Classification and Regression Models of Audio and Vibration Signals for Machine State Monitoring in Precision Machining Systems

Seulki Han[a], Nasir Mannan[b], Daryl C. Stein[c], Krishna R. Pattipati[d], George M. Bollas[a,*]

[a]Department of Chemical & Biomolecular Engineering, University of Connecticut, 159 Discovery Dr., Storrs, CT, 06269, USA
[b]Connecticut Center for Advanced Technology, Inc. 222 Pitkin St., Suite 101, East Hartford, CT, 06108, USA
[c]Gerber Technology, Inc. 24 Industrial Park Rd., Tolland, CT, 06084, USA
[d]Department of Electrical & Computer Engineering, University of Connecticut, 371 Fairfield Way, Storrs, CT, 06269, USA

Abstract

We present a data-driven method for monitoring machine status in manufacturing processes. Audio and vibration data from precision machining are used for inference in two operating scenarios: a) variable machine health states (anomaly detection); and b) settings of machine operation (state estimation). Audio and vibration signals are first processed through Fast Fourier Transform and Principal Component Analysis to extract transformed and informative features. These features are then used in the training of classification and regression models for machine state monitoring. Specifically, three classifiers (K-nearest neighbors, convolutional neural networks and support vector machines) and two regressors (support vector regression and neural network regression) were explored, in terms of their accuracy in machine state prediction. It is shown that the audio and vibration signals are sufficiently

*Corresponding Author: George M. Bollas, george.bollas@uconn.edu

rich in information about the machine that 100% state classification accuracy could be accomplished. Data fusion was also explored, showing overall superior accuracy of data-driven regression models.

Keywords: Machining, Machine Learning, Signal processing, Fault detection, machine state monitoring

---

1. Introduction

Machine state monitoring is receiving increased attention in smart manufacturing systems because it has the potential to increase productivity and efficiency by minimizing wastage and scrap [1]. Faults in machining, such as tool wear, could affect the durability and dimensional accuracy of the machined part. For instance, the gradual increase in tool wear during machining leads to overall failure of the process and damage to the machine tools in addition to high scrap yields and product quality issues. Mourtzis et al. [2] reported that production downtime in a machining process due to tool change and maintenance makes up to 75% of the total production downtime. Kothuru et al. [3] calculated that maintenance cost of critical components of machine tools accounts for 12% of the total production cost. Thus, monitoring of machine state is vital to improve the efficiency of the entire machining process, which, by extension, leads to a better and predictable production schedule for the entire manufacturing floor [4].

A machine state monitoring system typically involves three stages: (a) sensing; (b) signal processing; and (c) decision making [5]. Sensing can be direct or indirect. For example, tool wear can be measured directly using a visual inspection camera and a laser beam. However, machines are rarely equipped with such sensors [6]. Tool wear can also be inferred indirectly from other measured process variables, such as vibration, audio, cutting power, force and temperature. While physics-based, semi-empirical, and analytical models can be developed for precision machining [7], data-driven models are appropriate for audio and vibration signals. Application of such indirect methods are becoming increasingly common because of their lower cost in

hardware requirements and machine modifications.

After the sensing stage, time-domain and/or frequency-domain signal processing are required to generate useful features from the collected signals. In the time domain, the low-level representation of the noisy signal can distort results, thus restricting the quality of decision-making [3]. In addition, filtering the effects of noise in the time domain involves substantial computational cost, whereas it is straightforward in the frequency domain. Fast Fourier Transform (FFT) is a widely used method to perform feature generation through frequency-domain analysis [3]. Several studies have applied FFT to audio and vibration signals for fault detection in machinery systems. A standard framework involves identification of the frequencies rich in fault information and use of this information in the detection of machine abnormal conditions. Glowacz [8] performed FFT to process acoustic signals gathered from a commutator motor and detected four fault states of the motor from essential frequency components. Atoui et al. [9] detected and diagnosed faults in rotating machinery through Wavelet Transform and FFT of vibration measurements obtained from accelerometer sensors. FFT has also been used to monitor the tool condition of various types of machinery by investigating frequency ranges that correlate with tool wear. Bhuiyan et al. [10] employed FFT on the audio and vibration signatures of turning, and investigated tool wear, chip formation and surface roughness of the workpiece under different machining conditions. Charoenprasit and Seemuang [11] inferred tool wear progression in drilling by using the audio signature processed through FFT and identified a relationship between the frequency of spindle noise and tool wear. Lu et al. [12] applied FFT to the audio signal obtained from a micro-

milling process for micro tool wear monitoring. Jauregui et al. [13] proposed FFT analysis of cutting force and vibration signals for estimating the tool condition of a micro-milling process.

The decision-making stage is the most important in machine state monitoring systems and can be categorized into two classes: model-based and data-driven [6]. Model-based methods require in-depth knowledge of the physical mechanisms by which a fault would affect the sensed information of the machine. In practice, it is very difficult to predict tool wear and other minor machine faults with a physics-based model with high accuracy, because of the difficulty in quantifying the impacts of environmental uncertainty, process noise, and modeling errors on sensed signals. Moreover, our understanding of the physics responsible for machine vibration and sound is very limited. To supplement model-based methods, data-driven methods apply machine learning techniques to predict tool wear and detect machine abnormalities. Recently, deep learning is receiving significant attention in signal processing and machine state monitoring. Painuli et al. [14] used the K-star algorithm (K-nearest neighbor algorithm using an entropy-based distance measure) to predict the healthy and faulty tool states in a CNC machine, using statistical features such as standard deviation, kurtosis, and skewness obtained from vibration signals. Shi and Gindv [15] developed a least squares support vector machine (LS-SVM)-based tool wear prediction model by learning correlations between the actual tool wear and the extracted features (e.g., average cutting force, the length of major axis, and inclination angle of major axis) on a Cincinnati vertical broaching machine. Gonzlez-Muiz et al. [16] proposed a system that uses 1-D deep convolutional

neural network (1D DCNN) to determine seven possible operating conditions for fault detection in a rotating machine. Wang et al. [17] used DCNN in the health monitoring of a gearbox, using features of the time-frequency domain graphical image data learned with different fault types. Li et al. [18] employed CNN to detect tool breakage of CNC machines from signals of spindle current. Ghosh et al. [19] estimated the average flank wear of a face milling machine from four features (cutting forces, spindle vibration, spindle current, and sound pressure), using a feed-forward neural network. Jia et al. [20] built a deep neural network (DNN) model to diagnose faults of rotating machinery from vibration signals. Corne et al. [21] developed neural network models for real-time tool wear prediction during drilling from spindle power data.

