Int J Adv Manuf Technol
DOI 10.1007/s00170-009-2110-z

ORIGINAL ARTICLE

Design of multisensor fusion-based tool condition monitoring system in end milling

Sohyung Cho · Sultan Binsaeid · Shihab Asfour

Received: 5 December 2008 / Accepted: 11 May 2009

Abstract Recent advancement in signal processing and information technology has resulted in the use of multiple sensors for the effective monitoring of tool conditions, which is the most crucial feedback information to the process controller. Interestingly, the abundance of data collected from multiple sensors allows us to employ various techniques such as feature extraction, selection, and classification methods for generating such crucial information. While the use of multiple sensors has improved the accuracy in the classification of tool conditions, design of tool condition monitoring system (TCM) for reduced complexity and increased robustness has been rarely studied. Therefore, this paper studies the design of effective multisensor-based TCM when machining 4340 steel by using a multilayer-coated and multiflute carbide end mill cutter. Multiple sensors tested in this paper include force, vibration, acoustic emission, and spindle power sensor for the time and frequency domain data. In addition, two feature selection methods and three classifiers with a machine ensemble technique are considered as design components. Importantly, different fusion methods are evaluated in this paper: (1) decision level fusion and (2) feature level fusion. The experimental results show that the design of TCM based on the feature level fusion can significantly improve the accuracy of the tool condition classification. It is also shown that the highest accuracy can be achieved by using force, vibration, and acoustic emission sensor together with correlation-based feature selection method and majority voting machine ensemble.

Keywords Multisensor fusion · Tool condition monitoring · Machine ensemble

S. Cho (✉)
Industrial and Manufacturing Engineering,
Southern Illinois University Edwardsville,
Edwardsville, IL 62026, USA
e-mail: scho@siue.edu

S. Binsaeid · S. Asfour
Department of Industrial Engineering, University of Miami,
Coral Gables, FL 33146, USA

1 Introduction

In today's fierce global competition, on-time delivery of highly diversified products with reduced manufacturing lead time has become a key determinant for the survival of manufacturing enterprises. Manufacturing lead time can be significantly reduced by effective control of disruptive events such as machine breakdown, material absence, and demand fluctuations. Among those disruptive events, machine breakdown is directly related to the increased manufacturing lead time and may result in reduced customer satisfaction. It should be emphasized that considerable portion (7–20%) of machine downtime results from tool failure [1, 2]. The tool failure can be prevented by efficiently monitoring conditional changes in the tool, and hence, tool condition monitoring (TCM) has been of great interest to both academia and industry. It has been reported that successful implementation of TCM can save up to 40% of production costs [3]. In general, there are three categories of tool condition, particularly for end-milling cutters: (1) tool breakage, (2) tool chipping, and (3) tool wear. These categories are different in their nature such that tool breakage occurs abruptly in an observable and random manner, tool chipping has the same characteristics as tool breakage except it is hardly detected for a considerable duration, whereas tool wear develops gradually and can be predicted to a certain extent.

Published online: 26 May 2009

With recent advancement in signal processing technology and information technology, a wide range of online sensors has been employed to retrieve information relevant to tool conditions, which is the most crucial feedback information to the process controller. Specifically, force sensor [4, 5], vibration sensor [6, 7], acoustic emission [8, 9], and spindle power sensor [10] have been used as an individual sensor or a group of sensors, referred to as a multiple sensor [11, 12]. The employment of multiple sensors has improved the accuracy in the classification of tool conditions because it is intended to fuse the informational power of individual sensor, resulting in complementary and redundant information [13]. Interestingly, the abundance of data collected from multiple sensors allows us to make use of various techniques such as feature extraction, selection, and classification methods for generating the crucial feedback information [13, 14]. While many research works focused on the improved accuracy in the classification of tool conditions by employing multiple sensors, design of multisensor-based TCM for reduced complexity and increased robustness has been rarely studied.

The main goal of this paper is to study the design of effective tool condition monitoring system in a more systematic manner when machining 4340 steel by using a multilayer-coated and multiflute carbide end mill cutter. Specifically, we study decision making in the design process that includes determination of a multisensor combination, feature selection method, machine learning-based classifier, and machine ensemble technique. Importantly, two different fusion methods are evaluated in this paper: (1) decision level fusion and (2) feature level fusion. To achieve the aforementioned goal, this paper investigates the following three objectives as part of the analysis. The first one is to study the significance of reducing the input space dimension for the classification model and selecting the most significant subset of features with which higher level of information related to the tool condition classification can be achieved. The second one is to study the significance of different information fusion strategies to the classification model, i.e., no fusion with best single sensor model, feature level fusion with best multiple sensors model. The third objective is to investigate the effectiveness of several decision-making methods, which are multilayer perceptron neural network (MLP), radial base function neural network (RBF), and support vector machine (SVM). Furthermore, these three classifiers are studied with a machine ensemble approach, which is referred to as majority vote. The rest of the paper is organized as follows: Section 2 explains the design of multisensor fusion-based TCM and required components that are involved in the design. A detailed review is given in regards to data acquisition system, signal processing methods and their extracted features, and feature selection method. Then, machine learning (ML) algorithms and machine ensemble approach are introduced. In addition, feature level and decision level fusion are introduced. Section 3 outlines the experimental setup, design of experiment, and definition of tool condition classes. Specifically, three classes are defined to describe three different states of flank wear progression, while two extra classes are assigned for tool chipping and breakage. The discussion and performance of constructed TCM models are provided in Section 4.

2 Design of multisensor fusion-based TCM system

Design of multisensor fusion-based TCM system considered in this paper consists of four layers as illustrated in Fig. 1: (1) data acquisition through multiple sensors and

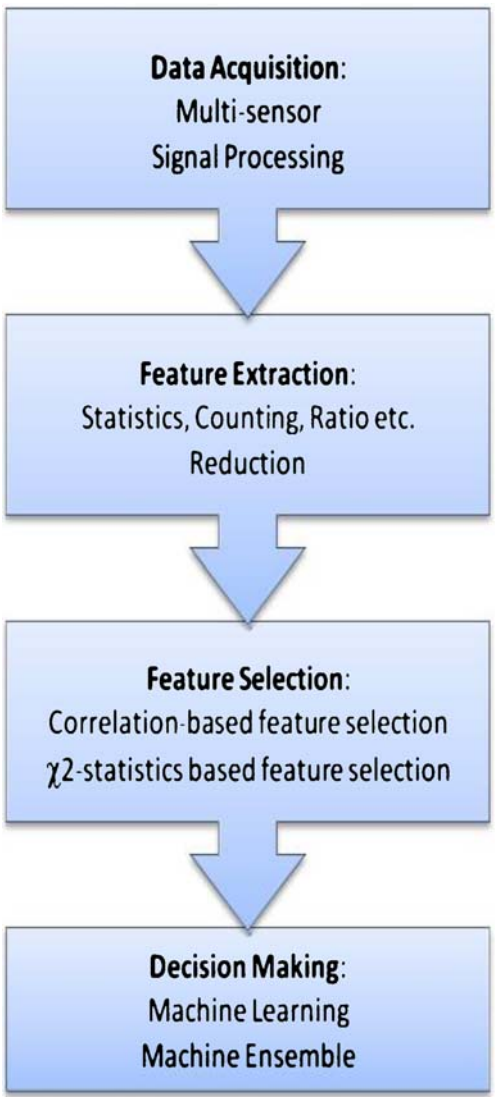

Fig. 1 Design of multisensor fusion-based TCM system

digital signal process, (2) feature extraction, (3) feature selection, and (4) ML and ensemble-based classification. Figure 1 also shows specific attributes associated with each layer. For example, selection of multiple sensors and signal processing techniques is the main attribute that is associated with data acquisition layer.

