Surface Roughness Prediction Based on CNN-BiTCN-Attention in End Milling

Guanhua Xiao[1], Hanqian Tu[2], Yunzhe Xu[3], Jiahao Shao[4], Dongming Xiang[5]*
Department of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China[1,2,4,5]
Department of QIXIN HONOR SCHOOL, Zhejiang Sci-Tech University, Hangzhou 310018, China[3]

***Abstract*—Surface roughness is a pivotal indicator of surface quality for machined components. It directly influences the performance and lifespan of manufactured products. Precise prediction of surface roughness is instrumental in refining production processes and curtailing costs. However, despite the use of identical processing parameters, the final surface roughness would be different. Thus, it challenges the effectiveness of traditional prediction models based solely on processing parameters. Current prevalent approaches for surface roughness prediction rely on handcrafted features, which require expert knowledge and considerable time investment. To address these challenges, we comprehensively consider the advantages of various deep learning methods and propose a novel end-to-end architecture. It synergistically integrates convolutional neural networks (CNN), bidirectional temporal convolutional networks (BiTCN), and attention mechanism, termed the CNN-BiTCN-Attention (CBTA) architecture. This architecture leverages CNN for automatic spatial feature extraction from signals, BiTCN to capture temporal dependencies, and the attention mechanism to focus on important features related to surface roughness. Experiments are conducted with popular deep learning methods on the public ACF dataset, which includes vibration, current, and force signals from the end milling process. The results demonstrate that the CBTA model outperforms other compared models. It achieves exceptional prediction performance with a mean absolute percentage error as low as 0.79% and an R^2 as high as 99.81%. This validates the effectiveness and superiority of CBTA in end milling surface roughness prediction.**

***Keywords*—Surface roughness prediction; end milling; CNN-BiTCN-Attention; deep learning**

I. Introduction

End milling is a machining process, which utilizes the cutting edges of a rotating cylindrical cutter to remove material from a workpiece. This method is extensively used for the production of parts with intricate shapes and stringent precision requirements. The control of surface roughness during end milling is essential for ensuring product quality, as it directly affects the appearance and wear resistance of parts as well as the compatibility with other components [1]. Hence, it is necessary to predict the surface roughness in end milling. This holds significant importance for optimizing machining parameters and enhancing both the efficiency of the process and the quality of the workpiece.

Methods for predicting surface roughness in end milling can be categorized into three main types: physical modeling, statistical analysis, and artificial intelligence based approaches. Firstly, the physical modeling methods are highly reliant on expert experience and prior knowledge, and they may struggle to accurately depict actual conditions due to the complex nonlinear characteristics of the end milling process [2]. Secondly, statistical analysis methods, which mostly consider only the influence of machining parameters, fail to address the situations where surface roughness differs despite identical machining parameters during the actual end milling [3]. Finally, the AI-based methods mainly focus on the impact on surface roughness from machining parameters, handcrafted features, or single signal. Moreover, their architectures are relatively simple, and handcrafted features also depend on the expertise of researchers [4].

In light of the above analysis, there is a need to propose an end-to-end deep learning method that fuses multiple signals for surface roughness prediction. After a thorough consideration of the advantages and disadvantages of various deep learning methods, we introduce the CNN-BiTCN-Attention (CBTA) architecture for the first time to predict surface roughness. This architecture employs convolutional neural network (CNN) for feature extraction, Bidirectional Temporal Convolutional Network (BiTCN) to capture long-term dependencies in signals, and incorporates an attention mechanism to allocate reasonable weights to different signals and features. Experiments are conducted on the public ACF dataset, which is collected during end milling of 45# steel and includes vibration, current, and force signals, each with three components. In this study, CBTA is compared with three popular deep learning models, and their performance is evaluated when using both multi-signal fusion and single-signal inputs, in order to assess the robustness and effectiveness of our proposed model.

The remainder of this paper is organized as follows: Section II presents related works. Section III introduces the surface roughness prediction model and data processing methods. Section IV details the experimental design, results, and discussion. The main conclusions of this study are presented in Section V.

II. Related Work

The traditional prediction of end milling surface roughness is primarily achieved through physical modeling methods. These methods analyze and establish mathematical models based on the physical phenomena occurring during the machining process. Feng et al. [5] proposed an analytical model to predict surface roughness during laser-assisted end milling of Inconel 718, based on tool movement and elastic response of the workpiece. Zhang et al. [6] developed an analytical model for surface roughness in particle-reinforced metal matrix composites. Their model utilized an undeformed chip thickness approach based on the Rayleigh probability distribution. Wang

*Corresponding authors.

et al. [7] developed a prediction model for surface roughness in milling. This model combined the effects of elastoplastic deformation, cutter parameters, microhardness, cutting force, and material properties with geometric and mechanical models. Jiang et al. [8] explored the influence of process parameters on cutting forces and surface roughness. They proposed a mathematical model to predict the milling forces and surface roughness of carbon fiber reinforced polymers, through a combination of theoretical and experimental approaches.

