Identifying optimal features for cutting tool condition monitoring using recurrent neural networks

Wennian Yu¹, Chris Mechefske² and Il Yong Kim²

Abstract
Identification of optimal features is necessary for the decision-making models such as the artificial neural network to achieve effective and robust on-line monitoring of cutting tool condition. Most feature selection strategies proposed in the literature are for pattern recognition or classification problems, and not suitable for prognostic problems. This paper applies three parameter suitability metrics introduced in previous similar studies for failure-time analysis and modifies them for failure-process analysis which allows for the unit-wise variation of the component in a population. The suitability of a feature used for cutting tool condition monitoring is determined by its fitness value calculated based on the three metrics. Two types of recurrent neural network are employed to analyze the prognostics ability of the features extracted from multi-sensor signals (acoustics emission, motor current, and vibration) collected from a milling machine under various operating conditions. The analysis results validate that the fitness value of a feature can depict its prognostic ability. It is found that adding more features which share abundant information does not increase the prediction performance but increases the burden on the decision-marking models. In addition, adding features with low fitness values may even deteriorate the prediction.

Keywords
Cutting tool wear, sensor fusion, features, recurrent neural network, prognostics

Date received: 27 August 2020; accepted: 4 December 2020
Handling Editor: James Baldwin

Introduction
On-line cutting tool wear monitoring plays a critical role in industry automation and has the potential to significantly increase productivity and improve product quality. In the literature, a significant amount of research work has been performed dealing with machine tool condition monitoring (TCM).¹–³ A variety of sensors can be employed to capture tool wear condition. Acoustic emission (AE), force, vibration, temperature and motor current sensors are the most commonly used sensors reported in the literature.⁴,⁵ However, different types of signals may respond differently to the tool condition during different machining applications (milling, turning, grinding, etc.).⁵ Thus, the so-called sensor fusion is becoming a prevailing concept in the sensor-based tool condition monitoring.

Although sensor signals contain sufficient tool condition information throughout a machining process, they are convoluted with noise from the cutting operation. Therefore, feature extraction becomes necessary in order to maximize the information utilization of sensor signals. A variety of typical monitoring indices can be obtained from the time-domain and frequency-domain

¹College of Mechanical Engineering, Chongqing University, Chongqing, China
²Department of Mechanical and Materials Engineering, Queen’s University, Kingston, ON, Canada

Corresponding author:
Wennian Yu, College of Mechanical Engineering, Chongqing University, 174 Shazheng Street, Chongqing 400044, China.
Email: wennian.yu@cqu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
features. Advanced signal processing techniques such as mel-frequency cepstral coefficient and S-transform have also been employed by some researchers to extract features that represent the characteristics of the cutting process (information) and to separate these features from various noise disturbances. However, features extracted from an independent data source are not equally informative: certain features may correspond to noise, not information; others may be correlated or not relevant to the target. Therefore, the successful selection of effective features describing the characteristics of the cutting process is a significant task for an accurate and robust on-line system for monitoring of tool conditions.

Compared with other commonly used data-driven prognostic methods, such as the support vector machine, Bayesian approaches, neuro-fuzzy techniques and stochastic degradation model, the artificial neural networks (ANNs) are the most popular data-driven models that are widely used as the decision-making systems for the cutting tool condition monitoring. The feasibility of using neural networks to integrate information from multiple sensors to monitor tool conditions in machining has been demonstrated by many studies in the literature. Optimization algorithms such as the genetic algorithm and the particle swarm optimization algorithm were used to optimize the neural network model for better predictions of the grinding parameters in some works. Since 2006, deep learning has become a rapidly growing research, and is adopted by many researchers as a bridge connecting multi-sensor time-series data and intelligent machine health monitoring. For instance, Aghazadeh et al. employed the convolutional neural network as a well-established deep learning algorithm for tool wear estimation. Yu et al. proposed a bidirectional recurrent neural network (BiRNN) based encoder-decoder scheme to obtain an unsupervised health index from the multi-sensor time-series data collected from a milling machine, based on which the degradation path of the cutting tool was established. Recently, Agarwal and Desai proposed a hybrid model that integrates the merits of a physics-based model and a data-driven model for the estimation of cutting forces during the end milling operation. The physics-based model establishes an explicit relationship between the process variables and attributes based on scientific knowledge whereas the data-driven model was used to learn their relationship which can be easily scaled to accommodate new variables and attributes without conducting numerous experiments.

It is well accepted that the prediction performance of a decision-making system is closely related to the features fed into it. Thus, identification of optimal features for an effective and robust on-line monitoring of tool wear is necessary.

The literature reports abundant feature selection strategies including partial least squares (PLS) method, euclidean distance evaluation technique, principle component analysis, and dominant feature identification. Basically, the key idea behind these methods is to find a mapping from the original higher dimensional feature space to a lower dimensional feature space. However, these methods are mainly for pattern recognition and general classification problems. For prognostic problems, the above methods become inapplicable as the features extracted inherently are not obtained under a fixed pattern or condition. Traditionally, identification of the optimal prognostic features (or monitoring parameters) is accomplished through visual inspection of the available information and engineering judgment. However, it is usually time-consuming and ineffective. Automated selection of the optimal features is therefore required. Some researchers have proposed several metrics to characterize the suitability of a feature used for prognostic problems. Scheffer and Heyns introduced two methods to select the most reliable features that correlate best with the cutting tool condition, which are based on the correlation between the feature and the objective, and the statistical overlap factor (SOF). Javed et al. proposed the cumulative features that have high monotonicity and trendability with the degradation process of the component, which were proved to surely enhance prognostics modeling. Guo et al. used the monotonicity and correlation metrics to select the most sensitive features, which were subsequently fed into a recurrent neural network (RNN) for the RUL prediction of bearings. Coble and Hines introduced three parameter suitability metrics to identify an ideal prognostic feature: monotonicity, prognosability, and trendability. A big step forward in this research is that the feature selection strategies were moved from failure-time analysis to failure-process analysis, which allows for the unit-wise variations in a population.