2.1 Data acquisition using multisensor

The following multiple sensors are used to collect data required for the subsequent design and analysis of TCM systems: a dynamometer to measure three-directional forces, an accelerometer to measure three-directional vibrations, an acoustic emission sensor, and a spindle power sensor. Note that these four sensors are most frequently used sensors in the literature [12] and account for eight sensory signals. In the subsequent design stages, data obtained from individual sensor is fused into each other at feature level or decision level depending on the design objectives. The advantage of fusing the outputs from one sensor with those from another independent sensors stems from redundancy being present in the information [13]. More specifically, if redundant sensors are employed, the overall uncertainty of the resulting measurement can be reduced, and thus, the performance of the system can be improved by averaging out the independent noise acting on the different sensors because the noise inherent in individual sensor measurement is not correlated with noise from other sensors to a large extent. In addition, complementary sensors provide extended and independent information about the process, which is difficult to be captured otherwise. On the other hand, signal processing attribute includes the selection of band pass filters, sampling rate, and gain of the coupler to improve the quality of the data. In this research, all the signals are properly filtered and analyzed by commercially available software—LabVIEW.

2.2 Feature extraction

The main purpose of feature extraction is to significantly reduce the dimension of raw data in time and frequency domain and at the same time maintain the relevant information about tool conditions in the extracted features. Many research works have studied various feature extraction methods, and most of these extraction methods can be found in [14–19]. In this paper, a comprehensive set of feature extraction methods that have been previously studied is established. Note that different extraction methods have different capabilities in extracting key information about tool conditions from multisensor signals. In this research, a program code that can automatically extract different features from incoming signals in both time and frequency domain has been constructed using the LabVIEW software. Specifically, thefollowingfeaturesareextractedfromthemultiplesensorsfor further analysis in the subsequent design stages. Note that amplitude values of a signal are expressed as $[x_1, x_2, \ldots x_n]$. Table 1 summarized features considered in this paper that are extracted from each sensor signal.

Table 2 provides the distribution of all extracted features in both time and frequency domain per sensor. Specifically, there are 135 extracted features from eight sensory signals, i.e., three force signals, one acoustic emission signal, three vibration signals, and one spindle power signal. For instance, there are 27 features from force sensor signals in the table (nine features from each force sensor × 3 force sensors of F_x, F_y, and F_z). In addition to the extracted features, machining parameters are also considered as a part of the feature space, which are axial depth of cut, cutting speed, and feed rate. Therefore, the total number of features considered in this paper is 138.

2.3 Feature reduction method

Training ML classifiers using the maximum number of features obtainable is not always the best option, as irrelevant and redundant features can negatively influence the performance of ML algorithms. In order to improve the accuracy of the classification model and increase the efficiency of the computational performance of TCM systems, inclusion of an optimal number of significant features in the final model is desirable. This can be achieved by reducing the number of features utilizing features selection techniques. Correlation-based feature selection method (CFS) and χ^2 statistics selection method are studied in this research to evaluate different feature subsets. Also, note that a greedy hill climbing search algorithm is employed to search for optimal subset size [20].

2.3.1 Correlation-based feature selection method

CFS measures the goodness of feature subsets by taking the followings into account:

- the level of correlation of individual features with the predicted class
- the level of inter-correlation among features

Importantly, high scores are assigned to subsets containing features that are highly correlated with the class, yet have low inter-correlation measure with each other. Entropy measures are utilized to obtain a measure of correlation between features and classes and also between features. All continuous features are discretized using the technique

Table 1 Features extracted from multiple sensors in time and frequency domain

Features	Description
Time domain	
Arithmetic mean (M)	$M = \frac{1}{n}\sum_{i=1}^{n} x_i$
Root mean square (RMS)	$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
Variance (V)	$V = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}$
Skewness (Sk)	$\mathrm{Sk} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}$
Kurtosis (Ku)	$\mathrm{Ku} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}$
Signal power (P)	$P = \frac{1}{n}\sum_{i=1}^{n} x_i^2$
Peak-to-peak amplitude (pp)	$\mathrm{pp} = \max(x_i) - \min(x_i)$
Crest factor (CF)	$CF = \frac{Peak}{RMS}$
Burst rate (Br)	Number of times the signal exceeds preset thresholds per second. This feature is only applied to vibration and AE signals. The preset threshold is set to 300 μV
Frequency domain	
Sum of total band power (STPB)	$\mathrm{STPB} = \int_{F_1}^{F_2} S(f)$ where $S(f)$ is the power at a specific frequency component and (F_1, F_2) is the frequency band
Mean of band power spectrum (MBP)	$\mathrm{MBP} = \frac{1}{n}\sum_{i=1}^{n} S(f)_i$
Variance of band power spectrum (VBP)	$\mathrm{VBP} = \frac{\sum_{i=1}^{n}\left(S(f)_i - \mathrm{MBP}\right)^2}{n-1}$
Skewness of band power spectrum (SkBP)	$\mathrm{SkBP} = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(S(f)_i - \mathrm{MBP}\right)^3}{\mathrm{VBP}^{3/2}}$
Kurtosis of band power spectrum (KuBP)	$\mathrm{KuBP} = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(S(f)_i - \mathrm{MBP}\right)^4}{\mathrm{VBP}^{4/2}}$
Maximum (peak) of band power (PBP)	Peak of power spectrum in a specific frequency band that is expressed by the energy level (W/Hz)
Frequency of maximum peak of band power (FPBP)	Relative frequency that corresponds to the highest amplitude
Relative spectral peak per band (RSPBP)	Ratio of peak of band power (PBP) over the mean of band power (MBP)
Total harmonic band power (THBP)[a]	$\mathrm{THBP} = \sum_{m=1}^{N} P(m), \quad m = 1, 2, ..., N$ where $P(m)$ is the power at the fundamental tooth frequency, body cutter, and their harmonics, and N is the largest integer for which N is the cut-off frequency for the sensor

[a] This feature is only applied to the three-directional force signals

Table 2 Distribution of time and frequency domain features

Sensor	Number of Features		
	Time domain	Freq. domain	
Force	24	27	51
AE	9	16	26
Vibration	27	24	51
Spindle power	8	0	8
Total	68	67	135

studied by Fayyad and Irani [21]. The entropy of a feature Y is given as follows:

$$H(Y) = -\sum_{y \in R_y} p(y)\log(p(y)) \tag{1}$$

where Y is a discrete random variable with respective range R_y. Then, the conditional entropy of any feature Y given the occurrence of feature X, which has range R_x, can be calculated as:

$$H(Y|X) = -\sum_{x \in R_x} p(x)\sum_{y \in R_y} p(y)\log(p(y)) \tag{2}$$

Therefore, a measure of correlation can be obtained for either two features or between a feature and a class X and Y

where a class of an instance is considered to be a feature. This measure is often called uncertainty coefficient of Y and is calculated as follows:

$$C(Y|X) = \frac{H(Y) - H(Y|X)}{H(Y)} \tag{3}$$

Now, the scores of the CFS subsets are obtained using the following heuristic:

$$\text{Merit}_S = \frac{k\overline{r_{\text{cf}}}}{\sqrt{k + k(k-1)\overline{r_{\text{ff}}}}} \tag{4}$$

where Merit_S is the heuristic of a feature subset S containing k number features, and $\overline{r_{\text{cf}}}$ and $\overline{r_{\text{ff}}}$ are the average feature–class correlation and average feature–feature intercorrelation, respectively. In Eq. 4, the numerator is an indication of the predictive power of the feature set, while the denominator measures redundancy among features.

