GAN-based feature representation and data augmentation for tool wear classification

Seulki Han[1] and George M. Bollas[1]

***Abstract*— Tool wear monitoring is critical for maintaining machining efficiency, product quality, and cost-effectiveness in manufacturing. Traditional data-driven approaches rely on manually engineered features, which require extensive preprocessing and domain expertise, and often lead to information loss and limited adaptability. This study proposes an alternative method that uses wavelet transform (WT)-based scalogram images and a conditional generative adversarial network (cGAN) for feature representation and data augmentation. Scalograms retain the full time-frequency structure of raw signals, eliminating the need for handcrafted features. The cGAN model synthesizes realistic scalograms, expanding dataset diversity and improving model generalization. Compared with FFT-based numerical features, the proposed method preserves more wear-related information and improves classification performance. Experimental results demonstrate that the cGAN-based approach effectively captures tool wear progression, enhances training robustness, and improves accuracy under limited data conditions. These findings highlight the potential of integrating cGAN-generated features into tool wear monitoring frameworks.**

I. Introduction

Tool wear monitoring is essential for maintaining machining efficiency, product quality, and cost-effectiveness in modern manufacturing [1]. Although data-driven methods have improved prediction accuracy, they face limitations such as insufficient data, complex feature engineering, and potential information loss. Building reliable models requires large datasets across diverse machining conditions, but high costs and time constraints often restrict data collection in real-world settings [2]. This scarcity hinders model generalization and increases the risk of overfitting. Feature engineering also presents challenges. Manual approaches often involve transforming signals into alternative domains and selecting representative features [3], [4]. These methods require domain expertise, are time-consuming, and may not adapt well to changing conditions [5]. Moreover, traditional transformations such as fast Fourier transform (FFT) overlook time-localized wear dynamics, while Wavelet transform (WT)-based statistical features may fail to capture key signal patterns [6], reducing monitoring accuracy.

To address these issues, this study proposes a scalogram-based representation using wavelet transform and conditional GANs (cGANs). This approach preserves time-frequency characteristics visually, eliminating the need for handcrafted features. A cGAN is used for data augmentation to improve dataset diversity and classification performance. For comparison, a traditional FFT-based method with band power features is also evaluated. The primary objectives of this study are: (i) to compare numerical and image-based features for tool wear representation, and (ii) to assess the effectiveness of GAN-based augmentation in enhancing model performance and generalization.

[1]Department of Chemical & Biomolecular Engineering, University of Connecticut, 159 Discovery Dr., Storrs, CT, 06269, USA george.bollas@uconn.edu

II. Generative Adversarial Networks (GANs)

GANs, introduced by Goodfellow et al. [7], are widely used for image synthesis, data augmentation, and domain adaptation. A GAN consists of two competing neural networks: a generator (G), which generates synthetic samples from random noise, and a discriminator (D), which distinguishes real samples from generated ones. The two networks are trained simultaneously in an adversarial setting, where the generator aims to produce realistic data that can deceive the discriminator, while the discriminator learns to correctly classify real and synthetic inputs.

The standard GAN objective is formulated as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (1)$$

where $p_{\text{data}}(x)$ represents the real data distribution, and $p_z(z)$ is the prior distribution of the noise input. $D(x)$ denotes the probability that the discriminator classifies a real sample x as real, while $D(G(z))$ denotes the probability that it classifies a generated sample $G(z)$ as real. The discriminator seeks to maximize this objective by assigning high probabilities to real samples and low probabilities to fake ones. Conversely, the generator minimizes the objective by producing samples that the discriminator cannot distinguish from real data, thereby improving the realism of its outputs [7].

A. Conditional GANs (cGANs)

Traditional GANs generate images without control over specific attributes. In many applications, however, it is beneficial to guide the generation process using class labels. Conditional GANs (cGANs) extend the original framework by incorporating label information into both the generator and discriminator, enabling label-conditioned generation [8]. This approach improves data relevance and helps the model capture structured variations across classes, potentially enhancing generalization.

In cGANs, the objective function is modified to include conditioning information y, representing the class

label—such as tool wear states in this study. The updated loss is defined as [9]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x|y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z|y)))] \quad (2)$$

As shown in Fig. 1, this formulation guides the generator to produce samples aligned with the given label, while the discriminator evaluates both authenticity and label consistency.