There has also been substantial interest in the application of various types of machine learning algorithms to machine state monitoring using audio and vibration signals. Cooper at el. [22] constructed CNN models to predict tool wear in a vertical milling machine using audio signals and compared the performance of 1D-CNN and 2D-CNN in terms of training time and accuracy. Yang and Rai [23] built and compared a back-propagation neural network (BPNN) and a CNN to classify the chatter caused by CNC machine vibrations, along with the machining settings such as speed, feed rate, and depth of cut. Kothuru et al. [24] performed a comparative analysis of SVM and CNN techniques to predict tool wear and the hardness variation of the workpiece, using audio signals generated in a milling machine. Madhusudana et al. [25] used decision trees to extract discrete wavelet transform features and then developed data-driven models using SVM, NN, Nave Bayes, and

K-star for audio signal classification in a face milling machine. Aghazadeh et al. [26] proposed a methodology that uses CNN to classify force and vibration signals under changing cutting parameters, and compared with other machine learning techniques such as SVM, Bayesian networks, and K-nearest-neighbors (KNN). Jedlinski and Jonak [27] evaluated the faults of a gearbox through SVM and multilayer perceptron (MLP), using vibration signals. Jing et al. [28] applied CNN to diagnose fault conditions of gearboxes from vibration signals and compared them with a fully-connected neural network (FNN), SVM and random forest (RF) algorithms.

Although these studies have shown good performance, there is still confusion with regards to the most effective tool chain and machine learning algorithms for machine state monitoring through audio and vibration signals. Most of the studies benchmarked various algorithms for the classification of discrete machine states. However, introducing degradation faults, such as tool wear, or machine settings require methods that can predict continuous variables. Therefore, it is important to explore regression models to predict faults and machining settings modeled as continuous variables. Indeed, regression models are seldom used and the fusion of data from multiple sources is often missing. This study proposes a machine learning-based framework for machine state monitoring in precision machining. The framework is tested on both classification and regression for the inference of machine faults and estimation of machine settings. We generated experimental data of audio and vibration signals for two different subtractive manufacturing processes and explored the performance of the proposed framework using various algorithms for classification and regression. First, signals are converted to the

frequency domain through FFT. Then, Principal Component Analysis (PCA) is used to reduce the number of features in the data by eliminating features of small variance. Finally, we employ various types of machine learning techniques for classification and regression to analyze the transformed features of audio and vibration signals and compare their performance on machine state monitoring in terms of their prediction accuracy.

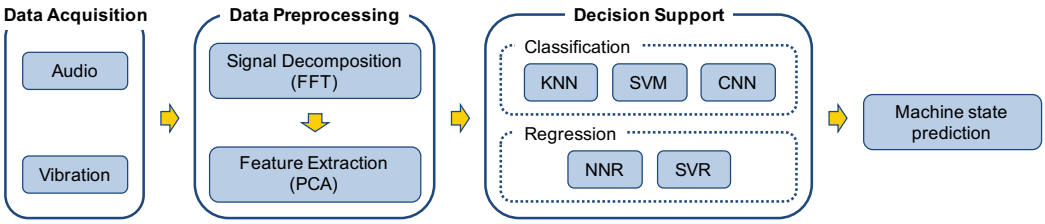

Figure 1: Machine learning-based framework of signal processing for machine state monitoring.

2. Problem Statement and Experimental Setup

The proposed framework for machine state monitoring is illustrated in Fig. 1. To demonstrate the application of the framework, we explored different applications in machine fault detection and machine state monitoring. In the following, we will simply refer to the application to machine fault detection (classification) as case 1 and the application to machine state prediction (regression) as case 2.

2.1. Case 1: Machine Fault Detection

Data from a Gerber Technology Paragon Cutter used in fabric cutting were analyzed to illustrate our approach to machine fault detection. Fabric

cutting in the Paragon Cutter is accomplished with a knife that is periodically sharpened. Sharpening and replacement of the knife are conducted at fixed time intervals and the process is agnostic to the actual wear of the knife. The machine generally deviates from its nominal operating mode due to issues that occur in the cutting head. Common faults of the machine are dull knife, loose knife, loose screws, loose swivel, belt issues, worn Lanchester balancer, to name a few. The main area of focus here was on the components that are related to the knife effectiveness. Each fault in the cutting head has different effects and behavior characteristics depending on the type and amount of material processed. Although the knives are at the same health state when they are first inserted into the cutting head, the inconsistency of materials being cut can cause each blades lifespan to differ substantially from others as time progresses.

During faulty operation, the machine produces a sound that is unique to what is causing the deviation of the machine from its designed operating conditions. We propose using the variability in the sound characteristics of the machine to detect the faulty condition and the consequent identification of degraded operating condition of the cutting head. For this purpose, we collected audio signals using a GRAS General Purpose Array Microphone (50 mV/Pa) from the National Instruments, a DataTranslation DT9837 Module, and QuickDAQ data logging software in very close proximity to the operating machinery. The microphone was attached to the cutting head to read the acoustic signals with good fidelity. Since the purpose our experiment was to classify the anomalies, the audio data was collected from known machine faults. Three faults that would be simple to model and that also happen

during operation were introduced; no knife, loose knife, and no belt. The no-knife scenario was operationalized by completely removing the knife from the Paragon Cutter while the loose-knife was defined as a knife loosened by three turns of the screw holding it in place. The no-belt scenario was operationalized by removing the belt that drives the knife. Finally, a baseline scenario that is representative of the Paragon Cutter in healthy operation (no faults) was included as well.