Statistical regression analysis is also a classic technique for predicting surface roughness, typically focusing on the relationship between machining parameters and surface roughness. Huang et al. [9] proposed a grey online modeling surface roughness monitoring system for end milling, based on grey theory and bilateral optimal fitting methods. Misaka et al. [10] employed the Co-Kriging method in conjunction with measurement data and analytical model to predict surface roughness in Computer Numerical Control (CNC) turning. The measurement data included cutting speed, feed rate, depth of cut, and acceleration. Gao et al. [11] developed an empirical model based on multiple regression analysis for dry face turning of AZ31B magnesium alloy. The model enabled the prediction of surface roughness through cutting speed, feed rate, and cutting depth. Sekulic et al.[12] utilized Response Surface Methodology (RSM) to predict surface roughness in ball-end milling of hardened steel, with spindle speed, feed per tooth, axial depth, and radial depth as input parameters.

In addition, machine learning and deep learning methods are also applied to surface roughness prediction especially with the rapid advancement of artificial intelligence. Chen et al. [13] proposed a backpropagation neural network (BPNN) to predict surface roughness in end milling, which utilizes spindle speed, feed rate, cutting depth, and milling distance as inputs. Li et al. [14] used cutting parameters and tool wear as input variables to propose a milling surface roughness prediction method which based on particle swarm optimization least squares support vector machine (PSO-LSSVM). Wu et al. [15] extracted time-domain and frequency-domain features from vibration signals through statistical calculations, and then input these features and cutting parameters into an Artificial Neural Network (ANN) to predict surface roughness of S45C steel. Shehzad et al. [16] introduced a CNN-LSTM model designed for the online monitoring of surface roughness in copper workpieces during ultraprecision fly cutting. This model employs vibration signals for its predictions and underwent a robustness assessment through validation cohort analysis. Guo et al. [17] collected an ISSA-optimized Deep Belief Network (DBN) model for surface roughness prediction. They gathered vibration and force signals during the milling process of P20 die steel, and subsequently extracted and filtered the time-frequency domain features of these signals to serve as inputs for the model.

In conclusion, the surface roughness prediction methods presented in the aforementioned literatures have achieved remarkable results in certain specific scenarios. Their research primarily relies on machining parameters or handcrafted features for prediction, while the use of fused signals and automatic feature extraction remains lacking. These methods suffer from issues such as neglect of dynamic factors, time-consuming processes, and an inability to comprehensively describe the entire machining process. To address these challenges, we propose an end-to-end deep learning architecture that fuses multiple signals.

III. Methodology

A. Surface Roughness Prediction Model

Deep learning models do not rely on handcrafted features, and directly utilize raw signals as input. Given the substantial volume and complex characteristics of vibration, current, and force signals, the model must possess capabilities for dimensionality reduction and feature extraction. Since these signals are time series data, the model requires strong temporal analysis capabilities. Additionally, these signals have different importance for surface roughness, and it is necessary to assign appropriate weights to the model. Based on these considerations, we propose the CBTA architecture, as illustrated in Fig. 1. Below is an introduction to each component of CBTA.

1) Convolutional Neural Network block: CNN is employed for dimensionality reduction and automatic feature extraction from signals in this paper. CNN is widely utilized in computer vision and speech recognition due to its robust spatial feature extraction capability. In the realm of intelligent manufacturing, applications of CNN include bearing remaining useful life prediction [18], tool wear estimation [19], and surface roughness prediction [20]. The CNN block comprises an input layer, convolutional layers, and pooling layers. The input layer receives the signal data. The convolutional layers contain a set of convolutional kernels, and one-dimensional convolution kernels are adopted in order to process time-series data. After the convolutional layer generates feature maps, the pooling layer uses the max pooling operation to reduce the number of network parameters and retain key features. By stacking multiple convolutional and pooling layers, CNN can extract deep features from signals and filter out redundant features. Furthermore, rectified linear unit (ReLU) [21] is utilized as the activation function to avoid overfitting and enhance convergence speed. The expressions for the convolution and pooling layers are as follows:

$$\mathrm{Y}_{m,k} = f\left(\sum_{i=1}^{n} X_i * w + b\right) \tag{1}$$

$$\mathrm{Z}_{m,l} = \max\left(\mathrm{Y}_{m,k}\right) \tag{2}$$

where $Y_{m,k}$ and $Z_{m,l}$ represent the outputs of the convolutional layer and the pooling layer, respectively; f denotes the activation function ReLU, X_i signifies the number of samples, w refers to the size of the convolutional kernel, and b represents the bias vector.

2) Bidirectional Temporal Convolutional Network block: TCN was proposed by Bai et al. [22], and has been proven superior to Long Short-Term Memory (LSTM) in various fields such as multivariate time series analysis and natural language processing [23]. Inspired by Bidirectional Long Short-Term Memory (BiLSTM) [24], we introduce BiTCN for surface roughness prediction. The deep features extracted by CNN are fed into BiTCN for analysis. As depicted in Fig. 1, BiTCN consists of a pair of parallel TCNs, termed Forward

