Journal of Materials Processing Technology 209 (2009) 4728–4738

Contents lists available at ScienceDirect

Journal of Materials Processing Technology

journal homepage: www.elsevier.com/locate/jmatprotec

Machine ensemble approach for simultaneous detection of transient and gradual abnormalities in end milling using multisensor fusion

Sultan Binsaeid [a], Shihab Asfour [a,*], Sohyung Cho [b], Arzu Onar [c]

[a] Department of Industrial Engineering, University of Miami, Coral Gables, FL 33146, United States
[b] Industrial and Manufacturing Engineering, Southern Illinois University Edwardsville, Edwardsville, IL 62026, United States
[c] Department of Biostatistics, St. Jude Children's Research Hospital, 332 North Lauderdale St., Mail Stop 768, Memphis, TN 38105-2794, United States

ARTICLE INFO

Article history:
Received 23 June 2008
Received in revised form
26 November 2008
Accepted 29 November 2008

Keywords:
Multisensor fusion
Tool condition monitoring
Machine ensemble
Feature selection

ABSTRACT

In a fully automated manufacturing environment, instant detection of the cutting tool condition is essential for the improved productivity and cost effectiveness. This paper studies a tool condition monitoring system (TCM) via machine learning (ML) and machine ensemble (ME) approach to investigate the effectiveness of multisensor fusion technique when machining 4340 steel with multilayer coated and multiflute carbide end mill cutter. In this study, 135 different features are extracted from multiple sensor signals of force, vibration, acoustic emission and spindle power in the time and frequency domain by using data acquisition and signal processing module. Then, a correlation-based feature selection technique (CFS) evaluates the significance of these features along with machining parameters collected from machining experiments. Next, an optimal feature subset is computed for various assorted combinations of sensors. Finally, machine ensemble methods based on majority voting and stacked generalization are studied for the selected features to classify not only flank wear but also breakage and chipping. It has been found in this paper that the stacked generalization ensemble can ensure the highest accuracy in tool condition monitoring. In addition, it has been shown that the support vector machine (SVM) outperforms other ML algorithms in most cases tested.

1. Introduction

Machining cutters suffer from various types of tool faults, which can degrade the quality of work produced, if not controlled. In general, there are three major tool faults associated with end-mill cutters: gradual tool wear, tool chipping and tool breakage. All of these faults are different in their nature: Tool breakage occurs abruptly, in a transient, observable, and random manner; tool chipping has the same characteristic as breakage except it is hardly detected for a long time and cannot be noticed by the operator, whereas tool wear develops gradually and can be predicted to a certain extent. If the machining process continues with a worn, chipped, or broken tool, the dimensional accuracy, surface quality of finished components and even process stability are deteriorated. Kegg (1984) reported that tool faults contribute to downtime of machining centers with average of 7%. Note that this estimate is more conservative than the one reported by Kurada and Bradley (1997) where the estimate is 20%. These notions of significant impact of tool faults on machine downtime have led to a conservative estimation of available tool life. Consequently, manufacturing industries have underutilized machining cutters in general and the frequency of tool changes has increased leading to increased total cost of machining as pointed by Dan and Mathew (1990).

Recently, the advances in process monitoring and signal processing have brought the usefulness of sensor systems closer to industrial realization. Specifically, a wide range of sensors that can retrieve information about machining process such as tool condition and surface roughness has been implemented. Rehorn et al. (2005) reported that by the utilization of sensor systems for the monitoring of metal cutting operations and machine tools, it is anticipated that significant reduction in machining errors, with better surface finish, and optimized machine system availability can be achieved. As an effort to use the wide range of sensors, Norman et al. (2007) studied a multisensor platform for monitoring the machined roughness of steel and aluminum at varying cutting parameters. Specifically, a multisensor platform consisting of force sensors, accelerometers and laser Doppler vibrometry (LDV) was used to correlate signal values with surface roughness values. The experimental setup of this research was based on their previous work, which was introduced in Norman et al. (2006) for more details. With these advances in process technology and monitoring,

* *Corresponding author at*: Department of Industrial Engineering, University of Miami, 1251 Memorial Drive, Rm. 268, Coral Gables, FL 33146, United States. Tel.: +305 607 7676; fax: +305 284 4040.
E-mail addresses: sultan1st@hotmail.com (S. Binsaeid), sasfour@miami.edu (S. Asfour), scho@siue.edu (S. Cho), arzu.onar@stjude.org (A. Onar).

0924-0136/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmatprotec.2008.11.038

an intelligent tool condition monitoring (TCM) using multisensor systems has attracted closer attention from academia and industry because successful application of TCM during machining can ensure high quality parts and preserve the machining systems.

The utilization of multisensor systems for TCM is intended to fuse the informational power of each unique sensor to provide complementary and redundant information about conditional changes in cutting tools, which is referred to as sensor fusion. In these multisensor systems, signal processing techniques extract sets of features that are sensitive to the tool condition. In the last decade, various pattern classification methods have been applied in the field of multisensor TCM to ensure high level of accuracy in prediction or classification results. For example, Reddy (1992) emphasized that pattern recognition can be an effective sensor fusion strategy in TCM. However, the level of complexity and robustness of the TCM model has been rarely part of the design objectives. The possibility of examining different model architectures in a more systematic manner has not been clearly investigated in the literature of multisensor TCM. Therefore, in this research various TCM systems employing different multisensor combinations for multiclass pattern recognition are compared in regard to their complexity of design, generalization as well as accuracy of classification. Specifically, in this research classification models for the aforementioned tool faults are studied, given their gradual and transient tool faults nature. Note that three classes are defined to describe three different states of flank wear progression, while two extra classes are assigned for the tool chipping and breakage. The overall aim of this paper is to draw a general conclusion about:

- The significance of extracted sensory signal features and, consequently, the minimum number of required features and type of sensors employed.
- The effectiveness of multisensor fusion and optimal sensor combination.
- The accuracy of proposed classification algorithms and ensemble schemes.

The following three important objectives are considered 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 that achieve higher level of accuracy for tool fault classification using correlation-based feature selection (CFS). The second objective is to study the significance of fusion strategies for 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 networks (MLP), radial base function neural networks (RBF), and multiclass support vector machine (SVM). Furthermore, these three classifiers are studied under two different ensemble approaches, which are majority vote and stacked generalizations.

The rest of the paper is organized as follows: Section 2 explains the components involved in the TCM system. A detailed review is given of the multisensory data acquisition system, signal processing methods and their extracted features, and feature selection method. Then, a general introduction is given of each ML algorithm and machine ensemble approach under study. Section 3 outlines the experimental setup and design of experiment details, and the definition of tool condition classes. A discussion and performance results of all constructed TCM models are provided in Section 4. Conclusion is provided in Section 5.