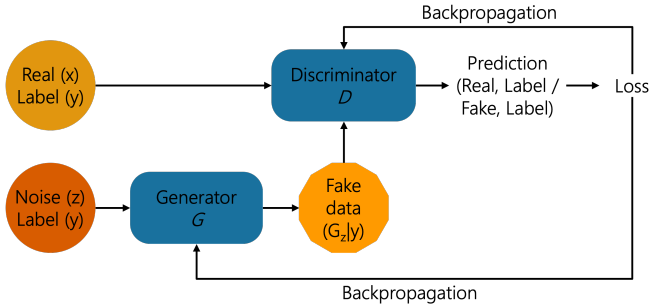

Fig. 1: Flow diagram of cGAN model

III. Tool Run-to-Failure Experiments

This study utilizes the NASA Milling Dataset, a widely used benchmark for tool wear analysis [10]. The experiments were performed on an MC-510V CNC milling machine using cast iron and stainless steel J45 workpieces (483 mm × 178 mm × 51 mm) with a 70 mm six-tooth face milling cutter fitted with KC710 carbide inserts. The dataset contains 16 cases, each representing different combinations of operating conditions and workpiece materials. Each case includes multi-sensor data such as acoustic emission (AE), vibration, and spindle motor current signals. Tool wear was measured via optical microscopy, with the maximum flank wear among the six inserts recorded as the wear value.

In this study, two cases with distinct operating conditions and different workpiece materials were selected for analysis due to their clear wear progression and variation across multiple test runs. By synthesizing scalogram images across these two cases, the proposed model is challenged to capture generalized wear patterns beyond a single experimental setting, demonstrating its potential for broader applicability in real-world monitoring scenarios. Table I summarizes the cutting parameters and workpiece materials for the two selected cases, while Fig. 2 presents the tool wear progression over time for each case.

TABLE I: Operating conditions and workpiece material of the selected cases

Cases	Depth of cut (mm)	Feed rate (mm/min)	Workpiece
1	1.5	0.75	Cast iron
2	0.75	0.5	Steel

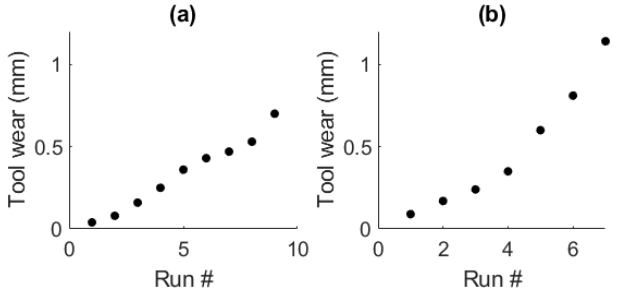

Fig. 2: Tool wear progression of the selected cases

IV. Methodology

This section introduces the feature representation approaches used in this study and their roles in tool wear monitoring. As illustrated in Fig. 3, the study compares a traditional FFT-based numerical feature extraction pipeline with a scalogram-based approach using deep generative modeling. It also presents additional analyses to evaluate the effectiveness of the generated features and their impact on tool wear classification.

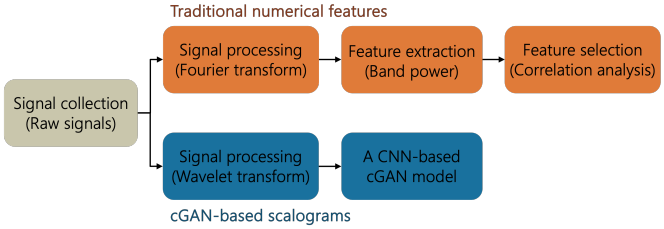

Fig. 3: Feature representation approaches: Manual numerical features and cGAN-based scalograms

A. FFT-based numerical feature extraction

Six signals were analyzed in the manual feature extraction approach, including AE, vibration, and spindle motor current (AC and DC) from both the spindle and table. FFT was applied to convert each signal to the frequency domain, and band power features were computed to capture signal energy across the full frequency range, avoiding potential information loss from selecting specific bands. Among the extracted features, the most representative features were selected based on the Spearman correlation coefficient (SCC), which is effective in detecting monotonic relationships with tool wear progression.