2.3.2 Greedy hill climbing search algorithm

Clearly, it is prohibitive to try all possible combinations of feature subset using the evaluation function of CFS. A simple yet effective search algorithm such as greedy hill climbing has demonstrated its efficiency in searching the feature space in reasonable time and provided good results [22]. Greedy search expands the current parent node and picks the child with the highest evaluation. Nodes are expanded by applying search space operators to them in which a single feature is added or deleted. A backward elimination strategy is employed where the search starts with the full set of features. Then, backward elimination will continue to delete features as long as child is not worse than its parents. This process is repeated until no more improvement can be achieved.

2.3.3 χ^2 statistics-based feature selection method

This method measures the rank of various features based on their statistical dependency relative to the class. The main objective of using χ^2 statistics is to maximize its relevancy. In statistics, the χ^2 test is applied to test the independence of two events, where two events, A and B, are defined to be independent if $P(AB)=P(A)P(B)$ or, equivalently, $P(A|B)=P(A)$ and $P(B|A)=P(B)$. Similarly, the χ^2 coefficient in feature selection application is given by Hong et al. [23]:

$$\chi^2(f, C) = \sum_{ij} \frac{\left(p\left(f = x_j, C_i\right) - p\left(f = x_j\right) \times p\left(C_j\right)\right)^2}{p\left(f = x_j\right) \times p\left(C_j\right)} \tag{5}$$

where $p(.)$ represents probability, C_j class j, and x_j feature j. Note that increasing values of χ^2 indicates higher dependency between feature values and class labels. It should be pointed out here that in the CFS method, the optimal number of features for each individual sensor is selected by using the greedy hill climbing search algorithm. However, χ^2 statistics method is a ranking method and thus requires setting a threshold value to include the specified number of features within the developed subset. Therefore, in this paper, to make a fair comparison between two feature selection methods, the number of features of the χ^2 statistics subset is set to be equal to the one achieved by the CFS method.

2.4 Machine learning classifiers

In this paper, three ML classifiers are used to classify tool conditions, and then, the ensemble techniques are applied for them to further improve the accuracy of the classification. In this paper, all classifiers and reduction methods have been implemented using WEKA ML suite, which provides a freeware environment supported by many machine learning authorities [24]. All of three ML algorithms employed in this study have proven to be effective in the pattern recognition communities: (1) SVM, (2) MLP, and (3) RBF.

2.4.1 Multiclass support vector machine

Since SVM, which is based on the statistical learning theory presented by [25], was proposed as a decision making method, SVM has received a lot of attention in the pattern recognition literature. While typical ML algorithms attempt to minimize the empirical risk that is the misclassification errors on the training set, the SVM attempts to minimize the structural risk that is the probability of misclassification of a previously unseen data point drawn randomly from fixed but unseen distribution. The SVM generates an efficient means of classification by condensing the relevant information and selecting the most important samples, called support vectors to the target. These support vectors achieve the maximal margin classification between classes. If linear separability of the data is not achieved, the training data are mapped into a higher dimensional feature space using a kernel function, which permits a higher level of linear separability.

In this paper, the SVM has been implemented using sequential minimum optimization algorithm. The selection of kernel function has influence on the decision boundary. In general, a RBF is favored instead of polynomial kernel functions because they are not sensitive to outliers and do not require inputs to have equal variances. Therefore, a RBF has been selected as a kernel function after preliminary analysis. The RBF kernel is defined as:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \qquad \gamma > 0 \tag{6}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ defines an inner product that maps the input vector $x \in \Re^{\text{d}}$ to a high dimensional space. Moreover, in

this research, a grid search has been performed on the training data in order to select the appropriate parameter for the width of the RBF function, γ, and the cost function parameter C. The grid search has resulted in optimal values of γ=0.25, and C=12.0.

2.4.2 Multilayer perceptron neural networks

MLP is the most widely used learning algorithm and is discussed at length in most neural network textbooks [26, 27]. The learning process of the MLP network is based on the data samples, composed of the N-dimensional vector $\mathbf{x}$ of the input layer and the M-dimensional desired output vector $\mathbf{c}$ of the output layer. By processing the input vector $\mathbf{x}$, the MLP creates the output vector $\mathbf{y}(\mathbf{x}, \mathbf{w})$, where $\mathbf{w}$ is the vector of modified weights. The error produced triggers a control mechanism of the learning algorithm. The corrective adjustments are designed to make the output signal y_k (k=1, 2,…, l) to the desired response c_k in an iterative manner, where l is number of classes in the output layer.

The learning algorithm of MLP is based on the minimization of the error function defined on the learning set (x_i,c_i) for i=1,2,…,p using a Euclidean norm, where p is the number of hidden nodes:

$$E(w) = \frac{1}{2}\sum_{i=1}^{P} \|\mathbf{y}(\mathbf{x}_i, \mathbf{w}) - \mathbf{c}_i\|^2 \tag{7}$$

The MLP used in this research consists of input layer, hidden layer, and output layer. The input layer has nodes representing the normalized features calculated from the sensory signals. There are various methods, both heuristic and systematic, to select the neural network structure and activation functions. In this paper, a heuristic that varies the number of input nodes depending on the number of features applied to the network has been established. Specifically, since the number of input changes depending on the number of sensors and features selected, the main search engine tries to set the hidden layer to be half of the total number of the input and output layers [24]. In addition, the number of output nodes represented the number of tool condition classes (five different tool condition classes in this paper). The target value of each output node produces confidence level that represents the classification probability of each class. In this research, the activation functions of sigmoid have been used in the hidden layers and in the output layer, respectively. Moreover, the MLP has been trained and implemented using back propagation (BPN) algorithm. Back propagation parameters include momentum and learning rate that affect the way the network is trained and, possibly, the performance of the learned classifier. Using all available features, a grid search has been performed to select the parameters of the BPN algorithm. As a result, a learning rate of 0.1 and momentum rate of 0.2 have been selected as optimal values. Interestingly, the MLP has been trained iteratively to minimize the performance function of mean squared error (MSE) between the network outputs and the corresponding target values. Specifically, the gradient of the performance function (MSE) has been used at each iteration to adjust the network weights and biases. In this study, a mean square error of 10^{-6}, a minimum gradient of 10^{-10}, and maximum iteration number (epoch) of 500 have been used. The training process terminates if any of these conditions are satisfied.

2.5 Radial basis function neural network

An RBF network generally consists of three layers: the input layer, the hidden layer, and the output layer. The input layer is the same as an MLP. The hidden layer consists of radial function neurons. A radial function has the form of $g(x) = g(\|\mathbf{x} - \mathbf{c}_j\|)$, which is symmetric with respect to $\mathbf{c}$. The $\mathbf{c}$ can also be called the center of the function. When a vector is feeding into an RBF network, each hidden neuron generates a value according to how close the input vector is from the center of the RBF. If the input space vector is close to the center, the hidden neuron generates a value that is close to 1. Then, the outputs of hidden neurons are combined linearly by vectors to generate the output based on the following equation.

$$y_k = w_0 \sum_{j=1}^{m} w_{jk}\, \varphi(\|\mathbf{x} - \mathbf{c}_j\|) \tag{8}$$

where $\varphi(\cdot)$is the radial basis function, w_{jk}, j=(1,2,…,m), and k=(1,2,…,l) are the output weights, w_0 is the bias, $\mathbf{x}$ are the inputs to the network, $\mathbf{c}_i$ are the centers associated with the basis function, m is the number of hidden neurons, and l is the number of classes. In this paper, the activation function $\varphi(\cdot)$is defined as follows:

$$\varphi\left(\|\mathbf{x} - c_j\| = \exp[\frac{(\mathbf{x} - c_j)^T(\mathbf{x} - c_j)}{2\sigma_j^2}\right. \tag{9}$$

where σ_j^2 is the dispersion or smoothing parameter of the jth basis function. K-means clustering algorithm is employed to provide the basis functions. The number of cluster of the k-mean that should be generated has been selected experimentally by a grid search using all available features.