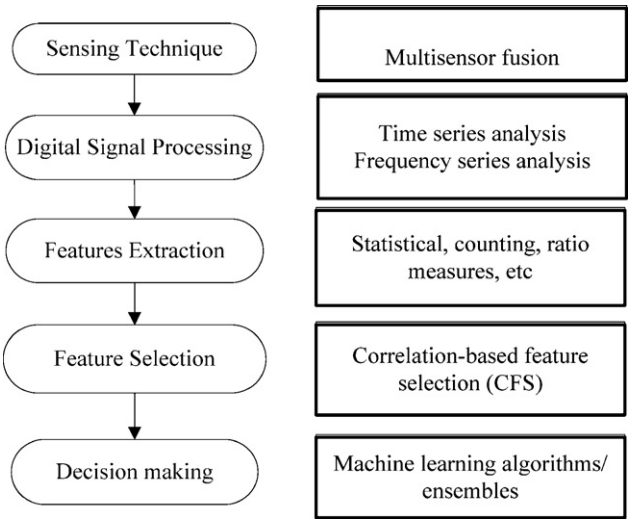

Fig. 1. General components of proposed TCM models.

2. TCM model architect

As illustrated in Fig. 1, the tool condition monitoring system proposed in this study consists of five components: (1) multisensor data acquisition system, (2) signal processing, (3) feature extraction, (4) feature selection, and (5) ML base\ensemble classifier.

2.1. Multisensor data acquisition system

In this research, a collection of machine learning (ML) techniques are proposed to classify the fault conditions including wear of a general purpose solid carbide end mill coated with a ground physical vapor deposited (PVD) during cutting. Tool condition is sensed by triaxial force, triaxial vibration, acoustic emission, and spindle power sensor. These four sensors account for eight sensory signals since the force and vibration sensors measure the machining process in three axes. Each piece of processed sensory information is fused into a pattern classification model which is constructed depending on the number of sensors involved, feature selection method, and classifier type.

Reddy (1992) has pointed out that the advantage in fusing the outputs from one sensor with those from another independent sensor stems from redundancy being present in the information. If redundant sensors are employed, by averaging out the independent noise processes acting on the different sensors, the overall uncertainty of the resulting measurement can be reduced, improving the performance of the system. This is possible, since, to a large extent, the noise inherent in individual sensor measurements will not correlate with noise from other sensors. In addition, complementary sensors provide extended and independent information about the process which cannot be captured otherwise.

2.2. Feature extraction and signal processing

Different research studies on TCM have adapted various feature extraction methods to capture sensory information and calculate its features. The purpose of feature extraction is to greatly reduce the dimension of the raw signal but at the same time maintain the relevant information of tool condition in the extracted features. The majority of these feature extraction methods are reported in Sick (2002) and Rehorn et al. (2005). In the study presented in this paper, various feature measurements that were previously reported by different TCM studies are exploited to assess their capability in extracting meaningful information about tool condition from sensory signals. These studies include Tansel and Mclaughlin (1993) where tool breakage in milling operations was detected by using force sensors and neural network approach, Lei et al. (1999) where tool wear length in finish milling operations was estimated by using

vibration sensors and a self-learning fuzzy inference algorithm, Brezak et al. (2004) where regression technique was employed for tool wear monitoring by using force, acoustic emission, feed current sensor and radial-basis function neural network approach, Cho et al. (2005) where tool breakage in end milling process was detected by using force and spindle power sensor and support vector regression technique, and Ghosh et al. (2007) where tool wear in CNC milling operations was estimated by using force, vibration, spindle current sensor and multilayer perceptron neural network approach. In this research, to assess various feature measurement and extraction methods features obtained from sensory signals are extracted by a constructed program by using LabVIEW, which analyzes the incoming signal from the sensors in the time and frequency domain automatically. Specifically, statistical measurements, counting, and other type of measurements are taken for both time and frequency domain methods as follows.

2.2.1. Time domain analysis

Arithmetic mean (M): The mean of amplitude values of raw data signal. The mean of n amplitude values of a signal $[x_1, x_2, \ldots, x_n]$ is

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

RMS: Root mean square of amplitude values. The RMS for a collection of n values in the raw data is defined as

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{2}$$

Variance (V): The variance of amplitude values. The variance for a collection of n value in the raw data is defined as

$$V = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1} \tag{3}$$

Skewness (Sk): The 3rd central moment and is a measure of the asymmetry of the probability distribution of the signal raw data. It is expressed as

$$Sk = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \mu)^3}{\sigma^3} \tag{4}$$

Kurtosis (Ku): Fourth central moment and is a measure of the "peakedness" of the probability distribution of the signal raw data:

$$\mathrm{Ku} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \mu)^4}{\sigma^4} \tag{5}$$

Power (P): Signal power is defined as the measured area under the rectified signal envelope. This is another measurement of the signal amplitude; however, it is sensitive to amplitude as well as duration, and it is less dependent on operating frequency. Power is defined as

$$P = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{6}$$

Peak-to-peak amplitude (pp): The highest peak value minus the lowest peak value attained by a signal. This peak-to-peak is expressed depending on the signal. For force it is expressed in newtons, for vibration and acoustics in microvolts, and for spindle power signal in kilowatts.

Crest factor (CF): The crest factor of a waveform is equal to the peak amplitude of a waveform divided by the RMS value. The use of the crest factor calculation is to provide an idea of the degree of impacting in a waveform. It is defined by the following formula:

$$\mathrm{CF} = \frac{\mathrm{peak}}{\mathrm{RMS}} \tag{7}$$

The burst rate (Br): Sometimes called the pulse rate, the number of times the signal exceeds pre-set thresholds per second. This feature is only applied to vibration and AE signals. The preset threshold is set to 300 μV.

2.2.2. Frequency domain analysis

The raw data of all signals are transformed into the frequency domain by a fast Fourier transform (FFT). Specifically, a Hanning window is applied as a windowing method to the raw data before FFT to prevent leakage. Then, the power spectral density (PSD) is obtained where it is formed by a plot of the frequency components on the x-axis and attendant power in that frequency on the y-axis. LabVIEW has been programmed to extract features from the PSD of all signals except for the non-stationary signal of spindle power.

Sum of total band power (STPB): The power spectrum does not directly give us the total power in the signal, only power in a particular spectral component. To obtain the total power in the signal or in a particular range, the integral of the PSD over the range of frequencies of interest must be obtained. The following formula defines STPB:

$$\mathrm{STPB} = \int_{F_1}^{F_2} S(f) \tag{8}$$

where $S(f)$ is the power at a specific frequency component and (F_1, F_2) is the frequency band.

