Check for updates

ASME Journal of Manufacturing Science and Engineering
Online journal at:
https://asmedigitalcollection.asme.org/manufacturingscience

L. Hu
Department of Mechanical and Aerospace Engineering,
Rutgers University-New Brunswick,
Piscataway, NJ 08854;
New Jersey Advanced Manufacturing Institute,
Rutgers University-New Brunswick,
Piscataway, NJ 08854
e-mail: lh660@scarletmail.rutgers.edu

R. Gao
Department of Mechanical and Aerospace Engineering,
Case Western Reserve University,
Cleveland, OH 44106
e-mail: Robert.Gao@case.edu

Y. B. Guo[1]
Department of Mechanical and Aerospace Engineering,
Rutgers University-New Brunswick,
Piscataway, NJ 08854;
New Jersey Advanced Manufacturing Institute,
Rutgers University-New Brunswick,
Piscataway, NJ 08854
e-mail: yuebin.guo@rutgers.edu

Time-Series Data-Driven Machine Learning of Milling Chattering

Milling is a critical manufacturing process to produce high-value components in aerospace, tooling, and automotive industries. However, milling is prone to chatter, a severe vibration that damages surface quality, cutting tools, and machines. Traditional experimental and mechanistic methods of chatter prediction have significant limitations. This study presents a data-driven machine learning (ML) model to predict and quantify milling chatter directly based on time-series vibration data. Three ML models, including hybrid long short-term memory (LSTM)—fully convolutional network (FCN) model, gated recurrent unit (GRU)—FCN model, and temporal convolutional network (TCN) models, have been developed and verified by incorporating milling parameters to enhance prediction accuracy and stability. Among the proposed models, the best-performing ML model (GRU-FCN) demonstrates strong performance in chatter prediction and severity quantification, providing actionable insights with improved computational efficiency. The integration of milling parameters into the ML model notably enhances the prediction accuracy and stability, proving particularly effective in real-time monitoring scenarios.
[DOI: 10.1115/1.4069196]

Keywords: milling, chatter, time-series data, machine learning

1 Introduction

1.1 Background. End milling is a critical manufacturing process extensively used to produce high-value components such as turbine blades, molds, and dies [1,2]. As a versatile process in the manufacturing sector, end milling has been integral to industrial development for decades, driving advancements in various industries including aerospace, automotive, tooling, and precision engineering. Compared to other manufacturing methods like electrical discharge machining, grinding, turning, and additive manufacturing, milling, in particular end milling, offers superior capabilities in producing complex part geometries with smooth surface finishes efficiently. However, the effectiveness of end milling depends on optimizing a variety of operational parameters. The performance and quality of milled components are influenced by a range of factors including milling parameters [3], material properties [4], milling tool conditions [5], workpiece geometry [6], and vibrations [7]. Despite its numerous advantages, end milling is particularly susceptible to chatter, a type of self-excited, unstable vibration that occurs during the cutting process [7,8]. Chatter not only degrades the surface quality of the machined components [9] but also accelerates tool wear [10] and increases production costs and downtime. Given the significant impact of chatter on both product quality and manufacturing efficiency, it is crucial to advance the basic understanding of this process phenomenon and develop effective strategies and tools to mitigate its effect. This understanding is the key to optimizing milling operations and ensuring that manufactured components consistently meet stringent quality standards and performance criteria.

1.2 Current Milling Chatter Prediction Methods and Limitations. To predict and mitigate the detrimental effects of chatter, various prediction methods have been developed over the years, broadly categorized into physics-based, experimental-based, and data-driven methods.

Physics-based methods utilize mathematical formulations to model the tool–workpiece interactions in the milling process, predicting the occurrence of chatter based on the dynamics of the machine–tool–workpiece system. Machining dynamics including chatter and its technical implications were extensively studied from the mid-20th century. During this period, some foundational theories and models have been developed for the basic understanding and prediction of chatter. For instance, the use of analytical methods to model the dynamic interaction between the milling cutter and the workpiece was widely studied [9,11–13]. This research helped formalize the concept and laid the groundwork for numerous strategies to mitigate chatter through tool design, cutting conditions, and machine tool design enhancements. While these models provide valuable theoretical insights and are beneficial during the design phase of milling setups, their accuracy is often limited by the assumptions and simplifications inherent in the models.

Experiment-based methods involve direct measurements using sensors to monitor the vibrations [14], forces [15], and acoustics signals [16] from the milling operations. Building on the theoretical

[1]Corresponding author.

Manuscript received March 27, 2025; final manuscript received July 9, 2025; published online August 4, 2025. Assoc. Editor: Andy Henderson.

Copyright © 2025 by ASME

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

foundations established by earlier physics-based models, research efforts increasingly turned toward empirical approaches to gather real-world data. Experiments have become crucial in validating and refining predictions made by physics-based methods, offering more practical solutions to the chatter problem [15,17]. For instance, Rafal et al. [18] analyzed cutting force from the milling experiment to verify and improve traditional stability diagram-based chatter prediction accuracy. Cao et al. [19] developed a method to analyze vibration signals collected from the spindle during the milling process to identify chatter occurrences. The spindle motor current can also be used to identify and predict chatter occurrences [20]. However, while experimental methods provide accurate and direct insights into chatter phenomena, they are often costly and difficult to adapt to different systems or setups in a production environment.

Data-driven methods include approaches that leverage historical or online data collected from milling operations. By applying machine learning (ML) algorithms, these methods predict chatter based on identified patterns and correlations within the data, offering better flexibility and adaptability across different systems compared to experiment-based and physics-based methods. Table 1 presents the recent work on data-driven methods for chatter analysis.

Rahimi et al. [21] implement an AlexNet-based convolutional neural network (CNN) model to analyze sound data collected during milling and make predictions for including air cutting, stable, and chatter conditions. Tran et al. [22] applied a continuous wavelet transform (CWT) for preprocessing milling force data before feeding it into a CNN to detect chatter occurrences. Similarly, Sener et al. [23] employed CWT-preprocessed spindle vibration data as input to a CNN for classifying milling states as chatter or stable. Cherukuri et al. [24] developed a multilayer perceptron (MLP) model, using a dynamic model to synthesize data corresponding to different spindle speeds and chip widths, proving the MLP's capability in detecting chatter during turning operations. Lamraoui et al. [25] utilized MLP to classify preprocessed spindle vibration signals into chatter or stable states. Sestito et al. [26] explored combining MLP and support vector machine (SVM) with acoustic emission signals to classify milling processes as chatter or stable. Similarly, Peng et al. [27] implemented an SVM model using milling parameters and milling forces measured during the cutting process as the input to identify stable cutting and chatter, and the results are compared with experimental results to confirm the validity. Karandikar et al. [28,29] employed the Bayesian-based learning algorithm and milling parameters (data size of 105 samples) to analyze milling stability and identify optimal milling parameters. Schmitz [30] introduced a method using K-nearest neighbors (KNN) to classify milling chatter/stable states based on generated stability lobe diagrams.