2.2. Case 2: Machine State Prediction

Data from a Mazak Variaxis 630-5X II T legacy CNC machine was used to predict machine settings based on audio and vibration signals. The Mazak Variaxis 630-5X II T Mill/Turn Machine is a 5-axis milling/turning machine with a tool power of 40 hp. A series of machining tests were performed in this machine at dry conditions (no coolant during milling). The tool used for cutting was a face mill cutter from Kennametal. Face milling was performed on a rectangular block of AISI 4340 steel with a hardness of 25 HRC. Face milling was performed on a rectangular boss in a linear fashion across the length of the piece. Each controlled linear pass of the face mill cutter was treated as an independent test from which machine measurements were collected.

To predict the machine state (its machining settings), we used audio and vibration signals collected from spindle vibration monitors and an audio signal capture device. We used a six-sigma design of experiments to generate audio and vibration signals under various cutting settings, resulting in combinations of cutting parameters that covered the operating envelope of the machine. In this particular case, tests were carried out under healthy con-

ditions (no tool wear). A tool with five teeth and a tool angle of 90° were used in all tests. The cutting parameters employed and the tests that were conducted are summarized in Table 1.

Table 1: Precision machining settings of the tests studied

	Depth of cut (mm)	Width of cut (mm)	Feed rate (mm/min)	Spindle speed (RPM)
Test 1	1.59	31.75	317.5	1299
Test 2	2.54	34.30	200.0	2000
Test 3	2.54	5.00	800.0	1500
Test 4	2.54	63.50	200.0	1500
Test 5	2.54	5.00	500.0	1000
Test 6	2.54	63.50	500.0	1000
Test 7	2.54	34.30	500.0	1500
Test 8	2.54	34.30	800.0	2000
Test 9	2.54	5.00	200.0	1500

3. Methodology

The presented framework was developed in Python and the algorithms used are identified in the following sections. The data used as well as the specific functions used in the machine learning model are provided as Supplementary Information to this manuscript.

3.1. Data acquisition

In the data acquisition stage, the sound and vibration signals from the machine are collected and labeled. Data are collected at specific time inter-

vals depending on the case analyzed. Then, repeatability of the data can be assessed by comparison of the signals at time intervals during which the machine performs identical operations. In case 1, audio data was collected with the cutter operating at 3,500RPM and at 10,000Hz sampling rate. Three fault scenarios and a healthy baseline scenario were recorded with four repeated trials for each scenario. The collected audio signals were labeled by scenarios, including the baseline (healthy operation). In case 2, audio and vibration signals were monitored. The audio signals were extracted at a sampling frequency of 48,000Hz from the videos recorded during each test. Each audio length was pre-processed to include only the actual cutting period, removing the first and last few seconds of machine tool noise. The collected audio and vibration signals were labeled with the cutting parameters (depth of cut, width of cut, feed rate and spindle speed) used to produce them.

3.2. Data Pre-processing

The signal pre-processing stage is necessary to extract the collected select class-sensitive informative features from the signals. In this step, time series data were transformed into the frequency domain using FFT. There exist more advanced methods such as Wavelet analysis and Hilbert-Huang transformation for signal processing. However, the small size of the collected audio and vibration signals, along with their unique and easily identifiable features in the frequency domain, did not necessitate complex signal processing approaches. FFT was shown sufficient to analyze signal data for the determination of machine status and settings. The raw audio and vibration signals collected are time discrete and finite, and therefore, the discrete Fourier transform is applicable. FFT separates the signal into its corre-

sponding spectral components, identifies salient frequencies and the weight each frequency contributes to the original signal energy [29]. We prioritize detection of a fault after a completed manufacturing process (piece by piece, as the manufacturing product being processed would not meet requirements after a manufacturing fault either way), and for that reason we used the entirety of the signal of one manufacturing operation transformed to the frequency domain.

After signal decomposition, a feature selection step aims to reduce the computational cost and increase accuracy by selecting the most informative features. One of the important steps in signal processing was feature selection, in that it was necessary to select features that are specific to machine state or settings. Despite its simplicity, PCA was found satisfactory in providing the most significant features for the cases studied here. In brief, PCA involves an orthogonal transformation that converts a set of features into a new feature space of equal or lower dimension, called principal components [30]. The feature with the largest variance has possibly the largest information content. The principal components are ordered by the amount of variation captured from the original data set and, therefore, the selection of a subset of principal components simplifies the database, reduces the number of dimensions, improves visualization of data, facilitates the explainability of the classification results, and typically increases accuracy in the decision-making step.

3.3. Decision Support

The audio and vibration data transformed through FFT and reduced through PCA were then used to train machine learning algorithms for de-

cision support. This stage aims to assist with the decision of the operator about the state of the machine. This can be in the form of data classification or model prediction through regression. Classification aims to predict a discrete class label, while regression predicts a continuous quantity. As discussed in the introduction, the decision support method depends on the type of machine state we are interested in.