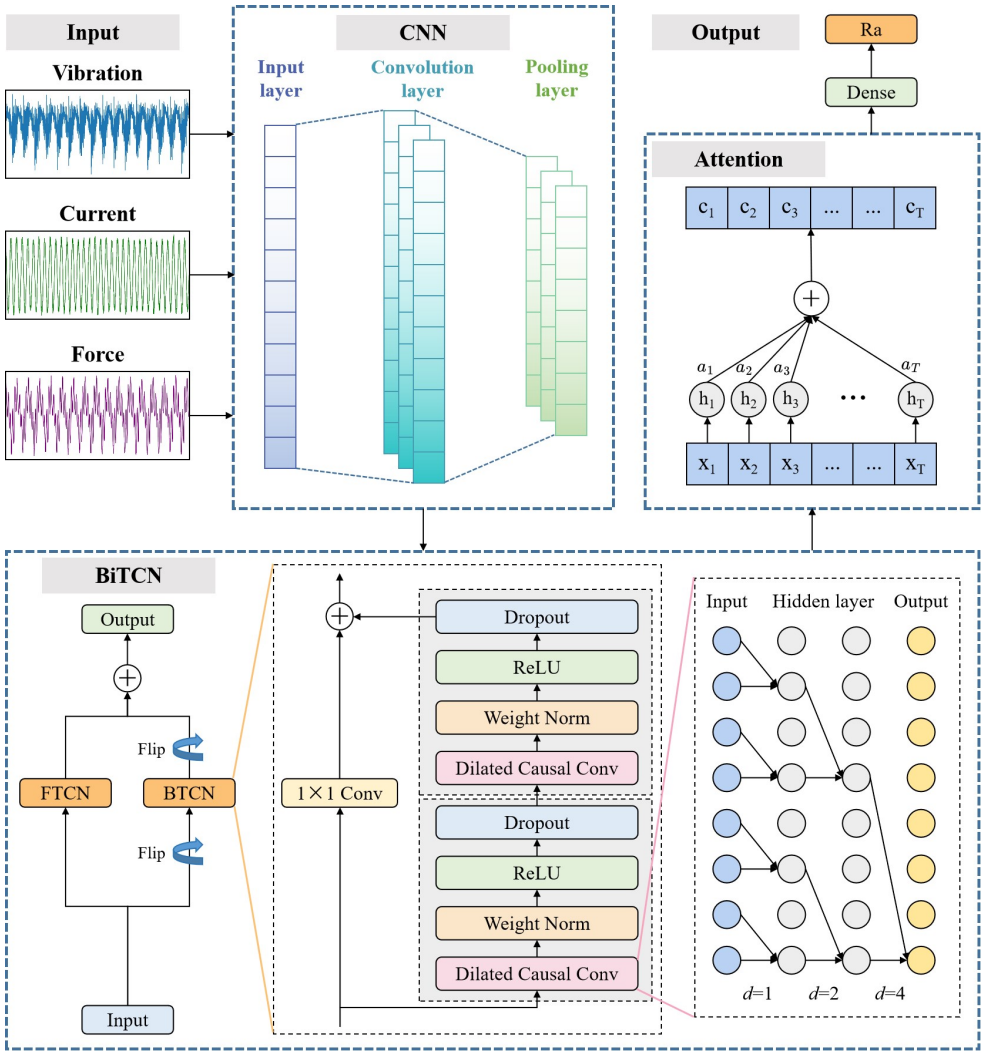

Fig. 1. CBTA architecture.

TCN (FTCN) and Backward TCN (BTCN), respectively. Their primary distinction lies in the order of input sequence, i.e. FTCN processes the sequence in the forward direction, while BTCN processes it in the reverse direction. Apart from this difference, the structures of FTCN and BTCN are identical.

TCN adheres to two principles: the output length of the network equals the input length, and future information cannot be leaked to the past. To satisfy these principles, 1D fully-convolutional network (1D-FCN) and causal convolution are introduced. The 1D-FCN ensures that the input and output lengths of each hidden layer remain identical. The causal convolution restricts the leakage of future information by simply padding the beginning and end of the time series, and then discards the excess padding values after convolution. It also limits the sliding direction solely from the past to the future. However, the modeling length of simple causal convolution is constrained by the size of the convolutional kernel. To establish long-range dependencies, numerous layers would need to be stacked, which can complicate the model and increase the risk of overfitting. To solve this issue, dilated convolution is introduced. The formula for dilated convolution is shown in Eq. (3).

$$F(s) = (X * df)(s) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i} \tag{3}$$

In Eq. (3), $X = (x_1, x_2, \ldots, x_l)$ represents the feature sequence output by the CNN, where l is the sequence length. The symbol $*$ denotes the convolution operation, d is the dilation factor, k is the filter size, and $s - d \cdot i$ indicates the direction of the past. The dilation factor can be viewed as the sampling stride between adjacent filters, typically recommended to be 2^n, where n is the number of hidden layers. When d is equal to 1, the dilated convolution degrades to a regular convolution. The increase of d allows the TCN to achieve a broader receptive field, as illustrated in Fig. 1. This mechanism enables the network to capture long-range dependencies efficiently, and enhance its ability to analyze complex temporal patterns in the data.

Additionally, as a deep neural network model, even with the use of dilated causal convolution, TCN may still encounter

some issues such as vanishing gradients or network degradation. To address them, TCN incorporates residual blocks [25] to enhance its stability and generalization ability. This residual block comprises two layers of dilated causal convolution, where each layer utilizes weight normalization [26] for normalization and employs the ReLU activation function to expedite convergence. Moreover, dropout is applied for regularization. The output of residual function F is then added to the input x of residual block, as shown in Eq. (4).

$$O = Activation(x + F(x)) \tag{4}$$

Since the number of channels in x and $F(x)$ may not be consistent, the TCN employs a 1×1 Conv. This ensures that x matches the shape of $F(x)$ after a simple transformation.