1.3 Research Objectives. Despite the advancements in chatter prediction methods, traditional methods face significant limitations. Physics-based methods, though foundational in understanding chatter dynamics, are constrained by theoretical assumptions and simplifications that often fail to capture the complexities of real-world scenarios. Similarly, experimental approaches, while providing accurate insights, are costly and inflexible across different setups, limiting their applications in dynamic production environments. While many data-driven methods analyze vibration signals, they often transform time-domain data into frequency-domain representations (e.g., STFT, CWT, spectrograms), adding complexity and latency. In contrast, the proposed method uses raw time-series signals directly, enabling a streamlined, real-time-capable ML pipeline. To address these gaps, this research aims to develop a method that utilizes raw time-series data directly to predict chatter, enhancing the speed and applicability of in situ monitoring. Furthermore, this study introduces a novel approach to quantifying chatter severity by correlating it with surface roughness factors, providing a detailed assessment method beyond the conventional binary chatter–stable classifications. These objectives aim to advance the understanding and prediction of milling chatter and the assessment of chatter severity and enhance model practicality and effectiveness, leading to more accurate and real-time monitoring and control in milling operations.

Table 1 Data-driven methods for chatter analysis

Model type	Input	Output	Ref
CNN	Milling sound	Milling states	[21]
CNN	Milling force (workpiece)	Chatter/stable	[22]
CNN	Spindle vibration	Chatter/stable	[23]
MLP	(Synthesized) spindle speed, chip width	Chatter/stable	[24]
MLP	Spindle vibration	Chatter/stable	[25]
MLP and SVM	Acoustic emission signal	Chatter/stable	[26]
SVM	Milling parameters, cutting forces	Chatter/stable	[27]
Bayesian inference	Milling parameters	Chatter/stable	[28,29]
KNN	Stability lobe diagram	Chatter/stable	[30]

2 Milling Experiment and Data Acquisition

2.1 Experimental Workflow. As illustrated in Fig. 1, this study develops a systematic approach aimed at the detection of chatter and assessment of chatter severity in milling operations using time-series analysis of vibration data. The research plan starts with planning of milling parameters, i.e., spindle speed (v), feed per tooth (f_z), and axial depth of cut (a_p), etc., which are designed and programmed into the G-codes to conduct milling experiments. An accelerometer is attached to the spindle to collect vibration signals. After milling, the machined surfaces are measured using the Keyence VR-3000 to obtain surface topography and surface roughness. Surface finish is used to identify milling chatter/stable condition and label the associated surfaces. On the other hand, the milling parameters and the vibration signals will be used as the input to train the ML model. The predicted milling conditions and labeled conditions will be compared to conduct a postanalysis to study the accuracy and efficiency of the ML model.

2.2 Milling Experiment Setup. The experiment was conducted using a Microlution 5100 CNC milling center (Fig. 2). A Kistler 8763B100 tri-axial accelerometer was mounted on the spindle and linked to a NI-9234 DAQ system to capture the milling vibration signals. These signals were then converted into a 5G-compatible format via a 5G node and transmitted to a 5G edge server for further processing.

A tapping test was first performed on the spindle to measure the system's natural frequency for each cutting tool. By tapping the milling tool from the X direction, the system will generate vibration. Such signals are collected by the accelerometer attached to the

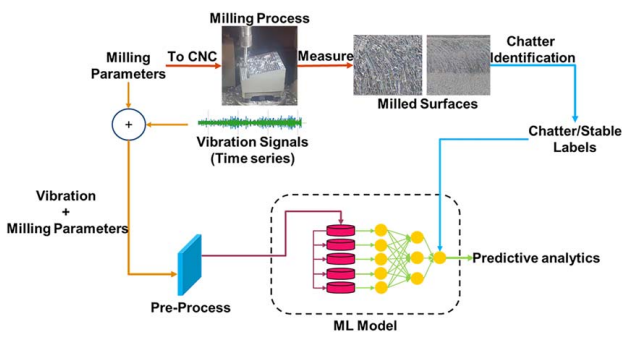


Fig. 1 Experiment workflow

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

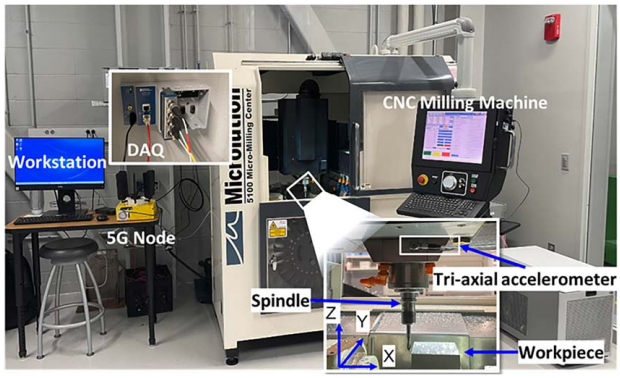


Fig. 2 Milling experiment setup

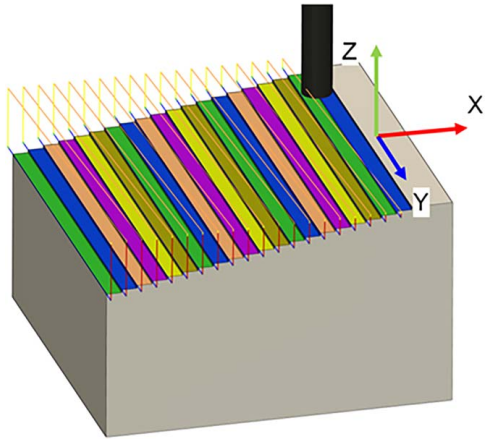


Fig. 4 Workpiece geometry

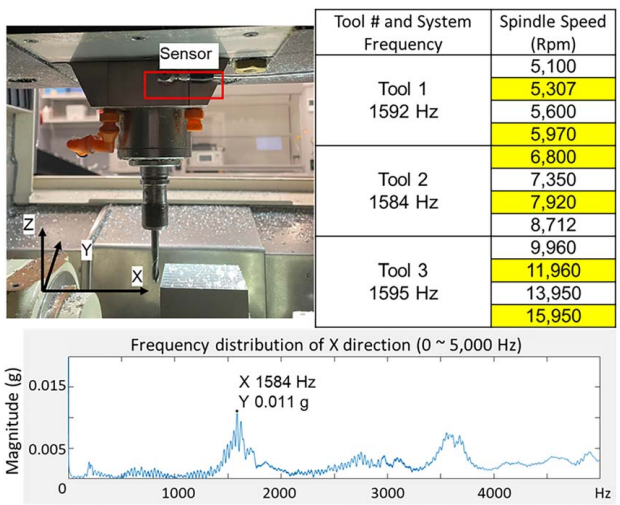
spindle as shown in Fig. 3. A fast Fourier transform (FFT) was applied to process these signals to obtain a frequency spectrum plot with frequency on the *x*-axis and magnitude on the *y*-axis. The peak in the FFT plot (in Fig. 3 at 1584 Hz) represents the natural frequencies of the tool that was tapped. It is the frequency at which the system naturally tends to vibrate when disturbed. The experiments use three tools with the same specifications (6.35 mm diameter 2-flute flat-end mill). A tapping test is performed on each tool to acquire the system's natural frequency. The highlighted spindle speeds in Fig. 3 are calculated based on the system's natural frequency, which theoretically will be easier to generate chatter.

Figure 4 illustrates the workpiece geometry and the simulated tool path. The milling process initiates from the left side of the workpiece, advancing in the *x*-positive direction. Throughout the experiments, the tool executes shoulder milling with a radial depth of cut of 3.175 mm and a feed per tooth of 0.0254 mm. For each spindle speed shown in Fig. 3, the milling experiment begins with an axial depth of cut of 6 mm and progressively reduces to 0.3 mm. The tool navigates each path with a step-over of 3.175 mm, decreasing the axial depth of cut by 0.3 mm for each subsequent path. Consequently, for each spindle speed, there are 20 paths, with each path's axial depth of cut decreasing by 0.3 mm relative to the previous one.