B. cGAN-based image feature representation

1) Scalograms generation: WT-based scalogram representations were explored as an alternative to manually engineered numerical features. Scalograms offer better time-frequency resolution compared to spectrograms and are particularly effective in capturing transient signal characteristics associated with tool wear. Based on the manual approach results, the most representative signal was selected and transformed into time-frequency images using wavelet transform. Raw signals were segmented using a sliding

window of 250 data points to ensure each segment captured meaningful wear dynamics and provided enough samples for cGAN training. For each test run, 40 scalogram images were generated. To maintain visual consistency, the global minimum and maximum WT coefficients across all segments were used to normalize the colormap scale, ensuring that intensity differences reflected actual wear progression across scalograms.

2) cGAN model development: A convolutional neural network (CNN)-based cGAN was developed to generate realistic scalogram images for data augmentation. The generator received a random noise vector and a class label, which was embedded and combined with the noise before being processed through a series of transposed convolutional layers with Leaky ReLU activations. Batch normalization was applied to stabilize training, and a final tanh activation function scaled output values between -1 and 1. The discriminator was designed to distinguish between real and generated scalograms, conditioned on the corresponding class label. It processed input images through convolutional layers with Leaky ReLU activations and batch normalization, with the final output obtained via a sigmoid function. The model was trained using the Wasserstein GAN with gradient penalty (WGAN-GP) loss function, which improves training stability and addresses issues of mode collapse [11]. The Adam optimizer was used for both networks, and the discriminator was updated more frequently than the generator to maintain balanced learning. The WGAN-GP loss functions for the generator and discriminator are defined as follows:

a) Discriminator Loss:

$$\mathcal{L}_D = \mathbb{E}_{\tilde{x}}[D(\tilde{x}, y)] - \mathbb{E}_x[D(x, y)] + \lambda \mathcal{L}_{GP} \tag{3}$$

b) Generator Loss:

$$\mathcal{L}_G = -\mathbb{E}_{\tilde{x}}[D(\tilde{x}, y)] \tag{4}$$

where x and $\tilde{x}$ denote real and generated samples, respectively, and y represents the class label. $\mathcal{L}_{GP}$ is the gradient penalty term that enforces the Lipschitz constraint, and λ is the penalty coefficient.

3) Quality evaluation of synthesized images: To evaluate the quality of the scalogram images generated by the cGAN model, two widely used image quality metrics were employed: the Fréchet Inception Distance (FID) and the Structural Similarity Index Measure (SSIM). These metrics capture different aspects of similarity. FID evaluates the distributional closeness between real and synthetic images in a deep feature space, while SSIM measures perceptual similarity based on local image structures.

FID calculates the distance between two multivariate Gaussian distributions estimated from the features of real and generated images [12]. Let μ_r, Σ_r and μ_g, Σ_g denote the means and covariances of features from real and generated image sets, respectively. Then, the FID score is given by:

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right) \tag{5}$$

Lower FID values indicate that the generated images have feature distributions closer to those of real images, suggesting higher generation quality.

SSIM, in contrast, compares two images x and y based on luminance $l(x, y)$, contrast $c(x, y)$, and structural similarity $s(x, y)$ [13]. It is defined as:

$$\text{SSIM}(x, y) = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma} \tag{6}$$

where α, β, γ are weighting parameters, typically set to 1. SSIM values range from 0 to 1, with higher values indicating greater similarity in visual structure.

C. Tool wear classification

To evaluate the effectiveness of the synthesized scalogram images, a tool wear classification task was conducted using a CNN model. A total of 640 real scalogram images were prepared, corresponding to 16 tool wear classes with 40 images per class. These classes represent distinct tool wear measurements extracted from two machining cases. In total, nine machining tests were performed for Case 1 and seven for Case 2, with 16 tool wear measurements collected from two cases and treated as discrete classes in the classification task.

The classification performance was assessed using 5-fold cross validation. Two training scenarios were designed: a baseline scenario using only real images, and an augmented scenario incorporating synthetic images generated by the cGAN model. In the augmented scenario, synthetic images were added to the real training set at varying ratios, from 10% to 100% of the real image count, increasing in 10% increments. All models were tested exclusively on real images to ensure consistent evaluation across scenarios. By comparing the baseline and augmented models, this experiment aims to investigate how GAN-based data augmentation influences training data diversity and improves classification performance.

