Effects of extrinsic noise factors on machine learning based chatter detection in machining

L. Lu[1], T. R. Kurfess[1], C. Saldana[1,*]
[1]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive, Atlanta, GA 30332 USA
*Corresponding author: christopher.saldana@me.gatech.edu

Abstract Unmitigated chatter can result in poor part quality, accelerated tool wear, and possible damage to spindle and machine. While various methods have been shown to effectively detect chatter, implementation of these methods in noisy environments, such as factory floors, has not been well studied. The present study seeks to explore the effects of extrinsic noise sources on threshold-based and machine learning-based chatter detection methods using audio signals of the machining process. To accomplish this, stable and unstable cuts were made on a milling machine and the audio signal was collected. Data augmentation using Gaussian white noise and periodic noise was conducted to simulate a range of noise levels and types. The performance of these techniques were then compared with respect to the increasing levels of noise. It was found that machine learning based approaches achieved satisfactory accuracies up to 98.6% under the presence of extrinsic noise. Conventional static threshold techniques, however, failed under most noise conditions and resulted in false positives depending on the threshold values used. Further, support vector machine approaches demonstrated an ability to classify noisy data despite limited training.

1 Introduction

Manufacturers are constantly balancing the demands of product quality, product variability, cost, and speed. As a result, manufacturers are increasingly turning to automation to meet these demands and remain competitive, globally. In recent decades, a vast amount of research has been focused on process monitoring to reduce the need for expert operators. Process monitoring allows machine tool users to migrate from scheduled maintenance operations and move towards condition-based maintenance approaches. Such process monitoring for machining-based processes involve monitoring of process variables such as cutting forces, vibrations, acoustic emission (AE), noise, temperature, surface finish, that are influenced by the cutting tool and cutting parameters inputted. These features potentially correlate with tool or process conditions of interest to the machine operator.

Milling-based processes are of particular importance for process monitoring. In milling, forces generated when the cutting tool engages with the work piece produce significant deflection of the tool-workpiece system. When the machining process becomes unstable, regenerative chatter occurs and can result in poor part quality, accelerated tool wear, and possible damage to the spindle and machine [1]. Unmitigated chatter results in poor surface finish, dimensional inaccuracy, excessive noise, machine tool damage, reduced tool life and reduced MRR material waste, energy waste, increased machining time, and increased costs [2]. Thus, chatter detection, avoidance, and suppression are all major areas of interest for researchers and machine operators. To avoid chatter, a common practice is to use conservative cutting parameters. Suboptimal parameters result in lower material removal rate (MRR) and often result in decreased productivity compared to a higher, stable MRR. Approaches that include model-based parameter selection are possible for a priori selection of machining parameters, but approaches for detection of chatter are still critical from a monitoring perspective. When machining-induced chatter occurs, the resulting vibration

has a frequency different from that of the cutting tooth passing frequency. For this reason, frequency-based techniques are generally effective at detecting chatter. Several threshold-based methods of analyzing audio signals in the frequency domain for detecting chatter have been developed by researchers over the past few decades [3-6]. Other methods involve analyzing characteristics of the time, frequency and time-frequency domains, as well as extracting relevant features to train machine learning (ML) classifiers such as support vector machines (SVM) or artificial neural networks (ANN) [7-11].

From a sensing perspective, researchers have been able to detect chatter with a variety of instruments, including dynamometers [12-14], accelerometers [11, 15-18], microphones [4-7, 18-19], as well as arrays of these sensors [20-22]. The use of microphones for monitoring the cutting process has been shown to be effective and inexpensive compared to other sensor approaches [4, 6, 16]. Accelerometers are acceptable as well, but placement can cause change in the apparent strength of different modes of vibration, this leading to sensitivity and noise problems. Schmitz et al. [5] sampled the audio signal of a cutting operation once per revolution using a microphone and was able to detect high variance in the measured audio signal under unstable cutting conditions. Ismail and Ziaei [23-24] used acoustic intensity to detect chatter in 5-axis machining by classifying signals as stable, moderate chatter, and severe chatter conditions. Their detection system was paired with a spindle speed ramping system that would move the operation into a stable cutting zone and suppress chatter based on classification. A similar approach is taken by Tsai et al. [6], where two microphone signals were averaged and converted into an acoustic chatter signal index. After this index exceeded a set threshold, chatter was said to be detected. This index was taken after background noises such as fluid flow and AC power have been filtered out properly.

One of the major disadvantages of using microphones in industrial environments, however, is the prevalence of noise coming from the factory floor. Stray noise sources in industrial settings could create false alarms in chatter detection systems [9]. False alarms have the potential to greatly slow down production, depending on the actions of the chatter detection system. If the system fails to operate properly in a real factory environment, either by missing chatter events or in misclassifying stable cutting, then the system would be ineffective. Sound isolation and filtering may be used to decrease the effects of noise on audio signal collection systems [4] and acoustic intensity can also be used to reduce the effects of background noise [4, 23-24]. In industrial environments, which have been found to have sound pressure levels of up to 90 dB, these false alarms could become a significant issue [25]. These factory sounds have a large range of frequencies they affect. Noise sources include, but are not limited to, motors, fans, other machining centers, talking, air flow, and fluid flow. Rarely do machining operations happen in complete isolation, so it is important to understand the capabilities of automated chatter detection methods in the presence of background noises. Many methods for detecting chatter require setting a constant noise threshold that once passed, raises an alarm for chatter [4-6, 19]. If the threshold is set high such that the system is less affected by noise, the chance for missed detection increases. Implementation of these audio signal chatter classification systems in industrial environments must consider the influence of non-process induced noise. Inaccurate classifications due to noise can lead to non-optimal tooling operations, or, worst-case catastrophic failure with unmitigated chatter. Understanding background noise can increase the reliability of future systems. ML approaches such as SVM-based classification may have the ability to function accurately in the presence of heavy factory noise, unlike standard thresholding methods. SVMs have been shown to be able to classify

accurately with limited training and fairly noisy data in other applications such as tool wear [26, 27].