3) Attention mechanism block: Traditional neural networks treat different features equally, which makes it difficult to distinguish important features. However, different features of different signals have varying degrees of correlation with surface roughness. To solve this problem, we introduced an attention mechanism. It assigns weights to enable the model to focus on features that are more important for surface roughness and reduce attention to less important information. For the computation of the attention mechanism, we adopt the commonly used *Bahdanau* algorithm with the following formulas:

$$e_{ij} = \tanh\left(W_1 * h_i + W_2 * h_j + b\right) \tag{5}$$

$$a_{ij} = softmax(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k=1}^{T} \exp(e_{ik})} \tag{6}$$

$$c_i = \sum_{j=1}^{T} a_{ij} * h_j \tag{7}$$

where, e_{ij} represents the degree of match between hidden layer states h_i and h_j, W is a learnable weight parameter, b is a bias vector, a_{ij} denotes the attention weight of the network, and c_i is the final output of the attention layer. By means of the attention mechanism, the computational efficiency of the model is enhanced, and the influence of noise and outliers is suppressed.

B. Data Processing Method

The public ACF dataset [27], derived from Shang et al.'s work, is employed for surface roughness prediction. This dataset captures end-milling operations on a 45# steel workpiece, which is performed with a four-flute carbide tool on a CNC machining center. It encompasses vibration, current, and force signals from the end milling process, along with corresponding surface roughness measurements. Each signal contains three directional components. All signals were sampled at a frequency of 20kHz to ensure synchronous and consistent data collection.

In this paper, our data processing is conducted on the ACF dataset, which involves two main steps: data segmentation and sample amplification, as shown in Fig. 2.

1) Data segmentation: Since the tool does not contact the workpiece before the actual milling process starts and after it ends, the raw signals consist of redundant data with values close to zero, as depicted in Fig. 2(a). Hence, it is necessary to identify and remove these redundant data segments and retain only valid segments from the actual machining process.

The method to extract valid data segments involves several steps. First, the original X-direction vibration data is divided into multiple segments through a fixed-length window. Secondly, the standard deviation for each segment is calculated according to Eq. (8). Thirdly, segments with standard deviation above a certain threshold are selected as valid data segments. Finally, valid segments from the other raw data are extracted based on the same positions. As shown in Fig. 2(b), the standard deviation of redundant data segments are significantly lower than those of valid data segments. Therefore, the threshold is set to a value slightly lower than the standard deviation of valid data segments. This method ensures that only the segments containing relevant machining activity are retained for further analysis, and enhance the precision and reliability of the data.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \tag{8}$$

2) Sample augmentation: Even after data segmentation, a single valid data segment still contains a large number of data points, which may hinder the learning speed and efficiency of deep learning models. Moreover, the quantity of samples can also impact the performance of these models. When the original sample size is limited, to improve the prediction accuracy of the model, sample augmentation can be employed to generate additional samples [28]. Specifically, the valid data segments are divided into N sub-signals using a fixed time step window without overlapping, as shown in Fig. 2(d). Each sub-signal is subsequently treated as a sample. By setting N to 50 and 100, we obtained two different datasets, namely ACF-50 and ACF-100, which divided the valid data segments into 50 and 100 equal parts respectively.

IV. Experiments

After data processing, we conduct experiments on the datasets ACF-50 and ACF-100, in order to evaluate the effectiveness of the CBTA model in predicting surface roughness. Each dataset is randomly split in a 4:1 ratio, with 80% used as the training set and 20% as the test set. In these experiments, CNN, BiTCN, and LSTM-Attention (LSTM-A) are employed as baseline models for comparison. The performances of models are evaluated by three common metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). Their formulas are shown in Eq. (9), (10), and (11). Smaller values of RMSE and MAPE, along with larger values of R2, indicate better model performance.

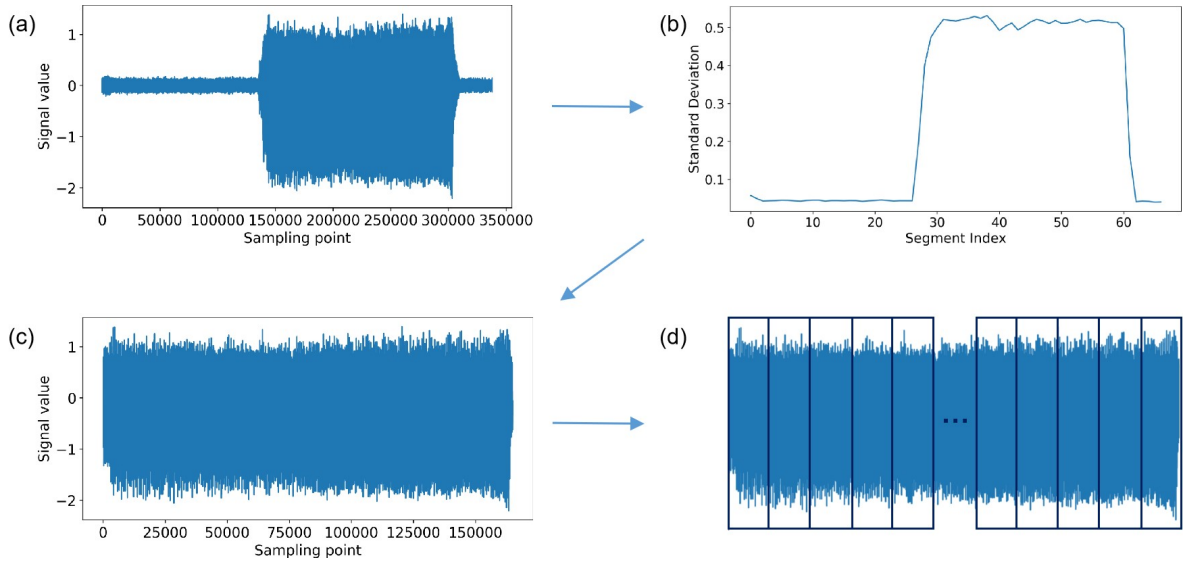

Fig. 2. Data segmentation and sample augmentation.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{9}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{11}$$