From the above literature review, most of the widely used feature selection strategies were aimed at pattern recognition and classification problems such as machine fault detection or diagnosis. For prognostic problems such as continually monitoring the degradation process of a system, or the prediction of remaining useful life (RUL) of a component, these strategies become inapplicable as the features extracted inherently are not obtained under a fixed pattern or condition. Although there were a few attempts in assessing the suitability of the monitoring features for prognostics of a component, different researchers preferred different parameter suitability metrics for different monitoring applications. Plus, most of the suitability metrics correspond to the failure-time analysis, which uses the failure times recorded during normal operation to estimate the RUL for a population of identical components. However, it
is more often that individuals from a population of identical components deviates slightly in the normal degradation path over time due to unit-wise variations.\textsuperscript{24,26} Thus, parameter suitability metrics based on the failure-process analysis using degradation measures are more practical and reliable to assess the prognostic ability of a feature.

This study proposed a method to directly calculate the fitness of a feature which assesses the prognostic ability of the feature based on three suitability metrics. The proposed method allows for the effect of unit-wise variations of the monitoring system because the three metrics are calculated based on degradation measures (e.g. tool wear, crack size) instead of degradation time. Two types of recurrent neural network are employed to analyze the prognostics ability of the features extracted from multi-sensor signals (acoustics emission, motor current, and vibration) collected from a milling machine under various operating conditions. The analysis results validate that the fitness value of a feature can depict its prognostic ability.

This paper is organized as follows: Section 1 reviews previous similar studies regarding the sensor-based tool condition monitoring system. Limitations of previous feature selection methods are pointed out, from which the objective of the current research is stressed. Section 2 presents three modified parameter suitability metrics based on the failure-process analysis, from which the fitness of a feature can be determined. The two types of recurrent neural network, which are used as the decision-making models to predict the tool wear in this study are also illustrated in this section. Section 3 first describes an experimental milling data set, based on which the fitness values of 10 features extracted from five sensors (AE, vibration, and current) are determined. After that, the RNNs are employed to predict tool wear based on the selected features. The prediction error distribution of each feature set is expressed by using the boxplot in order to holistically compare the prediction performance among different feature sets fed into the RNNs. The main conclusions of this study are provided in the last section.

**Methodology**

**Fitness of a feature**

Identification of appropriate prognostic features that are provided as inputs to decision-making models is key for reliable on-line condition monitoring of tool wear. There are several criteria that can be used to select the most reliable features which correlate best with tool condition.

**Correlation coefficient COR.** When using the failure-time analysis (usually for long-term degradation of a component such as bearings), the operating time of the individual component is used as the degradation measure. The linear correlation between the feature $q$ and operating time $t$ is calculated by:\textsuperscript{27}

$$\text{Cor} = \frac{\sum_{i=1}^{n} (q_i - \bar{q})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^{n} (q_i - \bar{q})^2 \sum_{i=1}^{n} (t_i - \bar{t})^2}}$$

where $q_i, t_i$ are the values of the feature and the corresponding operating time of the $i$th observed sample, respectively. $\bar{q}, \bar{t}$ are the average values of $q$ and $t$. $n$ is the number of samples.

When using the degradation measure (usually for short-term degradation such as cutting tool wear), the amount of tool flank wear ($VB$) in the cutting tool condition monitoring application, the correlation between the feature $q$ and the objective $\text{VB}$ (the amount of tool wear) can be calculated according to:

$$\text{COR} = \frac{\sum_{i=1}^{n} (q_i - \bar{q})(VB_i - \bar{VB})}{\sqrt{\sum_{i=1}^{n} (q_i - \bar{q})^2 \sum_{i=1}^{n} (VB_i - \bar{VB})^2}}$$

where $VB_i$ is the amount of tool wear of the $i$th observation sample. $\bar{VB}$ is the average value of $VB$. COR is the correlation coefficient that represents the linear relationship between $q$ and $\text{VB}$. The higher the value of COR, the stronger the linear relation between $q$ and $\text{VB}$. In tool condition monitoring, it is expected that the monitoring feature shows a clear relationship (either increasing or decreasing trendability) with the increase of tool wear. Therefore, COR can be used as a metric to select reliable features that are sensitive to the tool condition.

**Monotonicity coefficient MON.** The monotonicity characterizes the underlying increasing or decreasing trend of a time series with the gradual increase of the time values. This metric has been used to select the features indicating gradual degradation (usually long-term) of mechanical components, such as bearings and gears, as it is generally assumed that the monitoring components do not undergo self-healing.\textsuperscript{24,26,27} When using the failure-time analysis, the monotonicity metric evaluates the increasing or decreasing trend of the feature $q$ with the operating time as expressed by:

$$\text{Mon} = \frac{|NO(dq > 0) - NO(dq < 0)|}{n - 1}$$

where $dq$ is the differential of feature series $q$: $NO (dq > 0)$ and $NO (dq < 0)$ represent the numbers of $dq$ that are positive and negative, respectively.
When using the degradation measure, the monotonicity between the feature \( q \) and the objective \( VB \) (the amount of tool wear) can be calculated as:

\[
MON = \left| \frac{\text{NO}(dq \ast dVB > 0)}{n - 1} - \frac{\text{NO}(dq \ast dVB < 0)}{n - 1} \right|
\]  

(4)

where \( dVB \) is the differential of the objective \( VB \). The monotonicity between the feature \( q \) and the objective \( VB \) indicates the consistency of the variations of the feature (increasing or decreasing) with the variations of the objective. The higher the value of \( MON \), the more consistency of the variations of the feature with the variations of the objective.