While the impact of noise levels on threshold-based chatter detection methods has been explored elsewhere [4, 9], the effect of noise on classification performance for ML-based chatter detection approaches is not well understood. ML-based approaches may provide significantly more enhanced performance in the presence of noise when compared to threshold-based methods [26]. Thus, the present study seeks to address the question of whether ML-based approaches are more robust to high levels of noise in assessing chatter in machining signals. The purpose of this study is to examine the performance of various chatter classification methods under varying background noises. To accomplish this, stable and unstable cuts were made on a milling machine and the audio signal was collected during processing. The audio signal was then superimposed with varying levels of white Gaussian noise and periodic noise to simulate the effects of a noisy data collection environment. The accuracy of these techniques was then compared with respect to the increasing levels of noise. The objective is to understand the balance between sensitivity, speed of implementation, and performance of various classification systems under noisy conditions.

2 Experimental Methods

Machining experiments were performed on an EMCO E350 machining center with a Siemens Sinumerik 828d controller. Experiments were conducted when no other machines were operating in the area to ensure the clearest raw audio signal could be obtained. A 0.375 in. diameter, 1.75 in. length, 2-flute solid carbide end mill (Kennametal ABDF0375J2AS K600) held in an ER32 collet and an SK30 taper tool holder was used to cut slots in AA6061-T6. Machining experiments utilized spindles speeds of 3000, 5000, 6000, and 7000 RPM, a constant linear feed rate of 0.03 in/tooth/rev and 100% radial immersion (full-width slot). From a starting depth of 0.10 in., the depth of cut (DOC) was increased until chatter was induced at 0.60 in. maximum. The feed per tooth was kept constant at 0.03 in/tooth. Coolant was on for the slotting operations. A PCB Piezotronics 130F20 microphone with 45 mV/Pa sensitivity and signal conditioner was used to collect the audio signal during the experiments. The microphone was placed inside the machining center enclosure approximately 30 inches from the tool cutting edge. Audio signals were collected in LabVIEW at a sample rate of 48 kHz using a compactRIO-9014 with an NI-9215 module. 48kHz was chosen to achieve a Nyquist frequency of 24 kHz. Processing was done separately in MATLAB.

Artificial noise was superimposed on the data by adding noise signal components directly to the original signal. This superposition of noise on the original audio signal had two components, a white Gaussian noise component and a periodic component, the latter of which was modeled by a sine wave and its first 2 harmonics. 3 levels of white Gaussian noise were chosen to overlay to create signal to noise ratios (SNR) of 20, 15, and 10 dBW. The periodic noise had a base frequency at 310 Hz and harmonics at 620 and 930 Hz. The base frequency of 310 Hz was chosen to be identifiable from the tooth passing frequencies and harmonics of the experimental cutting conditions. The periodic noise frequencies are limited to below 1000 Hz, as the attenuation of sound through air is larger at high frequencies [50]. Doing so allows the noise to pass through the filter with some attenuation. The base frequencies had the highest amplitude of 2, 1 and 0.5 Pa. The first and second harmonics had amplitudes that were 30% and 20% of the base frequency amplitude. It is assumed that beyond the second harmonic, the amplitudes decreased rapidly and thus had little effect on the original signal. Tables 1 summarizes the periodic and white Gaussian

noise signals composition used for data augmentation into 3 noise levels. In the ensuing, these data were used for training a range of ML-based classifiers, including decision tree, SVM, kNN and bagged tree models with a range of training sets. When training the models, a cross-validation was done with 5 folds validation to tune the model parameters. The design of these experiments are described below.

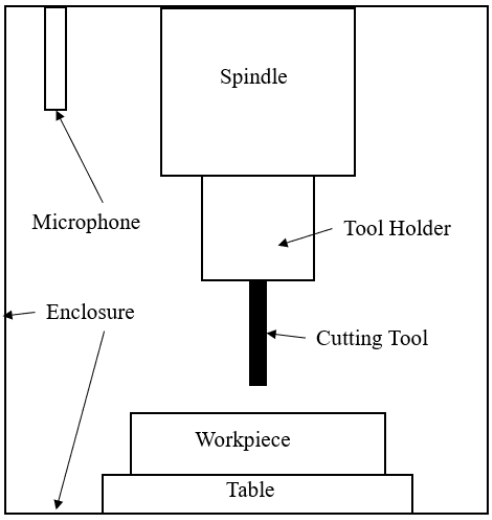

Figure 1. Audio collection setup.

Table 1. Noise signal composition at 3 levels

White Gaussian Noise Level	SNR (dBW)
1	20
2	15
3	10

Periodic Noise Level	Frequency (Hz)	Amplitude (Pa)
1	310	0.5
	620	0.15
	930	0.1
2	310	1
	620	0.3
	930	0.2
3	310	2
	620	0.6
	930	0.4

3 Results and Discussion

Audio signals were collected across the range of experimental conditions tested. Example full audio signals can be seen in Fig. 2, where Fig. 2(a) shows an example of a stable cut and Fig. 2(b) shows an example of an unstable cut. In the stable cut of Fig. 2(a), the amplitude of the cut

remained relatively stable. In the unstable cut of Fig. 2(b), the amplitude of the acoustic signal increased as the cut progressed due to the regenerative nature of chatter. For both sample audio signatures, when the tool entered and exited the workpiece, there was a sharp increase in the amplitude of the signal that quickly dissipated. This sharp increase in the audio signal is expected as the tool initiated the contact with the workpiece and the engagement reached steady state. Similarly, as the tool exited contact, the lack of stiffness in the system as the tool unloaded gave rise to similar increase in the audio signal. The transient behaviors in the audio signal at the start and end of each machining pass were truncated such that the final signal used for training would be indicative of the portion of the trace wherein the tool was fully immersed in the cut. Each tool path took approximately 7 seconds from entrance to exit. After truncation, approximately 5 seconds of audio data was gathered.