Mean of band power (MBP): Mean of power spectrum in a specific frequency band which is defined as

$$\mathrm{MBP} = \frac{1}{n}\sum_{i=1}^{n} S(f)_i \tag{9}$$

Variance of band power (VBP): Variance of power spectrum in a specific frequency band, which is defined as

$$\mathrm{VBP} = \frac{\sum_{i=1}^{n}(S(f)_i - \mathrm{MBP})^2}{n-1} \tag{10}$$

Skewness of band power (SkBP): Skewness of power spectrum in a specific frequency band, which is defined as

$$\mathrm{SkBP} = \frac{1}{n}\frac{\sum_{i=1}^{n}(S(f)_i - \mathrm{MBP})^3}{\mathrm{VBP}^{3/2}} \tag{11}$$

Kurtosis of band power (KuBP): Kurtosis of power spectrum in a specific frequency band, which is defined as

$$\mathrm{KuBP} = \frac{1}{n}\frac{\sum_{i=1}^{n}(S(f)_i - \mathrm{MBP})^4}{\mathrm{VBP}^{4/2}} \tag{12}$$

Maximum (peak) of band power (PBP): The peak of power spectrum in a specific frequency band. This peak is expressed by the energy level (W/Hz).

Frequency of maximum peak of band power (FPBP): The relative frequency that corresponds to the highest amplitude.

Relative spectral peak per band (RSPBP): The ratio of peak of band power (PBP) over the mean of band power (MBP).

Total harmonic band power (THBP): Cutter tooth frequency f_t is calculated by using the equation:

$$f_t = \frac{S}{60}n$$

where S = spindle RPM and n = the number of teeth on the cutter. The main frequency of the cutter body can also be found by

$$f_c = \frac{S}{60}$$

Table 1
Distribution of time and frequency domain features per sensor.

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

The total power of these frequencies and their harmonics can be measured by the following formula:

$$\text{THBP} = \sum_{m=1}^{N} P(m), \quad m = 1, 2, \ldots, N \tag{13}$$

where $P(m)$ is the power at the fundamental tooth frequency, body cutter, and their harmonics, N is the largest integer for which N is the cutoff frequency for the sensor. This feature is only applied to the triaxial force signals.

Table 1 provides the distribution of all extracted features in both time and frequency domain per sensor. There are 135 calculated features which correspond to eight sensory signals: three axial force signals, one acoustic emission signal, three axial vibration signals, and one spindle power signal. In addition to the calculated features, machining parameters are also considered part of the feature space; axial depth of cut, spindle speed, and feed rate. Therefore, the total number of features is 138.

2.3. Feature reduction method

Training a classifier by using the maximum number of features obtainable is not always the best option, as irrelevant and redundant features can negatively influence ML algorithm performance. In order to improve the accuracy of the classification model and increase the efficiency of the computational performance of the TCM system, only an optimal number of significant features should be included in the final model. This can be achieved by reducing the number of features by utilizing feature selection techniques. In this research, a correlation-based feature selection method (CFS) is employed to evaluate different feature subsets (Hall, 1999) and a greedy hill climbing search algorithm is employed to search for optimal subset size.

2.3.1. Correlation-based feature selection method

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

- the level of correlation of individual features with the predicted class;
- the level of intercorrelation among features.

Therefore, high scores are assigned to subsets containing features that are highly correlated with the class, yet have a low intercorrelation measure with each other. Entropy measures are used to obtain a measure of correlation between features and class and between features. All continuous features are discretized by using the technique of Fayyad and Irani (1993). The entropy of a feature Y is given as follows:

$$H(Y) = -\sum_{y \in R_y} p(y)\log(p(y)) \tag{13}$$

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

$$H(Y \mid X) = -\sum_{x \in R_x} p(x) \sum_{y \in R_y} p(y)\log(p(y)) \tag{14}$$

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 the uncertainty coefficient of Y and is calculated as follows:

$$C(Y \mid X) = \frac{H(Y) - H(Y \mid X)}{H(Y)} \tag{15}$$

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

$$\text{Merit}_s = \frac{k\overline{r_{\text{cf}}}}{\sqrt{k + k(k-1)\overline{r_{\text{ff}}}}} \tag{16}$$

where Merit_s is the heuristic of a feature subset S containing k number features, $\overline{r_{\text{cf}}}$ and $\overline{r_{\text{ff}}}$ are the average feature-class correlation and average feature–feature intercorrelation, respectively. In Eq. (16), 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 subsets by using the evaluation function of CFS. A simple yet effective search algorithm such as greedy hill climbing has demonstrated efficiency in searching the feature space in reasonable time and provided good results (Kohavi and john, 1997). 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 a full set of features. Then, backward elimination will continue to delete features as long as a child is not worse than its parents. This process is repeated until no more improvement can be achieved.

2.4. Machine learning classifiers

In this research, three ML classifiers are used to classify tool conditions and ensemble techniques are applied on them to improve classification accuracy. All applied classifiers and reduction methods have been implemented by using WEKA ML suite, which is an established freeware environment supported by many machine learning authorities (Witten and Frank, 2005). All of the three machine learning algorithms employed in this study have proven to be effective in the pattern recognition communities. The first ML algorithm employs the structural risk minimization approach, i.e., (1) support vector machine (SVM). The other two are artificial neural networks which employ an empirical risk minimization approach: (2) multilayer perceptron neural networks (MLP) and (3) radial basis function neural networks (RBF).

2.4.1. Multiclass support vector machine

Recently, the application of support vector machine (SVM), which is based on a statistical learning theory presented by Vapnik (1999), is proposed as a decision-making method. In the last decade, SVM has received a great deal of attention in the pattern recognition literature. While typical ML algorithms attempt to minimize the empirical risk, which is the misclassification errors on the training set, SVM attempts to minimize the structural risk, which is the probability of misclassification of a previously unseen data point drawn randomly from fixed but unseen distribution. 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 research, SVM was implemented by using a sequential minimum optimization (SMO) algorithm (Platt, 1998). The selection of kernel function has influence on the decision boundary. Usually, radial basis functions (RBF) are favored instead of polynomial kernel functions, because they are not sensitive to outliers and do not require inputs to have equal variances. Therefore, a radial basis function (RBF) was selected as a kernel function after preliminary analysis, not shown here. The RBF kernel is defined as

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2) \quad \gamma > 0 \tag{17}$$

where $K(x_i, x_j)$ defines an inner product that maps the input vector $x \in \Re^d$ to a high-dimensional space. In this research, a grid search was 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 resulted in optimal values $\gamma = 0.25$, and $C = 12.0$.

