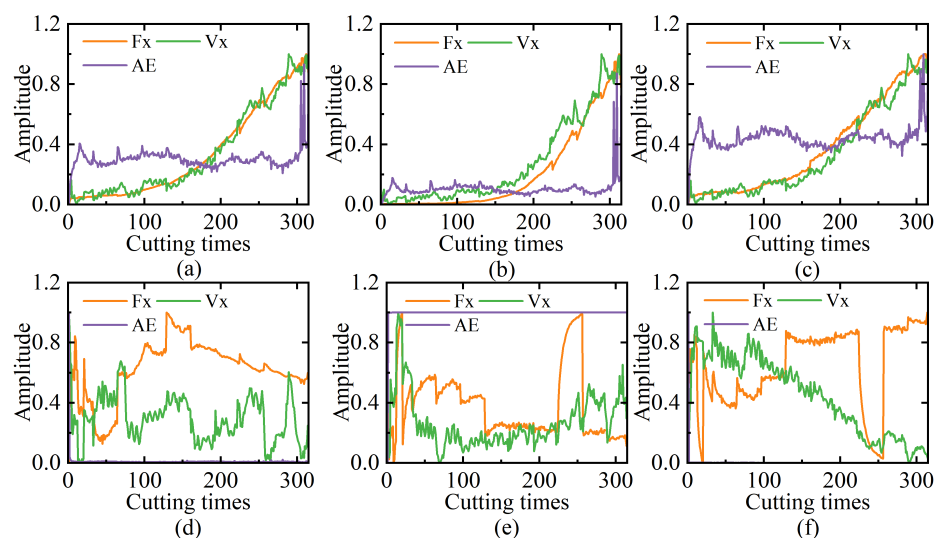


Figure 4. Curves of C1 in features from different domains: (a) standard deviation; (b) variance; (c) peak-to-peak; (d) frequency variance; (e) energy ratio 1; (f) energy ratio 2.

4.3. Ablation Experiment

The prediction accuracy of the proposed method for tool wear prediction demonstrates significant dependence on multi-domain feature selection and model architecture parameters. To quantify these influences, ablation studies systematically evaluate four critical variables: (1) sensor type integration, (2) feature-domain combinations, (3) position embedding effectiveness, and (4) number of attention heads.

4.3.1. Ablation Experiment on Multi-Sensor Signal Input

To systematically validate the superiority of multi-sensor fusion in tool wear monitoring, we conduct rigorous comparative experiments across three tool cases (C1/C4/C6) with four sensing configurations: individual cutting force, vibration, and AE sensors versus their fused combination. Table 3 shows the results of different sensor types, and Figure 5 displays the comparison of predicted tool wear versus the ground truth, including trend curves and prediction errors.

Table 3. Performance comparison of different sensor approaches.

Sensor Type	C1		C4		C6	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Cutting force sensor	17.43	22.62	27.74	30.27	13.80	16.82
Vibration sensor	10.81	12.30	10.75	15.79	12.36	15.04
AE sensor	20.77	25.26	26.90	34.14	24.77	33.52
Multi-sensor	4.47	6.35	8.27	12.06	6.06	7.19

Note: Bold values indicate better performance.

The experimental results reveal significant performance differences among different single-sensor signal inputs. It can be seen that the vibration sensor demonstrates superior

performance to the cutting force sensor, while the AE sensor yields the least favorable results. This phenomenon is also clearly reflected by the area size enclosed by the error curve shown in Figure 5. The vibration sensor achieves its optimal performance on C1 with an MAE of 10.81. By employing multi-sensor signals, this value decreases to 4.47, representing a performance improvement exceeding 50%. The results demonstrate that the multi-sensor input strategy consistently outperforms single-sensor approaches, effectively compensating for individual sensor limitations through complementary data integration, with substantial enhancements in both MAE and RMSE compared to single-sensor methods.