V. Results and Discussion

A. FFT-based numerical feature extraction

The frequency spectra of the six signals from the initial machining test in Case 1 are shown in Fig. 4. The current signals (Figs. 4(a) and (b)) show dominant low-frequency peaks, reflecting their sensitivity to cutting forces and the tool rotational motion. In contrast, the vibration and AE signals (Figs. 4(c)-(f)) exhibit broader frequency distributions. These signals capture more complex dynamics: vibration reflects structural interactions within the milling system, while AE signals originate from transient events such as chip formation and material deformation.

Fig. 5 shows the band power trends of the six sensor signals over time. The current signals (Figs. 5(a) and (b)) increase with tool wear, with the AC current (a) showing a near-linear rise due to increasing cutting resistance. The table vibration signal (c) decreases with fluctuations, possibly due to damping at the worn cutting edge, while the spindle vibration (d) displays irregular patterns from transient tool–workpiece interactions. The AE signals (e) and

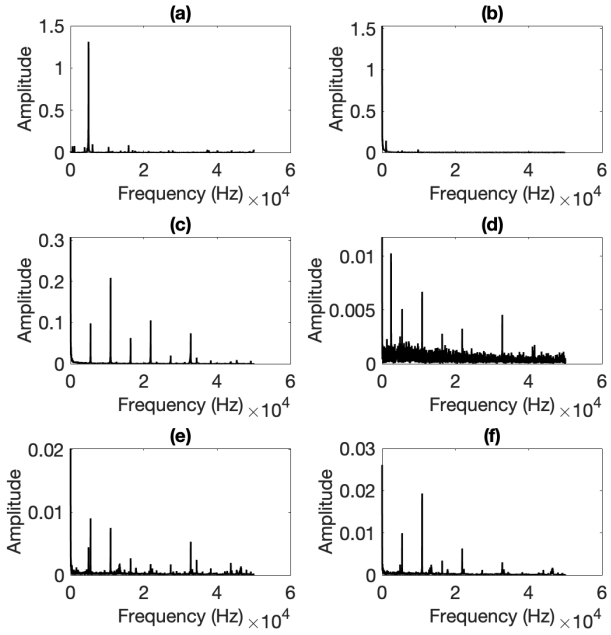

Fig. 4: Frequency spectrum of six signals collected from the first machining test of Case 10: (a) AC current, (b) DC current, (c) vibration at table, (d) vibration at spindle, (e) acoustic emission at table, and (f) acoustic emission at spindle

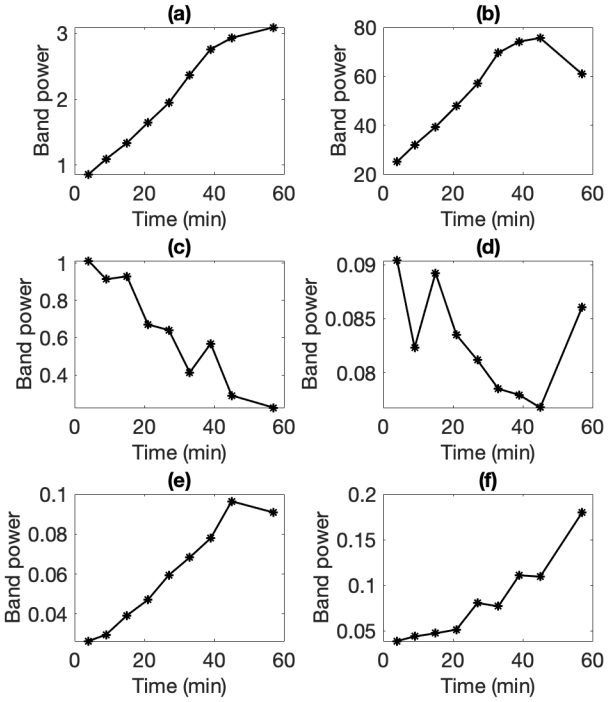

Fig. 5: Band power computed from six signals of Case 10: (a) AC current, (b) DC current, (c) vibration at table, (d) vibration at spindle, (e) acoustic emission at table, and (f) acoustic emission at spindle

(f) gradually increase, reflecting higher stress and energy release. Overall, current and AE signals show a strong correlation with tool wear, while vibration signals exhibit more complex variations.

Table II summarizes the correlation between band power and tool wear. As expected from Fig. 5, current and AE signals show strong positive correlations, with the AC current exhibiting the highest. In contrast, vibration signals show weaker or negative correlations, with the spindle vibration signal showing the lowest due to its irregular trend.