**Statistical overlap factor SOF.** The statistical overlap factor (SOF) is a metric which can determine the degrees of separation of a feature \( q \) under different operating conditions,\(^{25}\) for instance, the cutting speed \( S \), depth of cut \( D \), and feed \( F \) in the cutting tool condition monitoring application. Ideally, a good feature should show a high degree of separation under different operating conditions, and a low degree of separation due to noise. The SOF of a feature \( q \) obtained from two different operating conditions (condition \( i \) and condition \( j \)) can be calculated as:

\[
\text{SOF}_{ij} = \left| \frac{\bar{q}_i - \bar{q}_j}{(\delta_i + \delta_j)/\lambda} \right|
\]  

(5)

where \( \bar{q}_i \) and \( \bar{q}_j \) are the average values of the feature series \( q \) collected from condition \( i \) and condition \( j \), respectively; \( \delta_i \) and \( \delta_j \) are the standard deviations of the feature series \( q \) collected from condition \( i \) and condition \( j \), respectively; \( \lambda \) is a coefficient whose value can be adjusted to make the value of \( \text{SOF}_{ij} \) fall between 0 and 1. The SOF of a feature \( q \) obtained from various operating conditions (operating condition set \( \Omega \)) is the mean value of \( \text{SOF}_{ij} \) of all possible combinations of two different operating conditions:

\[
\text{SOF} = \text{mean}(\text{SOF}_{ij}), \forall i,j \in \Omega
\]  

(6)

The SOF of a feature \( q \) calculated by equation (6) indicates its ability to distinguish among different operating conditions. A higher value of SOF represents that the feature shows a stronger variation due to the change of operating conditions rather than the noise.

**Fitness FIT.** The suitableness of a feature used for cutting tool condition monitoring can be defined as a weighted sum of the above three mentioned metrics\(^{24,27}\):

\[
FIT = a \times \text{COR} + b \times \text{MON} + c \times \text{SOF}
\]  

(7)

where \( a, b, c \) are the weights of the correlation coefficient \( \text{COR} \), the monotonicity coefficient \( \text{MON} \) and the statistical overlap factor \( \text{SOF} \), respectively. In addition, the sum of their values should be 1. Obviously, the values of these weights determine the contributions of the three metrics on the final feature selection criteria \( \text{FIT} \). The exact weight coefficients for the three metrics are hard to theoretically determined. A good monitoring feature should not only have good correlation and monotonicity with the degradation measure, but also a stronger variation due to the change of operating conditions rather than noise. Assigning same weight coefficients to them seems a reasonable idea.\(^{27}\) However, unequal weights can also be assigned if a metric needs more consideration depending on the specific application.\(^{24}\)

**Decision-making models**

In this study, in order to demonstrate that the proposed methodology is not model-specific, we used two types of recurrent neural network as the decision-making models to predict the tool wear: Elman network and long short-term memory (LSTM). Compared with feedforward neural networks, RNNs have loops in them, which allows information to be passed from a previous time step of the unit to the next as shown in Figure 1. Therefore, unlike feedforward neural networks (FFNN), RNNs possess an internal memory that can make use of past information, which makes it especially suitable for processing sequential data or time series data. In fact, the FFNNs are normally used for pattern recognition or classification problems especially when the input data do not possess explicit time dependencies, whereas the RNNs are more powerful in dealing regression modeling or prognostic problems as the degradation of a mechanical component is usually a time-dependent, continuous, monotonical phenomenon.\(^{20,29}\)

An Elman network is typically a three-layer network (input, hidden, and output). The outputs from the hidden layer in the previous time step are also fed into the input layer in the current time step. It is usually trained using classic backpropagation through time.\(^{27}\) However, traditional RNNs are only capable of handling time series (sequences) with short-term...
dependencies (short time span between the point with the relevant information and the point to be predicted) due to the problem of gradient vanishing and/or exploding during training. The long short-term memory was first introduced by Hochreiter and Schmidhuber\textsuperscript{30} to overcome the vanishing gradient problem. It has a similar chain-like structure as shown in Figure 1. However, instead of having a simple unit inside the hidden neurons, the LSTMs have three gates (input gate, forget gate, and output gate) and one cell interacting inside each of their hidden neurons, as shown in Figure 2. These gates are used to control the passing of information along the sequences, which can be mathematically expressed as:\textsuperscript{18,27,31,32}

\begin{align}
  i_t &= \sigma(w_{ix}I_t + w_{ih}H_{t-1} + b_i) \\
  o_t &= \sigma(w_{ox}I_t + w_{oh}H_{t-1} + b_o) \\
  f_t &= \sigma(w_{fx}I_t + w_{fh}H_{t-1} + b_f) \\
  c_t &= f_t \odot c_{t-1} + i_t \odot \phi(w_{cx}I_t + w_{ch}H_{t-1} + b_c) \\
  H_t &= o_t \odot \phi(c_t)
\end{align}

where $i_t$, $o_t$, $f_t$, and $c_t$ are the input gate, the output gate, the forget gate, and the memory cell within units of the hidden layer respectively. $w_{ix}$, $w_{ox}$, $w_{fx}$, and $w_{cx}$ are the weighting matrices connecting the inputs with gates and cell of the hidden layer (the input gate, the output gate, the forget gate, and the cell, respectively). Similarly, $w_{ih}$, $w_{oh}$, $w_{fh}$, and $w_{ch}$ are the weighting matrices connecting the hidden state at the last time step $H_{t-1}$ with gates and cell of the hidden layer at the current time step. $b_i$, $b_o$, $b_f$, and $b_c$ are the corresponding bias vectors. $\sigma(\cdot)$ and $\phi(\cdot)$ are the sigmoid and the hyperbolic tangent functions respectively. $\odot$ is a mathematic operator representing the elementwise product.