2.6 Machine learning ensemble

In this research, a machine ensemble technique is introduced as another part of the information fusion approach. In

the approach considered in this paper, a multiple classifier model conducts an ensemble of generally weak and/or diverse classifiers. Then, a pool of opinion is made using a meta-classification decision that makes superior decision to the individual classifier. The diversity of classifiers allows different decision boundaries to be created. The intuition is that each classifier makes a different error, and strategically combining these classifiers can reduce the total error and at least reduces the variance of classification error [28, 29]. In order to improve the accuracy of classification, the base classifiers must have high disagreement between one another in mapping the solution space [30]. Otherwise, if all classifiers map the solution space in a similar manner, only little improvement can be achieved over simply using one of the base classifiers. The machine ensemble technique introduced in this paper is referred to as majority vote ensemble. This ensemble technique combines the classification power of three ML algorithms, which are SVM, MLP, and RBF. Figure 2 illustrates general architecture of TCM model using machine ensemble approach, which utilizes all extracted sensory features that has been selected by a feature selection method (CFS in this illustration).

More specifically, the majority vote ensemble is achieved by combining the aforementioned base classifiers: SVM, MLP, and RBF. Under this ensemble scheme, each classifier is trained with the same data set. When the testing set is applied to all base classifiers involved in the ensemble, the class with the most number of predictions is voted to be the final prediction. In general, a majority vote classifier is defined as:

$$C_{\text{meta}}(X) = \arg \max_{i}^{L} \sum_{j=1}^{B} I\left(C_j(X) = i\right) \tag{10}$$

where $I(\cdot)$is an indicator function, C_j are the classifiers where j=(1,..., B), and L is the number of target classes. It has to be noted that the confidence level of each prediction provided by each base classifier is not considered. Therefore, the resulting vote is unweighted with all base classifiers having equal input to vote.

3 Experimentation

3.1 Experimental setup

Figure 3 shows the experimental setup of this paper to study the design of TCM system. The experiment was conducted by using an OKUMA ES 3016 CNC vertical machining center for machining AISI 4340 steel using Kennametal (type: HEC500S2; 12.7 mm diameter) general purpose solid carbide two-flute end mill coated with a ground physical vapor deposited multilayer coating of titanium nitride/titanium carbo-nitride/titanium nitride (TiN/TiCN/TiN).

An acoustic emission (AE) sensor, manufactured by Physical Acoustic Corporation, (PAC-Wsα), was used to capture the AE signal generated during machining operation. The AE signal was divided into two frequency bins. The first one was created by a band pass filter of 100–300 kHz using a linear filtering with third-order Butterworth filters. The second frequency bin also has a third-order Butterworth band pass filter of 300–600 kHz. In addition, to avoid aliasing in AE signal, the sampling rate was set to 1.5 MHz, which is a little over the Nyquist sampling rate of 1.2 MHz. Gain was set to 40 dB. Also, a triaxial accelerometer manufactured by Kistler (Type 8692C50) that simultaneously measures vibration in three mutually perpendicular axes (x, y, and z) was mounted on the spindle used to measure the vibration during cutting operation. The sensor was connected to a Kistler (type 5134) coupler, which provides a DC power and a signal processing by

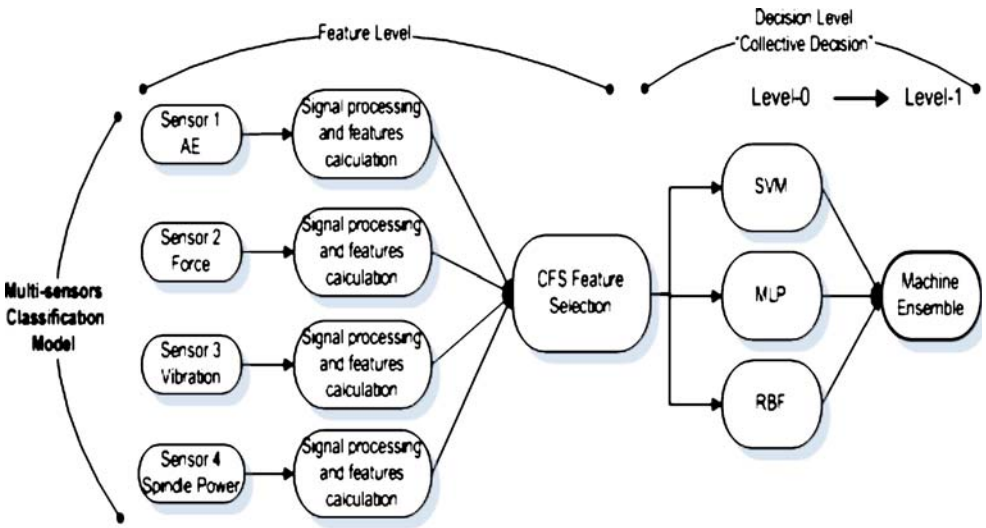

Fig. 2 Example of a multisensor fusion TCM system utilizing correlation-based feature selection method and a machine learning ensemble technique