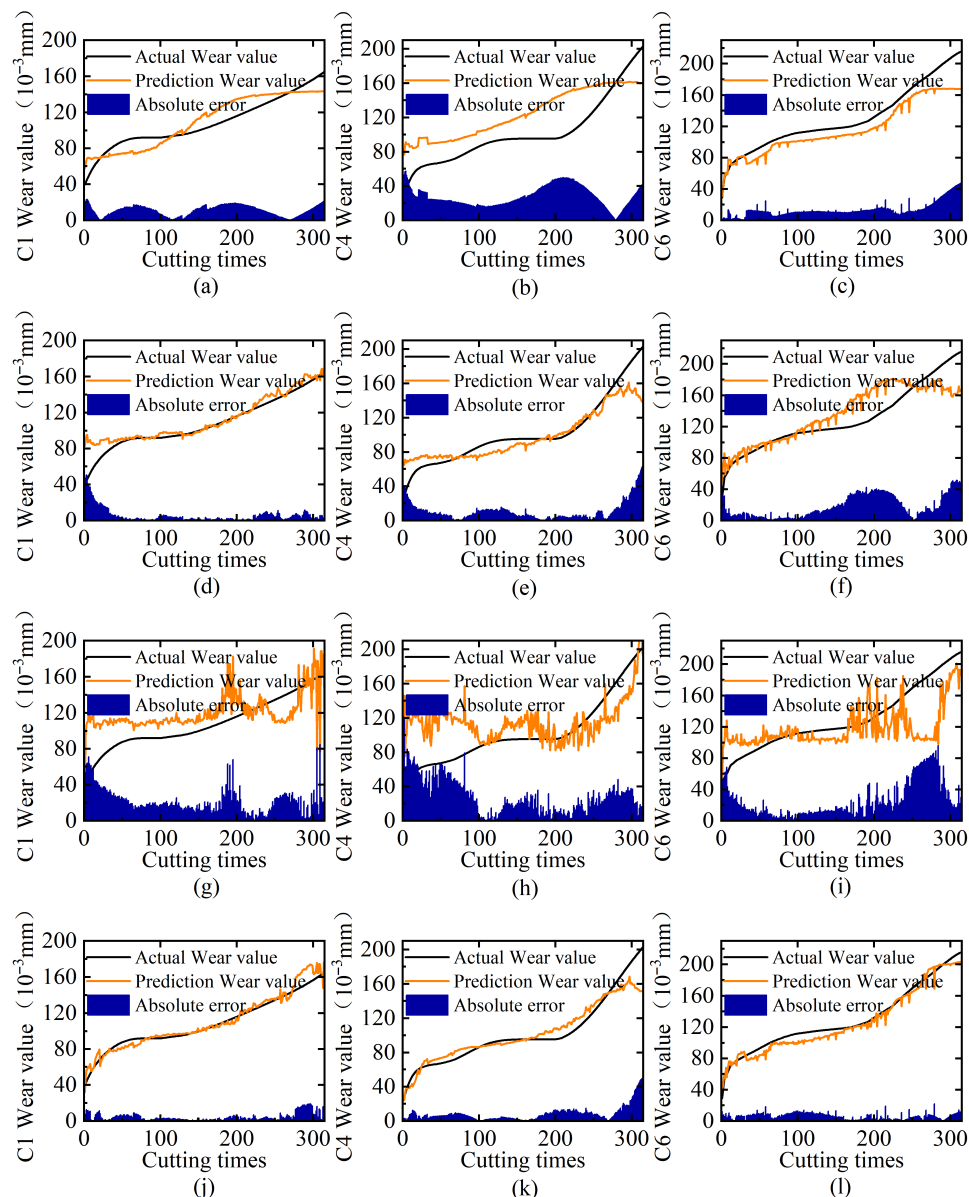


Figure 5. Ablation analysis of multi-sensor input: (a–c) prediction results on C1, C4, and C6 tool test sets using only cutting force sensor; (d–f) corresponding results using only vibration sensor; (g–i) corresponding results using only AE sensor; (j–l) corresponding results using multi-sensor fusion.

4.3.2. Ablation Experiment on Multi-Domain Feature Selection

To rigorously assess the complementary benefits of multi-domain feature fusion, we conduct controlled experiments across three tool wear cases (C1/C4/C6) under four configurations: individual time-domain, frequency-domain, and wavelet time–frequency-domain feature inputs versus their combined multi-domain fusion. All experiments maintain identi-

cal network architectures and hyperparameters, with performance quantified through MAE and RMSE metrics. Table 4 shows the results of different feature domains, and Figure 6 displays the comparison of predicted tool wear versus the ground truth, including trend curves and prediction errors.

Table 4. Performance comparison of different feature domains.

Feature Domain	C1		C4		C6	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Time domain	6.30	9.34	13.75	16.24	18.88	21.37
Frequency domain	5.60	8.13	14.34	17.45	14.29	15.38
Time–frequency domain	9.45	13.60	10.29	14.40	10.87	15.26
Multi-domain	4.47	6.35	8.27	12.06	6.06	7.19

Note: Bold values indicate better performance.