2.3 Data Curation. The milled surfaces were measured using a Keyence VR 3100 to obtain surface topography and roughness parameters (*Ra*), which are essential for evaluating surface quality. The experiment consisted of 12 batches, each utilizing a different spindle speed ranging from 5100 rpm to 15,950 rpm. Within each batch, 20 cutting paths were performed, each with a different axial depth of cut ranging from 0.3 to 6.0 mm. In total, 240 surfaces were measured and evaluated. The workpiece was aluminum 6061, and the cutting tool was uncoated carbide. To minimize the tool wear influence, the cutting tool was inspected after every 20 cutting paths, with no visible wear observed under 50× magnification using a digital microscope. In addition, all measurements were performed without removing the tool from tool holder to avoid recalibration errors. These close controls support the assumption that chatter is the dominant contributor to surface roughness variation.

Figure 5 presents the measured surface roughness in relation to spindle speed and axial depth of cut. The data reveal a general trend where surface roughness increases with an increase in both spindle speed and axial depth of cut. However, this trend is not consistent across all conditions. In some cases, even at the maximum axial depth of cut (6 mm), the surface roughness remains relatively low (valley in Fig. 5). This variability results in the observed peaks and valleys in the surface roughness profile rather than a continuous increase at the highest axial depth of cut. This phenomenon may be attributed to the spindle speeds producing tool path frequencies that do not excite dominant system modes (e.g., natural frequency or its harmonics), thus reducing the likelihood of chatter and leading to smoother surfaces.

When milling parameters such as spindle speed, radial depth of cut, and feed rate are constant, using a small axial depth of cut



Tool # and System Frequency	Spindle Speed (Rpm)
Tool 1 1592 Hz	5,100
	5,307
	5,600
	5,970
Tool 2 1584 Hz	6,800
	7,350
	7,920
	8,712
Tool 3 1595 Hz	9,960
	11,960
	13,950
	15,950

Fig. 3 Milling experiment setup

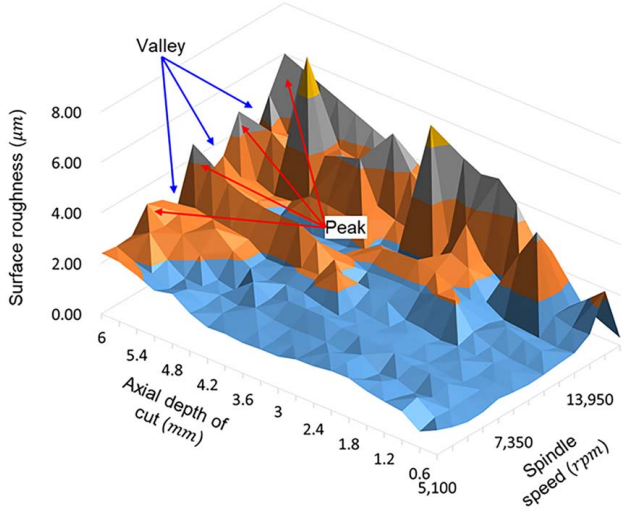


Fig. 5 Milling surface roughness versus spindle speed and axial depth of cut

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

(1 mm in this case) will result in a stable milling condition, which should produce a surface with low roughness. In this study, milling conditions with a 1 mm axial depth of cut are considered as stable, and use this baseline to identify chatter. If the surface roughness produced under the same spindle speed is significantly larger than that of the stable condition, then it is classified as chatter.

Chatter surfaces are identified based on surface roughness using Eq. (1). For each spindle speed, S_0 stands for the average surface roughness for the axial depth of cut less than 1 mm (0.3 mm, 0.6 mm, and 0.9 mm), while S_n represents the surface roughness for each specific axial depth of cut at that spindle speed. Several thresholds between 1.30 and 2.00 were tested; 1.3 labeled nearly all surfaces as chatter, while 2.0 labeled most as stable. A threshold of 1.6 was empirically derived after testing multiple values (1.3–2.0), as it provided the most consistent alignment with ground truth, visual inspection, and are consistent across samples under controlled conditions. While not intended as a universal standard, it serves as a practical criterion for labeling chatter in this experimental setup. A milling condition is identified as chatter if the calculated value exceeds 1.6. In this study, 160 surfaces are identified as chatter conditions, and 80 surfaces are produced under stable conditions.

$$\text{Chatter if } \frac{S_n}{S_0} > 1.6,\ n = 1,\ 2,\ 3, \ldots,\ 20 \tag{1}$$

The vibration signals were collected throughout the milling process. With the feed-per-tooth constant at 0.0254 mm for all experiments, an increase in spindle speed results in longer milling times for the same path length. Figure 6 illustrates two vibration signals under different milling conditions. In Fig. 6(*a*), the vibration signal has an amplitude ranging from 30 to 40 g and a milling time of approximately 9 s, including air cutting. In contrast, Fig. 6(*b*) shows an amplitude range of 5–10 g with a milling time of about 20 s, including air cutting.

To ensure data integrity for the ML model development, representative data segments were selected from the middle of each milling path's vibration signals. This selection avoids the initial and final stages of the milling process, such as tool engaging and air cutting. Figure 6 highlights these segments within a pink box, illustrating the data of interest. Each segment was then further divided into smaller units, each containing 2048 data points (no overlap), equivalent to approximately 0.04 s of vibration data, to enhance the precision of the ML model's predictions.

3 Time-Series Machine Learning Models for Chatter Prediction

This section introduces the primary challenges addressed in this study: predicting chatter labels and quantifying chatter severity from the vibration data. Two ML tasks are defined based on the characteristics of the chatter phenomenon in milling processes.

Task 1: Chatter label prediction (binary classification). Given a sequence of vibration signals, the objective is to determine whether milling condition is stable or indicates chatter. Formally, a function f is defined that map input sequences X into a binary label y:

$$y = f(X; \theta) \tag{2}$$

where $y \in \{0, 1\}$ represents the binary chatter label, with 0 denoting stable conditions and 1 indicating chatter. The function f is parameterized by θ, encompassing all the learnable parameters required to perform this classification task.

Task 2: Chatter severity prediction (regression). Beyond detecting chatter, quantifying chatter severity is critical for operational and safety decisions. The severity prediction task involves estimating a continuous value z that reflects the chatter severity based on the same input sequence X.

$$z = f(X; \theta) \tag{3}$$

where $z \in \mathbb{R}$ represents the severity of the chatter condition. The function f is parameterized by θ, which models the relationship between the input signals and the severity outcome, aiming to provide actionable insights into the severity of the chatter.

3.1 Time-Series Machine Learning Model Types. This study utilizes three time-series ML model types to analyze and predict chatter from vibration data collected during milling operations. Each model type was chosen for its specific strengths in handling sequential data, which is critical for accurately modeling the milling dynamics.

3.1.1 Long Short-Term Memory—Fully Convolutional Networks. This hybrid LSTM-fully convolutional network (FCN) model combines the strengths of LSTM networks with FCNs [31,32]. LSTM is one type of recurrent neural networks and are highly effective at capturing long-term dependencies within time-series data, making it ideal for modeling the continuous and temporal nature of milling processes where past conditions significantly influence future outcomes. The FCN component enhances the model's spatial processing capabilities, allowing for a more detailed understanding of the patterns that indicate the onset of chatter.

