Pergamon

International Journal of Machine Tools & Manufacture 40 (2000) 381–400

INTERNATIONAL JOURNAL OF
MACHINE TOOLS
& MANUFACTURE
DESIGN, RESEARCH AND APPLICATION

Data fusion neural network for tool condition monitoring in CNC milling machining

Shang-Liang Chen *, Y.W. Jen

Institute of Manufacturing Engineering, National Cheng-Kung University, Tainan 701, Taiwan, ROC

Received 10 July 1998; received in revised form 5 May 1999; accepted 22 June 1999

Abstract

Several data fusion methods are addressed in this research to integrate the detected data for the neural network applications of on-line monitoring of the tool condition in CNC milling machining. One dynamometer and one accelerometer were used in the experiments. The collected signals were pre-processed to extract the feature elements for the purpose of effectively monitoring the tool wear condition. Different data fusion methods were adopted to integrate the obtained feature elements before they were applied into the learning procedure of the neural networks. The training-efficiency and test-performance of the data fusion methods were then analyzed. The convergence speed and the test error were recorded and used to represent the training efficiency and test performance of the different data fusion methods. From an analysis of the results of the calculations based on the experimental data, it was found that the performance of the monitoring system could be significantly improved with suitable selection of the data fusion method.

Keywords: Tool wear; Data fusion; Neural networks; Feature elements; Milling

1. Introduction

In an automatic manufacturing system, machining quality is significantly affected by tool condition. Tool breakage or tool wear may significantly decrease machining quality or even create an obstacle in the production line. Therefore, on-line monitoring of tool condition, and replacement of the cutting tool at the right time are very important to assure cutting quality and system reliability. Many algorithms or methods have already been applied by previous researchers to effectively monitor tool condition [1–5]. Using different operating parameters, many different response sig-

* Corresponding author. Tel.: +886-6-2757575, ext. 63921; fax: +886-6-2085334.
E-mail address: slchen@mail.ncku.edu.tw (S.-L. Chen)

0890-6955/00/$ - see front matter
PII: S0890-6955(99)00066-8

nals may be generated, including acoustic emission signal [1], vibration signal [2], or force variation of cutting processes [3,4]. Technology for fast and accurate recognition, differentiation, or grouping of the various response signals collected at the proper time intervals during the machining process is then essential for monitoring the machining process. Several assumptions are usually required for fast pattern recognition [5,6]. These assumptions often mean loss of accuracy of prediction for complicated machining processes. This loss of accuracy is reduced by the use of neural networks.

In the last decade, neural networks have been applied to monitor many complicated machining processes, including lathe machining, milling machining, and grinding machining [7]. Neural network technology is a mathematical method for simulating the information transfer mechanism of human neurons. Using the learning and training process, knowledge is memorized by the neural networks. The highly non-linear relationship between tool wear condition and the detected signal makes analysis or tool wear recognition very difficult using traditional methods. Neural networks may overcome the non-linear difficulty. In 1990, Rangwala and Dornfeld [7] applied a neural network algorithm to CNC lathe machining processes for monitoring tool wear condition. Neural networks were found to have the advantages of reducing the effects of noise in the collected signal and increasing the pattern recognition ability in a wide parameter range. In recent years, the technology of neural networks has been rapidly developed. Back-propagation neural networks, recurrent neural networks, and competition neural networks are several network architectures which are well developed and can be applied in various cases [8,9].

Sensor fusion technology was first used in the system control of robots in a flexible machining environment. In 1988, Chryssolouris and Domroese [10] suggested that the application of multi-sensors in monitoring could give a better prediction performance and reduce the effects of noise. They compared three different sensor fusion methods: (1) least-squares regression; (2) group method of data handling (GMDH); (3) the data fusion neural networks method. Each method had its merits and drawbacks. However, the data fusion neural network method had much better test performance for monitoring the machining processes. This is especially significant with a highly non-linear relationship between the operating conditions and the detected signals. In 1990, Rangwala and Dornfeld [7] integrated information from multiple sensors (acoustic emission and force) using neural networks in order to recognize the occurrence of tool wear in a turning operation [6]. They announced that the superior learning and noise suppression abilities of neural networks enabled high success rates for recognizing tool wear under a range of machining conditions.

Some data fusion methods which preprocessed collected data were designed in this research and applied into the neural networks architecture in an attempt to increase learning efficiency and test performance. The results from the data fusion methods have been compared. Fig. 1 shows the flow chart used in this study.

2. Basic theory

Even in single sensor detection, the detected signal is usually preprocessed, for selecting the data characters, and then applied into the decision mechanism. The detected signal from single sensor detection is usually much simpler than the detected signals from multi-sensors. The relationship between the detected signals from multi-sensors will be very complicated [7]. Direct

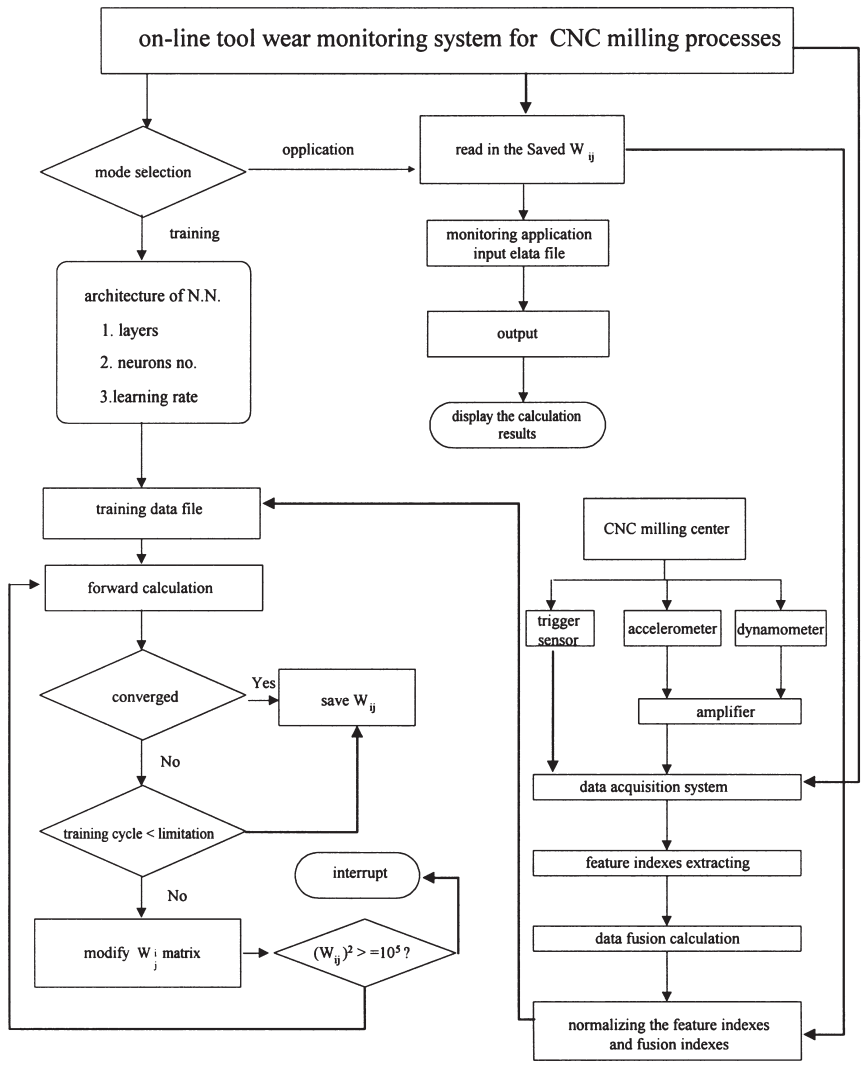


Fig. 1. The flow chart of the monitoring system.