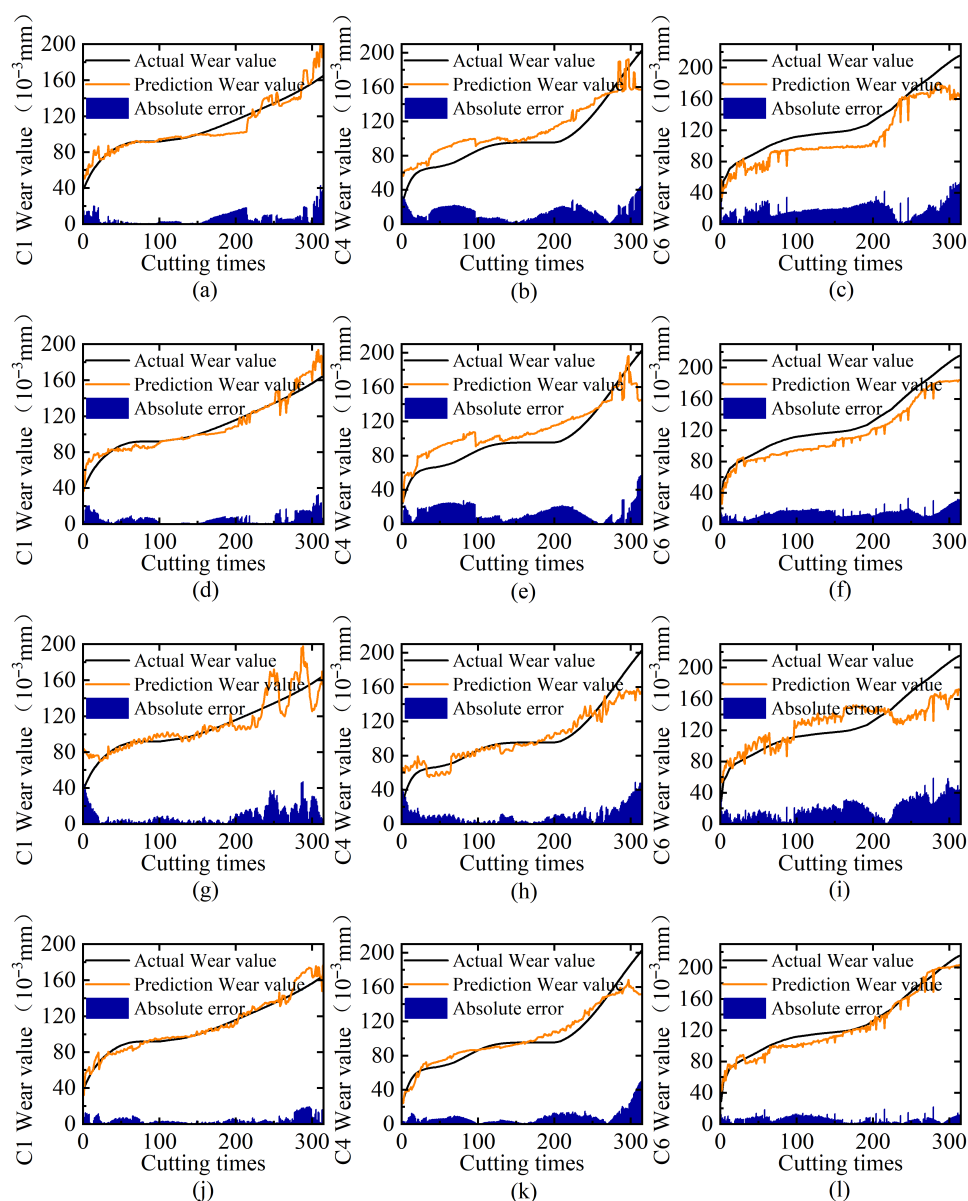


Figure 6. Ablation analysis of multi-domain features: (a–c) prediction results on C1, C4, and C6 tool test sets using only time-domain features; (d–f) corresponding results using only frequency-domain features; (g–i) corresponding results using only wavelet time–frequency-domain features; (j–l) corresponding results using multi-domain feature fusion.

The experimental results demonstrate distinct performance characteristics among features from different domains. The single time-domain and frequency-domain features achieve optimal performance for tool C1 but exhibit significant degradation for tools C4 and C6, indicating their limited capability in capturing complex wear patterns. In contrast, the wavelet time–frequency features demonstrate superior stability across different tools, maintaining MAE values around 10, which suggests their sensitivity to signal noise and transient events. However, the proposed multi-domain approach in this study yields the highest prediction accuracy, achieving an average MAE reduction of 3.74 across all three tools, with the lowest MAE of 4.47 observed for tool C1. Furthermore, the visualization results in Figure 6 demonstrate that the tool wear curve predicted using the multi-domain feature fusion strategy exhibits consistent variation trends with the actual wear while showing the smallest area enclosed by the error curve. These findings confirm that multi-domain feature fusion effectively integrates complementary advantages: time-domain trends capture gradual wear progression, frequency components identify periodic variations, and wavelet coefficients localize abrupt transitions.