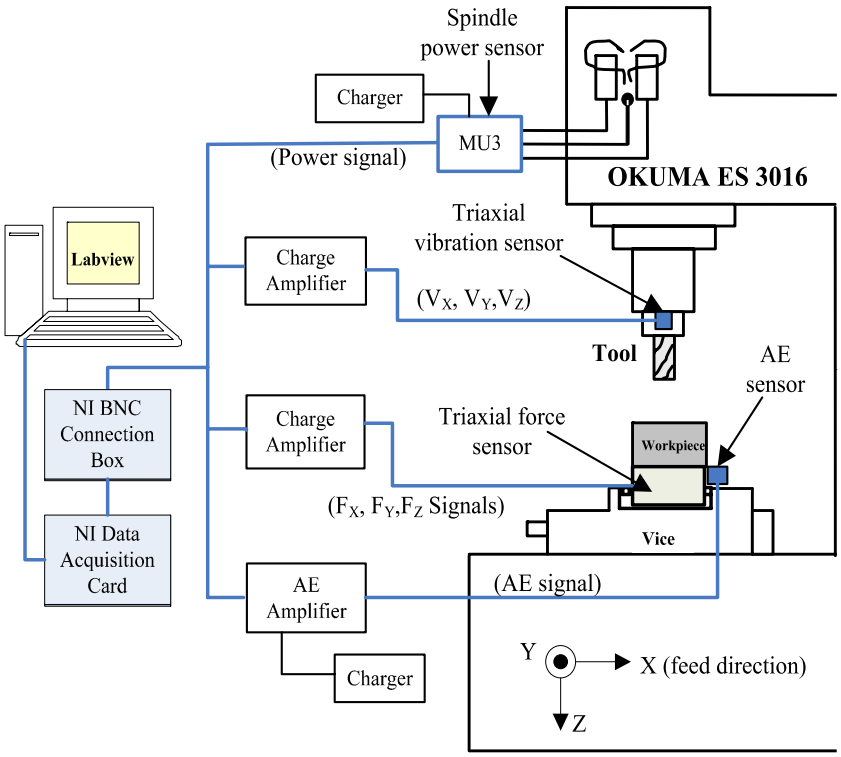

Fig. 3 Schematic diagram of experimental setup

adjustable gains and cut-off frequencies. The gain of the coupler was selected to be 10× for (*X*, *Z*) signals and 5× for (*Y*) signal. The filtering was digitally accomplished by using LabVIEW software. An IIR filter with an order of 29 and a cut-off frequency of 3,000 kHz was selected for all *x*, *y*, and *z* vibration signals. In addition, a quartz three-component dynamometer manufactured by Kistler (type: 9257B) was connected to a charge amplifier (Kistler, type: 5010B) and mounted on the machining table under the job to measure the three orthogonal components of force. A band pass filter of (30–3,000 Hz) was applied on each of the axial force signals. The last sensor used in this experiment was a true power measuring transducer MU3 manufactured by Artis Systems, which was used in combination with two hall sensors (model: LT-100S) to measure true power of spindle motor. All sensors were connected through BNC cables to a National Instrument noise rejecting shielded BNC connection box that acts as a gateway for all the eight signals. The connection box then sends all of the eight signals to a National Instrument data acquisition card (model: NI PCI-6133), which has the ability to convert the signals from analog to digital with a high sampling rate of 3 MB/s/channel. Finally, all the digital signals are properly filtered and analyzed by LabVIEW software. The software extracts predefined features in the time and frequency domains as defined in Section 2.2. These features have been set as the input features (predictors) to the TCM classification model. An identical sampling time of all the sensors signals, which is 200 ms, was used. The flank wear was measured by a microscope (Carl Zeiss Axioskop 2 Mat), which has a high-resolution digital camera (Axiocam MRC™). The combination of microscope, digital camera, and Axiovision software was used to acquire, edit, measure, and store images in conjunction to measure the flank wear of the tool and any abnormalities such as edge chipping.

3.2 Design of experiment

It is desirable that TCM model reflects the conditional changes in cutting tools under diverse cutting conditions such as different level of cutting speed, feed rate, and depth of cut. Therefore, in this study a 2^3 full-factorial design with three replications was selected in order to effectively capture the relationship between the milling process parameters (independent variables) and the calculated signal features (dependent variables). Specifically, three factors (independent variables) used for the design of experiment in this study were surface speed, chip load, and axial depth of cut. Each factor has two levels, i.e., High (H) and Low (L). These levels are provided in Table 3, and in total, there are eight cutting conditions or treatments per replication. In Table 3, different cutting conditions of an experiment are represented using three-letter notations such that the first letter refers to the depth of cut, the second letter defines the cutting speed, and the third letter refers to the feed rate. The radial depth of cut, also referred to as immersion, was kept constant at 11.1 mm throughout the

Table 3 Machining parameters and their levels

Depth of cut (mm)	Cutting speed (m/min)	Feed rate (mm/tooth)	
2.54	122	0.08	0.13
		LLL	LLH
	152	LHL	LHH
3.56	122	0.08	0.13
		HLL	HLH
	152	HHL	HHH

experiments. This radial depth of cut is about 80% of tool diameter (12.7 mm), and this insures that all teeth of the milling cutter (two-flute tool cutter in this experiment) are in contact with the workpiece during end milling.

3.3 Tool condition classes

In this research, a tool condition is classified based on the following three criteria: (1) if the tool is worn due to flank wear, (2) if the tool is worn due to chipping, and (3) if the tool breaks. For each criterion, a measurement was established to verify the level of failure. The reason behind differentiating between chipping and breakage is that chipping phenomena cannot be detected for a considerable amount of time and is hardly noticed by the operator. Measurements of wear and chipping were collected as the tool condition progresses during cutting. However, breakage was obtained artificially through the grinding of the tool since it is hard to observe breakage for each treatment within the design of experiment. Therefore, experiments were conducted for five different tool conditions, namely breakage (B), chipping (C), and three states of wear, which are defined as slight (LW), medium (MW), and severe (SW) wear. Table 4 summaries the specification of tool condition classes according to the level of maximum flank wear, i.e., VB_{max} that is illustrated in Fig. 4.

The flank wear (VB_{max}) is the average of the two flank wear readings recorded for each flute. Reading of the flank wear was taken at the end of every third cut until the tool reached its wear criterion of 0.6 mm per each treatment. Chipping of the tool edge was considered valid if there were a chipped area of over 0.04 mm^2. Otherwise, the tool was classified by its flank wear level. Tool breakage class was defined for a tool that has a breakage area over 0.36 mm^2. Figure 5 provides the distribution of classes per cutting condition after the experiment was conducted. This distribution clearly shows that at LHL condition, tool life increased with no chipping occurred so that various data points can be collected for further analysis.

4 Result and discussion

4.1 Classifier training and evaluation

The notion of training and testing a dataset is fundamental to the performance of ML algorithms. In this paper, the training set contains examples of signals features (135 features) and/or machining parameters features, i.e., speed, feed, and depth of cut, from different classes (tool conditions), and this is used to build the classification model. The testing set represents the unknown sensory information that can be classified. Both testing and training sets are labeled with appropriate class a priori. To improve training process of a classifier, normalization was applied to all training features as a preprocessing step. To test and compare algorithms, we used ten times repeated 10-fold stratified cross-validation (CV), where accuracy results were averaged across replications to minimize type I error [31]. This means that each classification model was trained on nine tenths of the total data and tested on the remaining tenth. This process is repeated ten times, each with a different partitioning seed, in order to account for the variance between partitions. Also, stratification was applied to every testing set in order to count for the non-uniform distribution of tool condition classes within the collected dataset as seen in Fig. 4. The accuracy of classification results in this research stands for the percentage of correctly classified instances over the total number of instances. For various accuracy measures and the relationship between the test conditions and the TCM accuracy, refer to the literature [32]. Each result represents the average of 100 runs (10×10 fold CV). A paired t test was applied for pairwise comparisons of classifications algorithms [33].

4.2 Classification under no sensor fusion

In this section, we compare the performance of base classifiers that are based on a single-sensor model for which no fusion is applied. With such restriction, a comparison of features per sensor, which are defined in the Section 2.2, is investigated, and the significance of each individual sensor on tool condition classification is determined. First, 28 features were extracted from AE sensor,

Table 4 Definition of the tool conditions classes

Tool condition class	(LW) mm	(MW) mm	(SW) mm	(C) mm^2	(B) mm^2
Tool features	0< wear< 0.25	0.25< wear< 0.4	0.4< wear< 0.6	0.04<C. area< 0.36	Brk. area> 0.36

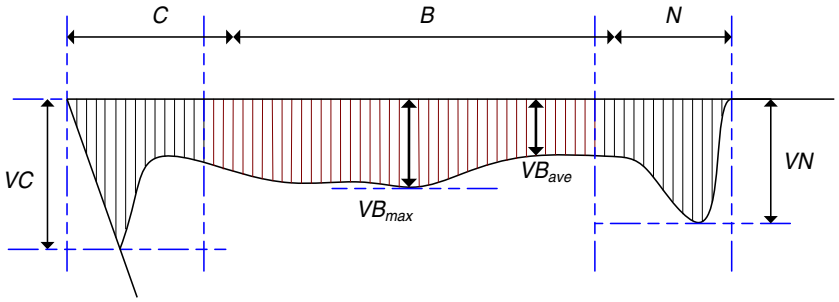

Fig. 4 Flank wear boundaries at the tool-cutting edge

and only 13 features were selected by CFS and χ^2 statistics. These features were used to compare the classification accuracy of different TCM architectures using different ML classifiers. The classification results are shown in Fig. 6, where AE_All, AE_CFS, and AE_Chi2 represent accuracy with all feature selected, reduced features by CFS and χ^2 statistics, respectively. It is observed that SVM outperforms both types of neural networks (MLP, RBF). Figure 6 also shows that CFS can increase the accuracy significantly. Interestingly, it is shown that the accuracy is not improved by using machine ensemble, and it is conjectured as follows: To improve the accuracy by using machine ensemble, individual classifiers must have high disagreement-maintaining diversity. The result shown in Fig. 6 implies that in case of features obtained from AE sensor, the decision boundaries generated by different classifiers are similar and thus produce insignificant improvement.