We employed three different algorithms for the audio and vibration signal classification; specifically: KNN, CNN and SVM. KNN is a simple algorithm to classify a data point based on how its known neighbors are labeled in a neighborhood of k-nearest neighbors (k is a hyper-parameter and is typically an odd number) [31]. SVM is another simple algorithm that finds a linear hyperplane in a much higher dimensional space, called kernel space, that when projected into the feature space can provide any nonlinear classification boundary among classes [15]. Among many possible hyperplanes that could separate classes, SVM searches for one that has the maximum margin (i.e., maximum distance between data points of different classes) with minimum error. We used SVM using a linear Kernel, which separates data using a single line. Lastly, CNN are suitable for signal processing because it contextualize opaque patterns in data by detecting features in 1-D (e.g., time series), or 2-D (e.g., image) data. CNN consist of convolution layers to learn the temporal and spectral structures of feature maps, along with pooling layers to reduce the size of the feature maps, while preserving the information that is needed for the model to learn effectively and carry out efficient prediction of a class [22, 23]. We trained the CNN with two blocks of 2 convolutional layers and 1 max pooling layer, followed by a single fully

connected layer. Results of these blocks were flattened and fed into the fully connected layer. The fully connected layer was connected to the final layer, which reduces the feature map to the number of classes for machine fault detection and tool condition assessment. We used rectified linear units (ReLU) as activation functions for the intermediate layers, a softmax function in the output layer, and categorical cross-entropy as the loss function. We employed the accelerated stochastic gradient algorithm Adam, for optimizing the loss function. For the regression models, two different techniques were explored: support vector regression (SVR) and neural network-based regression (NNR). SVR and NNR are regression functions for mapping the input data through a nonlinear function to predict the outputs. [32].

4. Results

Results obtained from each stage in the presented framework are communicated in the following sections in the order of tasks presented in Fig. 1.

4.1. Data-driven model for fault detection

4.1.1. Signal decomposition for case 1

Audio signals collected were first transformed using FFT. We applied FFT to the audio data from all four trials of each operating condition (i.e., baseline, loose knife, no knife and no belt). After FFT, we identified key signal features: the frequencies and amplitudes at which a prominent peak occurs. Prominent peaks were defined as those with amplitude higher than 0.25. The transformed signals at the four operating conditions are presented

in Fig. 2. Note that in Fig. 2, frequencies were shifted so that each trial is offset by 100 Hz to avoid overlap of signals and improve the visual quality of the plot. The solid lines are the signals in the frequency domain and the circle marks represent the key signal features. Eight frequencies in every trial for the four operating conditions were selected as the key features. The eight key frequencies show clear differences in the peak position and its magnitude depending on the type of operation, while the repetitions show a small standard deviation, indicating good data quality. This means that a unique sound signature is generated when the operating conditions deviate from the baseline operation of the Paragon Cutter. The detailed decomposed audio data for case 1 are summarized in Table S.1 of the Supplementary Information.

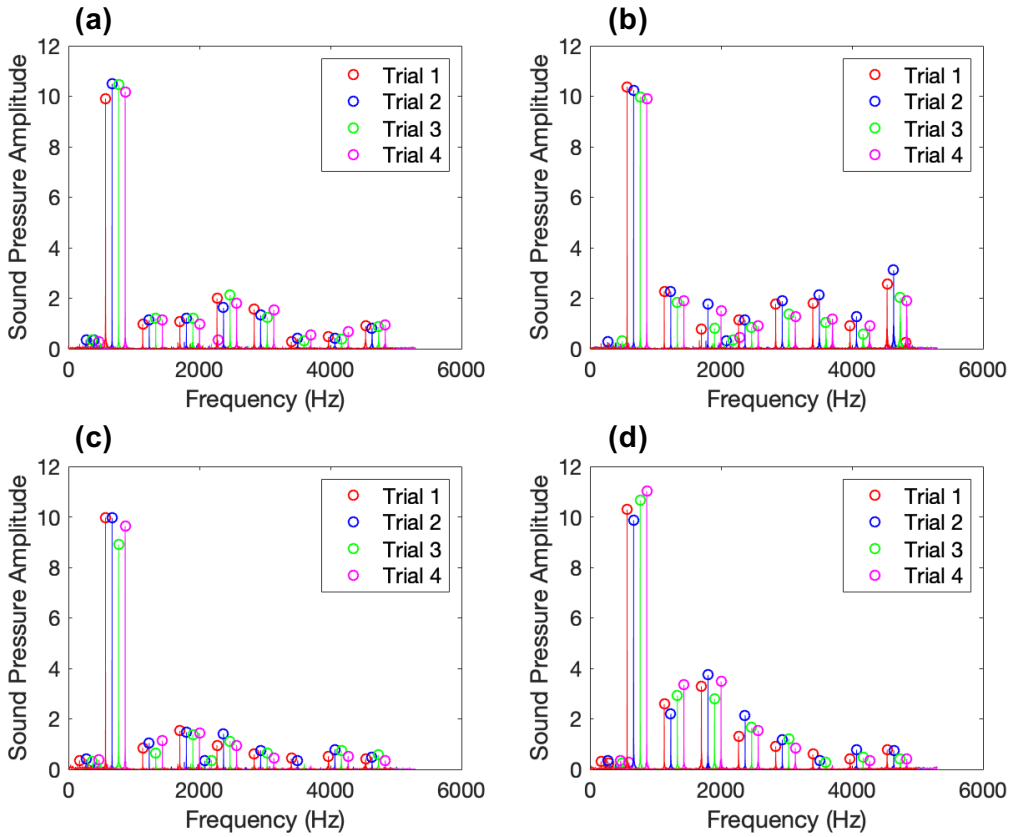

Figure 2: Decomposed signals with mains peaks identified for case 1: (a) baseline, (b) loose knife, (c) no knife, and (d) no belt operation. Frequencies were shifted by 100 Hz to avoid signal overlap and improve the visual quality of the plot.

4.1.2. Data generation for case 1

Despite the lean audio signals of the tests in case 1, we should expect more noisy data from the standard machine operation and, of course, richer data sets. Due to the scarcity of available data in this case, we used the collected audio signals to create new synthetic data points. Specifically, the key frequency amplitudes from FFT were used for synthetic data generation

on the basis of their mean value. We computed a range for new data points in which synthetic data points would lie. Within this range, we produced 75 new random data points. We applied this process to each of the eight key frequencies at the four operating conditions. Thus, a total 300 synthetic data points were produced for machine fault classifier. The average and standard deviations of each operating condition are summarized in Table 2.