4.3.3. Ablation Experiment on Position Embedding

In natural language processing (NLP), Transformers require position embeddings to encode word order dependencies. However, for tool wear monitoring, multi-sensor systems capture temporally synchronized signals from different sources without inherent sequential relationships. In other words, tool wear progression represents a continuous physical process monitored through simultaneously sampled sensor signals. Therefore, we eliminate the use of position embedding in this study. To validate our proposed position embedding elimination strategy (Section 3.3.1), we conduct rigorous comparative experiments across three tool cases (C1/C4/C6) using two configurations: one with position embeddings and one without. Table 5 shows the ablation results of position embedding, and Figure 7 displays the comparison of predicted tool wear versus the ground truth, including trend curves and prediction errors.

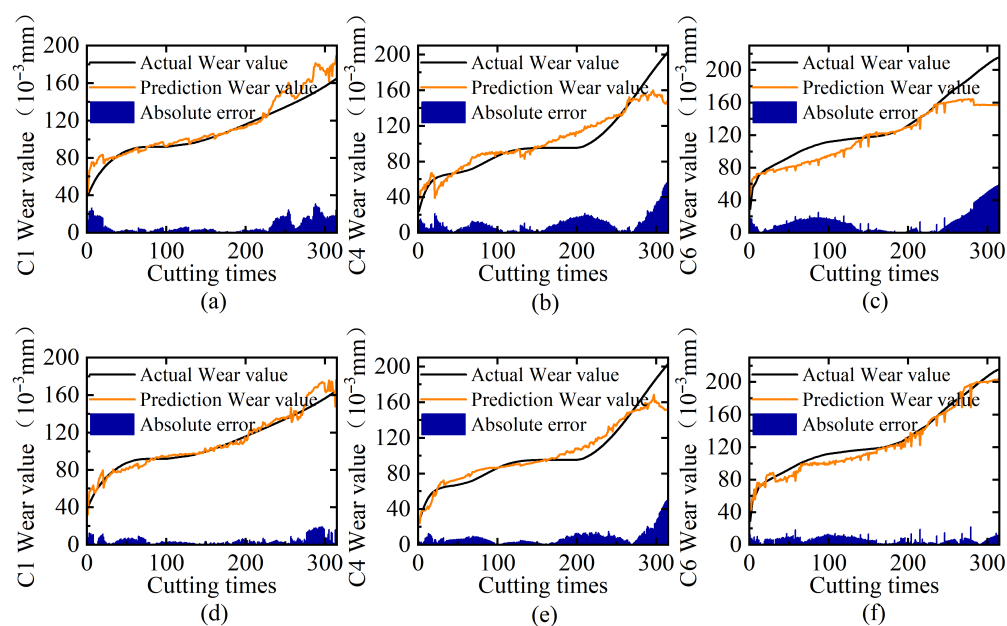


Figure 7. Ablation analysis of position embedding: (a–c) prediction results on C1, C4, and C6 tool test sets with standard position embedding; (d–f) corresponding results without position embedding.

Table 5. Performance comparison with and without position embedding.

	C1		C4		C6	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
With position embedding	7.39	10.21	10.46	14.61	13.33	19.46
Without position embedding	4.47	6.35	8.47	12.06	6.06	7.19

Note: Bold values indicate better performance.

The experimental results demonstrate the significant advantages of eliminating position embedding for tool wear state prediction. Comparative analysis reveals that the prediction model without position embedding outperforms the fixed position embedding model, with MAE reductions of 39.15% for C1 and 19.02% for C4 and even a remarkable 54.54% improvement for the most challenging case C6. The visualization results in Figure 7 further confirm high consistency between the wear values predicted by the position-embedding-free MSMDT and the ground truth measurements. These findings indicate that our framework maintains good performance across different sensor configurations through its self-attention mechanism. The results validate that removing position embedding prevents artificial spatial bias injection, enables pure correlation learning through attention weights, and ensures architectural extensibility for practical deployment.

4.3.4. Ablation Experiment on Multi-Head Attention

The number of heads in multi-head attention significantly impacts model performance, as it determines the capacity for learning diverse feature representations. To systematically investigate this relationship, we evaluate prediction performance across three distinct tool cases (C1/C4/C6) using varying numbers of attention heads (1, 4, 7, 14, and 28). Quantitative comparisons are presented in Table 6.

Table 6. Performance comparison of different multi-head attention configurations.

Heads	C1		C4		C6	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	5.17	6.49	8.19	12.34	14.80	17.92
4	7.20	9.26	11.50	13.19	7.81	9.43
7	4.73	5.97	11.70	14.22	13.13	15.61
14	4.47	6.35	8.27	12.06	6.06	7.19
28	5.82	9.37	7.90	12.11	6.05	8.44

Note: Bold values indicate better performance.