application of the detected signals into the neural networks is very simple but may be inefficient. An adequate data fusion technology was then obviously worth investigating for increasing the performance of the application from a practical point of view. For the monitoring of the milling process using a multi-sensor, extracting the feature element that is sensitive to tool condition and then applying that into neural networks would be very helpful for increasing the performance of the decision mechanism. Two kinds of sensors were used in this research for on-line detection of the signal from the milling processes. A strain gauged dynamometer on the strain was used for the on-line detection of the cutting force. The three independent directions (x, y, z) of the cutting forces could be detected at the same time using this dynamometer. The cutter was arranged along the z direction in all the experiments. The cutting direction can then be decomposed along

the *x*- or *y*-directions. From the pilot experiments, the force variation in the *z* direction was found to be much smaller than that in the *x*- or *y*-directions for various operating parameters. Therefore, only *x*- and *y*-direction force signals were considered and adopted for the tool wear monitoring application. As the vibration signal may also contain useful information about tool wear condition, an accelerometer was used for the on-line detection of the vibration signal. The detected raw signal data were pre-processed first to find the feature elements of each independent signal data set. Several data fusion methods were used in this research for further calculating the fusion indices from the obtained feature elements. The purpose of extracting the feature elements from the independent data set was to extract the useful tool wear information. In this research, the calculation of feature elements was based on the average value, variance and frequency of the detected signals, as there are three original indices of a detected signal. The calculation of feature elements is further explained in the following sections.

2.1. Average of the magnitude of the detected signal $A_{s,i}$

$$A_{s,i}=\left(\sum_{j=1}^{N}|X_j|/N\right) \tag{1}$$

In Eq. (1), X_j is the *j*th data value of the collected raw data set *i*. The subscript 's' in $A_{s,i}$ represents the type of independent signal data that comes from different sensors. In other words, $A_{s,i}$ represents the total of $A_{x,i}$, $A_{y,i}$, and $A_{sh,i}$; where 'x' means the cutting force along the *x*-direction, 'y' means the cutting force along the *y*-direction; and 'sh' means the detected vibrating amplitude from the accelerometer. Here, a data set means the data collected within the time period of a cutting cycle. A cutting cycle was defined as the time required for the cutting edge to complete a 360° cut. *N* is the total data number of data set *i*, including the response data within a cutting cycle. As it was planned that the data would be sampled at every degree of rotation of the cutter and that the size of the data set was to be fixed at *N*=360, the sampling rate had to vary as the spindle speed varied as given in the following equation,

$$\text{Data per cycle}=\frac{\text{sampling rate}}{\text{spindle speed}\times\frac{1}{60}}=360$$

thus,

$$\text{Sampling rate}=\frac{360\times\text{spindle speed}}{60}=60\times\text{spindle speed} \tag{2}$$

where spindle speed is the number of revolutions per minute. This shows that each data set represented a complete cutting cycle. From Eq. (1), the absolute values of the obtained results were then added to obtain an average value of the magnitude of the cutting force. For a normal cutting process, this average value will remain within a reasonable range. When the cutting tool is worn or damaged, this value will be significantly varied. The variation of the operating parameters in the real cutting process will also affect the results. Therefore, operating parameters will

be used as one of the input data for training the neural networks. A trigger sensor was used to initiate the recording of the data within a cutting cycle (see Figs. 1 and 2).

A feature index of element, $C_{as,i}$, was defined based on the result of the average value of a single cutting cycle, $A_{s,i}$, and the average value of the standard cycle $A_{s,std}$. $A_{s,std}$ was the index from an experiment with a new cutter and for parameters F=0.016 mm/cutting edge and L_z=0.3 mm. F is the feed rate per cutting edge and L_z is the depth of the cut. The $A_{s,std}$ was obtained by repeating the experiment ten times and averaging the results. Use of F=0.016 mm/cutting edge, L_z=0.3 mm, and a new cutter was one of the limiting conditions of this research. It was a very light cutting process. The definition is given as follows:

$$C_{as,i}=\left(\sum_{i=1}^{10}A_{s,i}\right)/10A_{s,std} \tag{3}$$

In Eq. (3), i represents the ith data set. Therefore, three new indices were obtained: $C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$. For each feature index $C_{ax,i}$, the cutting experiment was repeated ten times. The average of the ten repeated feature indices was used to represent the real feature index $C_{ax,i}$. This will

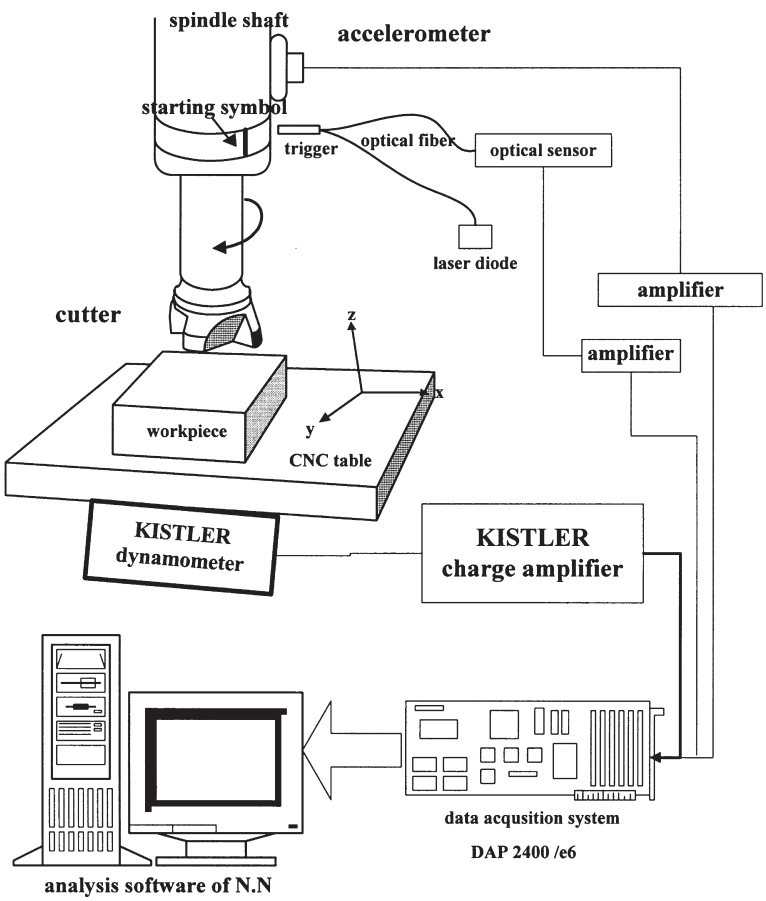


Fig. 2. The experiment set-up for on-line monitoring of the tool condition of the CNC milling machine.

make the data more reasonable. It is worth mentioning that $C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$ were approximately 1 for the light milling conditions with a new cutter. This was because, for example $C_{ax,i}$ is a relative value for the purpose of showing the variation between a monitored cutting cycle and a standard cutting cycle (F=0.016 mm/cutting edge, L_z=0.3 mm, new cutter). This design will reduce the complexity of the variation of the threshold value.