Table 2: Mean values and standard deviation of peak amplitudes for the key peaks of the four trials of case 1

	Baseline		Loose Knife		No Knife		No Belt	
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
567Hz	10.26	0.27	10.13	0.22	9.62	0.50	10.46	0.49
1134Hz	1.12	0.09	2.07	0.24	0.92	0.22	2.77	0.49
1701Hz	1.11	0.12	1.21	0.50	1.45	0.06	3.33	0.40
2269Hz	1.88	0.22	1.00	0.15	1.11	0.21	1.66	0.35
2836Hz	1.43	0.16	1.58	0.30	0.61	0.12	1.04	0.18
3403Hz	0.38	0.13	1.55	0.51	0.29	0.13	0.35	0.18
3969Hz	0.49	0.12	0.92	0.29	0.64	0.14	0.50	0.19
4536Hz	0.88	0.05	2.40	0.55	0.46	0.09	0.59	0.20

4.1.3. Principal component analysis for case 1

After signal transformation, the amplitudes of the eight key frequencies were analyzed with PCA. The original matrix has eight columns with rows representing each trial tested (i.e., 16 trials for four different operating conditions) and the amplitudes of the key frequencies populated the columns. Eight principal components were generated as new variables, and a new ma-

trix of eight dimensions that has columns representing the principal components was created. Given that the new matrix has the same dimension with the original matrix, PCA seems have no impact in terms of dimensionality reduction.

Table 3 and Fig. 3 show the variance of each principal component and the clustering of the first three principal components, respectively. In Table 3, the variances are expressed as percentages of total variation. The first three components capture over 90% of the total variation of the original data set. The last three components account for less than 1% of the variance, making them practically irrelevant. Therefore, we can conclude the variation that was originally contained within the eight principal components can be reduced to five dimensions. As shown in Fig. 3, the first three principal components led to clearly separable clusters for each of the four operating conditions. These results indicate that any random point tested can be labeled with high certainty to the right category. To elucidate the significance of feature reduction through PCA in case 1, we compare the classification accuracy using PCA and the original feature set, and discuss results in the next section.

Table 3: Variances obtained from PCA performed for the key features of case 1

	Variance (%)
PC1	45.99
PC2	38.11
PC3	8.41
PC4	5.24
PC5	1.33
PC6	0.47
PC7	0.30
PC8	0.14

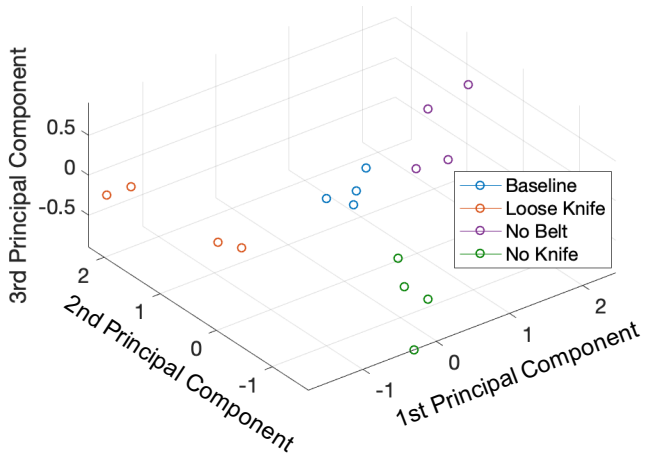

Figure 3: Clustering of the first three principal components of the key features of case 1.

4.1.4. Machine fault classification

Three classifiers (KNN, CNN and SVM) were employed for fault classification, and were evaluated using their resulting confusion matrices, and the average correct classification metric obtained from those (the sum of occurrences in the diagonal over the sum of all occurrences in the confusion matrix). The data were distributed between training and testing sets in the ratio of 80% to 20%, using random shuffling to avoid inductive biases. The goal of the fault classification was to predict unknown labeled faults (i.e., baseline, loose knife, no knife and no belt), and compare the accuracy when using PCA versus when using original features to investigate the effect of PCA. The confusion matrices of the three classifiers using PCA and original features, respectively, are presented in Fig. 4. Classes 1, 2, 3, and 4 represent the scenarios of baseline, loose knife, no knife and no belt, respectively. All classifiers had 100% accuracy regardless of the type of data used between the

PCA and the original feature sets. It might be concluded that PCA is not useful in case 1. However, the main purpose of PCA is to find the subset of features that best capture information in the noisy signal data and reduce dimensions with minimal loss of information. Given the same classification accuracy of using PCA and original feature sets, PCA found the best subset of features without loss of information. It could also be concluded that using features reduced by PCA achieves 100% accuracy regardless of the classification algorithm. However, it should be noted that, when principal components are used, the number of features has less impact to the accuracy of most classifiers. The classification accuracy based on PCA is more relevant to the type of data than to the number of features reduced or the variance captured in the principal components [33]. Consequently, this means that irrespective of the machine learning algorithm, the audio signature was so clear and unique in each of the scenarios tested, that the complexity of the classifier did not matter. The classification results on data sets, including the synthetically generated data and the on-site collected data, suggest that the algorithms will work well as long as new unknown data are within the precision ranges of the original data.

4.2. Data-driven model for machine state prediction

4.2.1. Signal decomposition for case 2

In the case of precision machining, we transformed audio and vibration signals of nine tests under different milling settings using FFT. As in case 1, we identified key signal features (i.e., frequencies and amplitudes of the prominent peak, defined as a frequency with amplitude higher than 0.015 for audio and 0.003 for vibration). Fig. 5 shows the transformed audio and