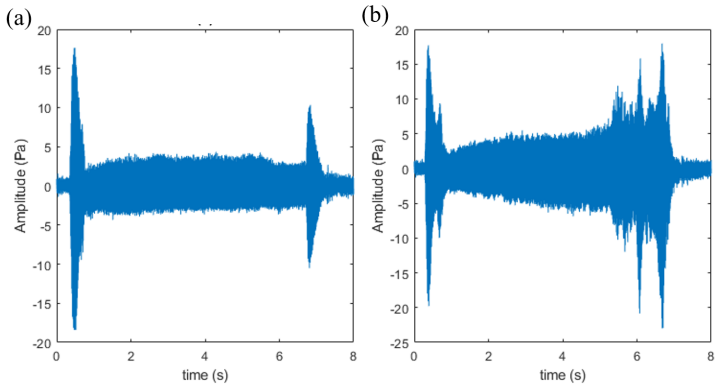

Figure 2. (a) A stable cut at 3000 RPM and 0.20 in. DOC compared with (b) an unstable cut done at 3000 RPM and 0.25 in. DOC.

From the full audio signal, the data was segmented into 0.1-second segments with a 50% overlap to increase the dataset size. Examples of data segments for the stable and unstable cuts are shown in Fig. 3. In both waveforms it is clear that the audio signal exhibited sinusoidal-type periodic character, this well attributed to the tooth-passing frequency, which was approximately 0.01 s for the 2-flute tool at 3000 RPM. However, in the unstable cut of Fig. 3(b), the waveform more clearly exhibited additional erratic frequencies resulting from chatter on the base signal. Figure 3(c) and Fig. 3(d) show transformation of the data segments to the frequency domain with the FFT and a Hanning window. In these plots, the tooth-passing and chatter signals are apparent, where the tooth passing frequency was 100 Hz, with first two harmonics being represented at 200 and 300 Hz. In addition, there were observed frequencies from spindle runout and harmonics at 50 Hz and 150 Hz. In Fig. 3(d), the unstable cut exhibited chatter and the chatter harmonics centered at 1510 Hz, where its harmonics were spaced apart by the tooth passing frequency. The dominant harmonics are shown at 1410 Hz and 1310 Hz. These chatter frequencies and harmonics were not apparent in the stable cut, though the tooth passing frequency and harmonics have the same mode of excitation.

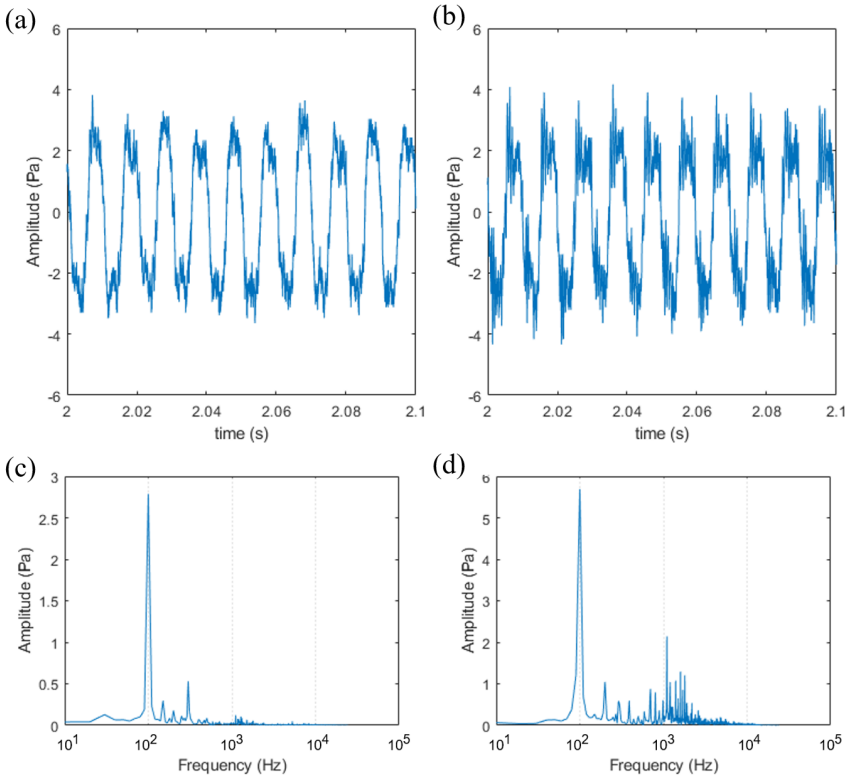

Figure 3. 0.1-second segment for (a) stable cut at 3000 RPM and 0.20in DOC and (b) an unstable cut done at 3000 RPM and 0.25in DOC. Corresponding FFTs for (c) stable and (d) unstable cuts.

To understand the impact of noise on classification of chatter signals, 3 levels of white Gaussian noise (i.e., low, medium, high) and 3 levels of periodic noise (i.e., low, medium, high) were added to the set of stable and unstable audio signals. Figure 4 shows an example of time series signals with the addition of the white Gaussian (W) and periodic (P) noise components, where a label of 1 corresponds to low, 2 to medium and 3 to high for the respective noise components. From the figure, it is clear that the addition of increasing periodic noise has added increasingly visible higher frequency components to the sinusoidal character of the baseline audio signal. Similarly, increasing levels of white Gaussian noise amplifies the relative magnitude of the waveform making the original sinusoidal character of the baseline signal less distinguishable from the background noise. Similarly, Fig. 5 shows an example of FFT plots for the corresponding samples in Fig. 4. As can be seen in the figure, addition of the periodic noise components increases the frequency content especially in the vicinity of the chatter frequencies, with higher amplitudes visible with increasing levels of periodic noise. The entire set of audio signals were modified according to this same approach and the total set of data were used to investigate the impact of white Gaussian noise and periodic noise on model classification performance using threshold-based and ML-based chatter detection approaches.