2.2. Variance of amplitude of the detected signals $V_{s,i}$

If the cutting edge was worn or broken, the amplitude and frequency of the detected signals varied significantly. The amplitude and frequency of the detected signals were chosen to represent tool vibrating conditions, and were analyzed separately. The definition of variance of amplitude if the true average value of the data set is zero is given as follows:

$$V_{s,i}=\frac{\sum_{j=1}^{N-1}(X_{j+1}-X_j)^2}{N-1} \tag{4}$$

In Eq. (4), $(X_{j+1}-X_j)$ is the difference between two sequential data values. There are a total of $N-1$ terms for data set i. The meaning of the subscript 's' is the same as in Eq. (2). The value can be obtained for ten continued cutting cycles. The feature index of the feature element, $V_{s,i}$, is defined as follows:

$$C_{vs,i}=\left(\sum_{i=1}^{10}V_{s,i}\right)/10V_{s,std} \tag{5}$$

where $V_{s,std}$ are the feature indices of the experiment with the parameters F=0.016 mm/cutting edge, L_z=0.3 mm, new cutter. Again, three new indices were obtained: $C_{vx,i}$, $C_{vy,i}$ and $C_{vsh,i}$. The cutting experiment was repeated ten times for each feature index $C_{vs,i}$. The average of the ten feature indices was used to represent the real feature index $C_{vs,i}$.

2.3. Local frequency of the detected signal

From the pilot study, the fluctuating frequency of the detected signal significantly varied with tool wear. The numbers of local minima and local maxima appearing within a cycle period were used as an index to display the degree of vibration frequency. The variation of local minimum and local maximum points, $f_{s,i}$, may be used to evaluate tool condition. A simple method to quickly obtain the variation of the numbers of local minima and local maxima was available on-line using the following two commands supplied by the DAP2400/e6 data acquisition processor [11]:

Command 1: FINDMAX (P1, 360, INSIDE, 0, 360, P2, P3)
Command 2: PCOUNT (P2, V1)

Command FINDMAX reads blocks of 36 values from channel P1 and sends the index of the

largest value between locations 0 and 360 to channel P3 and the largest value to channel P2. Command PCOUNT reads values from channel P2 and increments the contents of V1 each time a value is read. The channel values are then discarded. The above two task commands were adopted and spliced into the software of the data acquisition system [11], which then made the acquisition process very simple and very fast (near on-line). The feature indices of $f_{s,i}$ were defined as follows:

$$C_{fs,i}=\left(\sum_{i=1}^{10} f_{s,i}\right)/10f_{s,std} \tag{6}$$

In Eq. (6), i represents the ith calculation. Three new indices were obtained: $C_{fx,i}$, $C_{fy,i}$ and $C_{fsh,i}$. Again, the cutting experiments were repeated ten times for each feature index $C_{fs,i}$. The average of ten feature indices was used to represent the real feature index $C_{fs,i}$. $f_{s,std}$ were the feature indices of the experiment with the parameters F=0.016 mm/cutting edge, L_z=0.3 mm, new cutter.

Several pilot studies were performed, and the detected cutting force, vibration signals, and the related grade of wear condition were recorded. The feature indices from data fusion calculation obtained from the previous section were compiled as an input data file for the neural networks learning algorithm. Before the training process was begun, a diagram showing the relationship between the feature indices and the wear condition was plotted and analyzed. It was found that a distinguishable feature index for different grade of tool wear condition improved neural networks training efficiency and monitoring performance. This implied that data fusion methods significantly affect the efficiency and performance of neural networks. Further discussion is given in the following sections.

The feature indices of average value ($C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$), variation of amplitude ($C_{vx,i}$, $C_{vy,i}$, $C_{vsh,i}$) and variation of fluctuating local frequency ($C_{fx,i}$, $C_{fy,i}$, $C_{fsh,i}$) were found to be closely related to tool condition. Generally speaking, the magnitudes of the feature indices increased as the tool wear increased. Nine independent variables were considered in this research: $C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$, $C_{vx,i}$, $C_{vy,i}$, $C_{vsh,i}$, $C_{fx,i}$, $C_{fy,i}$ and $C_{fsh,i}$. A further three fusion indices for neural networks applications were derived from these nine variables through data fusion calculation. An input file for the training process was composed of these three fusion indices, the feature indices, the operating parameters, and the grade of tool wear condition. Five different data fusion methods were considered.

2.3.1. Comparison group (CG)

The nine feature indices without data fusion ($C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$, $C_{vx,i}$, $C_{vy,i}$, $C_{vsh,i}$, $C_{fx,i}$, $C_{fy,i}$ and $C_{fsh,i}$) and the three operating parameters were directly used for the input file of the neural networks. The training efficiency and the test performance of this fusion method were used as a standard for comparing the performance of the other four methods. After carefully considering the operating parameters and the tool specification, three operating parameters were selected as the input data for the neural networks. They were feed rate/cutting edge (F), depth of cut (L_z), and effective cut width (L_e). The depth of cut or effective cut width can be obtained from the parameter design or direct measurement from the cut results. In this research, the L_z was designed as 0.3, 0.5, 0.7 and 0.9 mm. L_e was designed as 50 mm which means the cutter was completely embedded in the workpiece during the cutting process. The feed rate/cutting edge (F) was obtained

by dividing the feed rate of the CNC table by the rpm speed of the spindle shaft. The feed rate of the CNC table was set at one of four levels, 28, 42, 56 or 70 mm/min, and the rpm of spindle shaft at 900, 1200, 1500 or 1800 rpm (Table 1). The calculated results show that F varied from 0.016 to 0.078 mm/cutting edge (see Table 1).

2.3.2. Index multiplication group (IMG)

Three new fusion indices were obtained through the following equation

$$\begin{aligned} C_{a,i} &= C_{ax,i} C_{ay,i} C_{ash,i} \\ C_{v,i} &= C_{vx,i} C_{vy,i} C_{vsh,i} \\ C_{f,i} &= C_{fx,i} C_{fy,i} C_{fsh,i} \end{aligned} \tag{7}$$

In Eq. (7), C_a, C_v and C_f represent the fusion indices of the three detected signals. A multiplication fusion method was adopted to enhance the effects of individual signal variation. As $C_{as,std}=C_{vs,std}=C_{fs,std}=1$ was applied in a limited condition of this research, the obtained $C_{as,i}$, $C_{vs,i}$ and $C_{fs,i}$ results will be larger than 1 for normal steady state milling processes. Therefore, the results of $C_{a,i}$, $C_{v,i}$ and $C_{f,i}$ were expected to be larger than 1. The variation of $C_{a,i}$, $C_{v,i}$, and $C_{f,i}$ will be very sensitive to the variation of $C_{as,i}$, $C_{vs,i}$ and $C_{fs,i}$.