3.1.2 Gated Recurrent Unit—Fully Convolutional Networks. Gated recurrent unit (GRU)-FCN has a similar model structure to LSTM-FCN, but GRUs have a simpler architecture that offers a streamlined architecture compared to traditional LSTMs, utilizing a gating mechanism that helps maintain information over longer sequences while reducing computational complexity. This efficiency makes GRUs particularly suitable for processing large datasets quickly, which is advantageous for real-time chatter detection systems in industrial settings.

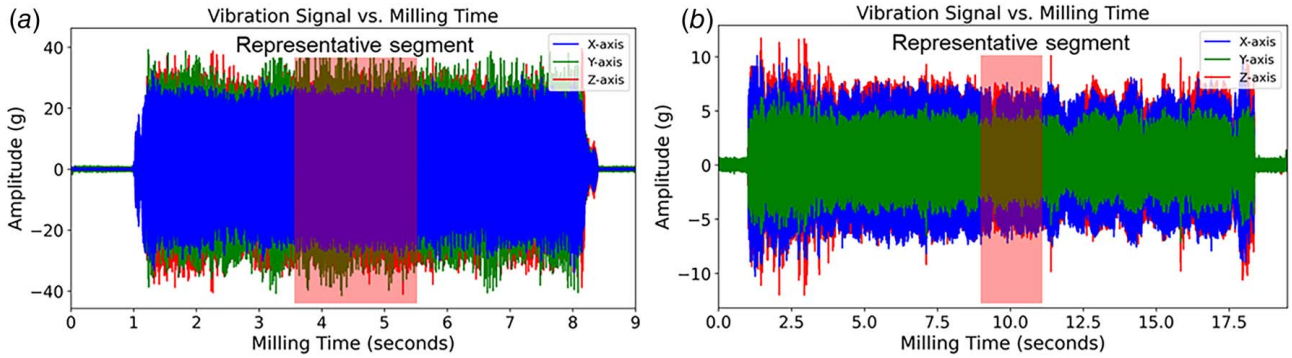


Fig. 6 Examples of vibration signals under different milling conditions: (*a*) spindle speed at 7200 rpm, axial depth 3.0 mm, and (*b*) spindle speed at 9960 rpm, axial depth 0.6 mm

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

3.1.3 Temporal Convolutional Networks. TCNs provide a robust alternative to both LSTMs and GRUs by enabling more efficient training through parallelizable operations [33,34]. Unlike recurrent networks, TCNs use convolutional layers to handle sequential data. This allows TCNs to expand their receptive field exponentially, capturing longer sequences without the limitations of vanishing gradients often encountered in LSTMs and GRUs. TCNs can process sequences in parallel rather than sequentially, leading to faster training times. Their ability to handle long-range dependencies without the need for recurrent connections makes them a strong candidate for detecting early signs of chatter in vibration signals.

By deploying these models, this study aims to rigorously evaluate their effectiveness in predicting chatter. The comparative analysis will focus on predictive accuracy and computational efficiency, which are key factors for their potential integration into real-time milling operations. The goal is to identify the optimal model for application in industrial settings, enhancing the precision and reliability of chatter prediction during milling processes.

3.2 Machine Learning Model Architectures. This section elaborates three ML model architectures and discusses the rationales behind their specific configurations for chatter prediction in milling operations. ML models are designed to leverage the rich temporal information embedded within the vibration signals and optimized to detect patterns indicative of chatter.

3.2.1 LSTM-FCN. The architecture of the LSTM-FCN model is a combination of three LSTM layers and three FCN layers (shown in Fig. 7), designed to harness both temporal and spatial data dimensions effectively. The vibration signals enter the LSTM layers, which capture long-term dependencies in the time-series data, essential for understanding the progression of milling conditions. These layers model the temporal dynamics, identifying patterns and trends indicative of potential chatter. Simultaneously, the FCN layers apply convolution operations to enhance the model's spatial analysis capabilities. The FCN layers can identify specific patterns and features within the data that correlate with chatter conditions while maintaining the spatial resolution of the input for detailed feature extraction at each time-step. Optionally, milling parameters such as axial depth of cut and spindle speed may be integrated into the model through a linear layer. This integration is flexible, allowing for the parameters to be included or excluded during model training based on specific requirements.

The features extracted from the LSTM layers, FCN layers, and optionally from the linear layer are then concatenated. This combined feature set is fed into either a sigmoid or a linear layer, depending on the desired output. The sigmoid layer outputs a probability value indicating the likelihood of chatter, resulting in a binary classification (chatter or stable). The linear layer outputs a value quantifying the chatter severity, providing a continuous measure essential for monitoring and control.

3.2.2 GRU-FCN. The GRU-FCN model integrates three GRU layers with three FCN layers, as depicted in Fig. 8, and is engineered to effectively manage both temporal and spatial data dimensions. The model is similar to the LSTM-FCN model architecture, but the temporal processing components are replaced with GRU. GRUs are known for their efficiency in modeling temporal sequences, similar to LSTMs but with a simpler gating mechanism. This simplicity allows GRUs to perform faster and often with less computational overhead compared to LSTMs.

Vibration signals are simultaneously fed into the GRU and FCN layers. The GRU layers focus on capturing essential temporal dependencies and dynamics within the time-series data, crucial for pinpointing evolving milling conditions and potential chatter. Similar to LSTM-FCN, optional milling parameters such as axial depth of cut and spindle speed can be integrated through a linear layer, allowing flexibility in the model's training to suit specific needs. The features extracted by the GRU layers, FCN layers, and optional linear layers are concatenated together. The concatenated features are then directed to a sigmoid/linear layer similar to the LSTM-FCN model.

3.2.3 Temporal Convolutional Network. The temporal convolutional network (TCN) model utilizes a series of temporal convolution blocks, as illustrated in Fig. 9, effectively handling both temporal and spatial dimensions of data. The process starts with vibration signals entering the temporal blocks. Each temporal block consists of two temporal convolution layers, each followed by a ReLU activation function to introduce nonlinearity and aid in complex pattern recognition within the time-series data. The model consists of three such blocks, where the output of one block serves as the input to the next, creating a deep, layered structure that enhances the model's ability to capture a wide range of temporal dependencies. After temporal blocks, the extracted features will be sent to a linear layer. Optionally, milling parameters such as axial depth of cut and spindle speed may be integrated into the model through a separate linear layer. The milling parameters' feature after linear layer can be concatenated with the temporal blocks' output. This modular integration allows for flexibility in including or excluding these parameters during model training. Finally, the processed features, postlinear layer(s), are directed to either a sigmoid or a linear output layer, similar to the previous models (LSTM-FCN and GRU-FCN).

The models were developed and implemented using PyTorch and trained over 100 epochs with an 80%/20% training/testing data split on the surface level to avoid data leak. Binary cross-entropy and mean squared error (MSE) were utilized as loss functions, and the Adam optimizer was used with a learning rate of 0.001 and a batch size of 128.