The experimental results demonstrate that the number of attention heads in the multi-head attention mechanism differentially affects prediction accuracy across different tool cases. As shown in Table 6, the best performance metrics for each tool are highlighted in bold. Specifically, the model demonstrates varying optimal configurations depending on both the tool case and evaluation metric: it achieves its lowest MAE of 4.47 on tool C1 with 14 attention heads while attaining minimum MAE values of 7.90 on C4 and 6.05 on C6 using 28 heads. However, for RMSE evaluation, the model shows superior performance on both C4 and C6 with the 14-head configuration. These variations reflect inherent engineering variability in industrial settings (e.g., tool installation tolerances) rather than methodological limitations. Based on this comprehensive analysis of prediction accuracy across all metrics and tool cases, the 14-head configuration is ultimately adopted due to its global advantages: achieving the optimal MAE on C1 (4.47) and optimal RMSE on both C4 (12.06) and C6 (7.19) while balancing computational efficiency and stability.

In summary, after sufficient ablation experiments, it is shown that multi-sensor signal inputs combined with a unified multi-domain feature extraction approach can effectively

utilize the complementary information between sensors, and the use of the Transformer structure without positional embedding enables the model to autonomously learn the correlation features between different signals and tool wear state, thus achieving better prediction results.

4.4. Comparison with Other Methods

To comprehensively evaluate the effectiveness of the proposed MSMDT framework, we conduct systematic comparisons with various baseline methods across three distinct tool cases (C1/C4/C6). The compared methods include the following: (1) traditional machine learning methods such as Linear Regression (LR) and Support Vector Regression (SVR); (2) deep learning architectures such as Multi-layer Perceptron (MLP) [40], CNN, RNN, LSTM, and Deep LSTMs; (3) specialized hybrid LSTM-based methods for tool wear monitoring including CNN-LSTM [41], CABLSTM [42], BiLSTM [32], and CNN-BiLSTM [20]; and (4) the recent Sparse Stacked Autoencoder (SSAE) approach [43]. Table 7 presents detailed MAE and RMSE metrics for all compared methods under identical experimental conditions.

Table 7. Performance comparison with other methods.

Method	C1		C4		C6	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LR	24.4	31.1	16.3	19.3	24.4	30.9
SVR	15.6	18.5	17.0	19.6	24.9	31.5
MLP	24.5	31.2	18.0	20.0	24.8	31.4
CNN	9.31	12.19	11.29	14.59	34.69	40.48
RNN	13.1	15.6	16.7	19.7	25.5	32.9
LSTM	19.6	23.9	15.6	20.8	25.3	32.4
Deep LSTMs [40]	8.3	12.1	8.7	10.2	15.2	18.9
CNN-LSTM [41]	11.18	13.77	9.39	11.85	11.34	14.33
CABLSTM [42]	7.47	8.17	-	-	-	-
BiLSTM [32]	12.8	14.6	10.9	14.2	14.7	17.7
CNN-BiLSTM [20]	5.53	6.93	7.70	10.10	8.66	11.84
SSAE [43]	-	6.66	-	11.59	-	8.49
MSMDT (Ours)	4.47	6.35	8.27	12.06	6.06	7.19

Note: Bold values indicate better performance. ‘-’ indicates that the result is not reported in the study.

The experimental results demonstrate that the proposed MSMDT framework demonstrates superior overall performance compared to conventional machine learning and deep learning approaches. In addition, the performance of MSMDT is outstanding compared to most LSTM variants. While CNN-BiLSTM shows slightly better results for the C4 tool case, our method achieves the lowest MAE and RMSE values for both the C1 (MAE: 4.47, RMSE: 6.35) and C6 (MAE: 6.06, RMSE: 7.19) tool cases. This represents significant improvements (19.17% and 8.34% for C1 and 30.02% and 39.27% for C6, respectively) over CNN-BiLSTM. This improvement validates the Transformer architecture’s enhanced capability in capturing long-range wear progression patterns compared to recurrent networks. The experimental results collectively demonstrate that the proposed MSMDT framework achieves optimally balanced performance across all evaluation metrics, significantly outperforming conventional approaches. This superior performance stems from its effective utilization of complementary multi-sensor multi-domain features coupled with an inherently adaptable network architecture that ensures robust generalizability and operational extensibility under diverse machining conditions.

5. Conclusions

This study proposes the MSMDT model based on the Transformer architecture for tool wear prediction. This method extracts multi-domain features, including time-domain, frequency-domain, and wavelet time–frequency-domain features, from multi-sensor signals and performs deep fusion. It fully exploits the complementarity of multi-sensor heterogeneous data and effectively captures key information highly relevant to tool wear states. The MSMDT adopts a feature fusion encoder without position embedding, which can efficiently process and integrate multi-sensor multi-domain features in parallel, thereby improving the accuracy and robustness of tool wear prediction. Experimental results demonstrate that the proposed method outperforms other state-of-the-art methods, offering a robust solution for tool wear monitoring.