2.3.3. Indices summation groups (ISG)

There were three new fusion indices were obtained through the following summation calculations

$$\begin{aligned} C_{a,i} &= C_{ax,i} + C_{ay,i} + C_{ash,i} \\ C_{v,i} &= C_{vx,i} + C_{vy,i} + C_{vsh,i} \\ C_{f,i} &= C_{fx,i} + C_{fy,i} + C_{fsh,i} \end{aligned} \tag{8}$$

In Eq. (8), $C_{a,i}$, $C_{v,i}$, $C_{f,i}$ represent the fusion indices of the three detected signals. The summation fusion method was adopted to enhance the effects of individual signal variation. Results of $C_{a,i}$, $C_{v,i}$ and $C_{f,i}$ larger than 1 were also expected. However, the variation of $C_{a,i}$, $C_{v,i}$, and $C_{f,i}$ with the summation fusion method will not be as sensitive as the variation of C_a, C_v and C_f with the multiplication fusion method.

Table 1
Feed rate per cutting edge used in this study

Feed rate of CNC table (mm/min)	Speed of spindle shaft (rpm)			
	900	1200	1500	1800
28	0.031	0.023	0.019	0.016
42	0.047	0.035	0.028	0.023
56	0.062	0.047	0.037	0.031
70	0.078	0.058	0.047	0.039

2.3.4. Indices multiplication and division group (IMDG)

From a theoretical viewpoint, the cutting direction will significantly affect the response signals. The response signal along the cutting direction will be especially affected by the tool wear condition, and the cutting force variation along the *x*- and *y*-directions will vary significantly. For example, if the *x*-direction cutting force increases significantly more than the *y*-direction cutting force, this may imply that the cutting tool may already be worn to some degree. Eq. (9) was created based on this concept, producing three new fusion indices

$$\begin{aligned} C_{a,i}&=C_{ax,i}C_{ash,i}/C_{ay,i}\\ C_{v,i}&=C_{vx,i}C_{vsh,i}/C_{vy,i}\\ C_{f,i}&=C_{fx,i}C_{fsh,i}/C_{fy,i} \end{aligned} \tag{9}$$

2.3.5. Vectors in mapping space group (VMSG)

The nine basic feature indices were divided into three groups. Therefore, three index vectors R_a, R_v, and R_f can be obtained as

$$\begin{aligned} R_{a,i}&=[C_{ax,i},C_{ay,i},C_{ash,i}]\\ R_{v,i}&=[C_{vx,i},C_{vy,i},C_{vsh,i}]\\ R_{f,i}&=[C_{fx,i},C_{fy,i},C_{fsh,i}] \end{aligned} \tag{10}$$

A flagging vector U_F is assumed here to be given by U_F=[1.0,1.0,1.0]. This assumption is convenient for calculation and also reasonable from a practical application viewpoint. The three new fusion indices were defined as the cosine component of $R_{a,i}$, $R_{v,i}$, $R_{f,i}$ related to the flagging vector U_F=[1.0, 1.0, 1.0]. The obtained cos α, cos ß, cos γ were used as the input data for the neural networks.

$$\begin{aligned} C_{a,i}&=\cos\alpha=\frac{R_aU_F}{|R_a||U_F|}=\frac{C_{ax,i}+C_{ay,i}+C_{ash,i}}{\sqrt{3(C_{ax,i}^2+C_{ay,i}^2+C_{ash,i}^2)}}\\ C_{v,i}&=\cos\beta=\frac{R_vU_F}{|R_v||U_F|}=\frac{C_{vx,i}+C_{vy,i}+C_{vsh,i}}{\sqrt{3(C_{vx,i}^2+C_{vy,i}^2+C_{vsh,i}^2)}}\\ C_{f,i}&=\cos\gamma=\frac{R_fU_F}{|R_f||U_F|}=\frac{C_{fx}+C_{fy}+C_{fsh}}{\sqrt{3(C_{fx,i}^2+C_{fy,i}^2+C_{fsh,i}^2)}} \end{aligned} \tag{11}$$

3. Experiments

The experiments were performed on a CNC milling machine center (LEADWELL, MCV-610). Fig. 2 shows the configuration of the experimental set-up. The response signals detected by the cutting dynamometer and accelerometer were amplified and acquired by a data acquisition system.

A program was developed for on-line calculation of the related feature indices. The data fusion of the different feature indices was also carried out on the same computer. The feature indices obtained were applied to the learning process of the neural networks. The KISTLER 9253A23sp dynamometer and the KISTLER 5019A multi-channel charge amplifier were used to detect the cutting force in this study. The dynamometer was placed under the CNC table as this arrangement allows the detection of the forces exerted on the workpiece. The high sensitivity accelerometer (KEYENCE AH-313) (using KEYENCE GA-245 amplifier) was attached to the frame of the spindle shaft. The response of the detected signal was affected by the place of attachment. Attaching the sensor close to the bearing increased the amplitude of the response signals. Although the noise may also have increased, this was still helpful for the study.

Fig. 3 shows an example of the typical calculated feature indices of the detected signals from the dynamometer and the accelerometer. The feed rate/cutting edge was varied from 0.016, through 0.023, 0.031, 0.037, 0.047 to 0.062 mm/cutting edge which is shown on the *x*-axis in Fig. 3. The effects of tool wear on the calculated C_{ax}, C_{ay}, C_{ash} are also shown in Fig. 3. The wear condition was rated as one of four levels: light wear, middle wear, severe wear, and tool breakage. The wear grade was considered as:

1. wear grade=0.1 (light wear), with the total wear amount lower than 0.5 mg;
2. wear grade=0.5 (middle wear), with the total wear amount between 2.0 and 3.0 mg;
3. wear grade=0.9 (severe wear), with the total wear amount between 5.0 and 6.0 mg;
4. wear grade=1.0 (tool breakage), with the total wear amount more than 6.0 mg.

The wear amount was obtained by calculating the weight difference of the cutter before and after the cutting experiments. This wear grade gives an evaluation of the tool wear condition. Flank wear or crater wear were not considered separately in this research.

A MICROSTAR DAP 2400/e6 data acquisition system was used in this study. Its sampling rate may reach 3.2 ms. HITACHI TPMN322-CH550 milling cutters were used. S45C carbon steel specimens of 100 mm×100 mm×50 mm were used as the workplaces for the cutting experiments. This size was pre-cut on the CNC milling machine for the purpose of making the depth of cut (L_z) and effective cut width (L_z) very accurate.

The experimental steps were as follows.

1. Start the data acquisition system: the program for data acquisition is ready to engage in response to the trigger.
2. Trigger the data acquisition system by the trigger sensor.
3. Collect the detected signal and save in the computer.
4. Vary the operating parameters and repeat steps 2 and 3.
5. Calculate the feature indices and fusion indices.
6. Normalize the feature indices and fusion indices.
7. Apply indices into neural networks for the learning process.
8. Test the performance of the monitoring system.