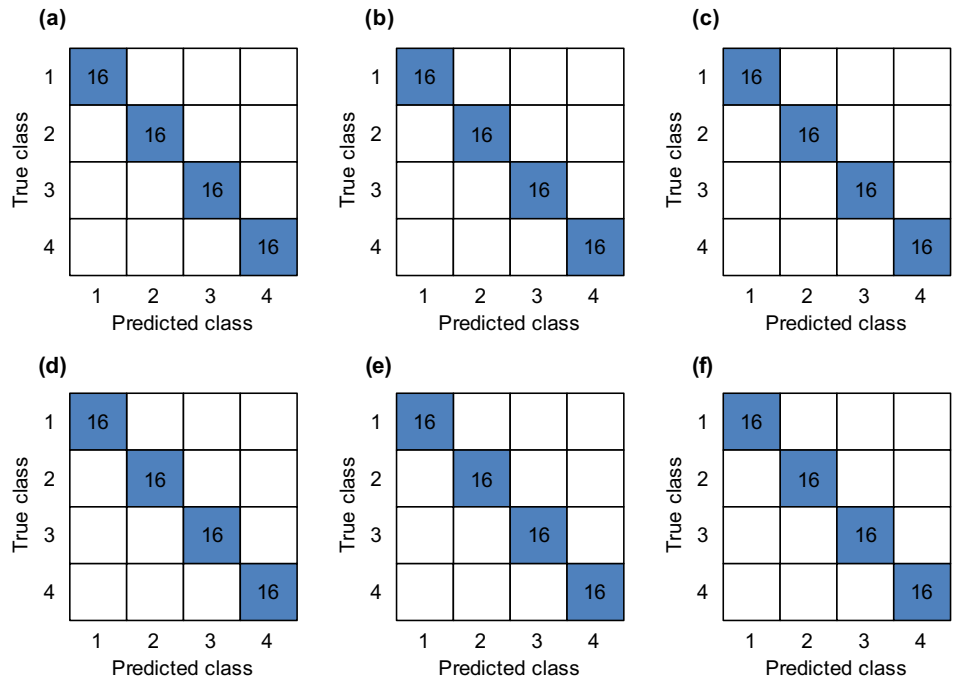

Figure 4: Confusion matrices for machine fault detection for case 1: (a) KNN with original features, (b) CNN with original features, (c) SVM with original features, (d) KNN with PCA, (e) CNN with PCA, and (f) SVM with PCA. Classes 1, 2, 3, and 4 represent the baseline scenario and the three fault scenarios of loose knife, no knife and no belt, respectively.

vibration signals for the nine tests. A total of 57 key audio frequencies and 9 salient vibration frequencies were selected. As shown in Fig. 5(a), the 57 key frequencies selected from the audio signals exhibit significant shifts in peak position and changes in peak amplitudes as a result of different milling parameter settings. In the vibration signals of Fig. 5(b), it can be seen that the 9 key frequencies have considerable change in peak amplitudes despite little shift in peak positions with varying test settings. This means that changes in milling parameter settings result in unique audio and vibration signatures.

The detailed data of the transformed audio and vibration signals of this case study are tabulated in Table S.2 of the Supplementary Information.

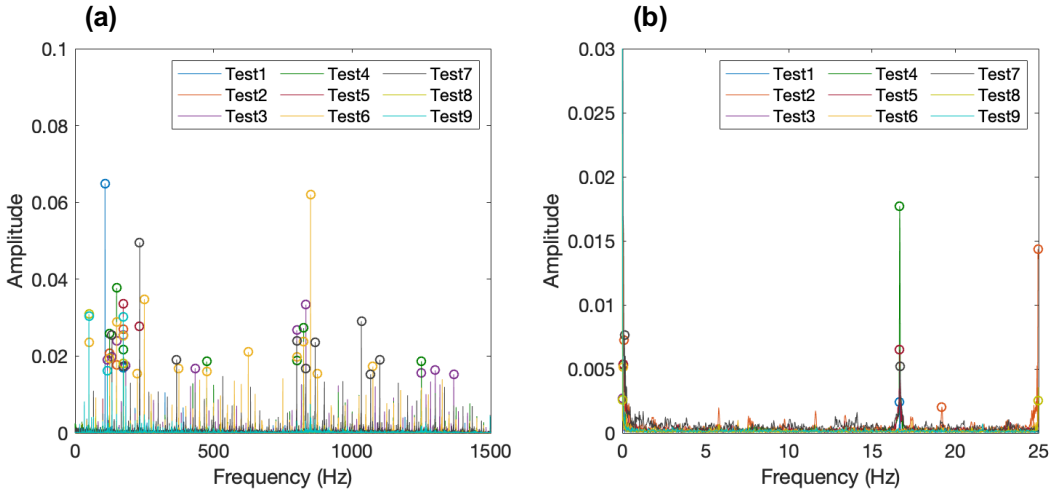

Figure 5: Decomposed signals with mains peaks identified for case 2: (a) audio signals and (b) vibration signals. Frequencies were shifted by 100 Hz to avoid signal overlap and improve the visual quality of the plot.

4.2.2. Data generation for case 2

To enrich the collected data and create noisy responses, we added white noise to the key frequencies in this case. 50 new data points were generated for each of the 66 key frequencies at each milling parameter setting with a standard deviation of 5%. In total, 450 additional data points were produced for machine state prediction.

4.2.3. Principal component analysis for case 2

The amplitudes of the 66 key frequencies obtained by the signal transformation (i.e., 57 frequencies of audio and 9 frequencies of vibration) were

analyzed with PCA. The variance of each principal component and the clustering of the first three principal components are shown in Table 4 and Fig. 6, respectively. The first five components capture over 90% of total variation found in the original 66-dimensional data set. The cluster plot of the first three principal components shows that tests 2, 4 and 9 are easy to classify, while other tests are more difficult to discern. When we enlarge the cluster in the middle of Fig. 6(a), six data points are also simple to classify as shown in Fig. 6(b). This clear clustering of the first three principal components indicates a good separation of the features extracted from for the signals recorded. This highlights the richness of information about machine settings in the audio and vibration signals and are promising indicators of an accurate regression model (shown later). To clarify the necessity for PCA in case 2, results of comparing regression accuracy using PCA and the original features, respectively, will be discussed in the next section.