**Discussions and results**

**Milling data set**

The experiment was done on a milling machine which describes the tool wear process under various operating conditions.\textsuperscript{33} The workpiece material is cast iron. The cutting speed was constant at 200 m/min (826 rev/min). Two different depths of cut (1.5 mm and 0.75 mm), and two feed rates (0.5 mm/rev and 0.25 mm/rev) were investigated. The experimental matrix is $2 \times 3$. Hence there are four different cutting conditions, as shown in Table 1. The experiment under each condition was done a second time with a second set of inserts. Therefore, there are eight cases with a variable number of runs per tool life. The number of runs was dependent on the degree of flank wear that was measured between runs at irregular intervals up to a set wear limit. There is a total of 108 data samples.

Three types of sensor (two AE sensors, two accelerometers and a current sensor) were adopted to acquire the AE signals, the vibration signals at different
positions, and the AC spindle motor current. Since the data were collected from a real milling machine (instead of simulation models or experimental platforms) working under operational conditions close to industrial applications, it provides important information for researchers to study the relationship between health states and measurements.9,34,35 In addition, we propose to use the leave-one-out cross-validation method to cope with the overfitting problem caused by the limitation regarding the number of samples. By doing so, we deem the milling data set is sufficient for an effective study as conducted in this paper. In this study, we select 10 common features used in cutting tool condition monitoring as listed in Table 2, which includes four time-domain dimensional statistical features (peak to peak, root mean square, standard deviation, peak), two time-domain dimensionless features (kurtosis, crest factor), two frequency-domain features (amplitude of the first harmonic, average of the first six harmonic amplitudes), and two statistical features based on the $\beta$-distribution (skewness and kurtosis based on the $\beta$-distribution). The first eight are classic features used in the condition monitoring of mechanical system, whereas the last two were introduced by Kannatey-Asibu and Dornfeld36 to monitor the tool wear. Based on equations (2), (4), and (6), the three metric values of the ten selected features (as shown in Table 2) from the five sensors are shown in Figure 3. The final fitness for each feature from each sensor is calculated based on equation (7) by using the same weight coefficient for the three metrics, and listed in Table 3. It can be observed that the features from the motor current have generally higher fitness values than the features from the other sensors, followed by the features of the table AE. The spindle vibration provides the features with the lowest fitness values. This indicates that the features from the motor current and the table AE may provide better prognostic ability for the tool flank wear, whereas the features from spindle vibration offer the poorest prognostic ability. This phenomenon is consistent with the findings concluded from previous similar research activities: AE signals and cutting forces generally offer much richer information18,36,37; the less costly and easily mountable motor current sensor can be a fairly good substitute of the dynamometer (force sensor) as the spindle motor current is well correlated with the cutting force.6 Besides, it can also be found that, generally, the four time-domain dimensional features (1, 2, 3, 4 in Table 2) show higher fitness values, followed by the two $\beta$-distribution moments (9, 10 in Table 2), and the two frequency-domain features (7, 8 in Table 2). The two time-domain dimensionless features (5, 6 in Table 2) have the lowest fitness values for most sensors. Predictions based on the RNNs All 108 samples from the milling data set are arranged sequentially36 and shown in Figure 4: the first tool set operates under the four cutting conditions sequentially, and then the second tool set undergoes the same order. Since the samples are limited, we use the leave-one-out cross-validation method to train the network in order to remedy issues like overfitting and selection bias, as

Table 2. Features used in this study.

| Symbol | Feature                       | Symbol | Feature                      |
|--------|-------------------------------|--------|------------------------------|
| 1      | Peak to peak (P2P)            | 6      | Crest factor (CF)            |
| 2      | Root mean square (RMS)        | 7      | Amplitude of first harmonic (PKS1) |
| 3      | Standard deviation (STD)      | 8      | Average of the first six harmonic amplitude (PKSA) |
| 4      | Peak (MAX)                    | 9      | Skewness based on $\beta$-distribution (SKWB) |
| 5      | Kurtosis (KUR)                | 10     | Kurtosis based on $\beta$-distribution (KURB) |

Table 3. Fitness of the ten features of the five sensors.

| Sensors                        | Features | P2P | RMS | STD | MAX | CF  | KUR | PKS1 | PKSA | SKWB | KURB |
|--------------------------------|----------|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| Motor Current (MC)             |          | 0.77| 0.75| 0.75| 0.77| 0.34| 0.23| 0.50| 0.54 | 0.71 | 0.69 |
| Table Vibration (TV)           |          | 0.36| 0.57| 0.41| 0.49| 0.22| 0.37| 0.36| 0.42 | 0.19 | 0.17 |
| Spindle Vibration (SV)         |          | 0.27| 0.52| 0.32| 0.36| 0.23| 0.19| 0.28| 0.34 | 0.26 | 0.38 |
| Table AE (TA)                  |          | 0.59| 0.66| 0.59| 0.64| 0.20| 0.21| 0.22| 0.38 | 0.30 | 0.43 |
| Spindle AE (SA)                |          | 0.32| 0.49| 0.33| 0.41| 0.24| 0.21| 0.19| 0.20 | 0.37 | 0.34 |
suggested by Malhotra et al. This is done by defining a moving window with a given width which is about $\frac{1}{4}$ of the total samples (i.e. 27 samples) as the validation set, whereas the rest 81 samples as the training set for model training to predict those 27 validation points. To cover all possible combinations, the window is slid along the entire data (as shown in Figure 4) so there are 108 times of training and validation. The average of the validation errors is used to assess the prognostic abilities of different input sets. This strategy ensures the decision-making models we developed generalize well on the milling datasets. We further run the leave-one-out cross-validation 10 times for each input set to cope with the stochastic process caused by the randomly set initial parameters of the neural networks. Comparisons among the error distributions yielded by neural