A. Model Parameters Settings

The experiments are conducted with Python 3.8 programming language and TensorFlow 2.9.0 framework. All models share the same common hyperparameters, i.e. a batch size of 32, a learning rate of 0.0005, the Adam optimizer, and a training epoch count of 200. All data are standardized before being input into the models, in order to eliminate the impact of different scales on the training process.

Table I presents the network structure of LSTM-A, while the structure of CBTA is detailed in Table II. The core network structures of the CNN and BiTCN models are identical to the CNN block and BiTCN block in CBTA, respectively. In our study, the number of stacks of residual blocks in TCN is set to 1, and the kernel initializer uses glorot uniform. According to the receptive field calculation formula (Eq. (12)) of TCN, the dilations in CBTA are set to [1, 2, 4, 8, 16] to ensure that the receptive field can fully cover the input sequence. As for BiTCN model, since there is no CNN block for data dimensionality reduction, the dilations are set to [1, 2, 4, 8, 16, 32, 64].

$$R_{field} = 1 + 2 \cdot (K_{size} - 1) \cdot N_{stack} \cdot \sum_{i} d_i \tag{12}$$

TABLE I. NETWORK STRUCTURE OF LSTM-A

Block	Layer	Units	Activation
LSTM block	LSTM	32	Tanh
Attention block	Attention	16	-
Output block	Dense	-	-

B. Training Loss

Training loss plot is crucial for evaluating model performance. Fig. 3 illustrates the training losses of various models on ACF-100. To prevent overfitting, an early stopping strategy was employed during the training process in our experiments. This strategy halts iterations when the training loss of models ceases to decrease. Both CNN and BiTCN demonstrate rapid convergence rates, and achieve convergence within 50 epochs. CNN stops training early due to the activation of the early stopping mechanism, while BiTCN's loss continues to decline slightly. LSTM-A has a slower convergence rate, and stabilizes around 90 epochs. It also triggers early stopping. CBTA exhibits the fastest convergence rate, and it achieves convergence within 30 epochs with the lowest loss.

C. Prediction Results

To investigate the effectiveness of various models in utilizing fused multi-process signals for prediction, we employed vibration, current, and force signals as inputs. Fig. 4 showcases the performance of different models on the ACF-50 and ACF-100 test sets. The prediction results and absolute errors of each model on ACF-50 are illustrated in Fig. 5, with only 50 samples displayed due to space limitations. Evidently, CBTA achieves the best performance on both datasets, followed by LSTM-A, while CNN and BiTCN exhibit similar effectiveness. Moreover, CBTA achieves RMSE, MAPE, and R^2 values of 0.0844, 2.76%, and 98.48% respectively on the ACF-50 dataset, while 0.0284, 0.79%, and 99.81% on the ACF-100 dataset. These results indicate that our proposed method

TABLE II. NETWORK STRUCTURE OF CBTA

Block	Layer	Filters	Kernel size	Units	Strides	Padding	Activation
CNN block	Conv1D	64	3	-	1	Same	ReLU
	MaxPool1D	2	-	-	2	-	-
	Conv1D	32	3	-	1	Same	ReLU
	MaxPool1D	2	-	-	2	-	-
BiTCN block	FTCN	16	16	-	-	Causal	ReLU
	BTCN	16	16	-	-	Causal	ReLU
Attention block	Attention	-	-	16	-	-	-
Output block	Dense	-	-	-	-	-	-

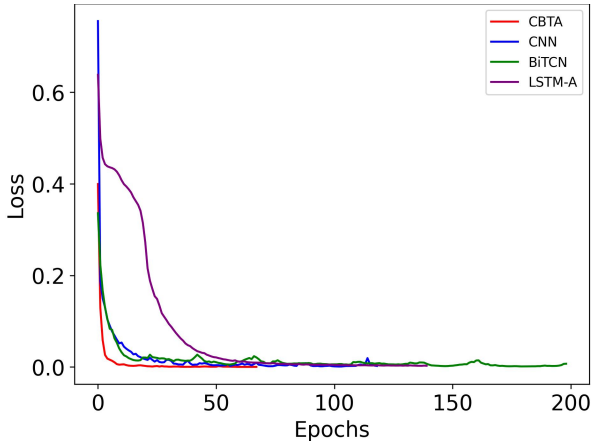

Fig. 3. Training loss plot for all models.

achieves satisfactory prediction accuracy, and demonstrate outstanding performance and robust prediction ability.

D. Comparison Under Single Signal Input

In order to evaluate the performance of each model with a single signal input, we conducted experiments on the ACF-100 dataset for predicting surface roughness with only vibration data. The results are shown in Fig. 6. It is evident that the performance of models declined to varying degrees when solely vibration signal is used, compared to the performance when employing a fusion of vibration, current, and force signals as inputs. However, CBTA is the least affected, and still achieves excellent results, with MAPE of 1.80% and R^2 value of 99.09%. This demonstrates that the fusion of multiple process signals contributes to improve accuracy, and also validates the effectiveness and robustness of CBTA in surface roughness prediction.