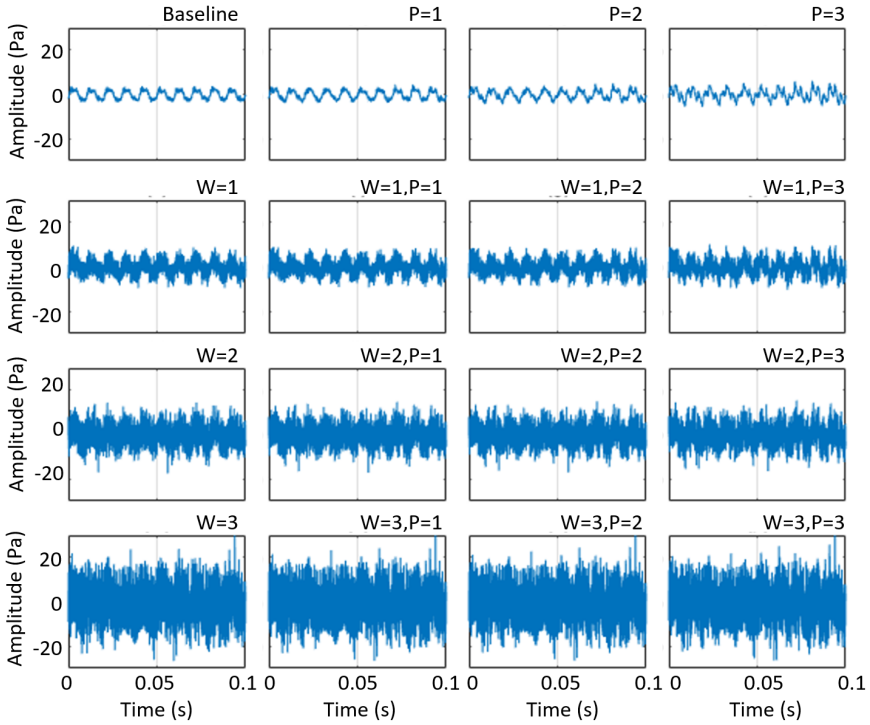

Figure 4. Example of time-series signals for unstable chatter samples, including baseline and increasing noise levels of white Gaussian noise W = {1, 2, 3} and periodic noise P = {1, 2, 3}.

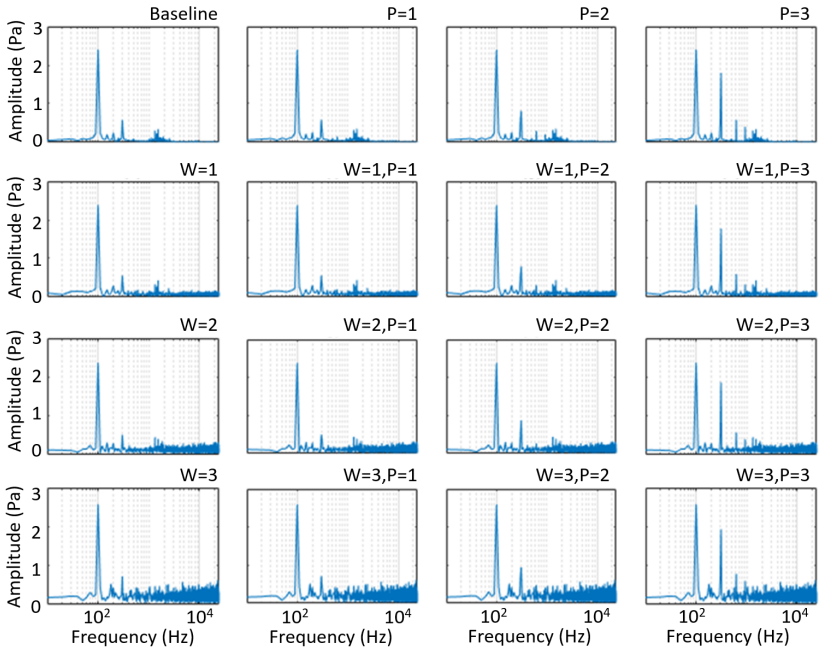

Figure 5. Example of FFT for unstable chatter samples, including baseline and increasing noise levels of white Gaussian noise W = {1, 2, 3} and periodic noise P = {1, 2, 3}.

A threshold-based chatter detection method was tested on the original signals and the augmented signals at threshold amplitude values of 0.05, 0.1, and 0.15 Pa. First, a comb filter was created to filter out the tooth passing frequency, spindle runout frequency, and subsequent harmonics. This comb filter attenuated all chosen frequencies to 0. Any remaining frequencies with amplitudes above the set threshold were classified as chatter. Figure 6 shows an example of the threshold-based chatter detection approach. The main tooth-passing and spindle runout frequency and harmonics were seen at 100, 150 and 200 Hz in Fig. 6(b). In this example, the chatter occurred in the 1000-1500 Hz range, with a max amplitude of 0.6 Pa. Performance of the threshold-based approach for the baseline audio signals is shown in Table 2. As can be seen in the table, a low threshold of 0.05 Pa resulted in significant amounts of false positive (FP) classifications due to a heightened sensitivity of the algorithm, as well as zero false negative (FN) classifications. As the threshold was increased to an ideal setting of 0.10 Pa, this resulted in zero FP and zero FN classifications. Increasing the threshold to 0.15 Pa resulted in lessened sensitivity of the algorithm and 0 FP classifications and 71 FN classifications. In this case, chatter frequencies that may not have fully developed will not be registered because the threshold is higher than their amplitude. A higher threshold makes a system more robust to noise but can lead to many missed detections which can be critical.

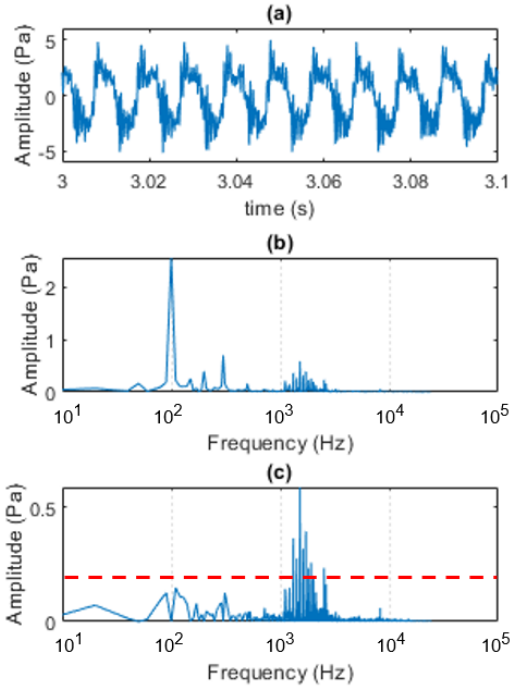

Figure 6. (a) Audio signal for machining at 3000 RPM and 0.25 in. DOC, (b) FFT for this audio signal, and (c) filtered audio signal with a threshold amplitude limit set at 0.2 Pa.