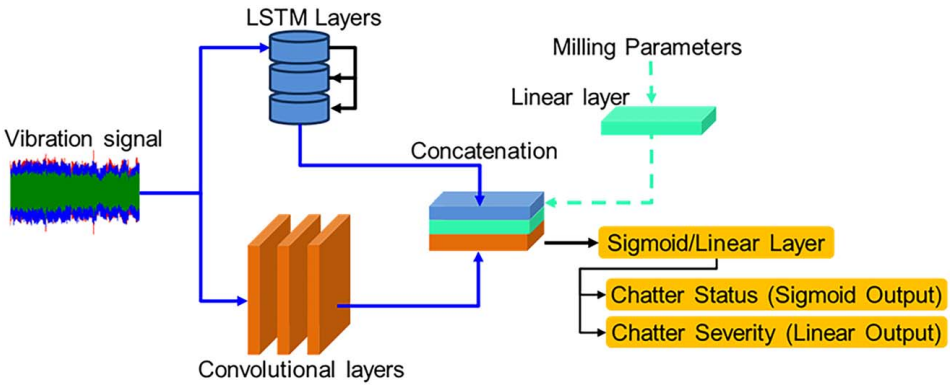


Fig. 7 The LSTM-FCN architecture

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

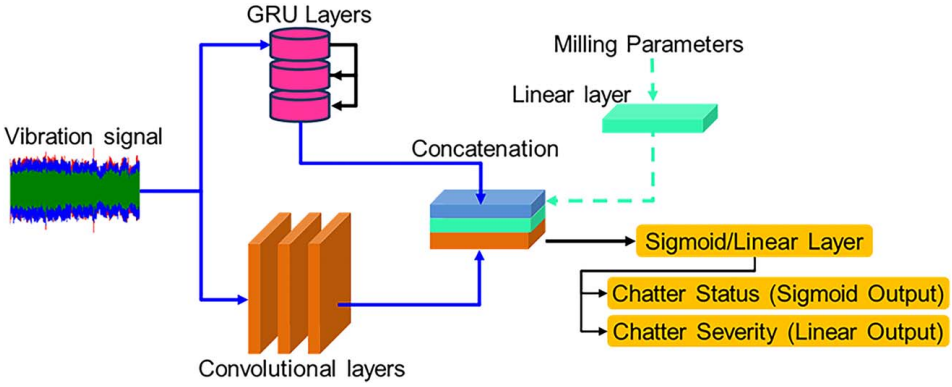


Fig. 8 The GRU-FCN architecture

4 Results and Discussion

This section evaluates and compares the ML models' performance in predicting milling stability (chatter/stable) and chatter severity under different configurations. ML models incorporating milling parameters alongside vibration data are compared against those using only the vibration data. The analysis includes binary predictions of stability via a sigmoid output layer and quantitative severity assessments through a linear output layer.

4.1 Prediction of Milling Chatter. This section discusses the performance of three ML models in terms of stability prediction, and the ML models are trained with two different setups: with milling parameters and without milling parameters. The models' sigmoid output is transformed into binary predictions based on the threshold of 0.5. If the output is larger than 0.5, it will be marked as chatter, otherwise stable. The prediction accuracy is calculated using Eq. (4), where N is the total number of samples in the test set, $\check{y}_i$ represents the predicted label for the ith sample, and y_i represents the actual label for the ith sample. This equation calculates the percentage of predictions that the model got right out of the total number of predictions made.

$$\text{Accuracy} = \left(\frac{\sum_{i=1}^{N} 1\left(\check{y}_i = y_i\right)}{N} \right) \times 100\% \tag{4}$$

Figure 10 shows the prediction accuracy of the models across the training epochs. The architecture of TCN (with milling parameters), GRU-FCN (with milling parameters), and LSTM-FCN (with milling parameters) includes an optional linear layer that incorporates milling parameters during both training and testing. In contrast, the TCN, GRU-FCN, and LSTM-FCN do not include the optional linear layer as discussed in Sec. 3.1. The plots show that the LSTM-FCN (with milling parameters) outperforms other models, achieving the highest prediction accuracy (99.94%), achieving the highest accuracy among the models tested. In contrast, the TCN model exhibits the lowest accuracy, reaching 93.39% by the end of training.

The ML models incorporating milling parameters generally achieve higher accuracy compared to their counterparts without these parameters. The plots also highlight the stability improvements for models trained with milling parameters. For instance, the GRU-FCN model (with milling parameters) displays smoother accuracy trends and less fluctuation compared to the GRU-FCN model, underscoring the benefits of integrating additional contextual data into the training process. This inclusion appears to enhance model robustness and performance, as shown by the consistently higher accuracy for these models trained with milling parameters.

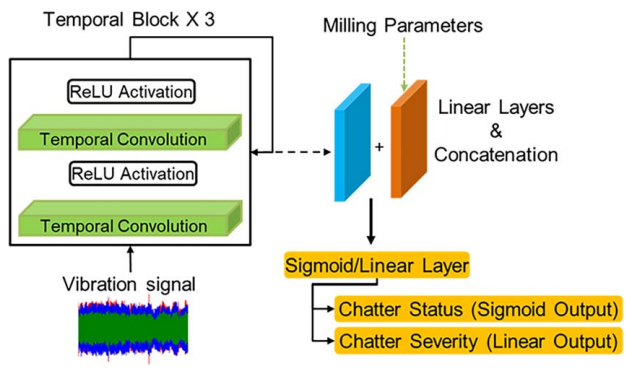


Fig. 9 The TCN architecture

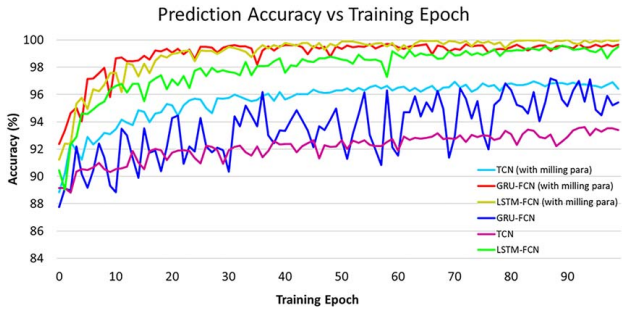


Fig. 10 Milling stability prediction accuracy versus training epoch

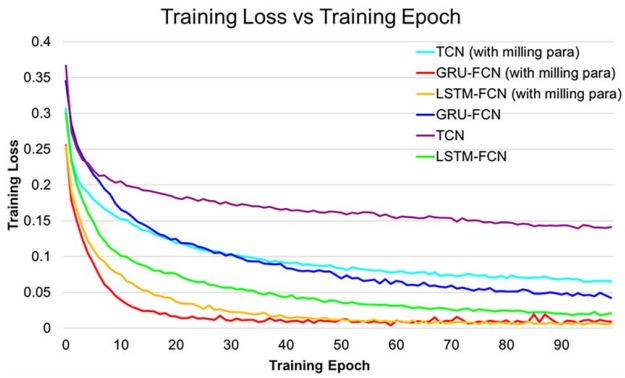


Fig. 11 Milling stability model training loss versus training epoch

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

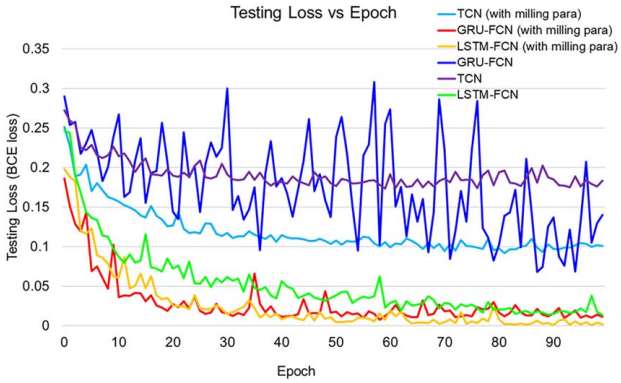


Fig. 12 Milling stability model testing loss

Figure 11 presents the training loss across training epochs for six different models, which demonstrate converging loss by the final epoch. Notably, the LSTM-FCN model (with milling parameters) and the GRU-FCN model (with milling parameters) achieve the lowest training losses of 0.006 and 0.009, respectively, by the end of the training period. In contrast, the TCN model exhibits the highest training loss of 0.141. This also highlights that models with milling parameters tend to converge more rapidly than those without these parameters.