E. Result Discussion

Given the above experimental results, along with different model frameworks, the superiority of CBTA is analyzed from the following aspects:

- The CNN block possesses remarkable capabilities in feature extraction. In the milling process, the generated vibration, current, and force signals are intricate and multidimensional. They encompass not only spatial features, such as waveforms and frequency distributions, but also temporal features, like trends and periodicity. Compared with BiTCN and LSTM-A models, the CNN block, through its unique convolutional and pooling layer structure, can automatically extract local features related to surface roughness from the raw data. These features serves as input variables in subsequent blocks.
- The BiTCN block enables the model to comprehend temporal dependencies within the data. In comparison to CNN and LSTM-A models, TCN expands the network's receptive field through dilated causal convolutions, which allows it to more effectively capture long-term dependencies across time steps. Moreover, the bidirectional architecture of the BiTCN block enables the model to harness information from the entire signal sequence, encompassing both antecedent and subsequent data. This capability not only bolsters the precision of predictions but also amplifies the model's robustness.
- The attention mechanism enhances CBTA to concentrate on the critical segments of the input sequence. During the milling process, certain fluctuations in the vibration signal may arise from external disturbances rather than actual changes in surface quality. Compared to CNN and BiTCN models, the attention mechanism helps CBTA discern those irrelevant fluctuations by assigning weights to different time steps or features. This allows the model to concentrate on the information most relevant to surface roughness. This targeted method not only enhances the reliability of the model but also equips it with greater resilience against noise and irrelevant data.
- Each component offers complementary advantages. The integration of CNN, BiTCN and attention mechanism equips our model with the ability to handle complex nonlinear relationships between input signals and surface roughness, as well as to model temporal dependencies. This hierarchical approach enables CBTA to conduct a profound and meticulous analysis of the milling process and ultimately achieve precise roughness predictions.

V. CONCLUSION

This study investigates a deep learning-based approach for surface roughness prediction in end milling with the fusion of multiple process signals. Initially, through data segmentation and sample augmentation, two distinct datasets are developed

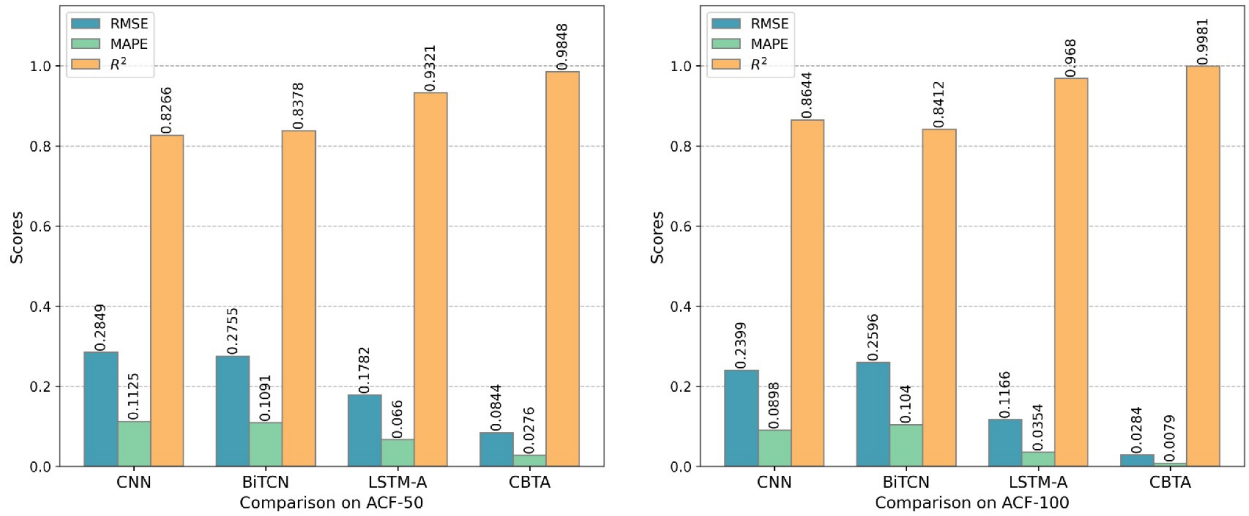

Fig. 4. Comparison of different models on the ACF-50 and ACF-100 test sets.