Table 2. Evaluation of threshold method at 3 thresholds while classifying original audio samples

Threshold	TP	TN	FP	FN	Precision	Recall	F1	Accuracy (%)
0.05 Pa	109	224	115	0	0.49	1.00	0.65	73
0.10 Pa	224	224	0	0	1.00	1.00	1.00	100
0.15 Pa	224	153	0	71	1	0.76	0.86	82.4

The impact of white Gaussian and periodic noise on the performance of the threshold-based chatter detection is summarized in Fig. 7, which shows model accuracy performance. In this case, the threshold-based approach was evaluated using the augmented dataset according to each noise level combination. From the figure, when the threshold was set to 0.10 Pa, an accuracy of 100% is obtained on the baseline audio signals. At increasing levels of noise, whether it include Gaussian white noise and/or periodic noise, model accuracy drops to 50%. When the threshold is set to 0.15 Pa, model accuracy on the baseline audio signals is lower than the lower threshold, however it is clear that the higher threshold performs better at a Gaussian white noise level of 1 (low), most likely due to less sensitivity to the increased levels of noise amplitude. Beyond low Gaussian white noise levels, model accuracy converges on the performance in the 0.1 Pa threshold case. It should be noted that all of these selected noise threshold levels classify the audio signals poorly at higher Gaussian white noise and periodic noise levels. Further, a 50% accuracy is due to the methods classifying every sample as chatter. Since the data are evenly distributed between stable signals and chatter signals, the methods have a 50% accuracy and a high false positive count.

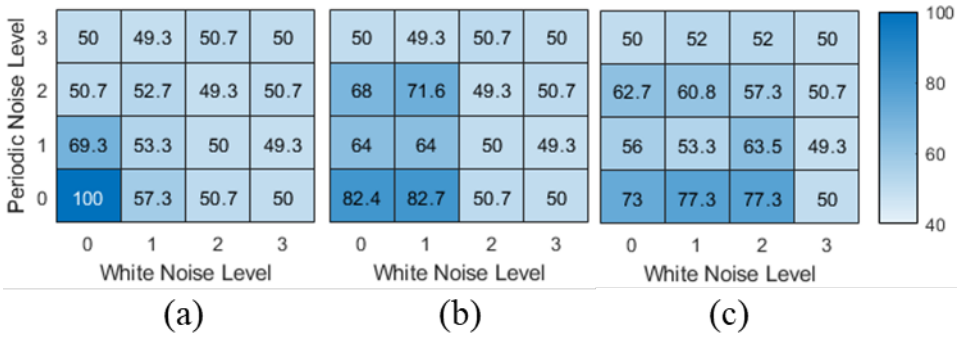

Figure 7. Accuracy of threshold-based chatter detection method in the presence of Gaussian white noise and periodic noise at thresholds of (a) 0.10 Pa, (b) 0.15 Pa and (c) 0.20 Pa.

Features were extracted from the stable and chatter audio signals to train ML classification models. 16 features from both the time and frequency domains were extracted from the segments of audio data. In the time domain, 9 features were extracted and included root mean square (RMS), variance (V), skewness (Sk), kurtosis (Ku), peak value (Pk), crest factor (CF), shape factor (SF), impulse factor (IF), and clearance factor (ClF). The time series signals were transformed into the frequency domain with an FFT and a Hanning window was used on the time series data before the FFT to prevent spectral leakage. 7 features were extracted from the frequency domain and included mean amplitude (M_f), variance (V_f), skewness (Sk_f), kurtosis (Ku_f), peak amplitude (Pk_f), relative peak (RPk), and total harmonic bandpower (HBP). The equations for these features are shown in Table 4 where x_i is the amplitude of a time signal at data point i and s_i is the amplitude of a frequency at i. The calculation of the HBP requires the tooth passing frequency, f_t.

A total of 7,168 samples were used for training and validation of the ML classifiers. The number of samples were equally distributed between stable and chatter conditions. The samples were also evenly distributed between the 16 combinations of Gaussian white noise and periodic noise augmentation. From the extracted features, each sample had 16 features for classification. The samples were split into a training set of 5,974 samples and a testing set of the remaining 1,194 samples. Both sets contained even distributions across chatter conditions, and noise variations. When training the models, a cross-validation was done with 5-folds validation to tune the

parameters. Decision tree, SVM, kNN, and bagged tree models were trained, each with 4 different training sets. Training sets were varied from 50% to 100% size to determine effects of training set size on model performance. After the models were fully trained, the common testing set was used to measure their performance.

The results of the model training are shown in Fig. 8. From the figure, the k-Nearest Neighbors (kNN) approach had the highest accuracy of the 4 models regardless of training set size. The performance of the kNN ranged from 91.8% to 94.1%. The performance of the Bagged Trees and the SVM remained consistent at all training set sizes. kNN, SVM, and Bagged Tree models show capabilities of acceptable accuracy even with limited training. The decision tree model had the lowest overall accuracy at all training set sizes. The decision tree's performance declined as it was exposed to more training data. This is an indicator that the decision tree model is overfitting on the training data, resulting in poor generalization.