2.4.2. Multilayer perceptron neural networks (MLP)

MLP is the most widely used learning algorithm, due originally to Rumelhart and McClelland (1986) and discussed at length in most neural network textbooks (Bishop, 2006). The learning process of an 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 ($k = 1, 2, \ldots, l$) 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 $(\mathbf{x}_i, \mathbf{c}_i)$ for $i = 1, 2, \ldots, p$ using an Euclidean norm, where p is the number of hidden nodes:

$$E(w) = \frac{1}{2} \sum_{i=1}^{P} \left\| y(x_i, w) - c_i \right\|^2 \tag{18}$$

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. To cope with our multisensor TCM analysis, we have selected a heuristic that varies the number of input nodes depending on the number of features applied to the network. Since the number of inputs changes depending on the number of sensors and features selected, a rule of thumb is to set the hidden layer to be half of the total number of the input and output layers (Witten and Frank, 2005). In addition, the number of output nodes represents the number of tool condition classes: five tool condition classes in this paper. The target value of each output node produces a confidence level that represents the classification probability of each class. In this research, the activation functions of sigmoid were used in the hidden layers and in the output layer, respectively. Moreover, the MLP was created, trained and implemented with a back propagation (BPN) training 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 was performed to select the parameters of the BPN training algorithm. As a result, a learning rate of 0.1 and momentum rate of 0.2 were selected as optimal values. The MLP was trained iteratively to minimize the performance function of mean squared error (MSE) between the network outputs and the corresponding target values. At each iteration a gradient of the performance function (MSE) was used 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 were used. The training process would stop if any of these conditions were met.

2.4.3. Radial basis function neural network (RBF)

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 that have the form of $g(x) = g(\left\| x - c_j \right\|)$, which is symmetric with respect to $\mathbf{c}$. The $\mathbf{c}$ can also be called the center of the function. When a vector is input to 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. The hidden neurons' outputs are combined linearly by vectors to generate the output based on the following equation:

$$y_k = w_0 \sum_{j=1}^{m} w_{jk} \varphi(\left\| x - c_j \right\|) \tag{19}$$

where $\varphi(\cdot)$ is the radial basis function, $w_{jk}, j = (1, 2, \ldots, m)$, and $k = (1, 2, \ldots, l)$ 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 research, the activation function $\varphi(\cdot)$ is described as follows:

$$\varphi(\left\| x - c_j \right\| = \exp[\frac{(x - c_j)^T (x - c_j)}{2\sigma_j^2} \tag{20}$$

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 clusters that the k-mean should be generating has been selected experimentally by a grid search using all available features.

2.5. Machine learning ensemble

In this research, a machine ensemble is introduced as another part of the data fusion approach and as another level in the complexity of the TCM modeling. In this approach, a multiple classifier model combines an ensemble of generally weak and/or diverse classifiers. Then, a pool of opinions is made by using a meta-classification decision that should be superior to any of the individual classifiers.

The diversity of the classifier 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 reduce the variance of classification error as pointed out in Kittler et al. (1998). In order to improve the accuracy of classification, the base classifiers must have high disagreement between one another in mapping the solution space in order to maintain diversity (Dietterich, 1998). On the other hand, if all classifiers map the solution space in a similar manner, then evidently little improvement can be achieved over simply using one of the base classifiers. Therefore, any ensemble should provide significantly higher accuracy level to outweigh the complexity it exerts in the classification model; otherwise the base classifier should be a better alternative.

In this research two types of ensemble schemes are introduced to assess their impact on improving classification accuracy: majority vote ensemble and generalized stack ensemble. Each ensemble

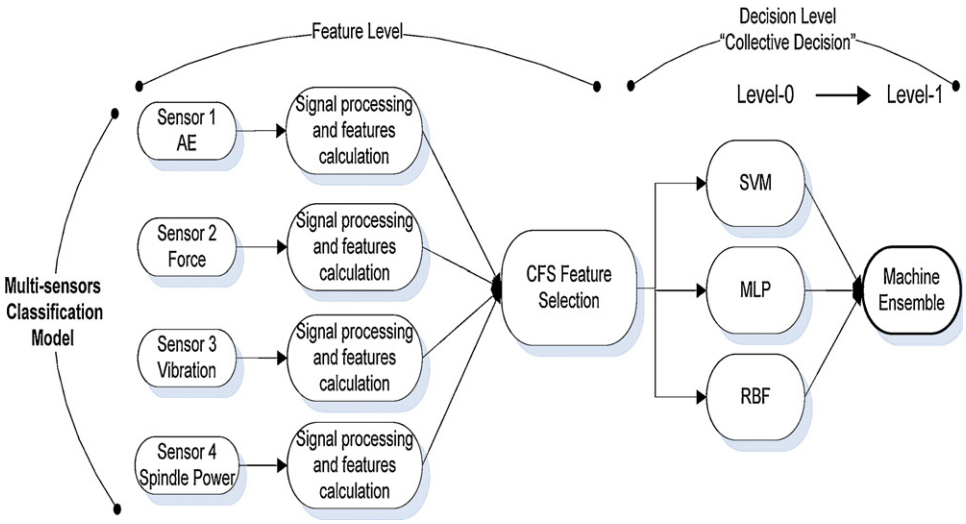

Fig. 2. Multisensor fusion model utilizing correlation based feature selection method and a machine learning ensemble scheme.

combines the classification power of three ML algorithms, which are SVM, MLP, and RBF. Fig. 2 illustrates the general schematic of the ensemble TCM model, which utilizes all extracted sensory features that have been selected by the CFS method. In the ensemble, each ML algorithm (base classifier) is trained based on the fused features from all combined sensors. Then, a collective decision is made by the ensemble method, i.e., majority vote or stack generalization.