The integration of a unified multi-domain feature extraction strategy for multi-sensor heterogeneous signals with a position-embedding-free Transformer architecture demonstrates exceptional adaptability to integrated multi-sensor monitoring systems in practical industrial scenarios, while providing significant advantages across multiple critical dimensions for deployment. By systematically integrating complementary time-domain, frequency-domain, and time–frequency-domain representations, the MSMDT framework adapts to diverse sensor configurations without requiring architectural modifications, making it suitable for real-world manufacturing environments where sensor setups may vary. Feature fusion based on physically interpretable characteristics facilitates deep learning-driven predictions and enhances the understanding of tool wear mechanisms. Additionally, dimension-optimized feature representations and parallelizable Transformer operations ensure computational efficiency, achieving superior performance while maintaining high model extensibility. These attributes make MSMDT suitable for future real-time tool wear monitoring systems.

While the current study validates the MSMDT framework on a publicly available dataset, future research will focus on deploying and testing the proposed method on real CNC machining platforms to evaluate its performance under dynamic industrial conditions. This implementation study will specifically examine the framework’s robustness against real-world variabilities in tooling, workpiece materials, and machining parameters. Moreover, we will further explore additional discriminative features and advanced network architectures to enhance prediction accuracy while maintaining real-time processing capabilities. Efforts will also be devoted to optimizing computational efficiency for edge device implementation, enabling seamless integration with smart manufacturing systems. These developments aim to bridge the gap between laboratory validation and industrial application, ultimately improving the practicality and reliability of tool wear monitoring in production environments.

Author Contributions: Data curation, P.Y.; formal analysis, J.Z.; investigation, J.Z. and X.L.; methodology, J.Z. and L.W.; visualization, Q.L.; writing—original draft, J.Z. and X.L.; writing—review and editing, T.W. All authors have read and agreed to the published version of the manuscript.

Funding: We wish to acknowledge the support received from the Postdoctoral Research Startup Fund Project of Shenzhen Polytechnic University (6024331018K), the Special Projects in Key Fields of General Universities in Guangdong Province (No. 2022ZDZX3070), and the Shenzhen Strategic Emerging Industry Support Plans (Grant XMHT20230115002).

Data Availability Statement: Data were derived from public domain resources. The data presented in this study are available in the 2010 PHM Society Conference Data Challenge at <https://iee-dataport.org/documents/2010-phm-society-conference-data-challenge> (accessed on 5 June 2025), reference number [39]. These data were derived from the following resources available in the

public domain: <https://ieee-dataport.org/documents/2010-phm-society-conference-data-challenge> (accessed on 5 June 2025).