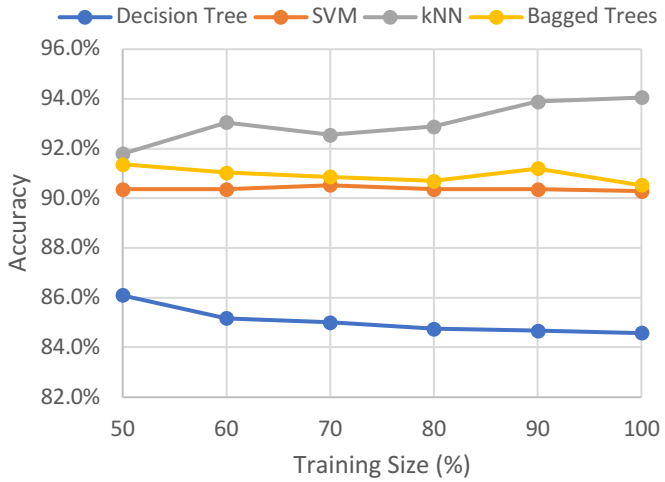

Figure 8. Accuracy of classifiers against the testing set while increasing the number of training data samples from 50% to 100% of all training data.

The performance of the ML classifiers when trained on the entire training dataset and evaluated on specific noise conditions is shown in Table 3, which shows performance of these models against: (1) the baseline signal data, (2) subsets of the testing data according to Gaussian white noise and periodic noise levels, and (3) the entire training dataset. As is evident in the table, the SVM, kNN and bagged trees models outperform the threshold-based approach with regard to the baseline signal data, while the decision tree approach performed worse. In the presence of noise, the performance of the ML classifiers decreased in general, but the performance of the threshold-based approach decreased more substantially in comparison. In terms of specific performance, the kNN model once again had the highest accuracy through all the noise level subsets, as well as the lowest occurrence of false positives. The bagged tree model had the lowest occurrence of false negatives. The SVM has the highest rate of false positives, and the decision tree had the highest rate of false negatives. As noise increased, all models suffered from decreased precision, but the recall rates remained consistent. From noise level 1 to noise level 3, the F1 scores did not change, indicating a balanced performance between false positives and negatives.

Table 3. Evaluation of fully trained ML models compared to threshold technique against subsets of the testing data with specific combinations of Gaussian white noise W = {1, 2, 3} and periodic noise P = {1, 2, 3}, as well as the complete testing set.

Original Signal

Model	TP	TN	FP	FN	Precision	Recall	F1	Accuracy
Threshold = 0.15	24	37	0	13	1.00	0.65	0.79	82.4%
Decision Tree	19	37	0	18	1.00	0.51	0.68	75.7%
SVM	37	37	0	0	1.00	1.00	1.00	100.0%
kNN	37	37	0	0	1.00	1.00	1.00	100.0%
Bagged Trees	36	37	0	1	1.00	0.97	0.99	98.6%

Level 1 (W=1, P=1)

Model	TP	TN	FP	FN	Precision	Recall	F1	Accuracy
Threshold = 0.15	28	20	17	10	0.62	0.74	0.67	64.0%
Decision Tree	27	37	0	11	1.00	0.71	0.83	85.3%
SVM	36	34	3	2	0.92	0.95	0.94	93.3%
kNN	36	35	2	2	0.95	0.95	0.95	94.7%
Bagged Trees	37	36	1	1	0.97	0.97	0.97	97.3%

Level 1 (W=2, P=2)

Model	TP	TN	FP	FN	Precision	Recall	F1	Accuracy
Threshold = 0.15	37	0	38	0	0.49	1.00	0.66	49.3%
Decision Tree	34	29	9	3	0.79	0.92	0.85	84.0%
SVM	32	30	8	5	0.80	0.86	0.83	82.7%
kNN	32	32	6	5	0.84	0.86	0.85	85.3%
Bagged Trees	32	30	8	5	0.80	0.86	0.83	82.7%

Level 1 (W=3, P=3)

Model	TP	TN	FP	FN	Precision	Recall	F1	Accuracy
Threshold = 0.15	37	0	37	0	0.50	1.00	0.67	50.0%
Decision Tree	36	28	9	1	0.80	0.97	0.88	86.5%
SVM	36	31	6	1	0.86	0.97	0.91	90.5%
kNN	33	36	4	1	0.89	0.97	0.93	93.2%
Bagged Trees	36	29	8	1	0.82	0.97	0.89	87.8%

All Data

Model	TP	TN	FP	FN	Precision	Recall	F1	Accuracy
Threshold = 0.15	554	142	455	43	0.55	0.93	0.69	58.3%
Decision Tree	495	515	82	102	0.86	0.83	0.84	84.6%
SVM	567	511	86	30	0.87	0.95	0.91	90.3%
kNN	570	553	44	27	0.93	0.95	0.94	94.1%
Bagged Trees	575	515	82	22	0.88	0.96	0.92	91.3%

To further understand the impact of limited training on model classification performance, the SVM model was trained on subsets of the training data. In this case, the model was trained on data from specific noise level categories and tested against the entire dataset. Table 4 presents the training sets used in this case. For training set 1, data included in model training only included the baseline signal data. For training set 2, data also included low Gaussian white noise data and low periodic noise data. Similarly, sets 3 and 4 included increasing amounts of data for model training. This experimental design enabled exploration of the capability of the SVM classifier to identify stable and chatter audio signals for data not included in training, including those of increasing levels of Gaussian white noise and periodic noise.

Table 4. Training set noise level composition.

Training Set	White Noise Levels	Periodic Noise Levels
1	0	0
2	0, 1	0,1
3	0,1,2	0,1,2
4	0,1,2,3	0,1,2,3

The results of the model testing are summarized in Fig. 9 and can be directly compared against the results of a threshold-based approach as shown earlier in Fig. 7. In Fig. 9(a), the SVM model was trained only on baseline data with no white noise or periodic noise added to the signal. As is clear in the figure, model performance in the presence of noise greatly reduced with increasing levels of noise. Figure 9(b) and Fig. 9(c) show similar model results with increasing levels of training to include low and medium white/periodic noise, respectively. From these tests, model performance increased substantially, however performance for noise categories not included in the training set was notably poorer in the case of high Gaussian white noise levels. The effect of training the SVM model using data encompassing all noise classes is shown in Fig. 9(d). In this case, the model performance ranged from 78.7% to 100% based on the specific Gaussian white noise and periodic noise in the test dataset.