TABLE II: Correlation between band power computed from each signal and tool wear

Signal type	Correlation
AC current	1.00
DC current	0.90
Table vibration	-0.96
Spindle vibration	-0.58
Table acoustic emission	0.98
Spindle acoustic emission	0.96

B. cGAN-based scalogram images

AC current signals, which were found to be the most representative of tool wear progression, were used for the cGAN-based approach. Fig. 6 shows example scalogram images generated from AC current signals at different tool wear states, corresponding to Run #1, 4, and 9 in Case 1 and Run #10, 13, and 16 in Case 2. As observed in Fig. 6, the scalograms consistently exhibit strong energy concentration in the lower frequency region, aligning with the dominant peak observed in the FFT spectrum of the AC current signal (see Fig. 4(a)). This indicates that the current signal carries most of its wear-related information in the low-frequency band. Furthermore, as tool wear progresses from early (Run #1 and #10) to late stages (Run #9 and #16), the intensity of the low-frequency pattern becomes more pronounced, with the color gradually shifting from blue to red. This trend suggests an increase in signal energy and cutting resistance as wear accumulates, highlighting the potential of scalograms to visually reflect wear progression.

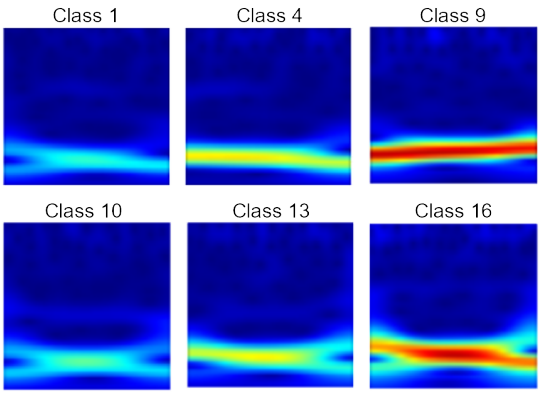

Fig. 6: Wavelet-based actual scalogram images corresponding to different tool wear states

Fig. 7 presents example scalogram images synthesized by the cGAN model for six representative tool wear classes: Run

#1, 4, 9 (Case 1) and Run #10, 13, 16 (Case 2). The generated images display strong visual similarity to the corresponding real scalograms shown in Fig. 6, particularly in the dominant low-frequency regions where wear-related signal energy is concentrated. The synthetic scalograms successfully captured the key spectral patterns observed in the real data. For instance, as tool wear progresses (e.g., from Run #1 to Run #9), the intensity of the low-frequency components gradually increases, which is reflected in the enhanced red coloration in the corresponding regions. This pattern aligns well with the physical interpretation of tool wear causing increased current and thus elevated energy in lower frequencies. Similar progression is observed in the synthesized images for Case 2 (Run #10 to Run #16), indicating that the generator has learned to model wear-related transitions across different operating conditions. Overall, these results suggest that the cGAN model is capable of producing realistic scalogram images that not only resemble the visual structure of real data, but also preserve class-specific signal characteristics relevant to tool wear progression.

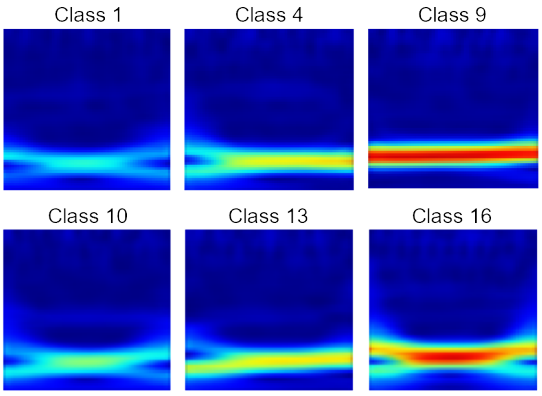

Fig. 7: Scalogram images, corresponding to different tool wear states, generated by the cGAN model

C. Training stability and image quality assessment

To monitor the training progress of the cGAN model, generator and discriminator losses, along with FID scores, were tracked throughout training. Figs. 8 (a) and (b) present both the loss curves and the FID scores evaluated at regular intervals, respectively. As shown in the figure, the generator loss decreased rapidly during early epochs while the discriminator loss increased, indicating effective adversarial learning. After approximately 50 epochs, both losses stabilized, suggesting that the model reached equilibrium. Similarly, the FID score dropped from over 90 to around 30 within the first 50 epochs and gradually converged to values near 20, indicating progressive improvement in the quality of generated scalograms.