Next, Fig. 7 shows the classification accuracy of different designs of TCM systems when only force sensors are used. It is observed that when the force sensor is used, 17 features were selected out of 54 features extracted and SVM outperforms MLP and RBF. It is also observed that like the case of AE sensor, CFS improves the accuracy. In addition, Fig. 7 shows that machine ensemble for feature subset F_CFS results in the highest accuracy with 91.98%. This result implies that base classifiers for F_CFS ensemble have high diversity in representing the solution space, which allows improved accuracy and stability in the classification. Improved stability can be captured by standard deviation that is shown in the right side graph. It should be pointed out here that force sensor-based TCM can provide higher accuracy than AE sensor-based TCM.

Figure 8 shows the classification accuracy of different designs of TCM system when only vibration sensors are used. In this case, 54 features were extracted, and they were reduced to 16 by selection methods. Like the case of AE and force sensor, the application of SVM and CFS improves the accuracy with 88.84%. While the application of machine ensemble does not improve the accuracy, it has improved the stability as shown in the right side figure.

When spindle power is measured, 11 features were extracted, and they were further reduced to six features by selection methods. Figure 9 shows that SVM outperforms MLP and RBF. It also shows that the application of feature selection methods and machine ensemble does not improve accuracy. We conjecture that the dimension of extracted features is not sufficiently large in this case.

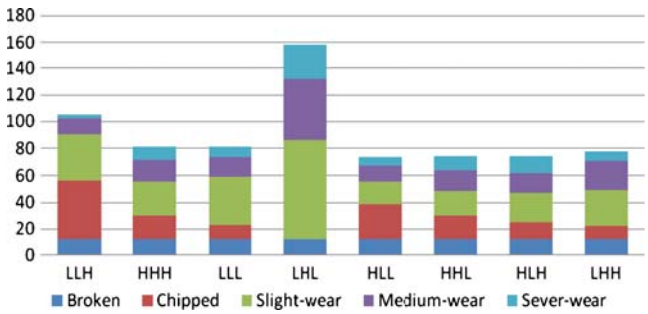

Fig. 5 Distribution of classes per cutting condition for 758 experimentally collected instances (first stack on bottom, *broken*; second stack, *chipped*; third stack, *slight wear*; fourth stack, *medium wear*; fifth stack, *severe wear*)

4.3 Classification using multisensor fusion

In this section, no fusion constraint in Section 4.2 is relaxed so that sensor fusion is applied to the information provided by each sensor. Specifically, in this paper, we test two different designs for multisensor fusion: (1) decision level fusion and (2) feature level fusion.

4.3.1 Decision level fusion

In this fusion method, individual TCM model using single sensor that is explained in previous section acts as an expert within its own feature space. Then, a pool of opinion is made by using a majority vote rule to provide a meta-classification decision that would be superior to individual classifiers. In this analysis, we need to find sensor combinations that can give us the best accuracy under a specific classifier as all the analysis required has been conducted under no fusion assumption in the previous section. The rank of single sensors is force (*F*), vibration

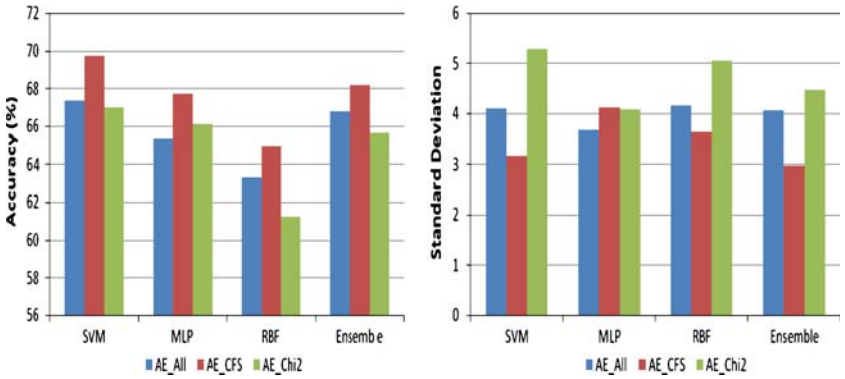

Fig. 6 Classification with different feature selection and machine learning methods when using AE sensor (*left column* AE_All, *middle column* AE_CFS, *right column* AE_Chi2): accuracy (*left*), standard deviation (*right*)

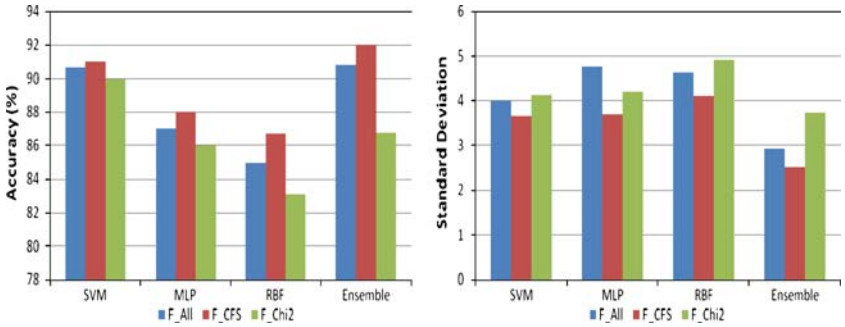

Fig. 7 Classification with different feature selection and machine learning methods when using force sensor (*left column* F_All, *middle column* F_CFS, *right column* F_Chi2): accuracy (*left*), standard deviation (*right*)

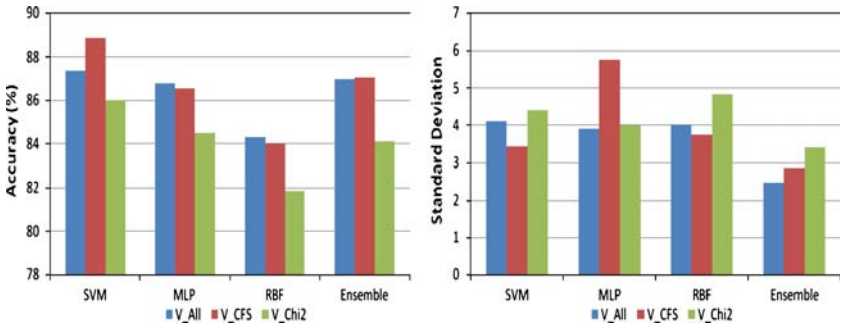

Fig. 8 Classification with different feature selection and machine learning methods when using vibration sensor (*left column* V_All, *middle column* V_CFS, *right column* V_Chi2): accuracy (*left*), standard deviation (*right*)