Figure 12 shows the testing loss correlated with training epochs for the same models. The training loss trends show all ML models converge by the end of training epoch. The LSTM-FCN model (with milling parameters) and GRU-FCN model (with milling parameters) again report the lowest testing losses of 0.001 and 0.011, respectively. Conversely, the TCN model shows the highest testing loss of 0.183. Notably, the testing loss for the

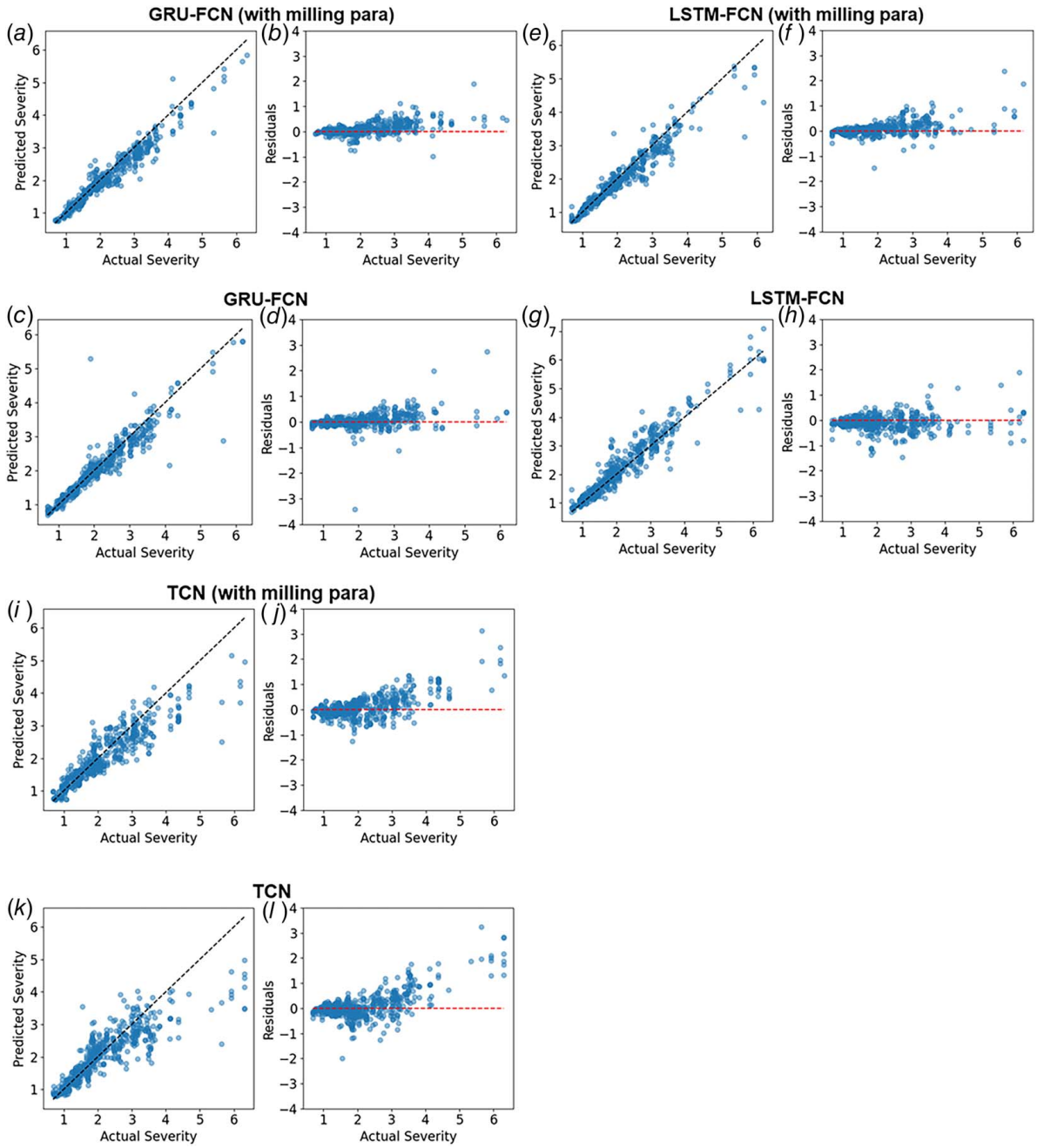


Fig. 13 Actual severity versus predicted severity of 500 predictions

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

GRU-FCN model displays significant fluctuations throughout the epochs compared to the other models.

The comprehensive evaluations of prediction accuracy and training and testing loss for all six model configurations clearly highlight the inherent performance differences between the models. The analysis from Figs. 10–12 reveals a consistent hierarchy in model performance, with LSTM-FCN outperforming GRU-FCN and both significantly surpassing TCN. This pattern is observed across both training and testing phases, with LSTM-FCN achieving lower losses and higher accuracies more efficiently than the other models. Upon incorporating milling parameters, there is a noticeable improvement in model performance, particularly for LSTM-FCN and GRU-FCN models. This inclusion leads to even lower training and testing losses and stabilizes prediction accuracy across training epochs.

4.2 Prediction of Chatter Severity. This section evaluates the ML models' performance in predicting chatter severity, utilizing a linear output layer for quantitative assessment. As with chatter predictions, the models were configured in two ways: with and without the input of milling parameters. As discussed in Sec. 3, the severity predictions are made through the linear layer, and the resulting values are compared against actual severity scores using MSE as shown in Eq. (5):

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\check{y}_i - y_i\right)^2 \tag{5}$$

where N is the number of samples, y_i is the actual value of the sample, and $\check{y}_i$ is the predicted value.

Additionally, the models predict chatter based on severity scores. Chatter labels are determined from these scores, classifying outputs as chatter if severity exceeds a threshold of 1.6 (as discussed in Sec. 2), and otherwise as stable. These labels are then used to compute prediction accuracy, applying Eq. (1).