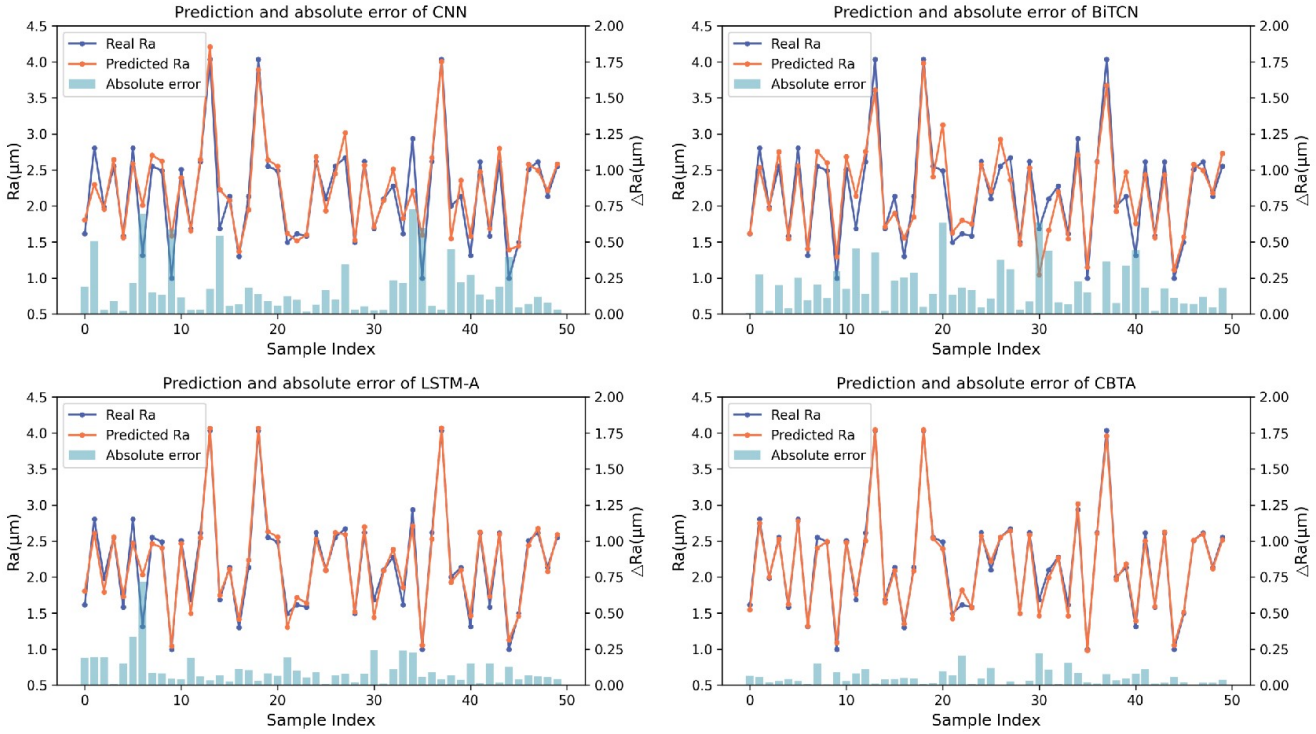

Fig. 5. Surface roughness prediction results and absolute errors of each model on the ACF-50 test set.

from the ACF dataset for model training and evaluation. Subsequently, the complementary strengths of CNN, BiTCN and attention mechanism lead to the proposal of an end-to-end CNN-BiTCN-Attention architecture (CBTA). This architecture possesses the capabilities to extract spatial features, capture temporal dependencies and allocate appropriate weights.

Three different deep learning architectures are utilized in experiments for comparison with CBTA. The results show that CBTA achieves high prediction accuracies of 98.48% and 99.81% on the two datasets, respectively, along with low MAPE values of 2.76% and 0.79%. These metrics indicate that CBTA outperforms other models in terms of precision. Additionally, the effectiveness of surface roughness prediction is examined by using only vibration signal. The results show that CBTA maintains excellent prediction results although the performance of other models decline, which highlight its robustness and effectiveness.

Despite its excellent performance on public datasets, CBTA still has some limitations. For instance, as a deep learning architecture, it lacks interpretability, which hinders fault diagnosis and process optimization in industrial applications. Additionally, as the performance of machining equipment may

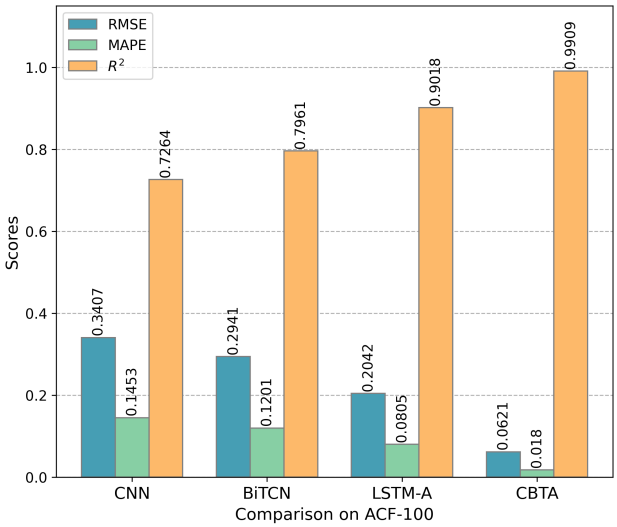

Fig. 6. Comparison of model prediction performance on the ACF-100 dataset using only vibration signal as input.

change over long-term operation, the prediction accuracy of the pre-trained model could decline. In such cases, it becomes necessary to adjust model parameters or retrain the model to maintain its effectiveness.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 62002328), and the Fundamental Research Funds of Zhejiang Sci-Tech University (Grant No. 24232123-Y).

REFERENCES

[1] K. Vipindas and J. Mathew, "Wear behavior of tialn coated wc tool during micro end milling of ti-6al-4v and analysis of surface roughness," *Wear*, vol. 424, pp. 165–182, 2019.

[2] A. M. Zain, H. Haron, and S. Sharif, "Application of ga to optimize cutting conditions for minimizing surface roughness in end milling machining process," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4650–4659, 2010.