To further evaluate the quality of the scalogram images generated by the trained cGAN model, both FID and SSIM metrics were assessed after training convergence. For each tool wear class, 10 synthetic images were generated using the final cGAN model. To mitigate the variability introduced

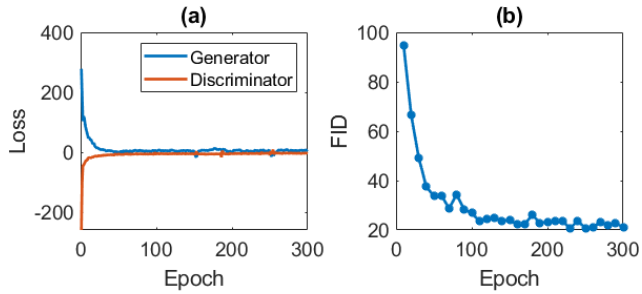

Fig. 8: Training progress of the cGAN model: (a) Generator and discriminator loss curves over 300 epochs, (b) FID scores evaluated during training progresses.

by the random noise vector, this generation process was repeated three times, and the average FID score across all classes was computed. The resulting mean FID was 22.98, indicating a high level of similarity between real and generated scalograms. SSIM was also calculated between real and synthetic images for each class. As shown in Table III, mean SSIM values ranged from 0.8756 to 0.9202, with most above 0.89, indicating strong structural resemblance. These results confirm that the trained cGAN model produces realistic and class-consistent scalograms.

TABLE III: Mean and standard deviation of SSIM between real and generated scalograms for each tool wear class

Class	Mean	STD	Class	Mean	STD
1	0.8959	0.0121	9	0.8756	0.0161
2	0.8985	0.0142	10	0.8944	0.0128
3	0.9202	0.0052	11	0.8939	0.0139
4	0.8935	0.0119	12	0.8956	0.0160
5	0.8939	0.0134	13	0.8941	0.0137
6	0.8884	0.0148	14	0.8839	0.0139
7	0.8905	0.0124	15	0.8863	0.0119
8	0.8911	0.0123	16	0.8893	0.0147

D. Tool wear classification

To evaluate the impact of cGAN-based data augmentation on classification performance, tool wear classification was conducted using training datasets augmented with different ratios of synthetic images. Fig. 9 illustrates the mean test accuracy over five cross-validation folds across various synthetic-to-real data ratios. The baseline model, trained with real images only, achieved an average accuracy of approximately 90%. When synthetic images were added, classification performance consistently improved, with the highest accuracy observed when the synthetic data accounted for 40–60% of the training set. Beyond this range, accuracy began to slightly decrease, though it generally remained comparable to the baseline. These results suggest that incorporating a moderate amount of high-quality synthetic data can enhance model generalization and performance, while excessive reliance on generated data may lead to diminishing returns. The improvement confirms the effectiveness of cGAN-based augmentation in expanding training diversity without sacrificing prediction accuracy.

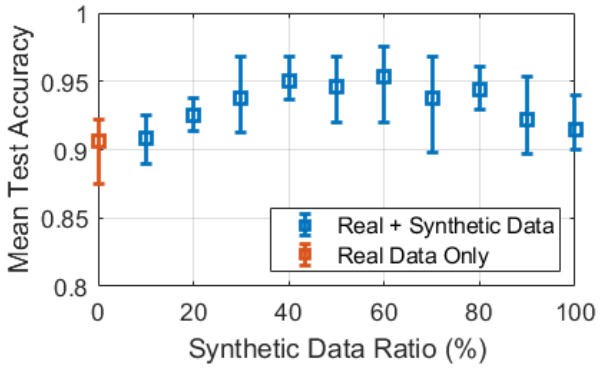

Fig. 9: Mean test accuracy across five folds for different synthetic data ratios in training.