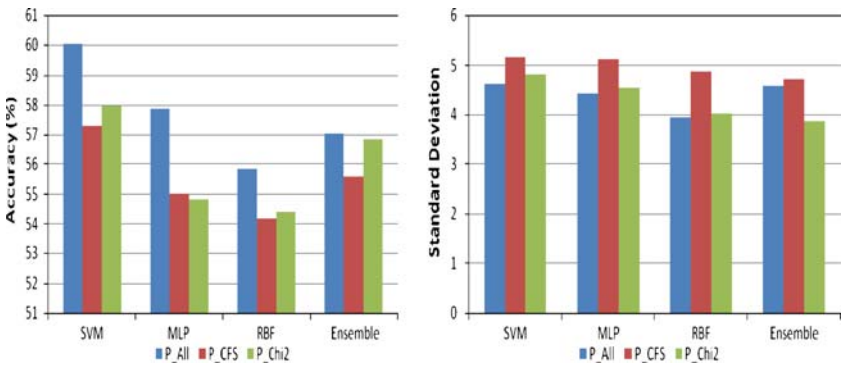

Fig. 9 Classification with different feature selection and machine learning methods when using spindle power sensor (*left column* P_All, *middle column* P_CFS, *right column* P_Chi2): accuracy (*left*), standard deviation (*right*)

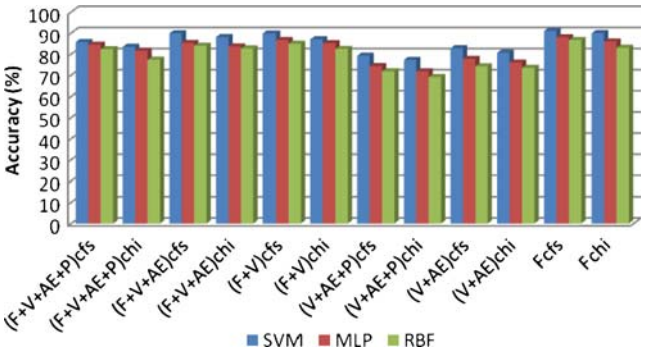

Fig. 10 Classification accuracy of different sensor combinations and feature selection methods (*left column* SVM, *middle column* MLP, *right column* RBF)

Table 5 Percent accuracy of tool condition classification obtained by using two-sensor combined TCM models with feature level fusion (all: all features; cfs: reduced by cfs; chi2: reduced by χ^2 statistics)

	All	CFS	Chi2
AE + *V*			
SVM	89.82±3.11	90.01±2.98	87.13±2.66
MLP	88.39±2.93	86.46±2.73	83.83±3.51
RBF	86.02±2.83	86.52±2.48	84.67±3.03
Ensemble	90.92±2.59	89.57±2.43	86.21±2.81
AE + *F*			
SVM	91.94±2.97	91.63±2.67	90.18±2.79
MLP	89.74±3.03	89.56±2.91	87.29±2.93
RBF	88.32±3.01	86.94±2.34	83.16±2.89
Ensemble	92.10±2.68	92.81±2.70	91.39±2.83
F + *P*			
SVM	90.11±3.97	89.79±3.59	89.52±3.50
MLP	87.12±3.83	87.73±3.42	86.19±3.83
RBF	85.05±4.24	85.97±3.90	84.15±4.71
Ensemble	89.46±2.99	88.85±2.70	86.57±3.18
AE + *P*			
SVM	71.96±3.16	71.46±3.72	70.63±3.57
MLP	70.51±3.97	69.47±3.50	66.31±4.20
RBF	69.44±3.92	64.31±3.61	64.19±3.80
Ensemble	71.47±2.91	71.08±2.34	70.37±2.97
F + *V*			
SVM	94.48±2.24	94.63±2.24	93.26±2.21
MLP	92.66±2.21	91.41±3.01	90.41±2.74
RBF	91.61±2.91	90.97±2.71	88.17±3.06
Ensemble	95.71±2.03	96.22±2.23	94.16±2.41
V + *P*			
SVM	88.20±3.73	89.17±3.50	88.56±3.61
MLP	87.98±3.32	88.15±4.50	85.61±3.97
RBF	86.19±3.83	86.03±3.59	84.92±4.50
Ensemble	90.15±2.41	89.63±2.27	87.21±2.57

Table 6 Percent accuracy of tool condition classification obtained by using three-sensor combined TCM models with feature level fusion (all: all features; cfs: reduced by cfs; chi2: reduced by χ^2 statistics)

	All	CFS	Chi2
AE + *F* + *V*			
SVM	96.23±2.21	95.89±2.03	93.99±2.49
MLP	94.83±2.36	93.98±2.29	92.42±2.88
RBF	93.21±2.20	93.72±2.27	91.38±2.63
Ensemble	96.97±1.73	97.67±1.39	94.51±2.73
AE + *V* + *P*			
SVM	91.08±2.94	89.78±2.56	87.46±2.67
MLP	89.96±2.87	88.79±2.97	86.03±3.35
RBF	87.01±3.27	86.15±3.01	84.73±3.61
Ensemble	93.55±2.60	91.58±2.53	90.09±2.94
AE + *F* + *P*			
SVM	92.53±2.67	91.10±2.97	89.13±3.05
MLP	89.91±2.53	89.04±2.80	87.62±3.01
RBF	87.85±2.41	87.19±2.37	85.54±3.49
Ensemble	93.51±2.33	94.27±2.19	92.94±2.69
F + *V* + *P*			
SVM	94.99±2.19	94.44±2.19	93.01±2.46
MLP	92.43±2.91	92.19±2.79	90.52±3.12
RBF	91.50±3.53	90.53±2.91	89.09±3.88
Ensemble	95.57±2.04	95.82±1.91	94.38±2.59

(*V*), acoustic emission (AE), and spindle power (*P*) in terms of their order of accuracy. Therefore, we test the following sensor combinations: (*F*+*V*+AE+*P*), (*F*+*V*+AE), (*F*+*V*), (*V*+AE+*P*), and (*V*+AE) with different ML classifiers (SVM, RBF, and MLP) and selection methods (CFS and χ^2 statistics). Figure 10 shows the classification accuracy of these sensor combinations with two different feature selection methods (CFS and χ^2 statistics). It is shown in the figure that none of those sensor combinations based on decision level fusion outperforms single-sensor TCM model with force sensor. Nonetheless, we can observe that SVM outperforms other neural networks (MLP and RBF) as the base classifier.

Table 7 Percent accuracy of tool condition classification obtained by using four-sensor combined TCM models with feature level fusion (all: all features; cfs: reduced by cfs; chi2: reduced by χ^2 statistics)

	AE + *F* + *V* + *P*		
	All	CFS	Chi2
SVM	95.91±2.34	95.56±2.40	93.91±2.67
MLP	94.23±2.31	93.59±2.35	92.34±3.10
RBF	93.04±2.14	93.49±2.19	91.49±2.59
Ensemble	97.32±1.98	97.28±1.70	94.43±2.21

4.3.2 Feature level fusion

In this fusion method, features from multiple sensors are combined into a single set, and then their informational combinatory power is fed to the classifiers for training purpose. This implies that individual classifier experiences a larger input space and higher volume of information and thus may increase the accuracy and robustness of the classification. It should be pointed out that when features from different sensors are fused, the performance of ML algorithms may not be as predictable as in the decision level fusion because in decision level fusion the performance of base classifiers is known from no fusion analysis a priori. Therefore, in feature level fusion, a total of 11 combinations of sensors are tested, which are one four-sensor combination, four three-sensor combinations, and six two-sensor combinations. The number of features is again reduced by CFS and χ^2 statistics. Tables 5, 6, and 7 summarize the classification accuracy achieved by two-, three-, and four-sensor combinations for feature fusion-based TCM system. It is observed from these tables that the best performance achieved in case of two-sensor combination is 96.22%, particularly by force and vibration sensor combination with CFS selection method and machine ensemble. It is also observed that in the case of three-sensor combination, the best performance achieved is 97.67% by force, vibration, and acoustic emission sensor combination with CFS selection method, and the best performance is 97.32% in case of all sensor combination without feature reduction. From the accuracy and robustness perspectives, we suggest to use force, vibration, and acoustic sensor combination with CFS feature selection method and machine ensemble technique for multisensor fusion TCM system. If this combination of sensors is not available, the alternative is to use a force and vibration sensor combination with CFS and machine ensemble technique. It should be emphasized here that overall, SVM outperforms other machine learning classifiers (MLP, RBF neural network) in all sensor combinations tested.