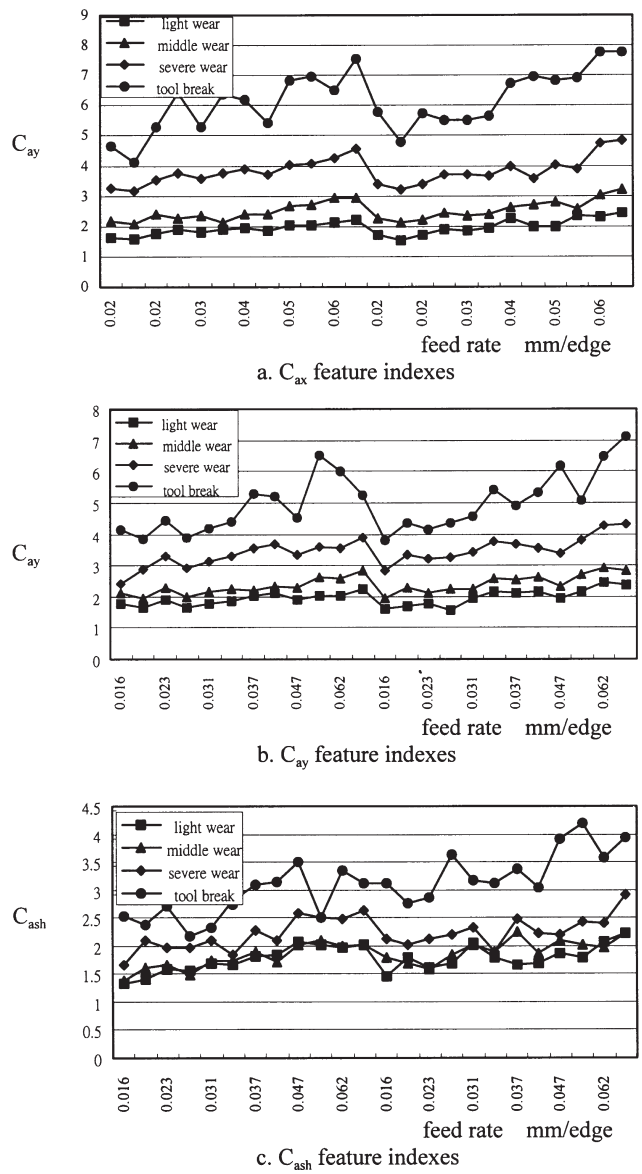


Fig. 3. Typical calculated feature indices of the detected signals from the dynamometer and accelerometer against feed rate.

4. Results and discussions

Developing a generalized neural networks architecture was found to be difficult. This was because the milling process is very complicated and many parameters interact. Therefore, different neural networks architectures may have different training results, necessitating a case by case design and evaluation. In the pilot study, three-, four-, and five-layer neural networks were

designed and investigated. In general, the three-layer architecture was found to have lower training efficiency and test performance than the four- and five-layer architectures. The three-layer architecture required more training time before converging to the allowable error. Also, the three-layer architecture was found to be much more likely to have divergent results than the other two architectures. Variation of the neurons in each layer of the three-layer architecture did not improve efficiency. Therefore, only the training results of the four- and five-layer architectures are compared here to evaluate training efficiency and test performance.

The input file was designed to have 13 input variables for all the calculations. The architecture of the input files was designed to be the same as for the CG. Three new fusion indices were generated in each of the IMG, ISG, IMDG, and VMSG data sets. The three fusion indices combined with six of the nine basic feature indices and three operating parameters and the tool condition to form a neural networks input file with 13 variables. It is worth noting that six of the feature indices remained in the input data file. This was advantageous for identifying which indices should be replaced by the fusion indices to improve monitoring. There were many possible ways to replace three of the nine feature indices by the three fusion indices. The following were the six typical cases:

1. (C_{ax}, C_{ay}, C_{ash}; C_{vx}, C_{vy}, C_{vsh}; $\underline{C_{a,i}}$, $\underline{C_{v,i}}$, $\underline{C_{f,i}}$). This implies that the results of the frequency fluctuation were replaced by the fusion indices.
2. ($C_{ax,i}$, $C_{ay,i}$, $C_{ash,i}$; $\underline{C_{v,i}}$, $\underline{C_{v,i}}$, $\underline{C_{f,i}}$; $C_{fx,i}$, $C_{fy,i}$, $C_{fsh,i}$). This implies that the results of the amplitude variation of the detected signal were replaced by the fusion indices.
3. ($\underline{C_{v,i}}$, $\underline{C_{v,i}}$, $\underline{C_{f,i}}$; $C_{vx,i}$, $C_{vy,i}$, $C_{vsh,i}$; $C_{fx,i}$, $C_{fy,i}$, $C_{fsh,i}$). This implies that the average results of the detected signal were replaced by the fusion indices.
4. ($\underline{C_{a,i}}$, $C_{ay,i}$, $C_{ash,i}$; $\underline{C_{v,i}}$, $C_{vy,i}$, $C_{vsh,i}$; $\underline{C_{f,i}}$, $C_{fy,i}$, $C_{fsh,i}$). This implies that the feature indices of the *x*-direction cutting force were replaced by the fusion indices.
5. ($C_{ax,i}$, $\underline{C_{a,i}}$, $C_{ash,i}$; $C_{vx,i}$, $\underline{C_{v,i}}$, $C_{vsh,i}$; $C_{fx,i}$, $\underline{C_{f,i}}$, $C_{fsh,i}$). This implies that the feature indices of the *y*-direction cutting force were replaced by the fusion indices.
6. ($C_{ax,i}$, $C_{ay,i}$, $\underline{C_{a,i}}$; $C_{vx,i}$, $C_{vy,i}$, $\underline{C_{v,i}}$; $C_{vx,i}$, $C_{fy,i}$, $\underline{C_{f,i}}$). This implies that the feature indices of the accelerometer signals were replaced by the fusion indices.

In cases 1–3, one of the feature indices (average, amplitude variation, or frequency fluctuation) from all the three sensors (*x*-direction force, *y*-direction force, and accelerometer) were replaced by one of the fusion indices. As average variation, amplitude variation, and frequency fluctuation were the original properties of a detected signal, it is better to consider them together within one analysis case. Therefore, the results of cases 1–3 are not discussed in the following sections. In cases 4–6, the three feature indices from one of the three sensors were completely replaced by the fusion indices. The examples of the calculated features indices from each sensor are shown in Fig. 3. It was found that the feature indices of the *x*-direction force vary significantly compared with the feature indices from the *y*-direction force or the accelerometer. Therefore, the feature indices of the *x*-direction force were included in the input file in the following training calculation. As the accelerometer was attached to the housing of the spindle shaft, close to the bearing seat, values may include some noise from the bearing. Therefore, including the feature indices from the *y*-direction force signals into the input file and replacing the accelerometer feature indices by the fusion indices for the following training calculation was a reasonable approach.