Figure 3. Values of the three metrics for the ten selected features from different sensors: (a) motor current, (b) table vibration, (c) spindle vibration, (d) table AE, and (e) spindle AE.
networks fed with different inputs can provide us a holistic impression on the prognostic ability of each input feature set.

As to the network parameters, their values should be kept same for different input sets so that the comparisons are effective. The learning rate for training LSTM should be much smaller than that for training Elman network to avoid the gradient exploding possibly due to its much more complex architecture (note that the LSTM can only handle the gradient vanishing problem). The details of the RNN parameters chosen for comparisons are summarized in Table 4.

Prediction accuracy of each feature from each sensor. In this section, we will compare the prediction errors (which is defined as the root mean square error between the predicted values and actual values of the wear for the validation set as shown in Figure 4) among the networks fed with different feature vectors, in order to find out if the fitness value of a feature truly reflects its prognostic ability of a feature. On the other hand, comparisons between the prediction errors from the Elman network and LSTM network show that the LSTM slightly outperforms the Elman network in yielding smaller and more stable prediction errors in most cases. However, the training time for the LSTM is considerably longer than that of the Elman network (considering that LSTM has 4 units in each of its hidden neurons to be trained in comparison to the one unit of that of Elman).

Figure 5 shows the boxplots of the prediction errors for all cases (10 features × 5 sensors × 2 RNNs). It can be noticed that the features with higher fitness values shown in Table 3 generally yield smaller prediction errors from both the Elman network and LSTM network. For instance, the first four features for each sensor generally yield lower prediction errors compared with others. The last two features of the motor current (Figure 5(a)) also yield comparatively lower prediction errors. The features of the spindle vibration (Figure 5(c)) yield poorest prediction performance. All these phenomena are largely consistent with the fitness values shown in Table 3. Thus the fitness value calculated from equations (1) to (7) truly reflects the prognostic ability of a feature.

Figure 6 shows the boxplots of the prediction errors based on the 10 features extracted from each sensor. The spindle vibration still performs the worst, followed by the table vibration. This indicates that the vibration signals are truly less sensitive to the cutting tool condition compared with the motor current signal and acoustic emission signals. The features extracted from the motor current yield the lowest prediction error among all the features, which is reasonable as the fitness values of the features extracted from the motor current are generally higher than those from other sensors. The features extracted from the acoustic emission at the table (TA) also show good prediction accuracy. However, the features extracted from the acoustic emission at the spindle seems to have less correlation with the tool condition compared with the abovementioned two sensors. Thus, it can be concluded that the motor current signal and the acoustic emission at the table contain more information regarding the tool condition. These two signal sources should be the focus for effective and reliable on-line monitoring of tool wear. It should be further noted that all these phenomena are

Prediction accuracy of each sensor. In the section, we will investigate which sensor signal contains sufficient information of the tool condition. This is conducted by setting the feature vector $I$ as:

$$I = \{S_1 \ldots S_{10}, D, F\}$$

(10)

where $S_1 \ldots S_{10}$ represent the 10 features of a sensor signal (MC, TV, SV, TA, SA).

Figure 6 shows the boxplots of the prediction errors based on the 10 features extracted from each sensor. The spindle vibration still performs the worst, followed by the table vibration. This indicates that the vibration signals are truly less sensitive to the cutting tool condition compared with the motor current signal and acoustic emission signals. The features extracted from the motor current yield the lowest prediction error among all the features, which is reasonable as the fitness values of the features extracted from the motor current are generally higher than those from other sensors. The features extracted from the acoustic emission at the table (TA) also show good prediction accuracy. However, the features extracted from the acoustic emission at the spindle seems to have less correlation with the tool condition compared with the abovementioned two sensors. Thus, it can be concluded that the motor current signal and the acoustic emission at the table contain more information regarding the tool condition. These two signal sources should be the focus for effective and reliable on-line monitoring of tool wear. It should be further noted that all these phenomena are

Table 4. Details of the RNNs.

| Network structure | Elman network | LSTM |
|-------------------|---------------|------|
| Training epochs   | 1000          | 1000 |
| Learning function | gradient descent with momentum | gradient descent with momentum |
| Learning rate     | 0.1           | 0.01 |
| Momentum coefficient | 0.9         | 0.9  |

Figure 4. Sliding window on the sequence of the tool wear.