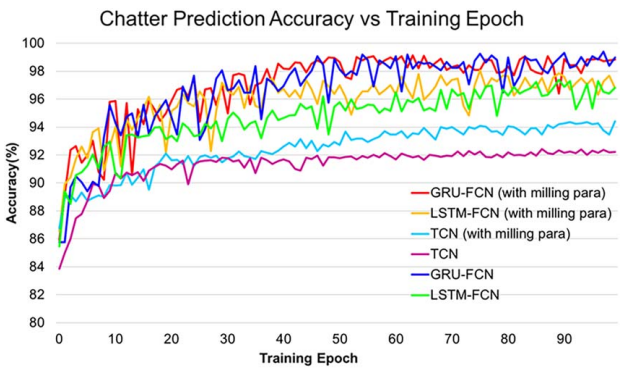


Fig. 14 Prediction accuracy of chatter severity based on severity score

Figure 13 shows the evaluation of six model configurations, the actual versus predicted severity of 500 predictions from the testing dataset by the end of training epoch and the corresponding residuals for each model. The subplots of Figs. 13(*a*), 13(*c*), 13(*e*), 13(*g*), 13(*h*), and 13(*j*) directly compare the predicted level with actual severity level. Points are plotted where the *x*-axis shows actual severity and the *y*-axis shows model predictions. The proximity of points to the diagonal line reflects the model's precision. Figures 13(*b*), 13(*d*), 13(*f*), 13(*h*), 13(*i*), and 13(*k*) display residuals, which are calculated as follows: Residual = Actual Severity − Predicted Severity. The *x*-axis shows actual severity, while the *y*-axis shows these residuals, with zero residual indicating perfect prediction. Positive residuals suggest overestimations, and negative ones indicate underestimations. These plots reveal error patterns and potential systematic biases or variances of the models. From those plots, the performance of GRU-FCN (with milling parameter) is similar to LSTM-FCN (with milling parameter) and is better than the others. Figures 13(*a*) and 13(*e*) show the points scattered closely around the perfect prediction line, suggesting effective model performance. Figures 13(*b*) and 13(*f*) reveal relatively low residuals across the range of severities compared with other models, indicating consistent prediction accuracy.

In general, models incorporating milling parameters generally exhibit superior predictive accuracy and consistency, as evidenced by tighter clustering around the perfect prediction line and smaller residuals.

Figure 14 shows the accuracy of predicted chatter severity along with the training epoch, and the chatter label is calculated based on the predicted severity. If predicted severity score is larger than 1.6, the corresponding milling process will be labeled as chatter by the model. Such labels will be compared with the experiment results from Sec. 2. Then the accuracy of chatter severity is calculated using Eq. (1). It is evident that the GRU-FCN (with milling parameters) and GRU-FCN achieve equivalent prediction accuracy (98.84% and 98.98%, respectively) by the end of the training. TCN has lowest prediction accuracy (i.e., 92.21%) by the end of the training epoch compared to all other model configurations. The inclusion of milling parameters in the models can improve the prediction accuracy, but not as much as observed in Sec. 4.1.

Figures 15(*a*) and 15(*b*) provide a detailed comparison of the MSE losses during (*a*) training and (*b*) testing phases for six different model configurations over 100 training epochs. Figure 15(*a*) shows all models generally improve and stabilize in terms of MSE as the training progresses, with those incorporating milling parameters exhibiting a slightly faster rate of improvement and lower overall loss levels compared to their counterparts without milling parameters. This indicates a more effective learning

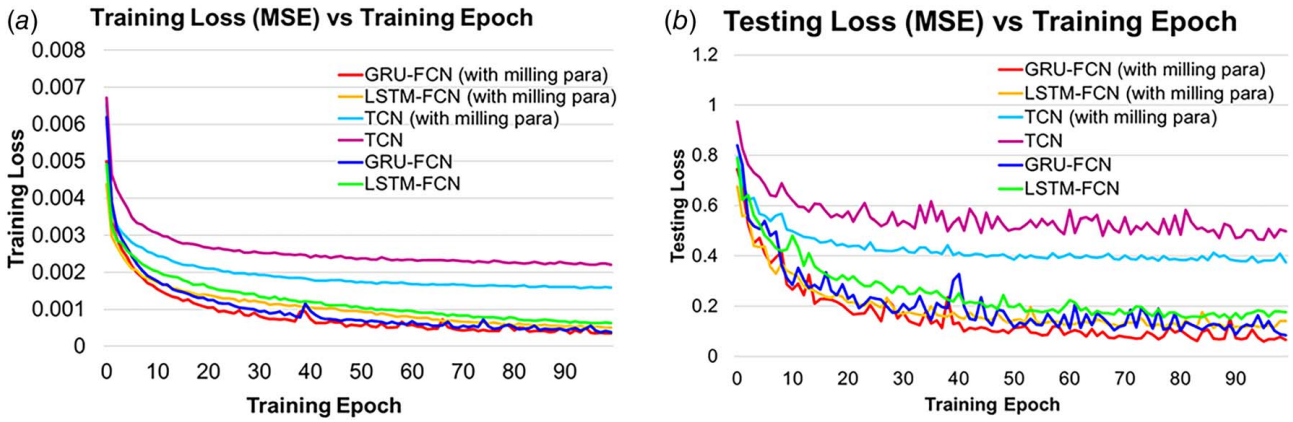


Fig. 15 (*a*) Training loss versus training epoch and (*b*) testing loss versus training epoch

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

process when contextual data are included in the model inputs. Similar to the training phase, the models with milling parameters again show lower levels of loss, suggesting better generalization and robustness. Notably, the GRU-FCN (with milling parameters) and GRU-FCN and LSTM-FCN (with milling parameters) consistently achieve the lowest testing losses, highlighting their superior performance in handling real-world data variability.

The analysis in this section highlights that incorporating milling parameters improves prediction accuracy for chatter detection and, to a lesser extent, for severity assessment. While GRU-FCN slightly outperforms LSTM-FCN in severity prediction due to its efficient sequential data handling, the inclusion of milling parameters yields more noticeable accuracy gains in LSTM-FCN and TCN compared to GRU-FCN. TCN, despite its strength in capturing extended historical contexts through dilated convolutions, shows the lowest performance, possibly due to weaker short-term responsiveness compared to GRU and LSTM. Integrating contextual data like milling parameters is crucial in improving the models' predictive accuracy, demonstrating more precise and consistent outputs.

5 Summary

This study has developed LSTM-FCN, GRU-FCN, and TCN models to predict chatter and its severity. Key results can be summarized as follows.

- Incorporation of the milling parameters as inputs into the ML models significantly improves prediction accuracy and model stability.
- GRU-FCN excels in severity prediction due to its efficient data processing, and LSTM-FCN performs well in chatter detection by effectively managing long-term data dependencies.
- The direct use of raw time-series data eliminates the need for costly transformations like STFT and CWT, which involve windowing and spectral analysis. This streamlines inference and support real-time monitoring.
- The ability to quantify chatter severity has provided an in-depth insight into milling chatter.

Future work may focus on deploying these models at the 5G-enabled milling testbed to validate their effectiveness for online chatter prediction and severity estimation in a live milling environment. Model-based process control could be further developed for real-time process correction if a chatter occurs by leveraging the low-latency and high-speed 5G network. Plans also include conducting a thorough comparison of time and computational efficiency between the proposed methods and traditional techniques that use preprocessed data, such as STFT. These efforts are expected to broaden the applicability of ML models in manufacturing process monitoring, operational optimization, and predictive maintenance.

Funding Data

- The National Science Foundation (Grants CMMI-2040358 and CMMI-2040288).