2.5.1. Majority vote ensemble scheme

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

$$C_{\text{meta}}(X) = \underset{i}{\overset{L}{\text{argmax}}} \sum_{j=1}^{B} I(C_j(X) = i) \tag{20}$$

where $I(\cdot)$ is an indicator function, C_j are the classifiers where $j = (1, \ldots, B)$, and L is the number of target classes. It has to be noted that the confidence level of each prediction provided by each base clas-

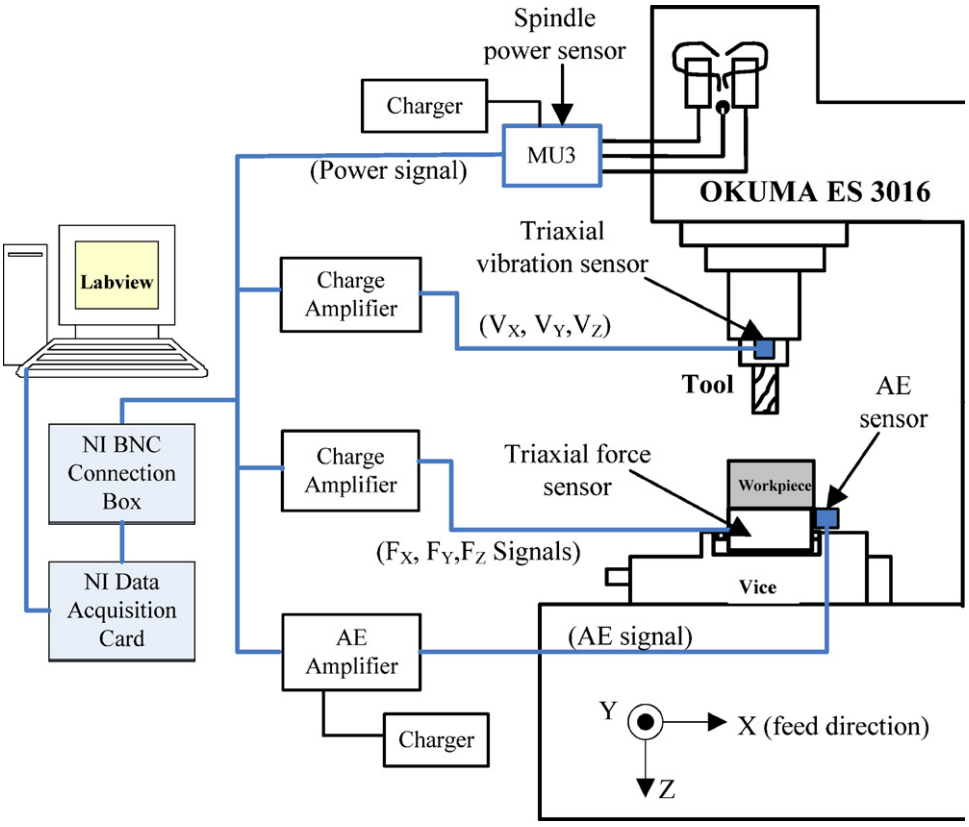

Fig. 3. Schematic diagram of experimental setup.

sifier is not considered. Therefore, the resulting vote is unweighted with all base classifiers having equal input to vote.

2.5.2. Stacked generalization ensemble scheme

Stacked generalization, or stacking, is an approach to combining predictions from multiple classifiers, introduced by Wolpert (1992). In stacking, the predicted target class of each base classifier is considered as a training input to a meta-learner or level 1-classifer where all base classifiers are considered to be level-0. A series of linear models such as multi-response linear regression (MLR) are generally used as meta-learners. In this research, an improved version of stacked generalization introduced by Seewald (2002), called StackingC, is employed. Instead of using actual target predictions that are associated with the target class during training and testing, StackingC uses prediction probabilities. This has been shown to improve the performance on multi-classed data. In this paper, StackingC is designed with the three ML algorithms introduced earlier in Section 2.4. The multi-response linear regression classifier is then trained on the output confidence predictions of the level-0 classifiers.

3. Experimentation

3.1. Experimental setup

Fig. 3 shows the experimental setup of the TCM system of this study. The experiments were carried out on 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 (PVD) 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 by using linear filtering with 3rd order Butterworth filters. The second frequency bin also has a 3rd order Butterworth band pass filter of 300–600 kHz. In addition, to avoid aliasing in the 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 the cutting operation. The sensor is connected to a Kistler (Type 5134) coupler which provides DC power and signal processing by 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 done digitally by LabVIEW software. An IIR filter with an order of 29 and a cut-off frequency of 3000 kHz was selected for all X, Y, 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 (30–3000 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 Inc., 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/second/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.

Table 2
Machining parameters and their levels.

Depth of cut (mm)	Cutting speed (m/min)	Chip load (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

The flank wear was measured by a microscope (Carl Zeiss Axioskop 2 Mat) which has a high-resolution digital camera (Axiocam MRC^{TM}). 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 (DOE)

To develop a viable TCM system, a TCM model must reflect tool fault under diverse cutting conditions such as different level of speed, feed, 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, axial depth of cut. Each factor has two levels: high (H) and low (L). These levels are provided in Table 2. In total, there were eight cutting conditions or treatments per replica. In Table 2, different cutting conditions of an experiment are represented by using three notations such that the first letter refers to the depth of cut, the second letter defines the surface speed, and the third letter refers to the chip load. The radial depth of cut was kept constant at 1.11 cm throughout the experiments. This width of cut is about 80% of tool diameter (12.7 mm), which insures that all teeth of the milling cutter are in contact with the work piece during cutting. It should be pointed out that this experimental setup is limited to three-axis machining center. In case of fives machines, achieving quasi-plane motion conditions without spindle motion would be important to eliminate disturbances related to the fast movement of the spindle head that may affect the accelerometers to measure vibrations (Norman et al., 2007).

3.3. Tool condition classes

Tool failure is identified in this research based on the following three criteria: (i) if the tool is worn due to flank wear; (ii) if the tool is worn due to chipping; and (iii) if the tool is broken. For each criterion, a measurement is 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 are hardly noticed by the operator. Measurements of wear and chipping were collected as the tool condition progressed 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, an experiment was conducted for five different tool conditions: breakage (B), chipping (C), and three states of wear

Table 3
Definition of the tool condition classes.

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

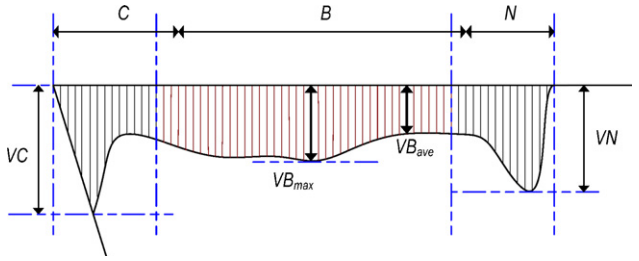

Fig. 4. Flank wear boundaries at the tool cutting edge.

which are defined as slight (LW), medium (MW), and severe (SW) wear. Table 3 summarizes the specification of each tool condition class 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. A reading of the flank wear was taken at the end of every 3rd 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 was 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. Fig. 5 provides the distribution of classes per cutting condition after the experiment was conducted. The optimal cutting condition was LHL where tool life increased dramatically and no chipping occurred.