There were three kinds of parameters included in the input data: operating parameters, fusion indices, and wear grade. The three operating parameters have already been mentioned in the last paragraph: feed rate/cutting edge (F), depth of cut (L_z), and effective cut width (L_e). The total input variables for the training process was N_{input}=number of fusion indices+number of feature indices+number of operating parameters+number of wear grade=3+6+3+1=13. Variation of the number of neurons in each layer may also affect the training efficiency to some degree, but this aspect was not investigated here. From the pilot study, the four-layer neural networks architecture (13–15–15–1), and five-layer architecture (13–15–10–5–1) were selected for the comparison of the effects of the fusion indices. The term (13–15–15–1) means that the neural network architecture had 13 neurons in the input layer, 15 neurons followed by another 15 neurons in the hidden layer and only one neuron in the output layer. The term (13–15–10–5–1) means that the neural network architecture had 13 neurons in the input layer, 15 neurons followed by 10 neurons and then five neurons in the hidden layer and only one neuron in the output layer. The usage of the representation of (13–15–15–1) and (13–15–10–5–1) for neural network architecture is very simple and useful and will be adopted in the following section. Fig. 4 shows the neural networks architecture used in this study. The training efficiency was quantitatively evaluated by comparing the following two results: (1) the required number of training cycles for approaching an allowable average error; (2) the average error being converged for the same number of training cycles.

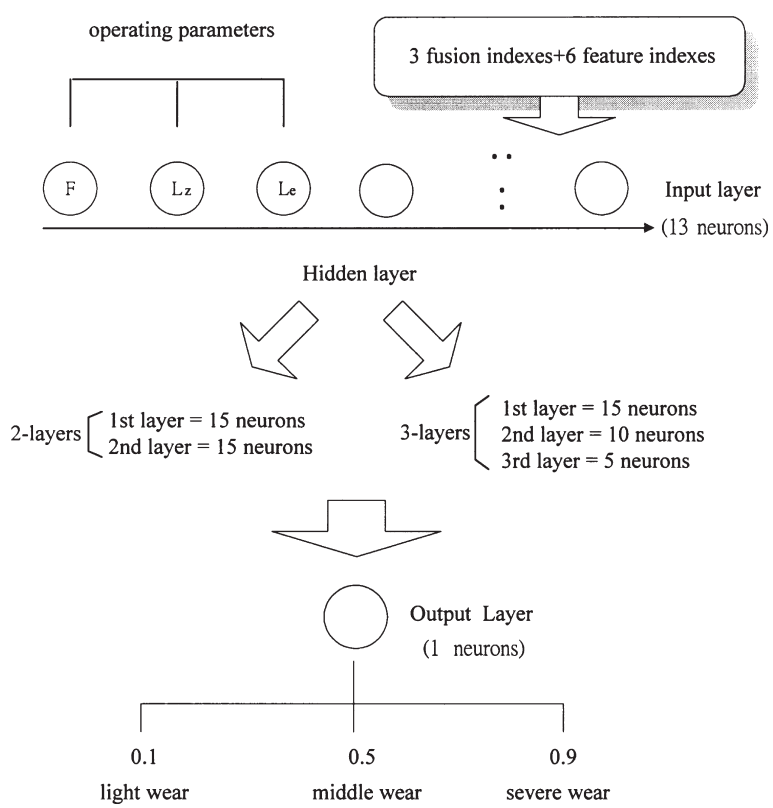
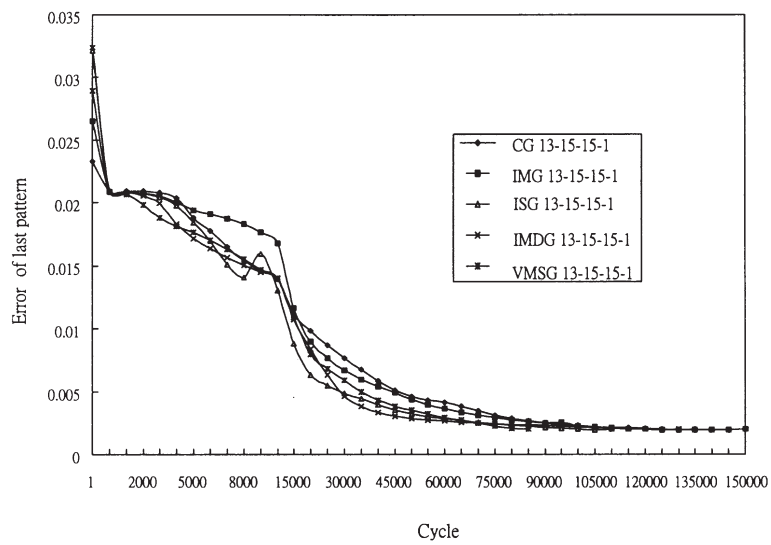
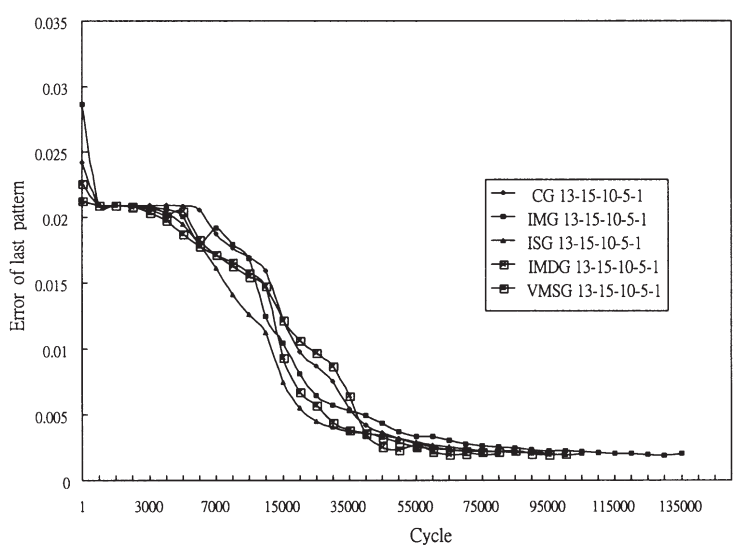


Fig. 4. The neural networks architecture ((15–15–15–1) and (15–15–10–5–1)) used in this study.

Fig. 5 shows the average error variation of the last pattern of each training cycle for different fusion methods. Fig. 5(a) and (b), show that the speed of convergence was typically very fast at the beginning of the training cycles and very slow at the end. The training efficiency (or converging route) was also found to be significantly different for training cycles up to 3000 cycles and for between 3000 and 60 000 cycles. Training efficiency did not vary significantly after a very



(a) 4-layers (12-15-15-1)

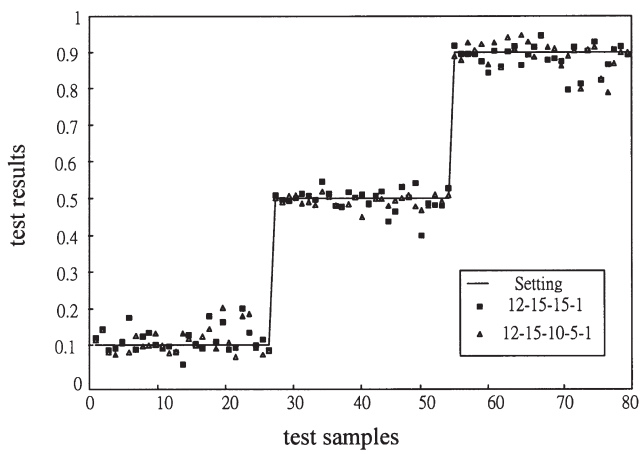


(b) 5-layers (12-15-10-5-1)