Figure 5 shows the boxplots of the prediction errors for all cases (10 features × 5 sensors × 2 RNNs). It can be noticed that the features with higher fitness values shown in Table 3 generally yield smaller prediction errors from both the Elman network and LSTM network. For instance, the first four features for each sensor generally yield lower prediction errors compared with others. The last two features of the motor current (Figure 5(a)) also yield comparatively lower prediction errors. The features of the spindle vibration (Figure 5(c)) yield poorest prediction performance. All these phenomena are largely consistent with the fitness values shown in Table 3. Thus the fitness value calculated from equations (1) to (7) truly reflects the prognostic ability of a feature. On the other hand, comparisons between the prediction errors from the Elman network and LSTM network show that the LSTM slightly outperforms the Elman network in yielding smaller and more stable prediction errors in most cases. However, the training time for the LSTM is considerably longer than that of the Elman network (considering that LSTM has 4 units in each of its hidden neurons to be trained in comparison to the one unit of that of Elman).

Prediction accuracy of each sensor. In this section, we will compare the prediction errors (which is defined as the root mean square error between the predicted values and actual values of the wear for the validation set as shown in Figure 4) among the networks fed with different feature vectors, in order to find out if the fitness value of a feature truly reflects its prognostic performance. The feature vector $I$ has three inputs: each individual feature $x$, the depth of cut ($D$) and the feed rate ($F$):

$$I = \{x, D, F\}$$

(9)

where $S_1 \ldots S_{10}$ represent the 10 features of a sensor signal (MC, TV, SV, TA, SA).

Figure 6 shows the boxplots of the prediction errors based on the 10 features extracted from each sensor. The spindle vibration still performs the worst, followed by the table vibration. This indicates that the vibration signals are truly less sensitive to the cutting tool condition compared with the motor current signal and acoustic emission signals. The features extracted from the motor current yield the lowest prediction error among all the features, which is reasonable as the fitness values of the features extracted from the motor current are generally higher than those from other sensors. The features extracted from the acoustic emission at the table (TA) also show good prediction accuracy. However, the features extracted from the acoustic emission at the spindle seems to have less correlation with the tool condition compared with the abovementioned two sensors. Thus, it can be concluded that the motor current signal and the acoustic emission at the table contain more information regarding the tool condition. These two signal sources should be the focus for effective and reliable on-line monitoring of tool wear. It should be further noted that all these phenomena are
nearly consistent with the fitness values shown in Table 2.

**Prediction accuracies of some fusing features.** It is traditionally believed that an increase in the number of sensor features generally improves the performance. In this section, we investigate the prediction errors from different feature sets extracted from the motor current and the table AE as shown in Figure 7. The labels in the horizontal axis of Figure 7 represent different feature sets as illustrated in Table 5. For example, MC17 means that the feature vector has four inputs: peak to peak of the motor current signal, first harmonic amplitude of the motor current, depth of cut, and feed rate.

It can be found from Figure 7 that for the motor current signal, the parameter P2P (MC1) alone can achieve slightly better prediction accuracy with that achieved by combining P2P and PKS1 (MC17), or even all 10 features (MC1-10). This indicates that adding more features from the motor current signal into the feature vector does not yield a more accurate prediction. In fact, adding features with lower fitness values may even deteriorate the predictions, as also confirmed

Figure 5. Boxplots of the prediction errors from each feature of each sensor: (a) motor current, (b) table vibration, (c) spindle vibration, (d) table AE, and (e) spindle AE.

Note: The left and right parts of each figure are the predictions from the Elman and LSTM, respectively.
by Rangwala and Dornfeld. Table 6 shows the cross-correlation chart of features from the motor current. It can be found that there are high correlations between the P2P with most of the other features from the motor current, which indicates that these highly correlated features contain abundant information with cutting process. This probably explains why only one typical time-domain statistical feature (normally RMS) of the cutting force signal or motor current signal is normally employed as the input of decision-making models in many previous similar studies. For the table AE signal, the combined feature set of P2P and PKS1 can achieve much better results than each alone does. This demonstrates that adding features with high fitness values but low correlations among each other can generally improve the prediction performance.

Tables 6 and 7 show the cross-correlation charts for the features from the motor current and table AE, respectively. It can be found that there are high correlations among most of the selected features from the motor current. This demonstrates that the selected features from the motor current have abundant information relating to the cutting process. On the other hand, the features from the table AE have comparatively lower correlations among each other, which means that these features may provide corroborative information. It can be concluded that an optimal feature set should contain features that have higher fitness values, as well as lower correlations with other features. Based

Table 5. Feature vectors \( I \) corresponding to the horizontal labels in Figure 7.

| Motor current | Feature vector \( I \) | Table AE | Feature vector \( I \) |
|---------------|-------------------|----------|-------------------|
| MC1-10        | \{MC1 ... MC10, D, F\} | TA1-10   | \{TA1 ... TA10, D, F\} |
| MC17          | \{MC1, MC7, D, F\}  | TA17     | \{TA1, TA7, D, F\}  |
| MC1           | \{MC1, D, F\}      | TA1      | \{TA1, D, F\}      |
| MC7           | \{MC7, D, F\}      | TA7      | \{TA7, D, F\}      |

Refer to Table 3 for the symbols in \( I \).
on the fitness values and the cross-correlation values, the following features are selected as the inputs to the networks:

\[ I = \{MC2, TA1, TA7, D, F\} \quad (11) \]