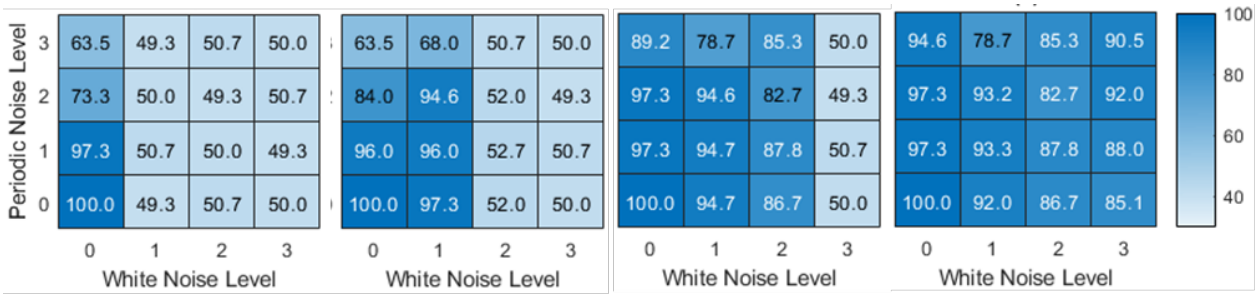

Figure 9. Accuracy of the SVM model on limited training with (a) training set 1, (b) training set 2, (c) training set 3 and (d) training set 4.

Discussion

When compared to the models that have been fully trained or partially trained on noisy data, the threshold method showed considerable weakness at higher Gaussian white noise and periodic noise levels. In Figure 7, the threshold-based method performance drops immediately with the introduction of noise, depending on the threshold selected. This drop in accuracy is the result of an increased number of false positives. In comparison, the SVM performance in Fig. 9 shows higher accuracies when classifying signals with all periodic noise levels despite no exposure to those specific noise levels in training. From Table 8, the F1 score and the precision of the threshold-based method decreased more significantly with increased noise when compared to the ML-based approaches. Overall, the threshold method showed acceptable performance at low noise levels, but setting the appropriate threshold based on the expected noise level will improve performance. Further, the SVM approach, when exposed to limited datasets still exhibited some robustness to classification of noisy data, as in Fig. 9. All methods showed deterioration in performance at the highest noise levels and an increased occurrence of false positives. At higher noise levels, the occurrence of missed detections (false negatives) was lower for all models except for the decision tree classifier.

Overall, all ML methods showed significant performance improvements compared to the threshold-based method, evident in Table 8. ML-based models have the capability to handle a noisier data set in comparison to the threshold-based approach when adequately trained. Across the different chatter detection approaches, several signals failed to be classified accurately by any of the used methods. Two examples of these samples are shown in Figure 10. One case is marginal chatter in Fig. 10(a). In this case, the frequency spectra contributions of the chatter frequency are very weak, and when increasing noise is added, the contributions of the chatter frequency to the extracted features becomes less apparent, as denoted in the figure. This essentially would increase the likelihood of missed detection. The other example signal to note is one frequently yielding a false positive diagnosis in Fig. 10(b). In this case, the false positive is caused likely due to the similarity in the chatter signal detected at a slightly higher axial depth of cut, presented in Fig. 6(a). With the introduction of noise to this audio signal, an increase would be present for the amplitude of the higher frequencies where chatter would occur, as denoted in the figure. Thus, the spectrum would more closely resemble a chatter-type condition.

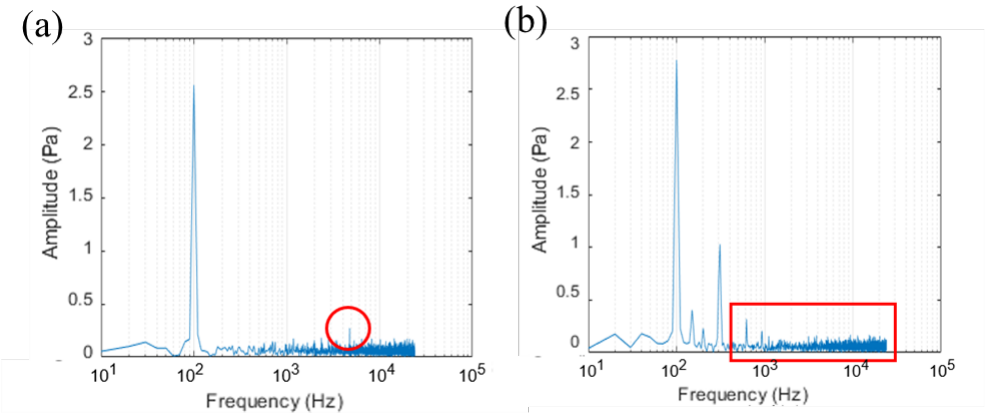

Figure 10. (a) FFT for audio signal at spindle speed of cutting at (a) 6000 RPM and DOC of 0.6 in. and (b) 3000 RPM and DOC of 0.2 in. Annotations highlight potential causes for missed detections in (a) and false positives in (b).

Conclusions

The present study sought to understand the effects of extrinsic noise on chatter classification for machining processes. Threshold-based and ML-based approaches for chatter detection were evaluated against a range of datasets that include Gaussian white noise and periodic noise components. Data augmentation was used to add this diversity to training sets. The results showed that performance of the threshold-based approach was dependent on the threshold parameter selected and generally was poorer when exposed to higher noise levels. The study showed that thresholding techniques had an increased false positive rate with the addition of periodic and white noise. Comparatively, ML-based approaches showed some levels of robustness by maintaining their performance even in the presence of excessive noise. The 4 ML classifiers chosen performed accurately when trained on the noisy data, with accuracies ranging from 84.6% to 98.6% depending on the specific noise level in the audio test signals. Further testing showed that the SVM approach demonstrated ability to classify noise in test sets that it was not exposed to during training. The SVM approach was also able to retain model accuracy on limited training sets. Both of these capabilities demonstrate that ML-based approaches, and specifically SVMs, have a robustness to unexpected noise.