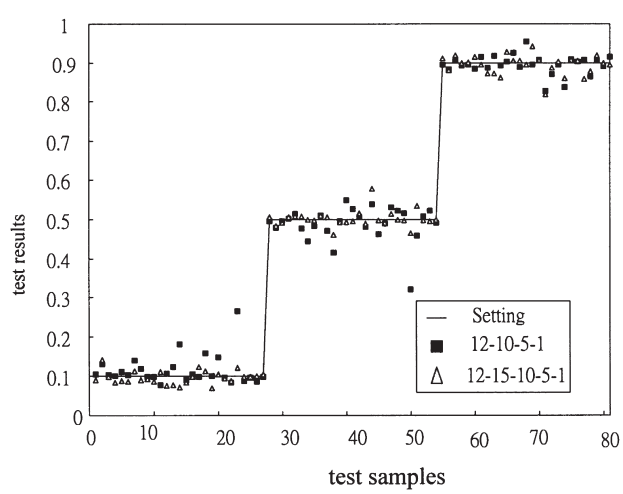
Fig. 5. The variation of the average error of the last pattern with training cycle for different fusion methods.

large number of training cycles (for example 100 000 cycles). In Fig. 5(a) and (b), the allowable error was set as less than 0.002 for the last pattern. To satisfy this requirement, the five-layer architecture (13–15–10–5–1) needed fewer training cycle than the four-layer architecture (13–15–15–1). Although the converging situations are very similar after a large number of training cycles using different fusion methods, the performace of the well-trained networks was of even greater interest.

Fig. 6 shows the test results using different fusion indices based on four- and five-layer networks. In these diagrams, the test samples were divided into three levels with wear grades 0.1, 0.5, and 0.9. From the test results shown in Fig. 6, the neural networks satisfactorily predicted

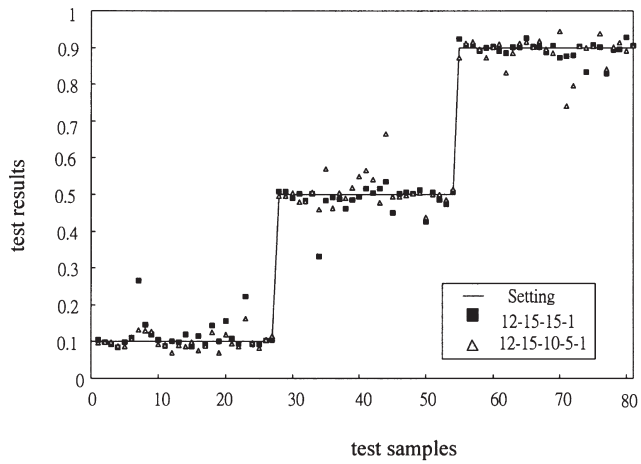


(a) Comparison group (CG group).

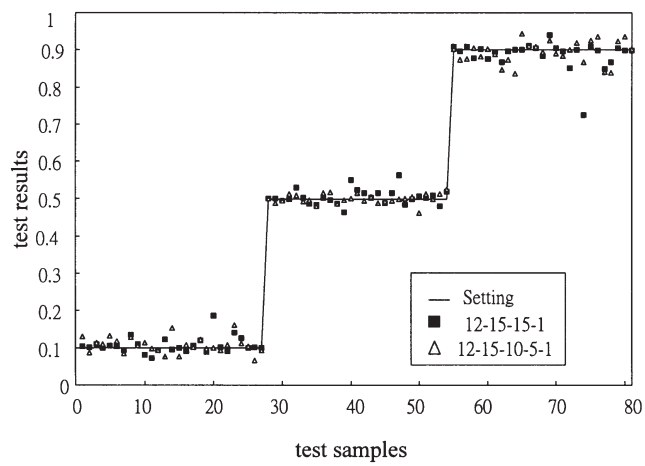


(b) Indices summation group (ISG).

Fig. 6. The neural networks test results.



(c) Indices multiplication group (IMG group).

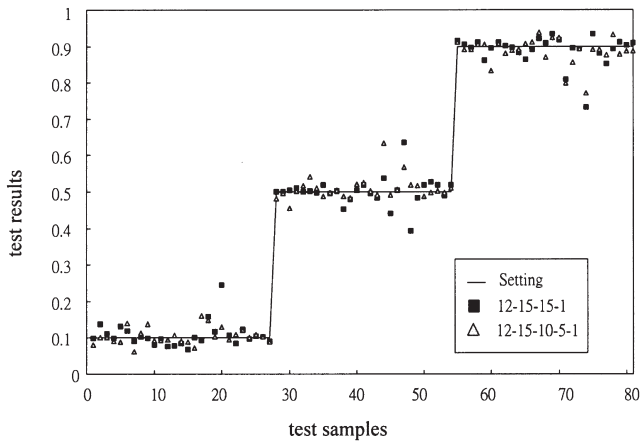


(d) Indices multiplication and division group (ISMG)

Fig. 6. (*continued*)

performance for all studied cases. To fairly compare the effects of different fusion groups on the test performance, some new factors were required. In total 80 test samples were used for each layer simulation. The test samples were distributed around the preset values. Therefore, the average and standard deviation of the test error can be used to represent a set of test error results from a statistical point of view, making a quantitative comparison between test results possible.

A quantitative comparison of test performance was helpful in understanding tool wear condition, as shown in Figs. 7 and 8. The items compared were the average and the standard deviation of the error of the test results. The definition of average error of the test results in Fig. 7 is defined as follows:



(e) Vectors in mapping space group (VMSG).

Fig. 6. (*continued*)

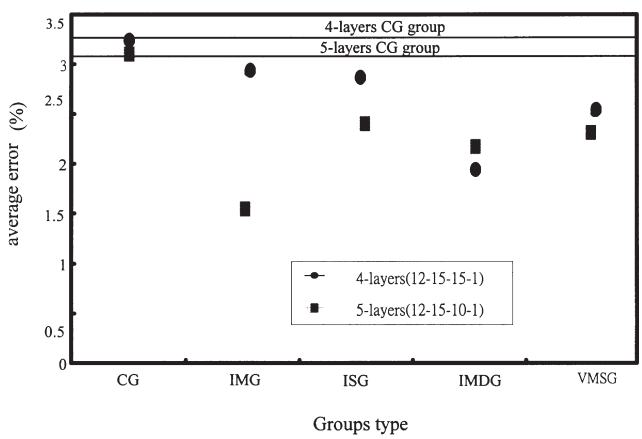


Fig. 7. A comparison of the standard deviation of the test results error using different fusion indices.