Acknowledgments: We express profound gratitude to the researchers involved in creating and maintaining the PHM2010. We express our gratitude to Jianguo Feng from Han's Laser Technology Industry Group Co., Ltd. for the software support in this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Banda, T.; Farid, A.A.; Li, C.; Jauw, V.L.; Lim, C.S. Application of machine vision for tool condition monitoring and tool performance optimization—a review. *Int. J. Adv. Manuf. Technol.* **2022**, *121*, 7057–7086. [\[CrossRef\]](#)
2. Zhang, X.; Lu, X.; Wang, S.; Wang, W.; Li, W. A multi-sensor based online tool condition monitoring system for milling process. *Procedia CIRP* **2018**, *72*, 1136–1141. [\[CrossRef\]](#)
3. Munaro, R.; Attanasio, A.; Del Prete, A. Tool Wear Monitoring with Artificial Intelligence Methods: A Review. *J. Manuf. Mater. Process.* **2023**, *7*, 129. [\[CrossRef\]](#)
4. Zhu, K.; Yu, X. The monitoring of micro milling tool wear conditions by wear area estimation. *Mech. Syst. Signal Process.* **2017**, *93*, 80–91. [\[CrossRef\]](#)
5. Čerče, L.; Pušavec, F.; Kopač, J. 3D cutting tool-wear monitoring in the process. *J. Mech. Sci. Technol.* **2015**, *29*, 3885–3895. [\[CrossRef\]](#)
6. Caesarendra, W.; Triwiyanto, T.; Pandiyan, V.; Glowacz, A.; Permana, S.D.H.; Tjahjowidodo, T. A CNN prediction method for belt grinding tool wear in a polishing process utilizing 3-axes force and vibration data. *Electronics* **2021**, *10*, 1429. [\[CrossRef\]](#)
7. Huang, Z.; Zhu, J.; Lei, J.; Li, X.; Tian, F. Tool wear monitoring with vibration signals based on short-time Fourier transform and deep convolutional neural network in milling. *Math. Probl. Eng.* **2021**, *2021*, 9976939. [\[CrossRef\]](#)
8. Bhaskaran, J. On-line monitoring the hard turning using distribution parameters of acoustic emission signal. *Appl. Mech. Mater.* **2015**, *787*, 907–911. [\[CrossRef\]](#)
9. de Farias, A.; de Almeida, S.L.R.; Delijaicov, S.; Seriacopi, V.; Bordinassi, E.C. Simple machine learning allied with data-driven methods for monitoring tool wear in machining processes. *Int. J. Adv. Manuf. Technol.* **2020**, *109*, 2491–2501. [\[CrossRef\]](#)
10. Jantunen, E. A summary of methods applied to tool condition monitoring in drilling. *Int. J. Mach. Tools Manuf.* **2002**, *42*, 997–1010. [\[CrossRef\]](#)
11. Zhu, G.; Hu, S.; Tang, H. Prediction of tool wear in CFRP drilling based on neural network with multicharacteristics and multisignal sources. *Compos. Adv. Mater.* **2021**, *30*, 1–15. [\[CrossRef\]](#)
12. Rodrigo Henriques Lopes da Silva, M.B.d.S.; Hassui, A. A probabilistic neural network applied in monitoring tool wear in the end milling operation via acoustic emission and cutting power signals. *Mach. Sci. Technol.* **2016**, *20*, 386–405. [\[CrossRef\]](#)
13. Shah, M.; Borade, H.; Dave, V.; Agrawal, H.; Nair, P.; Vakharia, V. Utilizing TGAN and ConSinGAN for improved tool wear prediction: A comparative study with ED-LSTM, GRU, and CNN models. *Electronics* **2024**, *13*, 3484. [\[CrossRef\]](#)
14. Rmili, W.; Ouahabi, A.; Serra, R.; Leroy, R. An automatic system based on vibratory analysis for cutting tool wear monitoring. *Measurement* **2016**, *77*, 117–123. [\[CrossRef\]](#)
15. Elangovan, M.; Devasenapati, S.B.; Sakthivel, N.; Ramachandran, K. Evaluation of expert system for condition monitoring of a single point cutting tool using principle component analysis and decision tree algorithm. *Expert Syst. Appl.* **2011**, *38*, 4450–4459. [\[CrossRef\]](#)
16. Lin, Y.R.; Lee, C.H.; Lu, M.C. Robust tool wear monitoring system development by sensors and feature fusion. *Asian J. Control* **2022**, *24*, 1005–1021. [\[CrossRef\]](#)
17. Rafezi, H.; Akbari, J.; Behzad, M. Tool Condition Monitoring based on sound and vibration analysis and wavelet packet decomposition. In Proceedings of the 2012 8th International Symposium on Mechatronics and its Applications, Sharjah, United Arab Emirates, 10–12 April 2012; pp. 1–4. [\[CrossRef\]](#)
18. Fu, Y.; Zhang, Y.; Gao, Y.; Gao, H.; Mao, T.; Zhou, H.; Li, D. Machining vibration states monitoring based on image representation using convolutional neural networks. *Eng. Appl. Artif. Intell.* **2017**, *65*, 240–251. [\[CrossRef\]](#)
19. Sun, H.; Zhang, J.; Mo, R.; Zhang, X. In-process tool condition forecasting based on a deep learning method. *Robot. Comput.-Integr. Manuf.* **2020**, *64*, 101924. [\[CrossRef\]](#)
20. Wang, C.; Wang, G.; Wang, T.; Xiong, X.; Ouyang, Z.; Gong, T. Exploring the Processing Paradigm of Input Data for End-to-End Deep Learning in Tool Condition Monitoring. *Sensors* **2024**, *24*, 5300. [\[CrossRef\]](#)
21. Gao, D.; Liao, Z.; Lv, Z.; Lu, Y. Multi-scale statistical signal processing of cutting force in cutting tool condition monitoring. *Int. J. Adv. Manuf. Technol.* **2015**, *80*, 1843–1853. [\[CrossRef\]](#)
22. Zhang, P.; Gao, D.; Lu, Y.; Ma, Z.; Wang, X.; Song, X. Cutting tool wear monitoring based on a smart toolholder with embedded force and vibration sensors and an improved residual network. *Measurement* **2022**, *199*, 111520. [\[CrossRef\]](#)