4. Result and discussion

4.1. Classifiers training and evaluation

The notion of training and testing a dataset is fundamental to the performance of ML algorithms. In our case, the training set contains examples of signal features (135 features) and/or machining parameters features, i.e., speed, feed, and depth of cut, from different classes (tool conditions) and is used to build the classification model. The testing set represents the unknown sensory information that would be classified. Both testing and training sets are labeled

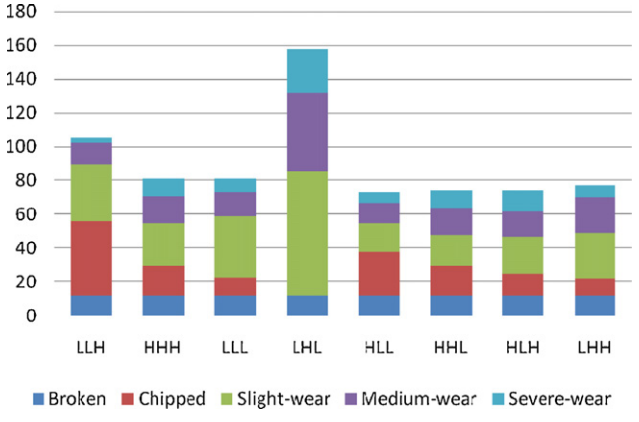

Fig. 5. Distribution of classes per cutting condition for 758 experimentally collected instances.

Table 4
Accuracy of SVM with and without feature selection (v: statistically better, *: worse).

Sensor	SVM	SVM-CFS	Sig.	#f/orig
AE	67.36 ± 4.19	69.72 ± 3.16	v	10/25
Force	90.62 ± 3.99	91.01 ± 3.65	–	19/51
Vibration	87.37 ± 4.10	88.84 ± 3.41	v	17/51
Power	60.04 ± 4.62	57.29 ± 5.20	*	3/8

Table 5
Accuracy of MLP with and without feature selection (v: statistically better, *: worse).

Sensor	MLP	MLP-CFS	Sig	#f/orig
AE	65.30 ± 3.67	67.71 ± 4.12	v	10/25
Force	87.03 ± 4.76	87.97 ± 3.68	v	19/51
Vibration	86.77 ± 3.89	86.55 ± 5.76	–	17/51
Power	57.86 ± 4.43	53.01 ± 5.04	*	3/8

with appropriate class a priori. To improve classifier training, normalization was applied to all training features as a preprocessing step.

To test and compare algorithms, we used 10 times repeated 10-fold stratified cross-validation (CV) where accuracy results were averaged across replications to minimize type I error (Bouckaert, 2003). This means that each classification model was trained on nine-tenths of the total data and tested on the remaining tenth. This process was 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 account for the non-uniform distribution of tool condition classes within the collected dataset as seen in Fig. 4.

The accuracy of classification results stands for the percentage of correctly classified instances over the total number of instances. Each result represents the average of 100 runs (10 × 10-fold CV). A paired *t*-test was applied for pairwise comparisons of classifications algorithms (Dietterich, 1998).

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 and thus no sensor fusion is considered. Specifically, a comparison of sensory information (features) per sensor is investigated and the significance of each individual sensor on tool fault classification is determined. For each investigated sensor (acoustic emission, force, vibration and spindle power), two types of feature sets are formed such that the first set includes all features and the second set includes selected features by CFS method. Both sets are applied to the three ML algorithms.

In Tables 4–6, we list the classification accuracy of each investigated machine learning algorithm per individual sensor. In addition, by applying the CFS method on the full feature set, the statistical significance for each ML algorithm is computed by using a paired t-test with 95% confidence level. Bold values show that one is significantly better than the other one. The last column shows the number of features selected by CFS versus the number of features in the original full feature set.

Clearly, CFS feature selection improved the performance of ML algorithms when it was applied on AE sensory features. Also, ML

Table 6
Accuracy of RBF with and without feature selection (v: statistically better, *: worse).

Sensor	RBF	RBF-CFS	Sig	#f/orig
AE	63.29 ± 4.15	64.94 ± 3.64	v	10/25
Force	84.93 ± 4.62	86.72 ± 4.09	v	19/51
Vibration	84.31 ± 4.01	83.98 ± 3.73	–	17/51
Power	55.83 ± 3.94	54.98 ± 5.37	*	3/8

Table 7
Accuracy of ML algorithms, majority vote and stacking ensembles across different multisensor combinations. *Note*: AE: acoustic emission, F: force, V: vibration, P: spindle power sensor (*t*-test: b/w SVM: Vote, and SVM: stacking; v: statistically better, *: worse).

Multisensor combination	SVM-CFS	MLP-CFS	RBF-CFS	Vote-CFS	Stacking-CFS	#f/org.
ALL	95.56 ± 3.15	94.16 ± 4.19	94.86 ± 3.90	96.38 ± 2.02	97.01 ± 2.01 ^{v}	26/138
AE + F + V	95.49 ± 2.32	94.28 ± 4.02	94.71 ± 3.42	96.52 ± 2.16^{v}	97.29 ± 1.98 ^{v}	25/130
AE + F + P	91.10 ± 3.09	89.04 ± 5.47	87.76 ± 4.79	91.27 ± 3.41	92.66 ± 2.71 ^{v}	21/88
AE + V + P	89.78 ± 3.55	88.79 ± 5.94	86.66 ± 3.81	89.38 ± 2.77	89.56 ± 2.61	19/88
F + V + P	94.54 ± 2.91	93.29 ± 3.97	92.53 ± 3.41	94.12 ± 2.67	96.23 ± 2.04 ^{v}	22/113
AE + V	90.01 ± 3.71	89.16 ± 4.11	86.52 ± 3.41	89.57 ± 3.32	88.12 ± 3.10*	19/80
AE + P	71.46 ± 3.63	69.47 ± 5.49	67.31 ± 3.69	70.08 ± 4.97*	71.21 ± 3.05*	11/36
AE + F	91.33 ± 3.74	89.55 ± 4.90	88.93 ± 4.04	90.01 ± 3.03*	90.93 ± 3.17	21/80
F + V	94.43 ± 2.76	93.14 ± 3.66	92.97 ± 2.73	95.22 ± 2.41 ^{v}	96.27 ± 2.13 ^{v}	22/105
F + P	89.50 ± 4.50	87.76 ± 3.93	87.60 ± 4.40	90.85 ± 3.73 ^{v}	90.10 ± 3.17	17/62
V + P	88.16 ± 3.79	86.95 ± 4.35	86.59 ± 4.01	87.63 ± 3.95	88.09 ± 3.73	16/62

algorithms applied on the reduced feature sets of force and vibration sensors improved the performance in many cases. It is evident that the force sensor is providing more explanatory information about the tool condition than any other sensors. With CFS applied to the force feature set, 91.01%, 87.97%, and 86.72% accuracy levels were achieved by SVM, MLP, RBF, respectively. Also, features extracted from the vibration sensor have given relatively high-accuracy results. Both acoustic emission and spindle power sensor have achieved accuracy levels lower than (60%).