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Koshy, P., Dewes, R. C., and Aspinwall, D. K., 2002, "High Speed End Milling of Hardened AISI D2 Tool Steel (~ 58 HRC)," J. Mater. Process. Technol., **127**(2), pp. 266–273.
[2] Biermann, D., Kersting, P., and Surmann, T., 2010, "A General Approach to Simulating Workpiece Vibrations During Five-Axis Milling of Turbine Blades," CIRP Ann., **59**(1), pp. 125–128.
[3] Sun, J., and Guo, Y. B., 2009, "A Comprehensive Experimental Study on Surface Integrity by End Milling Ti–6Al–4V," J. Mater. Process. Technol., **209**(8), pp. 4036–4042.
[4] Shokrani, A., Dhokia, V., and Newman, S. T., 2016, "Investigation of the Effects of Cryogenic Machining on Surface Integrity in CNC End Milling of Ti–6Al–4V Titanium Alloy," J. Manuf. Process., **21**, pp. 172–179.
[5] Nouri, M., Fussell, B. K., Ziniti, B. L., and Linder, E., 2015, "Real-Time Tool Wear Monitoring in Milling Using a Cutting Condition Independent Method," Int. J. Mach. Tools. Manuf., **89**, pp. 1–13.
[6] Seguy, S., Dessein, G., and Arnaud, L., 2008, "Surface Roughness Variation of Thin Wall Milling, Related to Modal Interactions," Int. J. Mach. Tools Manuf., **48**(3–4), pp. 261–274.
[7] Altintas, Y., Stepan, G., Merdol, D., and Dombovari, Z., 2008, "Chatter Stability of Milling in Frequency and Discrete Time Domain," CIRP J. Manuf. Sci. Technol., **1**(1), pp. 35–44.
[8] Altintas, Y., and Weck, M., 2004, "Chatter Stability of Metal Cutting and Grinding," CIRP Ann., **53**(22), pp. 619–642.
[9] Budak, E., and Altintas, Y., 1998, "Analytical Prediction of Chatter Stability in Milling—Part I: General Formulation," ASME J. Dyn. Sys., Meas., Control, **120**(1), pp. 25–31. .
[10] Lamraoui, M., Thomas, M., and El Badaoui, M., 2014, "Cyclostationarity Approach for Monitoring Chatter and Tool Wear in High Speed Milling," Mech. Syst. Signal. Process., **44**(1–2), pp. 177–198.
[11] Tlusty, J., Zaton, W., and Ismail, F., 1983, "Stability Lobes in Milling," CIRP Ann., **32**(1), pp. 309–313.
[12] Tlusty, J., and Ismail, F., 1983, "Special Aspects of Chatter in Milling," ASME J. Vib. Acoust., **105**(1), pp. 24–32.
[13] Budak, E., and Altintas, Y., 1998, "Analytical Prediction of Chatter Stability in Milling—Part II: Application of the General Formulation to Common Milling Systems," ASME J. Dyn. Sys., Meas., Control, **120**(1), pp. 31–36.
[14] Cao, H., Zhou, K., and Chen, X., 2015, "Chatter Identification in End Milling Process Based on EEMD and Nonlinear Dimensionless Indicators," Int. J. Mach. Tools Manuf., **92**, pp. 52–59.
[15] Kuljanic, E., Sortino, M., and Totis, G., 2008, "Multisensor Approaches for Chatter Detection in Milling," J. Sound Vib., **312**(4–5), pp. 672–693.
[16] Tsai, N.-C., Chen, D.-C., and Lee, R.-M., 2010, "Chatter Prevention for Milling Process by Acoustic Signal Feedback," Int. J. Adv. Manuf. Technol., **47**(9–12), pp. 1013–1021.
[17] Kuljanic, E., Totis, G., and Sortino, M., 2009, "Development of an Intelligent Multisensor Chatter Detection System in Milling," Mech. Syst. Signal. Process., **23**(5), pp. 1704–1718.
[18] Rafal, R., Pawel, L., Krzysztof, K., Bogdan, K., and Jerzy, W., 2015, "Chatter Identification Methods on the Basis of Time Series Measured During Titanium Superalloy Milling," Int. J. Mech. Sci., **99**, pp. 196–207.
[19] Cao, H., Lei, Y., and He, Z., 2013, "Chatter Identification in End Milling Process Using Wavelet Packets and Hilbert–Huang Transform," Int. J. Mach. Tools Manuf., **69**, pp. 11–19.
[20] Aslan, D., and Altintas, Y., 2018, "On-Line Chatter Detection in Milling Using Drive Motor Current Commands Extracted From CNC," Int. J. Mach. Tools Manuf., **132**, pp. 64–80.
[21] Rahimi, M. H., Huynh, H. N., and Altintas, Y., 2021, "On-Line Chatter Detection in Milling With Hybrid Machine Learning and Physics-Based Model," CIRP J. Manuf. Sci. Technol., **35**, pp. 25–40.
[22] Tran, M. Q., Liu, M. K., and Tran, Q. V., 2020, "Milling Chatter Detection Using Scalogram and Deep Convolutional Neural Network," Int. J. Adv. Manuf. Technol., **107**(3–4), pp. 1505–1516.
[23] Sener, B., Gudelek, M. U., Ozbayoglu, A. M., and Unver, H. O., 2021, "A Novel Chatter Detection Method for Milling Using Deep Convolution Neural Networks," Measurement, **182**, p. 109689.
[24] Cherukuri, H., Perez-Bernabeu, E., Selles, M., and Schmitz, T., 2019, "Machining Chatter Prediction Using a Data Learning Model," J. Manuf. Mater. Process., **3**(2), p. 45.
[25] Lamraoui, M., Barakat, M., Thomas, M., and El Badaoui, M., 2015, "Chatter Detection in Milling Machines by Neural Network Classification and Feature Selection," J. Vib. Control, **21**(7), pp. 1251–1266.
[26] Sestito, G. S., Venter, G. S., Ribeiro, K. S. B., Rodrigues, A. R., and da Silva, M. M., 2022, "In-Process Chatter Detection in Micro-Milling Using Acoustic Emission via Machine Learning Classifiers," Int. J. Adv. Manuf. Technol., **120**(11–12), pp. 7293–7303.
[27] Peng, C., Wang, L., and Liao, T. W., 2015, "A New Method for the Prediction of Chatter Stability Lobes Based on Dynamic Cutting Force Simulation Model and Support Vector Machine," J. Sound Vib., **354**, pp. 118–131.
[28] Karandikar, J., Honeycutt, A., Schmitz, T., and Smith, S., 2020, "Stability Boundary and Optimal Operating Parameter Identification in Milling Using Bayesian Learning," J. Manuf. Process., **56**, pp. 1252–1262.
[29] Karandikar, J., Honeycutt, A., Smith, S., and Schmitz, T., 2020, "Milling Stability Identification Using Bayesian Machine Learning," Procedia CIRP, **93**, pp. 1423–1428.

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026

[30] Schmitz, T., 2024, "Physics-Informed KNN Milling Stability Model With Process Damping Effects," J. Manuf. Process., **120**, pp. 1124–1129.
[31] Karim, F., Majumdar, S., Darabi, H., and Chen, S., Dec. 2017, "LSTM Fully Convolutional Networks for Time Series Classification," IEEE Access., **6**, pp. 1662–1669.
[32] Sainath, T. N., Vinyals, O., Senior, A., and Sak, H., 2015, "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, Apr. 19.
[33] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D., 2017, "Temporal Convolutional Networks for Action Segmentation and Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, July 21–26, pp. 156–165,
[34] Lea, C., Vidal, R., Reiter, A., and Hager, G. D., 2016, "Temporal Convolutional Networks: A Unified Approach to Action Segmentation," *Computer Vision—ECCV 2016 Workshops*, G. Hua, and H. Jégou, eds., Springer International Publishing, Cham, pp. 47–54.

Downloaded from http://asmedigitalcollection.asme.org/manufacturingscience/article-pdf/147/10/101005/7522416/manu-25-1218.pdf by Rutgers University user on 28 May 2026