E. Advantages of cGAN-based scalogram images for feature representation

The comparison between traditional handcrafted features and cGAN-based scalogram images reveals several key advantages of the image-based approach for tool wear monitoring. First, scalogram representations preserve the full time-frequency characteristics of sensor signals without requiring explicit feature engineering. In contrast, the manual approach involves multiple steps such as signal transformation, feature extraction and selection, and dimensionality reduction which demand extensive preprocessing and domain expertise. As shown in Section V-A, even though band power features can effectively reflect wear progression, the selection of relevant features often depends on machining conditions and sensor types, limiting generalizability. Scalogram images also reduce information loss. Handcrafted features condense complex signals into scalar values, which may discard important spectral patterns. For instance, band power features only represent total energy within a frequency band, ignoring temporal variations. Even with wavelet transforms, numerical feature extraction often leads to simplification. In contrast, scalograms maintain the full time-frequency distribution, allowing models to learn richer wear-related information directly from the data. A further advantage of the cGAN-based approach is its ability to perform data augmentation. By generating synthetic scalograms, the cGAN model increases dataset diversity and improves classification robustness. As demonstrated in Section V-D, adding synthetic images led to improved classification accuracy. While training a cGAN model requires careful tuning and architectural design, the benefits in preserving signal integrity, reducing manual preprocessing, and expanding data diversity and size make the approach highly promising. Overall, cGAN-based scalogram representations offer a more scalable and data-efficient solution for tool wear monitoring.

VI. Conclusions

This study explored the effectiveness of cGAN-based image generation and data augmentation as an alternative to traditional numerical feature extraction for tool wear monitoring. FFT-transformed band power features showed strong correlation with tool wear but involved extensive preprocessing and information compression. In contrast, WT-based scalogram images preserved full time-frequency information, enabling richer representation of wear-related patterns. The cGAN model successfully synthesized realistic scalogram images that captured the spectral structure and progression trends of tool wear. Quantitative evaluations confirmed image quality, with an average FID of 22.98 and mean SSIM values exceeding 0.89 for most classes. In a classification task, models trained with a moderate ratio (40–60%) of synthetic images achieved higher accuracy than the baseline, demonstrating the effectiveness of cGAN-based augmentation in improving performance under limited data conditions. Overall, the proposed approach reduces information loss, simplifies preprocessing, and expands training data, making it a promising direction for scalable tool wear monitoring. Future work will focus on enhancing generation quality and extending this framework to other monitoring tasks, such as remaining useful life (RUL) prediction and multi-sensor fusion.

ACKNOWLEDGMENT

References

[1] S. Han, N. Mannan, D. C. Stein, K. R. Pattipati, G. M. Bollas, Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems, Journal of Manufacturing Systems, 61 (2021) 45–53.

[2] Mohanraj, T., Sai Bharath, R. (2024). Real-time tool condition monitoring with the internet of things and machine learning algorithms. International Journal of Computer Integrated Manufacturing.

[3] S. Han, Q. Yang, K. R. Pattipati, G. M. Bollas, Sensor selection and tool wear prediction with data-driven models for precision machining, Journal of Advanced Manufacturing and Processing (2022).

[4] Munaro, R., Attanasio, A., del Prete, A. (2023). Tool Wear Monitoring with Artificial Intelligence Methods: A Review. Journal of Manufacturing and Materials Processing, 7(4), 129.

[5] Mohanraj, T., Shankar, S., Rajasekar, R., Sakthivel, N. R., Pramanik, A. (2020). Tool condition monitoring techniques in milling process — a review. Journal of Materials Research and Technology, 9(1), 1032–1042.

[6] Wang, Q., Wang, H., Hou, L., Yi, S. (2021). Overview of Tool Wear Monitoring Methods Based on Convolutional Neural Network. Applied Sciences 2021, Vol. 11, Page 12041, 11(24), 12041.

[7] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Networks. Science Robotics, 3, 2672–2680.

[8] Feng, Y. (2023). Improving Tool Wear Prediction with Synthetic Features from Conditional Generative Adversarial Networks. Authorea Preprints, XX, 1.

[9] Luo, J., Huang, J., Li, H. (2021). A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. Journal of Intelligent Manufacturing, 32(2), 407–425.

[10] A. Agogino, K. Goebel, Milling Data Set, BEST lab, UC Berkeley, NASA Ames Prognostics Data Repository, 2007.

[11] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A. (2017). Improved Training of Wasserstein GANs, Montreal Institute for Learning Algorithms.

[12] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.

[13] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, (2017)."Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, 13(4), 600-612.