[3] N. Karkalos, N. Galanis, and A. Markopoulos, "Surface roughness prediction for the milling of ti–6al–4v eli alloy with the use of statistical and soft computing techniques," *Measurement*, vol. 90, pp. 25–35, 2016.

[4] H. Yang, H. Zheng, and T. Zhang, "A review of artificial intelligent methods for machined surface roughness prediction," *Tribology International*, p. 109935, 2024.

[5] Y. Feng, T.-P. Hung, Y.-T. Lu, Y.-F. Lin, F.-C. Hsu, C.-F. Lin, Y.-C. Lu, X. Lu, and S. Y. Liang, "Surface roughness modeling in laser-assisted end milling of inconel 718," *Machining Science and Technology*, 2019.

[6] Z. Zhang, P. Yao, J. Wang, C. Huang, R. Cai, and H. Zhu, "Analytical modeling of surface roughness in precision grinding of particle reinforced metal matrix composites considering nanomechanical response of material," *International Journal of Mechanical Sciences*, vol. 157, pp. 243–253, 2019.

[7] B. Wang, Q. Zhang, M. Wang, Y. Zheng, and X. Kong, "A predictive model of milling surface roughness," *The International Journal of Advanced Manufacturing Technology*, vol. 108, pp. 2755–2762, 2020.

[8] J. Xiaohui, G. Shan, Z. Yong, H. Shirong, and L. Lei, "Prediction modeling of surface roughness in milling of carbon fiber reinforced polymers (cfrp)," *The International Journal of Advanced Manufacturing Technology*, vol. 113, pp. 389–405, 2021.

[9] P. B. Huang, H.-J. Zhang, and Y.-C. Lin, "Development of a grey online modeling surface roughness monitoring system in end milling operations," *Journal of Intelligent Manufacturing*, vol. 30, pp. 1923–1936, 2019.

[10] T. Misaka, J. Herwan, O. Ryabov, S. Kano, H. Sawada, N. Kasashima, and Y. Furukawa, "Prediction of surface roughness in cnc turning by model-assisted response surface method," *Precision Engineering*, vol. 62, pp. 196–203, 2020.

[11] H. Gao, B. Ma, R. P. Singh, and H. Yang, "Areal surface roughness of az31b magnesium alloy processed by dry face turning: An experimental framework combined with regression analysis," *Materials*, vol. 13, no. 10, p. 2303, 2020.

[12] M. Sekulic, V. Pejic, M. Brezocnik, M. Gostimirović, and M. Hadzistevic, "Prediction of surface roughness in the ball-end milling process using response surface methodology, genetic algorithms, and grey wolf optimizer algorithm," *Advances in Production Engineering & Management*, vol. 13, no. 1, pp. 18–30, 2018.

[13] C.-H. Chen, S.-Y. Jeng, and C.-J. Lin, "Prediction and analysis of the surface roughness in cnc end milling using neural networks," *Applied Sciences*, vol. 12, no. 1, p. 393, 2021.

[14] B. Li and X. Tian, "An effective pso-lssvm-based approach for surface roughness prediction in high-speed precision milling," *Ieee Access*, vol. 9, pp. 80 006–80 014, 2021.

[15] T. Wu and K. Lei, "Prediction of surface roughness in milling process using vibration signal analysis and artificial neural network," *The International Journal of Advanced Manufacturing Technology*, vol. 102, no. 1, pp. 305–314, 2019.

[16] A. Shehzad, X. Rui, Y. Ding, J. Zhang, Y. Chang, H. Lu, and Y. Chen, "Deep-learning-assisted online surface roughness monitoring in ultraprecision fly cutting," *Science China Technological Sciences*, pp. 1–16, 2024.

[17] M. Guo, J. Zhou, X. Li, Z. Lin, and W. Guo, "Prediction of surface roughness based on fused features and issa-dbn in milling of die steel p20," *Scientific Reports*, vol. 13, no. 1, p. 15951, 2023.

[18] D. Yao, B. Li, H. Liu, J. Yang, and L. Jia, "Remaining useful life prediction of roller bearings based on improved 1d-cnn and simple recurrent unit," *Measurement*, vol. 175, p. 109166, 2021.

[19] F. Aghazadeh, A. Tahan, and M. Thomas, "Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process," *The International Journal of Advanced Manufacturing Technology*, vol. 98, pp. 3217–3227, 2018.

[20] A. P. Rifai, H. Aoyama, N. H. Tho, S. Z. M. Dawal, and N. A. Masruroh, "Evaluation of turned and milled surfaces roughness using convolutional neural network," *Measurement*, vol. 161, p. 107860, 2020.

[21] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[22] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[23] S. Gopali, F. Abri, S. Siami-Namini, and A. S. Namin, "A comparison of tcn and lstm models in detecting anomalies in time series data," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 2415–2420.

[24] S.-H. Chien, B. Sencer, and R. Ward, "Accurate prediction of machining cycle times and feedrates with deep neural networks using bilstm," *Journal of Manufacturing Systems*, vol. 68, pp. 680–686, 2023.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," *Advances in neural information processing systems*, vol. 29, 2016.

[27] S. Suiyan, "Dataset for digital twin paper," Feb. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7643683

[28] X. Yan, Y. Liu, and M. Jia, "Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions," *Knowledge-Based Systems*, vol. 193, p. 105484, 2020.