As shown in Fig. 6, when ML algorithms are compared across all sensors, the accuracy result achieved by SVM (before and after feature reduction) is relatively higher than MLP and RBF. This result confirms other research studies which claim that SVM outperforms MLPs or RBF neural networks when considering the globality of solution (Bishop, 2006). This also suggests that training of the SVM results in a global solution for the problem under study, whereas MLP and RBF networks may have arrived at different local minima leading to an untrustworthy solution.

4.3. Classification under multisensor fusion

In this section, the limitation we had earlier with no fusion constraint is relaxed so that sensor fusion is applied to the information provided by each sensor. The fusion of information is created during the feature set development where features from multiple sensors are combined in a single set. Then, the feature set is reduced by CFS and the informational combinatory power of the remaining features is fed to the classifier for the training purpose. This means each individual classifier experiences a larger input space and a higher amount of information from multiple sensors, which ultimately increases its classification robustness, accuracy, and ability to generalize.

Three different multisensor combinations are tested from the four sensors and result in a four-sensor, 4 three-sensor, and 6 two-sensor combinations. Therefore, 11 multisensor combinations are applied to all ML algorithms and ensemble schemes. Table 7 provides the accuracy results of each multisensor combination. For each combination, SVM has given higher accuracy results than MLP and RBF. As a result, a paired *t*-test was only applied between SVM and each of the ensemble schemes to evaluate the significance in Table 7; values are in bold if one is significantly better than another.

Table 7 clearly shows how fusion of information can provide classifiers with higher explanatory power. This is evident when each sensor within the multisensor fusion combination provides complementary information that is not redundant to the information provided by other sensors. For example, the accuracy of the (*V* + *F*) multisensor combination for SVM has increased to 94.43% while it was 88.84%, 91.01% under no fusion constraint for individual vibration and force sensor models, respectively. On the other hand, (*F* + *P*) has not improved accuracy levels compared to their individual models. This is expected since the spindle power sensor provides information that is highly correlated with cutting forces. Fig. 7 graphically summarizes the results shown in Table 7.

In addition, Table 7 compares the performance statistics of the StackingC and majority vote ensemble. For stack generalization, the

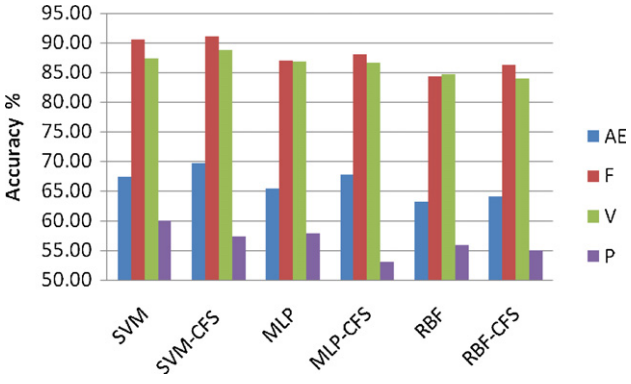

Fig. 6. ML algorithm classification performance with and without CFS.

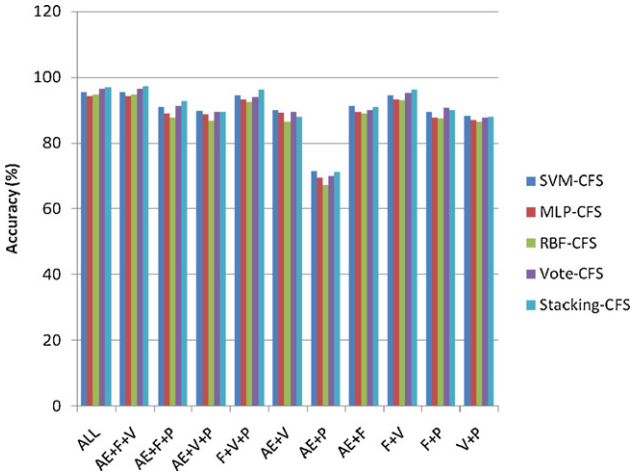

Fig. 7. Performance of different sensor fusions.

Table 8
Paired t-test of highest accuracy results (Not: not significant, v: better, *: worse).

Classifier (sensors)	Vote (AE + F + V)	StackingC (AE + F + V)	StackingC (F + V)
Vote (AE + F + V)	–	Not	Not
StackingC (AE + F + V)	Not	–	v
StackingC (F + V)	Not	*	–

Table 9
Confusion matrix for the ensemble classifiers applied to the (AE + force + vibration) multisensor combination CFS-reduced feature set.

Classifier as (%) →	Slight wear	Medium wear	Severe wear	Chipped	Broken
(a) Stacking					
Slight wear	99.22	0.39	0.00	0.39	0.00
Medium wear	4.61	92.10	3.29	0.00	0.00
Severe wear	0.00	8.62	90.61	0.77	0.00
Chipped	0.00	0.00	0.00	100.00	0.00
Broken	0.00	0.00	0.00	0.00	100.00
Correctly classified (%): **97.22**					
(b) Majority vote					
Slight wear	98.43	1.10	0.00	0.47	0.00
Medium wear	3.94	91.27	4.13	0.66	0.00
Severe wear	0.53	8.55	89.67	1.25	0.00
Chipped	0.00	0.54	1.09	97.24	1.13
Broken	0.00	0.00	0.00	0.00	100.00
Correctly classified (%)**: 96.52**					

final prediction is based on a meta-classifier which is trained on the class probabilities and targets for each training example. It should be noted that stack generalization attempts to predict when the base classifiers are incorrect. The majority vote simply sums the predictions of each class from the base classifiers and picks the most popular class. There exists some degree of diversity in mapping the predictions of tool conditions between base ML algorithms. This can be explained by the higher accuracy levels achieved by both ensemble schemes. For example, 5 of the 11 and 3 of the 11 multisensor combinations achieved statistically better result with StackingC and voting, respectively.

Table 8 provide paired *t*-tests of the highest accuracy results which correspond to the (AE + F + V) and (F + V) multisensor combinations. The performance of StackingC on (AE + F + V) is statistically better than (AE + V). This finding suggests that the 25 features selected by CFS from the three-sensor combination feature set (AE + F + V) comprise the optimal feature set when applied to either ensemble scheme.