Table 4: Variances obtained from PCA performed for the key features of case 2

	Variance (%)
PC1	32.63
PC2	24.48
PC3	18.79
PC4	7.91
PC5	6.96
PC6	4.90
PC7	3.48
PC8	0.84

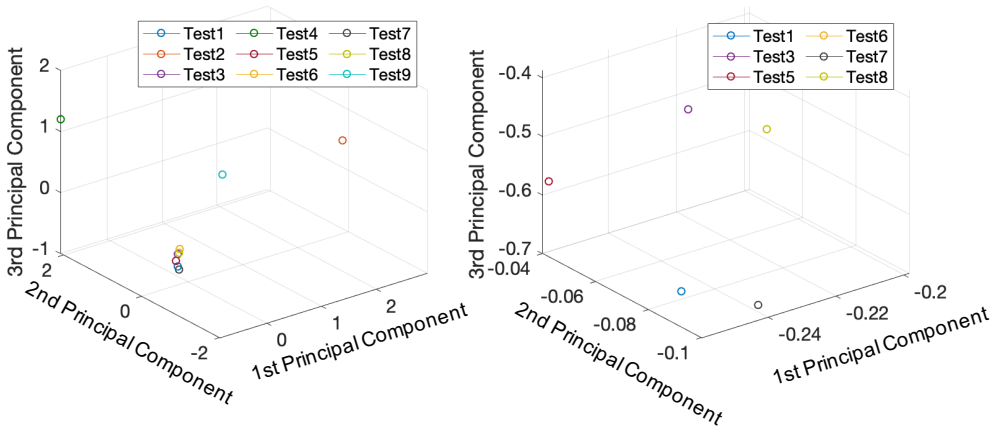

Figure 6: Clustering of the first three principal components of the key features of case 2.

4.2.4. Machine state prediction

Two regressors (SVR and NNR) were used for machine state inference. The root mean squared error (RMSE), shown in Table 5, was used to evaluate the regressor performance. The objective of regression was to predict unknown cutting parameters: depth of cut, width of cut, feed rate and spindle speed. Three scenarios were studied in which each data set of audio and vibration signals and a data set with both signals were used as inputs. For each scenario, regression accuracy using PCA and that using the original features were compared. The RMSE of regression using PCA was lower than that using the original features, regardless of the type of model and scenario. The RMSE of NNR was lower than that of SVR for all regression tasks. The RMSEs of vibration signal models were lower than that of audio signal models. In summary: i) regression accuracy achieved with PCA was significantly better than that with the original feature sets, ii) NNR showed better performance than SVR in estimating the unknown continuous cutting

settings using either audio or vibration signals as inputs; and iii) the models using vibration signals were more accurate. The use of fused audio and vibration signals as inputs further improved the prediction accuracy, indicating different information content in the two signals. The entirety of data, signal processing and regression algorithms and execution guides can be found in the Supplementary Information of this document.

Table 5: Root mean squared error of SVR and NNR models, trained with the original features and those reduced through PCA from the audio, vibration data and fused audio and vibration data

		Audio		Vibration		Audio+Vibration	
		SVR	NNR	SVR	NNR	SVR	NNR
Original features	Depth of cut	0.021	0.014	0.007	0.006	0.005	0.004
	Width of cut	0.023	0.012	0.006	0.004	0.005	0.003
	Feed rate	0.053	0.039	0.007	0.006	0.006	0.005
	Spindle speed	0.019	0.011	0.005	0.004	0.004	0.002
PCA	Depth of cut	0.0070	0.0049	0.0046	0.0031	0.0027	0.0019
	Width of cut	0.0094	0.0042	0.0043	0.0029	0.0017	0.0012
	Feed rate	0.0093	0.0059	0.0029	0.0024	0.0016	0.0013
	Spindle speed	0.0099	0.0035	0.0037	0.0028	0.0020	0.0010

5. Conclusions

This study proposed a machine-learning framework, coupled with signal processing, for machine state monitoring. We applied the framework for two purposes: machine fault detection and machine state prediction. First, we

collected labeled data of audio and vibration signals for the two applications. Then, we analyzed the collected signals through FFT and PCA to extract information from the signals as features. Finally, we performed classification and regression tasks using different machine learning algorithms. Consequently, it was demonstrated that machine faults and cutting settings can be recognized from the audio and vibration signals. In case 1, three machine learning classifiers, KNN, CNN, and SVM, were used to enable audio-based detection of machine faults. By comparing the results of the three machine learning models, it was concluded that all models correctly classified the four operating conditions from audio signals. In case 2, two regressors (SVR and NNR) were employed to predict the continuous cutting parameters from the audio and vibration signals. In these regression tasks, data reduction through PCA was shown to be superior to the original features. NNR had better performance than SVR in terms of RMSE, and the prediction was more accurate through vibration signals instead of audio signals. In short, the findings of this paper demonstrate that audio and vibration signals are rich in information about the machine state, and machine learning-based approaches can be beneficially employed to machine state monitoring. Despite the successes of the presented machine learning models, the number of signal data and precision of machine parameter determination for practical applications are limited. In the future, the much greater number of signal data and different types of input signal images, such as short-time Fourier Transform spectrogram or Mel frequency cepstral coefficients image, can be used to train machine learning models. In addition to machine learning models used in this study, other models such as a recurrent neural network and

least squared support vector machine can be used for processing signals in smart manufacturing. Finally, given that the method presented here depends on particularities of a machining process (e.g., machine, tooling and work piece), analysis of signals from different machines is underway to explore potential commonalities between the extracted features to provide generalized inference of machine state or machine settings for specific manufacturing operations.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0007613.

References

[1] J. Wang, Y. Ma, L. Zhang, R. X. Gao, D. Wu, Deep learning for smart manufacturing: Methods and applications, Journal of Manufacturing Systems 48 (2018) 144–156.

[2] D. Mourtzis, E. Vlachou, N. Milas, N. Xanthopoulos, A cloud-based approach for maintenance of machine tools and equipment based on shop-floor monitoring, Procedia CIRP 41 (2016) 665–660.