Acknowledgements This work was supported in part by DE-EE0008303 and NSF CMMI. This work was also partially supported by NSF CMMI-1646013, CMMI-1825640 and IIP-1631803.

References

1. Quintana, G., & Ciurana, J. (2011). Chatter in machining processes: A review. International Journal of Machine Tools and Manufacture, 51(5), 363-376.
2. Siddhpura, M., & Paurobally, R. (2012). A review of chatter vibration research in turning. International Journal of Machine tools and manufacture, 61, 27-47.
3. Tangjitsitcharoen, S. (2009). In-process monitoring and detection of chip formation and chatter for CNC turning. Journal of Materials Processing Technology, 209(10), 4682-4688.
4. Delio, T., Tlusty, J., and Smith, S. (May 1, 1992). "Use of Audio Signals for Chatter Detection and Control." ASME. J. Eng. Ind. May 1992; 114(2): 146–157.
5. Schmitz, T. L. (2003). Chatter recognition by a statistical evaluation of the synchronously sampled audio signal. Journal of Sound and Vibration, 262(3), 721-730.
6. Tsai, N. C., Chen, D. C., & Lee, R. M. (2010). Chatter prevention for milling process by acoustic signal feedback. The International Journal of Advanced Manufacturing Technology, 47(9-12), 1013-1021.
7. Thaler, T., Potočnik, P., Bric, I., & Govekar, E. (2014). Chatter detection in band sawing based on discriminant analysis of sound features. Applied acoustics, 77, 114-121.
8. Yoon, M. C., & Chin, D. H. (2005). Cutting force monitoring in the endmilling operation for chatter detection. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 219(6), 455-465.
9. Scheffer, C., & Heyns, P. S. (2004). An industrial tool wear monitoring system for interrupted turning. Mechanical Systems and Signal Processing, 18(5), 1219-1242.
10. Lamraoui, M., Barakat, M., Thomas, M., & Badaoui, M. E. (2015). Chatter detection in milling machines by neural network classification and feature selection. Journal of Vibration and Control, 21(7), 1251-1266.

11. Zhang, C. L., Yue, X., Jiang, Y. T., & Zheng, W. (2010). A hybrid approach of ANN and HMM for cutting chatter monitoring. In Advanced Materials Research (Vol. 97, pp. 3225-3232). Trans Tech Publications Ltd.
12. Tarng, Y. S., and Li, T. C. (1994). "Detection and Suppression of Drilling Chatter." ASME. J. Dyn. Sys., Meas., Control. December 1994; 116(4): 729–734.
13. Liao, Y. S., & Young, Y. C. (1996). A new on-line spindle speed regulation strategy for chatter control. International Journal of Machine Tools and Manufacture, 36(5), 651-660.
14. Tansel, I. N., Li, M., Demetgul, M., Bickraj, K., Kaya, B., & Ozcelik, B. (2012). Detecting chatter and estimating wear from the torque of end milling signals by using Index Based Reasoner (IBR). The International Journal of Advanced Manufacturing Technology, 58(1-4), 109-118.
15. Chen, J. C., & Huang, B. (2003). An in-process neural network-based surface roughness prediction (INN-SRP) system using a dynamometer in end milling operations. The International Journal of Advanced Manufacturing Technology, 21(5), 339-347.
16. Faassen, R. P. H., Doppenberg, E. J. J., van de Wouw, N., Oosterling, J. A. J., & Nijmeijer, H. (2006). Online detection of the onset and occurrence of machine tool chatter in the milling process. In CIRP 2nd International Conference on High Performance Cutting (pp. paper-no).
17. Suprock, C. A., Fussell, B. K., Hassan, R. Z., & Jerard, R. B. (2008, January). A low cost wireless tool tip vibration sensor for milling. In ASME 2008 International Manufacturing Science and Engineering Conference collocated with the 3rd JSME/ASME International Conference on Materials and Processing (pp. 465-474). American Society of Mechanical Engineers Digital Collection.
18. Marinescu, I., & Axinte, D. A. (2008). A critical analysis of effectiveness of acoustic emission signals to detect tool and workpiece malfunctions in milling operations. International Journal of Machine Tools and Manufacture, 48(10), 1148-1160.
19. Smith, S., & Delio, T. (1989, December). Sensor-based control for chatter-free milling by spindle speed selection. In Symposium on Control Issues in Manufacturing (Vol. 18, pp. 107-114).
20. Binsaeid, S., Asfour, S., Cho, S., & Onar, A. (2009). Machine ensemble approach for simultaneous detection of transient and gradual abnormalities in end milling using multisensor fusion. Journal of Materials Processing Technology, 209(10), 4728-4738.
21. Kuljanic, E., Sortino, M., & Totis, G. (2008). Multisensor approaches for chatter detection in milling. Journal of Sound and Vibration, 312(4-5), 672-693.
22. Subrahmanya, N., & Shin, Y. C. (2008). Automated sensor selection and fusion for monitoring and diagnostics of plunge grinding. Journal of manufacturing science and engineering, 130(3).
23. Ismail, F., & Ziaei, R. (2002). Chatter suppression in five-axis machining of flexible parts. International Journal of Machine Tools and Manufacture, 42(1), 115-122.
24. Ismail, F., & Ziaei, R. (2000). Monitoring machining chatter using acoustic intensity. In World Automation Congress Conference, Maui, Hawaii, USA.
25. Guarnaccia, C., Quartieri, J., & Ruggiero, A. (2014). Acoustical noise study of a factory: Indoor and outdoor simulations integration procedure. International Journal of Mechanics, 8(1), 298-306.

26. Yao, Z., Mei, D., & Chen, Z. (2010). On-line chatter detection and identification based on wavelet and support vector machine. Journal of Materials Processing Technology, 210(5), 713-719.
27. Ademujimi, T. T., Brundage, M. P., & Prabhu, V. V. (2017, September). A review of current machine learning techniques used in manufacturing diagnosis. In IFIP International Conference on Advances in Production Management Systems (pp. 407-415). Springer, Cham.