The confusion matrices for the ensemble classifiers on the (AE + F + V) multisensor combination of the CFS reduced feature subset are presented in Table 9. These two matrices provide a detailed breakdown of misclassifications between tool condition classes. The overall accuracy of both matrices represents the highest classification accuracy obtained from all sensor combinations and classifier types. Both ensembles have managed to classify transient faults with 100% classification accuracy for breakage and chipping. The classification of gradual tool fault, which is represented by flank wear, is highly accurate. However, majority vote has higher confusion rates than StackingC ensemble where it classifies some medium and severe wear instances as chipped.

Surprisingly, the feature list that was selected by the CFS in the (AE + V + F) multisensor feature set disregarded all machining parameter features, i.e., feed rate, depth of cut, and spindle speed. It only included sensor features. The features were classified by signal processing method as shown in Table 10. Matching the result of individual sensor models, CFS has selected most of the features from the force sensor which had the highest classification accuracy. In the second place were the vibration sensor features, then AE sensor features.

Table 10
The best feature subset (25 features) selected by CFS from (AE + F + V) feature set.

Sensor	Time domain features	Frequency domain features
Force	Fy-M, Fy-Sk, Fy-Ku, Fy-pp, Fy-CF, Fx-Ku, Fx-CF, Fz-CF	Fy-SkBP, Fy-kuBP, Fy-RSPBP, Fz-VBP, Fz-kuBP
AE	AE-RMS	AE1-STBP, AE2-SkBP
Vibration	Vx-P, Vy-V, Vy-Sk, Vz-M, Vz-Ku	Vx-STBP, Vx-MBP, Vx-VBP, Vy-STBP

5. Conclusion

This paper presented a benchmark study to compare various ML algorithms and ensemble schemes used in classifying tool conditions during end milling. We found in this study that SVM has the highest accuracy in terms of base classifiers. It has been also found that the introduction of machine ensemble techniques such as majority vote and generalized stacking ensembles has improved the accuracy in a certain degree, 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.

While most of existing research works on TCM focuses on a specific toll fault, e.g., estimation of wear level, breakage detection (normal or broken), our study in this paper is able to measure most of known fault conditions including chipping under one classification framework. Specifically, we were able to make classification of a multiflute end milling tool with great accuracy and optimal feature subset size and sensor combination. In addition, irrelevant sensors and features could be detected. This facilitates more efficient modeling of the TCM because a selection of a 25-feature subset size proves to be more accurate and robust than the inclusion of the entire set of 138 explanatory features, which would significantly increase the computational performance of the TCM system. In the future, real-time classification can be studied in great depth based on the groundwork this research offered.

References

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer Science and Business Media, LLC, New York, NY.
Bouckaert, R.R., 2003. Choosing between two learning algorithms based on calibrated tests. In: Proceedings of 20th International Conference on Machine Learning. Morgan Kauffman.
Brezak, D., Udiljak, T., Majetic, D., Novakovic, B., Kasac, J., 2004. Tool wear monitoring using radial basis function neural network. In: Proceedings of IEEE International Joint Conference on Neural Networks, vol. 3, pp. 1859–1862.
Cho, S., Asfour, S., Onar, A., Kaundinya, N., 2005. Tool breakage detection using support vector machine learning in a milling process. International Journal of Machine Tools and Manufacturing 45 (3), 241–249.
Dan, L., Mathew, J., 1990. Tool wear and failure monitoring techniques for turning: a review. International Journal of Machine Tools & Manufacture 30 (4), 579–598.
Dietterich, T.G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation 10 (7), 1895–1924.

Fayyad, U.M., Irani, K.B., 1993. Multi-interval discretization of continuous-valued attributes for classification learning. IJCA 93, 1022–1027.
Ghosh, N., Ravi, Y.B., Patra, A., Mukhopadhyay, S., Paul, S., Mohanty, A.R., Chattopadhyay, A.B., 2007. Estimation of tool wer during CNC machining using neural network-based sensor fusion. Mechanical Systems and Signal Processing 21, 466–479.
Hall, M.A., 1999. Correlation-based feature selection for machine learning. PhD Dissertation. The University of Waikato. New Zealand.
Kohavi, R., john, G.H., 1997. Wrappers for feature subset selection. Artificial Intelligence 97, 273–324.
Kegg, R.L., 1984. On-line machine and process diagnostics. Annals of the CIRP 32 (2), 469–473.
Kittler, J., Hatef, M., Duin, R., Matas, J., 1998. On combining classifiers. IEEE Trans. Pattern Analysis and Machine Intelligence 20 (3), 226–239.
Kurada, S., Bradley, C., 1997. A review of machine vision sensors for tool condition monitoring. Computers in Industry 34, 55–72.
Lei, M., Yang, X., Yang, S., 1999. Tool wear length estimation with a self-learning fuzzy inference algorithm in finish milling. International Journal of Advanced Manufacturing Technology 15, 537–545.
Norman, P., Backstrom, M., Rantatalo, M., Svoboda, A., Kaplan, A., 2006. A sophisticated platform for characterization, monitoring and control of machining. Measurement Science & Technology 17, 847–854.
Norman, P., Kaplan, A., Rantatalo, M., Svenningsson, I., 2007. Study of a sensor platform for monitoring machining of aluminum and steel. Measurement Science & Technology 18, 1155–1166.
Platt, J., 1998. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), Advances in Kernel Methods-Support Vector Learning. Cambridge MIT Press, pp.185–208.
Reddy, Y.B., 1992. Multisensor data fusion: state of the art. Journal of Information Science and Technology 2 (1), 91–103.
Rehorn, A.G., Jiang, J., Orban, P.E., 2005. State-of-the-art methods and results in tool condition monitoring: a review. International Journal of Advance Manufacturing Technology 26, 693–710.
Rumelhart, D.E., McClelland, J.L., 1986. Parallel Distributed Processing: Exploration in the microstructure of cognition. MIT Press, Boston, MA.
Seewald, A.K., 2002. How to make stacking better and faster while also taking care of an unknown weakness. In: Proceedings of the 19th International Conference on Machine Learning, San Francisco, California.
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. Mechanical Systems and Signal Processing 16 (4), 487–546.
Tansel, I.N., Mclaughlin, C., 1993. Detection of tool breakage in milling operations—II. The neural network approach. Int. J. Mach. Tool. Manuf. 33 (4), 545–558.
Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Transactions on Neural Networks 10 (5), 988–999.
Witten, I.H., Frank, E., 2005. Data Mining: Practical machine learning tools and techniques, 2nd ed. Morgan Kaufmann, San Francisco.
Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5, 241–259.