[3] A. Kothuru, S. P. Nooka, R. Liu, Application of deep visualization in cnn-based tool condition monitoring for end milling, Procedia Manufacturing 34 (2019) 995–1004.

[4] J. Abellan-Nebot, F. Romero Subiron, A review of machining monitoring systems based on artificial intelligence process models, The International Journal of Advanced Manufacturing Technology 47 (2010) 237–257.

[5] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, Advanced monitoring of machining operations, CIRP Annals - Manufacturing Technology 59 (2010) 717–739.

[6] Z. Li, R. Liu, D. Wu, Data-driven smart manufacturing: Tool wear monitoring with audio signals and machine learning, Journal of Manufacturing Processes 48 (2019) 66–76.

[7] U. Awasthi, K. A. Palmer, G. M. Bollas, Optimal test and sensor selection for active fault diagnosis using integer programming, Journal of Process Control 92 (2020) 202–211.

[8] A. Glowacz, Acoustic-based fault diagnosis of commutator motor, Electronics 7 (2018) 299–320.

[9] I. Atoui, H. Meradi, R. Boulkroune, R. Saidi, A. Grid, Fault detection and diagnosis in rotating machinery by vibration monitoring using fft and wavelet techniques, 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) (2013) 401–406.

[10] M. S. H. Bhuiyan, I. A. Choudhury, M. Dahari, Monitoring the tool wear, surface roughness and chip formation occurrences using multiple sensors in turning, Journal of Manufacturing Systems 33 (2014) 476–487.

[11] S. Charoenprasit, N. Seemuang, Monitoring tool wear in drilling process using spindle noise features, International Journal of Mechanical Engineering and Robotics Research 7 (2018) 564–568.

[12] M. C. Lu, B. S. Wan, Study of high-frequency sound signals for tool wear monitoring in micromilling, International Journal of Advanced Manufacturing Technology 66 (2012) 9–12.

[13] J. C. Juregui, J. R. Resndiz, S. Thenozhi, T. Szalay, . Jacs, M. Takcs, Frequency and time-frequency analysis of cutting force and vibration signals for tool condition monitoring, IEEE Access 6 (2018) 6400–6410.

[14] S. Painuli, M. Elangovan, V. Sugumaran, Tool condition monitoring using k-star algorithm, Expert systems with Applications 41 (2014) 2638–2643.

[15] D. Shi, N. N. Gindy, Tool wear predictive model based on least squares support vector machines, Mechanical Systems and Signal Processing 21 (2007) 1799–1814.

[16] A. Gonzlez-Muiz, I. Daz, A. A. Cuadrado, Dcnn for condition monitoring and fault detection in rotating machines and its contribution to the understanding of machine nature, Heliyon 6 (2020).

[17] P. Wang, Ananya, R. Yan, R. X. Gao, Virtualization and deep recognition for system fault classification, Journal of Manufacturing Systems 44 (2017) 310–316.

[18] G. Li, X. Yang, D. Chen, A. Song, Y. Fang, J. Zhou, Tool breakage detection using deep learning, 2018 IEEE International Conference on Big Data (2018).

[19] N. Ghosh, Y. B. Ravi, A. Patra, S. Mukhopadhyay, S. Paul, A. R. Mohanty, A. B. Chattopadhyay, Estimation of tool wear during cnc milling using neural network-based sensor fusion, Mechanical Systems and Signal Processing 21 (2007) 466–479.

[20] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mechanical Systems and Signal Processing 72-73 (2016) 303–315.

[21] R. Corne, C. Nath, M. El Mansori, T. Kurfess, Study of spindle power data with neural network for predicting real-time tool wear/breakage

during inconel drilling, Journal of Manufacturing Systems 43 (2017) 287–295.

[22] C. Cooper, P. Wang, J. Zhang, R. X. Gao, Convolutional neural network-based tool condition monitoring in vertical milling operations using acoustic signals, Procedia Manufacturing 49 (2020) 105111.

[23] R. Y. Yang, R. Rai, Machine auscultation: enabling machine diagnostics using convolutional neural networks and large-scale machine audio data, Advances in Manufacturing volume 7 (2019) 174187.

[24] A. Kothuru, S. P. Nooka, R. Liu, Audio-based tool condition monitoring in milling of the workpiece material with the hardness variation using support vector machines and convolutional neural networks, Journal of Manufacturing Science and Engineering 140 (2018) 111006–111015.

[25] C. K. Madhusudana, H. Kumar, S. Narendranath, Face milling tool condition monitoring using sound signal, International Journal of System Assurance Engineering and Management 8 (2017) 1643–1653.

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

[27] L. Jedliski, J. Jonak, Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform, Applied Soft Computing 30 (2015) 636–641.

[28] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.

[29] A. John, M. A. Putman, Signal Processing Techniques, EEG Info, 2007. URL: http://www.eeginfo.com/research/researchpapers/Signal_Processing.pdf.

[30] P. Jahnke, Machine learning approaches for failure type detection and predictive maintenance, Thesis, TU Darmstadt, 2015. URL: https://www.ke.tu-darmstadt.de/lehre/arbeiten/master/2015/Jahnke_Patrick.pdf.

[31] P. Junior, D. M. D'Addona, P. Aguiar, R. Teti, Dressing tool condition monitoring through impedance-based sensors: Part 2 - neural networks and k-nearest neighbor classifier approach, Sensors 18 (2018) 4453–4474.

[32] M. Awad, R. Khanna, Support Vector Regression, Apress, Berkeley, CA, 2015. doi:10.1007/978-1-4302-5990-9_4.

[33] A. Janecek, W. Gansterer, M. Demel, G. Ecker, On the Relationship Between Feature Selection and Classification Accuracy, in: Proceedings of the Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery at ECML/PKDD 2008, volume 4 of *Proceedings of Machine Learning Research*, PMLR, 2008, pp. 90–105. URL: http://proceedings.mlr.press/v4/janecek08a.html.