23. Morgan, J.; O'Donnell, G.E. Cyber physical process monitoring systems. *J. Intell. Manuf.* **2018**, *29*, 1317–1328. [\[CrossRef\]](#)
24. Alonso, F.; Salgado, D. Analysis of the structure of vibration signals for tool wear detection. *Mech. Syst. Signal Process.* **2008**, *22*, 735–748. [\[CrossRef\]](#)
25. Sorgato, M.; Bertolini, R.; Ghiotti, A.; Bruschi, S. Tool wear analysis in high-frequency vibration-assisted drilling of additive manufactured Ti6Al4V alloy. *Wear* **2021**, *477*, 203814. [\[CrossRef\]](#)
26. Danielak, M.; Witaszek, K.; Ekielski, A.; Żelaziński, T.; Dudnyk, A.; Durczak, K. Wear Detection of Extruder Elements Based on Current Signature by Means of a Continuous Wavelet Transform. *Processes* **2023**, *11*, 3240. [\[CrossRef\]](#)
27. Wang, K.; Wang, A.; Wu, L.; Xie, G. Machine tool wear prediction technology based on multi-sensor information fusion. *Sensors* **2024**, *24*, 2652. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Cai, W.; Zhang, W.; Hu, X.; Liu, Y. A hybrid information model based on long short-term memory network for tool condition monitoring. *J. Intell. Manuf.* **2020**, *31*, 1497–1510. [\[CrossRef\]](#)
29. Marani, M.; Zeinali, M.; Songmene, V.; Mechefske, C.K. Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling. *Measurement* **2021**, *177*, 109329. [\[CrossRef\]](#)
30. He, X.; Zhong, M.; He, C.; Wu, J.; Yang, H.; Zhao, Z.; Yang, W.; Jing, C.; Li, Y.; Chen, G. A Novel Tool Wear Identification Method Based on a Semi-Supervised LSTM. *Lubricants* **2025**, *13*, 72. [\[CrossRef\]](#)
31. Cheng, M.; Jiao, L.; Yan, P.; Jiang, H.; Wang, R.; Qiu, T.; Wang, X. Intelligent tool wear monitoring and multi-step prediction based on deep learning model. *J. Manuf. Syst.* **2022**, *62*, 286–300. [\[CrossRef\]](#)
32. Chan, Y.W.; Kang, T.C.; Yang, C.T.; Chang, C.H.; Huang, S.M.; Tsai, Y.T. Tool wear prediction using convolutional bidirectional LSTM networks. *J. Supercomput.* **2022**, *78*, 810–832. [\[CrossRef\]](#)
33. Li, X.; Qin, X.; Wu, J.; Yang, J.; Huang, Z. Tool wear prediction based on convolutional bidirectional LSTM model with improved particle swarm optimization. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 4025–4039. [\[CrossRef\]](#)
34. Bazi, R.; Benkedjouh, T.; Habbouche, H.; Rechak, S.; Zerhouni, N. A hybrid CNN-BiLSTM approach-based variational mode decomposition for tool wear monitoring. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 3803–3817. [\[CrossRef\]](#)
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; pp. 6000–6010. [\[CrossRef\]](#)
36. Nath, A.G.; Udmale, S.S.; Raghuwanshi, D.; Singh, S.K. Structural Rotor Fault Diagnosis Using Attention-Based Sensor Fusion and Transformers. *IEEE Sens. J.* **2022**, *22*, 707–719. [\[CrossRef\]](#)
37. Hou, Y.; Wang, J.; Chen, Z.; Ma, J.; Li, T. Diagnosisformer: An efficient rolling bearing fault diagnosis method based on improved Transformer. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106507. [\[CrossRef\]](#)
38. Zhong, S.; Ali, R. Joint Self-Attention Mechanism and Residual Network for Automated Monitoring of Intelligent Sensor in Consumer Electronics. *IEEE Trans. Consum. Electron.* **2024**, *70*, 1302–1309. [\[CrossRef\]](#)
39. PHM Society. 2010 PHM Society Conference Data Challenge. 2010. Available online: <https://phmsociety.org/2010-phm-data-challenge-comes-to-a-close/> (accessed on 18 May 2010).
40. Zhao, R.; Wang, J.; Yan, R.; Mao, K. Machine health monitoring with LSTM networks. In Proceedings of the 2016 10th International Conference on Sensing Technology (ICST), Nanjing, China, 11–13 November 2016; pp. 1–6. [\[CrossRef\]](#)
41. Qiao, H.; Wang, T.; Wang, P.; Qiao, S.; Zhang, L. A Time-Distributed Spatiotemporal Feature Learning Method for Machine Health Monitoring with Multi-Sensor Time Series. *Sensors* **2018**, *18*, 2932. [\[CrossRef\]](#)
42. Nie, L.; Zhang, L.; Xu, S.; Cai, W.; Yang, H. Remaining Useful Life Prediction of Milling Cutters Based on CNN-BiLSTM and Attention Mechanism. *Symmetry* **2022**, *14*, 2243. [\[CrossRef\]](#)
43. He, Z.; Shi, T.; Chen, X. An Innovative Study for Tool Wear Prediction Based on Stacked Sparse Autoencoder and Ensemble Learning Strategy. *Sensors* **2025**, *25*, 2391. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.