$$\text{Average error of the test} = \left[\sum_{j=1}^{a}\frac{|(X_j-0.1)|}{0.1a}\right] + \left[\sum_{j=1}^{b}\frac{|(X_j-0.5)|}{0.5b}\right] + \left[\sum_{j=1}^{c}\frac{|(X_j-0.9)|}{0.9c}\right] = \sum_{j=1}^{a}\frac{|e_j|}{0.1a} + \sum_{j=1}^{b}\frac{|e_j|}{0.5b} + \sum_{j=1}^{c}\frac{|e_j|}{0.9c} \tag{12}$$

In Eq. (12), X_j is the test result and e_j is the error of the jth test; 0.1, 0.5, and 0.9 were the pre-set values for testing; and a, b, and c separately represent the number of tests in the 0.1, 0.5, and 0.9 regions, so that $a+b+c$ equal the total number of tests. To avoid the effects of the pre-set values on the average errors, the test errors were divided by the pre-set values. Fig. 7 shows that the test error of the CG method (3.3%) with four-layer architecture was significantly larger than

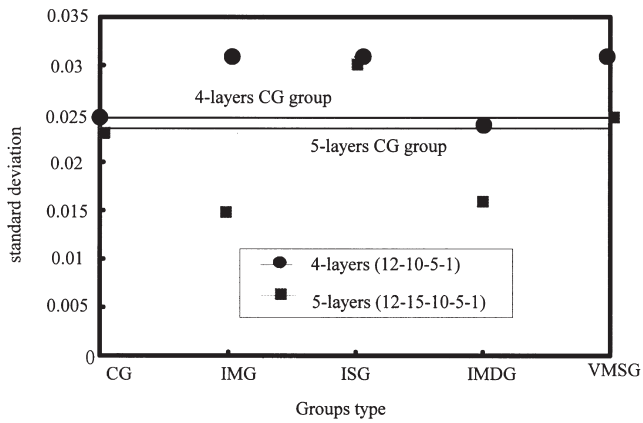


Fig. 8. A comparison of the standard deviations of the test results error using different fusion indices.

that of the other methods (IMG, 2.9%; ISG, 2.8%; IMDG, 1.9%; VMSG, 2.4%). For five-layer architecture, the test errors for the IMG fusion method (1.35%) and IMDG fusion method (2.2%) are significantly smaller than for the CG method (3.1%). Using the average error as a measurement of performance, the test performance of the IMG and IMDG methods are generally better after the data fusion process. The standard deviations of the error of the test results are shown in Fig. 8, and are defined follows:

$$\text{Standard deviation of the error of the test results } e = \sqrt{\frac{n\sum_{j=1}^{n} e_j^2 - \left(\sum_{j=1}^{n} e_j\right)^2}{n(n-1)}} \tag{13}$$

In Eq. (13), n is the number of the tests and e_j is the error of the test results of the jth test.

From Fig. 8, the IMG (0.03), ISG (0.031) and VMSG (0.032) fusion methods seem to have no significant advantages compared with the CG group (0.025) from the standard deviation of the error of the test results in four-layer neural networks architecture. However, the IMG (0.014) and IMDG (0.015) fusion methods were significantly better than the CG group (0.024) from the standard deviation of the error of the test results in five-layer neural networks architecture. The above calculations were repeated more than ten times with different four- and five-layer neural networks architectures. The test results differed little from the results in Figs. 7 and 8. The initial weight matrix was randomly generated for each calculation, which may also affect the training efficiency and test results. The initial weight matrix is a guessed matrix of the training result. The initial weight matrix was modified and updated after each training calculation and converging to the correct results. The initial weight matrix is usually randomly generated by computer in most applications and because of this, the number of training cycles required for converging to a good result may vary. However, the converged results should be the same. An example of the typical number of training cycles required for convergence to the allowable cycle average error of 0.02 is shown in Table 2. The required number of training cycles for convergence may vary for each training process with the same input data and neuron architecture, but the variation was not very significant (lower than 5%) especially with training cycles numbering about 1×10^5 times.

Table 2
Example of the required training cycles with allowable cycle average error of 0.002 and first weight matrix randomly generated

Fusion method	CG	IMG	ISG	IMDG	VMSG
Four layers (12–15–15–1)	139 955±5%	154 976±5%	109 906±5%	130 000±5%	84 941±5%
Five layers (12–15–10–5–1)	134 724±5%	79 835±5%	104 715±5%	69 300±5%	99 886±5%

Careful examination of Figs. 7 and 8 shows that both the IMG and IMDG fusion methods have a much better test performance (smaller average error and standard deviation) compared with the CG group in five-layer architecture. The feature indices were multiplied in the IMG fusion method to enhance the effects of each index. The performance of the IMG fusion method seems very promising. In the IMDG fusion method, the feature indices were multiplied or divided by another feature index to enhance the effects of each index. Since the two methods employ the same basic concept, good performance of the IMDG fusion method was also expected.

5. Summary

The effects of using several different data fusion methods on the training efficiency and test performance of neural networks were addressed. An accelerometer and a dynamometer were used in the experiments for monitoring the tool wear condition of a CNC milling machine. The feature elements were extracted from the detected signals and integrated before they were used as the input data of neural networks. The analysis of results based on the experimental data show that the training efficiency and test performance were significantly affected by different data fusion methods. In this research, the training efficiency and test performance of IMDG and IMG fusion methods were significantly improved after the data fusion processing.

Acknowledgements

This research was supported in part by the Industry Technology Research Institute of the Republic of China. The authors would like to express their deep appreciation. Thanks also are extended to I.T. Yaw for his help in the analysis of the experimental data.

References

[1] S.R. Hayashi, C.E. Thomas, D.G. Wildes, Tool break detection by monitoring ultrasonic vibrations, Annals of the CIRP 37 (1) (1998) 61–68.
[2] Y.S. Tarng, Measurement of quasi-mean resultant force using the vibrational signal of spindle in milling, Int. J. Mach. Tools & Manufac. 31 (3) (1991) 295–304.
[3] Y. Altintas, L. Yellowley, In-process detection of tool failure in milling using cutting force models, Journal of Engineering for Industry 111 (1989) 149–1574.

[4] Y.S. Tarng, Monitoring of tool fracture in milling, Int. J. Adv. Manuf. Technol. 8 (1993) 2–8.
[5] K.C. Vierck, J. Tlusty, Adaptive Thresholding for Cutter Breakage in Milling. PED-Vol. 55, Sensors and Signal Processing for Manufacturing, ASME, 1992, pp. 17–32.
[6] T.J. Ko, D.W. Cho, Tool wear monitoring in diamond turning by fuzzy pattern recognition, Journal of Engineering for Industry 116 (1994) 225–232.
[7] S. Rangwala, D. Dornfeld, Sensor integration using neural networks for intelligent tool condition monitoring, Journal of Engineering for Industry 112 (1990) 219–228.
[8] B. Kosko, Neural Networks and Fuzzy Systems, Prentice-Hall, New York, 1992.
[9] S.T. Welstead, Neural Networks and Fuzzy Logic Applications in C / C++, Wiley, 1994.
[10] G. Chryssolouris, M. Domroese, Sensor integration for tool wear estimation in machining, in: Sensors and Controls for Manufacturing, Proceedings of the Winter Annual Meeting, USA ASME, 1988, pp. 115–123.
[11] DAP 2400/e6 User Manual, Micro-star Laboratories, 1996.