5 Conclusion

This paper studied the design of multisensor fusion-based TCM system in end-milling process. Specifically, we focused on the accuracy and robustness of tool condition classification when different design components are considered for multisensor fusion-based TCM system, for example, different sensor combinations, ML algorithms (SVM, MLP, and RBF), and fusion methods (feature and decision level). The results from the experiment have shown that the SVM outperforms other neural network-based algorithms (MLP and RBF) due to its nature of structural risk minimization. It is also shown that the introduction of machine ensemble (majority vote) has relatively improved the accuracy of tool condition classification. This is true especially:

- where there exist more complementary features from each sensor within the fused feature set
- where there exists a high degree of diversity (in mapping of the solution space) within the ensemble base classifiers

It has also been shown from the experiment that CFS can improve the accuracy and robustness of the classification. This is important because CFS can reduce feature space considerably in general. It should be pointed out here that most of existing research on TCM systems for end-milling operations focuses on a specific condition of the cutting tools such as estimation of wear level and breakage detection (normal or broken), and thus, chipping is mostly ignored as a critical tool condition in the literature. As a conclusion, this study can measure every possible tool condition under one TCM system. The results from the experiments show that our method can classify a multiflute end-milling tool with great accuracy and optimally reduced features and specific sensor combination. In addition, irrelevant sensors and features have been observed. This facilitates more efficient modeling of the TCM because a selection of 25 feature subset size proves to be more accurate and robust than the inclusion of the entire set of 138 explanatory features, which can significantly reduce computational effort of the TCM system.

References

1. Kegg RL (1984) On-line machine and process diagnostics. Ann CIRP 32(2):469–473
2. Kurada S, Bradley C (1997) A review of machine vision sensors for tool condition monitoring. Comput Ind 34:55–72. doi:10.1016/S0166-3615(96)00075-9
3. Elbestawi MA, Papazafiriou TA, Du RX (1991) Process monitoring of tool wear in milling using cutting force signature. Int J Mach Tools Manuf 31(1):55–73. doi:10.1016/0890-6955(91)90051-4
4. Lee BY, Tarng YS (1999) Milling cutter breakage detection by discrete wavelet transform. Mechatronics 9:225–234. doi:10.1016/S0957-4158(98)00049-X
5. Bhattacharyya P, Senupta D, Mukhopadhaya S (2007) Cutting force based real-time estimation of tool wear in face milling using a combination of signal processing techniques. Mech Syst Signal 21(6):2665–2683. doi:10.1016/j.ymssp.2007.01.004
6. Chen JC, Chen W (1999) Tool breakage detection system using accelerometer sensor. J Intell Manuf 10:187–197. doi:10.1023/A:1008980821787
7. Yesilyurt I, Ozturk H (2007) Tool condition monitoring in milling using vibration analysis. Int J Prod Res 45(4):1013–1028. doi:10.1080/00207540600677781
8. Atlas L, Ostendorf M, Bernard GD (2000) Hidden Markov models for monitoring machining tool wear. IEEE International Conference on Acoustics, Speech, and Signal Processing 6:3887–3890

9. Tansel IN, Trujillo ME, Bao WY (2001) Acoustic emission-based tool breakage detector for micro-end milling operations. Int J Model Simul 21(1):10–16
10. Ghosh N, Ravi YB, Patra A, Mukhopadhyay S, Paul S, Mohanty AR, Chattopadhyay AB (2007) Estimation of tool wear during CNC milling using neural network based sensor fusion. Mech Syst Signal Process 21:466–479. doi:10.1016/j.ymssp.2005.10.010
11. Cho S, Asfour S, Onar A, Kaundinya N (2005) Tool breakage detection using support vector machine learning in a milling process. Int J Mach Tools Manuf 45(3):241–249. doi:10.1016/j.ijmachtools.2004.08.016
12. Norman P, Kaplan A, Rantatalo M, Svenningsson I (2007) Study of a sensor platform for monitoring machining of aluminum and steel. Meas Sci Technol 18:1155–1166. doi:10.1088/0957-0233/18/5/001
13. Reddy YB (1992) Multisensor data fusion: state of the art. J Inf Sci Technol 2(1):91–103
14. Rehorn AG, Jiang J, Orban PE (2005) State-of-the-art methods and results in tool condition monitoring: a review. Int J Adv Manuf Technol 26:693–710. doi:10.1007/s00170-004-2038-2
15. Elbestawi MA, Marks J, Papazafiriou T (1989) Process monitoring in milling by pattern recognition. Mech Syst Signal Process 3(3):305–315. doi:10.1016/0888-3270(89)90055-1
16. Silva RG, Reuben RL, Wilcox SJ (1998) Tool wear monitoring of turning operation by neural network and expert system classification of a feature set generated from multiple sensors. Mech Syst Signal Process 12(2):319–332. doi:10.1006/mssp.1997.0123
17. Sick B (2002) 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 16(4):487–546. doi:10.1006/mssp.2001.1460
18. Brezak D, Udiljak T, Majetic D, Novakovic B, Kasac J (2004) Tool wear monitoring using radial basis function neural network. Proc IEEE Int Jt Conf Neural Netw 3:1859–1862
19. Yuan S, Chu F (2006) Support vector machines-based fault diagnosis for turbo-pump rotor. Mech Syst Signal Process 20:939–952. doi:10.1016/j.ymssp.2005.09.006
20. Hall MA (1999) Correlation-based feature selection for machine learning. PhD dissertation, The University of Waikato, New Zealand
21. Fayyad UM, Irani KB (1993) Multi-interval discretization of continuous-valued attributes for classification learning. IJCA 93:1022–1027
22. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324. doi:10.1016/S0004-3702(97)00043-X
23. Hong SJ, Raman M, Wong YS (2002) Feature extraction and selection in tool condition monitoring system. Lecture notes in computer science. Springer, Berlin, pp 487–497
24. Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco 2005
25. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999. doi:10.1109/72.788640
26. Rumelhart DE, McClelland JL (1986) Parallel distributed processing: explorations in the microstructure of cognition. MIT Press, Boston
27. Bishop CM (2006) Pattern recognition and machine learning. Springer Science and Business Media, LLC., New York
28. Kittler J, Hatef M, Duin R, Matas J (1998) On combining classifiers. IEEE Trans Pattern Anal Mach Intell 20(3):226–239. doi:10.1109/34.667881
29. Hansen LK, Salamon P (1990) Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 12(10):993–1001. doi:10.1109/34.58871
30. Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach Learn 40(2):139–157. doi:10.1023/A:1007607513941
31. Bouckaert RR (2003) Choosing between two learning algorithms based on calibrated test. Proceedings of 20th International Conference on Machine Learning, Morgan Kauffmann
32. Palanisamy P, Rajendran I, Shanmugasundaram S (2008) Prediction of tool wear using regression and ANN models in end-milling operation. Int J Adv Manuf Technol 37:29–41. doi:10.1007/s00170-007-0948-5
33. Wolpert DH (1992) Stacked generalization. Neural Netw 5:241–259. doi:10.1016/S0893-6080(05)80023-1