Figure 8 shows the comparisons of the prediction errors based on four feature sets. The first feature set MC1-10 is all the 10 features of the motor current. The second feature set TA1-10 is all the 10 features of the table AE. The third feature set MC1-10_TA1-10 is all the 20 features from the motor current and the table AE. The forth feature set is the optimal feature set which has only three features from the sensors as shown in equation (11). Table 8 shows the prediction errors based on the feature sets.
performance in terms of the average training time and average prediction error after 10,000 training epochs. It can be found that fusing the motor current and table AE can generally improve the prediction performance. However, some features may provide abundant information, and some others may deteriorate the predictions. After deleting these features, the optimal feature set outperforms all the others. Since there are only 71 training samples available, the training time of the optimal feature set is only slightly smaller than those of the others. The training time saved by using the optimal feature set will be much more significant if hundreds or thousands of samples are trained. It can also be found that the LSTM network yields more accurate tool wear predictions (lower prediction error) than the Elman network does on the studied milling data set. This is due to the gating mechanism (input gate, forget gate, and output gate) introduced in the hidden neurons for LSTM as shown in Figure 2. This gating mechanism enables the LSTM to adaptively control the information passing through the hidden neurons (as to what to input, what to forget and what to output). However, since the LSTM has three more gates to be trained, the training time for LSTM is much longer than the Elman network. As shown in Table 8, the training time for LSTM is generally twice as longer as that for Elman network. Considering that the networks are usually trained off-line, the expensive computation caused by training LSTM doesn’t pose a significant problem. In conclusion, the LSTM network is better than the Elman network in predicting the milling tool wear.

**Conclusion**

Identification of the optimal features from a variety of sensors is necessary for reliable and effective on-line monitoring of cutting tool condition. This study introduces parameter fitness to determine the prognostic ability of a feature. It is calculated based on three modified parameter suitability metrics, which allows for the unit-wise variation of the component in a population. The publicly available milling data set provided in the literature are analyzed, from which ten common features from five sensors are extracted. Two types of recurrent neural network (Elman and LSTM) are employed as the decision-making models to predict the tool wear based on the selected features. The leave-one-out cross-validation method is adopted to cope with the limited number of training samples. Due to the usual random initialization of the neural networks’ parameters (weights and biases), the neural network prediction for each setting is done 10 times. The prediction error distribution for each feature set fed into the neural networks is expressed by the boxplot, so that a holistic comparison of prediction performance among different feature sets can be conducted. The main conclusions of this study include:

1. Three metrics (correlation, monotonicity, and the statistical overlap factor) can be used to determine the general fitness of a feature used for prognostics. A feature having a higher fitness value is proved to yield a smaller prediction error.
2. Features that have high correlation among each other may provide abundant information relating to the cutting tool condition. Adding more correlated features will not increase the prediction accuracy, but will increase the burden on the decision-making models. In addition, adding features that have lower fitness values may deteriorate the prediction performance. Generally, a feature that has a higher fitness, and lower correlations with others is preferred.
3. Different types of signal sources (motor current and table AE in this case) may provide corroborative information for tool condition monitoring. Thus, fusing features from different sensors can generally improve the prediction performance. The optimal feature set after deleting those unnecessary features can outperform the other feature sets in terms of prediction accuracy and training time.

It should be mentioned that the above three conclusions regarding the determination of optimal features for condition monitoring are adequately generic for sensor-based prediction of gradual damage in mechanical systems.

### Table 8. Comparisons in prediction performance among four feature sets.

| Label | Feature vector | Average training time (s) | Average prediction error (mm) |
|-------|----------------|--------------------------|-----------------------------|
|       |                | Elman | LSTM | Elman | LSTM |
| MC1-10 | (MC10, D, F)  | 10.41 | 19.63 | 0.1695 | 0.1158 |
| TA1-10 | (TA10, D, F)  | 10.96 | 19.84 | 0.1808 | 0.1512 |
| MC1-10_TA1-10 | (MC10, TA10, D, F) | 11.01 | 20.08 | 0.1938 | 0.1123 |
| MC2_TA17 | (MC2, TA1, TA7, D, F) | 9.57 | 19.06 | 0.1580 | 0.1066 |
Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge support provided by the Natural Sciences and Engineering Research Council of Canada (Grant number: RGPIN/05922-2014).

ORCID iD

Wennian Yu https://orcid.org/0000-0002-0077-0836

References

1. Sick B. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. Mech Syst Signal Process 2002; 16: 487–546.  
2. Sun H, Cao D, Zhao Z, et al. A hybrid approach to cutting tool remaining useful life prediction based on the Wiener process. IEEE Trans Reliab 2018; 67: 1294–1303.  
3. Aghazadeh F, Tahan A and Thomas M. Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. Int J Adv Manuf Technol 2018; 98: 3217–3227.  
4. Ahmed YS, Alam MS, Arif AFM, et al. Use of acoustic emission and cutting force signals to monitor built-up edge formation in stainless steel turning. Int J Adv Manuf Technol 2019; 103: 2257–2276.  
5. Bayraktar S and Turgut Y. Investigation of the cutting forces and surface roughness in milling carbon-fiber-reinforced polymer composite material. Mater Technol 2016; 30: 591–600.  
6. Ghosh N, Ravi YB, Patra A, et al. Estimation of tool wear during CNC milling using neural network-based sensor fusion. Mech Syst Signal Process 2007; 21: 466–479.  
7. Lei Y and Zuo MJ. Gear crack level identification based on weighted K nearest neighbor classification algorithm. Mech Syst Signal Process 2009; 23: 1535–1547.  
8. Frigieri EP, Campos PHS, Paiva AP, et al. A mel-frequency cepstral coefficient-based approach for surface roughness diagnosis in hard turning using acoustic signals and gaussian mixture models. Appl Acoust 2016; 113: 230–237.  
9. Rad JS, Zhang Y and Chen C. A novel local time-frequency domain feature extraction method for tool condition monitoring using S-transform and genetic algorithm. IFAC Proc Vol 2014; 47: 3516–3521.  
10. Benkedjouh T, Medjaher K, Zerhouni N, et al. Health assessment and life prediction of cutting tools based on support vector regression. J Intell Manuf 2015; 26: 213–223.  
11. Mosallam A, Medjaher K and Zerhouni N. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. J Intell Manuf 2016; 27: 1037–1048.  
12. Gajate A, Haber R, Toro R Del, et al. Tool wear monitoring using neuro-fuzzy techniques: a comparative study in a turning process. J Intell Manuf 2012; 23: 869–882.  
13. Quintana G, Garcia-Romeu ML and Ciurana J. Surface roughness monitoring application based on artificial neural networks for ball-end milling operations. J Intell Manuf 2011; 22: 607–617.  
14. Alizadeh A and Omrani H. An integrated multi response Taguchi-neural network-robotic data envelopment analysis model for CO2 laser cutting. Measurement 2019; 131: 69–78.  
15. Zhou H, Ding WF, Li Z, et al. Predicting the grinding force of titanium matrix composites using the genetic algorithm optimizing back-propagation neural network model. Proc Inst Mech Eng B J Eng Manuf 2019; 233: 1157–1167.  
16. Liu C, Ding W, Li Z, et al. Prediction of high-speed grinding temperature of titanium matrix composites using BP neural network based on PSO algorithm. Int J Adv Manuf Technol 2017; 89: 2277–2285.  
17. Zhao R, Yan R, Chen Z, et al. Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 2019; 115: 213–237.  
18. Yu W, Kim IY and Mechefske CK. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. Mech Syst Signal Process 2019; 129: 764–780.  
19. Yu W, Kim IY and Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. Reliab Eng Syst Saf 2020; 199: 106926.  
20. Agarwal A and Desai KA. Amalgamation of physics-based cutting force model and machine learning approach for end milling operation. Procedia CIRP 2020; 93: 1405–1410.  
21. Li S and Elbestawi MA. Tool condition monitoring in machining by fuzzy neural networks. J Dyn Syst Meas Control 1996; 118: 665–672.  
22. Deiab I, Assaleh K and Hammad F. On modeling of tool wear using sensor fusion and polynomial classifiers. Mech Syst Signal Process 2009; 23: 1719–1729.  
23. Zhou JH, Pang CK, Lewis FL, et al. Intelligent diagnosis and prognosis of tool wear using dominant feature identification. IEEE Trans Ind Inform 2009; 5: 454–464.  
24. Coble J and Hines JW. Identifying optimal prognostic parameters from data: a genetic algorithms approach. Proc Annu Conf Progn Heal Manag Soc 2009; 27: 1–11.  
25. Scheffer C and Heyns PS. An industrial tool wear monitoring system for interrupted turning. Mech Syst Signal Process 2004; 18: 1219–1242.  
26. Javed K, Gouriveau R, Zerhouni N, et al. A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling. In: 2013 IEEE conference on prognostics and health management (PHM), Gaithersburg, MD, 24–27 June 2013, pp.1–7. New York: IEEE.  
27. Guo L, Li N, Jia F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 2017; 240: 98–109.
28. Lu CJ and Meeker WQ. Using degradation measures to estimate a time-to-failure distribution. *Technometrics* 1993; 35: 161–174.

29. Yu W, Mechefske CK and Kim IY. Cutting tool wear estimation using a genetic algorithm based long short-term memory neural network. In: *Proceedings of the ASME 2018 international design engineering technical conference*, Quebec City, Canada, 26–29 August 2018.

30. Hochreiter S and Schmidhuber JU. Long short-term memory. *Neural Comput* 1997; 9: 1735–1780.

31. Yu W, Mechefske C and Kim IY. Time series reconstruction using a bidirectional recurrent neural network based encoder-decoder scheme. In: *ALAC - 18th Australian international aerospace congress (ALAC2018); HUMS - 11th defence science and technology (DST) international conference on health and usage monitoring (HUMS 2019); ISSFD - 27th international symposium on space flight dynamics (ISSFD)*, Melbourne, 2019.

32. Yu W, Yong I and Mechefske C. Analysis of different RNN autoencoder variants for time series classification and machine prognostics. *Mech Syst Signal Process* 2021; 149: 107322.

33. Agogino A and Goebel K. Milling Data Set. BEST lab, UC Berkeley, NASA Ames Prognostics Data Repository, https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/ (2007, 20 April 2019).

34. Hanachi H, Yu W, Kim IY, et al. Hybrid data-driven physics-based model fusion framework for tool wear prediction. *Int J Adv Manuf Technol* 2019; 101: 2861–2872.

35. Lei Y, Li N, Guo L, et al. Machinery health prognostics: a systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process* 2018; 104: 799–834.

36. Kannatey-Asibu E and Dornfeld DA. A study of tool wear using statistical analysis of metal-cutting acoustic emission. *Wear* 1982; 76: 247–261.

37. Lin JT, Bhattacharyya D and Keeman V. Multiple regression and neural networks analyses in composites machining. *Compos Sci Technol* 2003; 63: 539–548.

38. Heimes FO. Recurrent neural networks for remaining useful life estimation. In: *International conference on prognostics and health management (PHM 2008)*, Denver, CO, 6–9 October 2008, pp.1–6. New York: IEEE.

39. Malhotra P, TV V, Ramakrishnan A, et al. Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder. *arXiv preprint arXiv:1608.06154*, 22 August 2016.

40. Rangwala S and Dornfeld D. Sensor integration using neural networks for intelligent tool condition monitoring. *J Eng Ind* 1990; 112: 219–228.

41. Dutta RK, Paul S and Chattopadhyay AB. Fuzzy controlled backpropagation neural network for tool condition monitoring in face milling. *Int J Prod Res* 2000; 38: 2989–3010.