Int J Adv Manuf Technol
DOI 10.1007/s00170-012-4399-2

ORIGINAL ARTICLE

Pattern recognition for bivariate process mean shifts using feature-based artificial neural network

Ibrahim Masood · Adnan Hassan

Received: 1 July 2010 / Accepted: 16 July 2012

Abstract In multivariate quality control, the artificial neural networks (ANN)-based pattern recognition schemes generally performed better for monitoring bivariate process mean shifts and provided more efficient information for diagnosing the source variable(s) compared to the traditional multivariate statistical process control charting. However, these schemes revealed disadvantages in term of reference bivariate patterns in identifying the joint effect and excess false alarms in identifying stable process condition. In this study, feature-based ANN scheme was investigated for recognizing bivariate correlated patterns. Feature-based input representation was utilized into an ANN training and testing towards strengthening discrimination capability between bivariate normal and bivariate mean shift patterns. Besides indicating an effective diagnosis capability in dealing with low correlation bivariate patterns, the proposed scheme promotes a smaller network size and better monitoring capability as compared to the raw data-based ANN scheme.

Keywords Artificial neural networks · Bivariate correlated patterns · Process monitoring and diagnosis · Statistical features · Pattern recognition

I. Masood (✉)
Faculty of Mechanical and Manufacturing Engineering,
Universiti Tun Hussein Onn Malaysia,
86400 Parit Raja, Batu Pahat,
Johor, Malaysia
e-mail: ibrahim@uthm.edu.my
URL: http://www.uthm.edu.my

A. Hassan
Faculty of Mechanical Engineering,
Universiti Teknologi Malaysia,
81310 UTM Skudai,
Johor, Malaysia
e-mail: adnan@fkm.utm.my
URL: http://www.utm.my

Published online: 04 August 2012

Abbreviations

ANN	Artificial neural network
ARL	Average run length
ART	Adaptive resonance theory
BPR	Bivariate pattern recognition
CCPR	Control chart pattern recognition
CUSUM	Cumulative sum
EWMA	Exponentially weighted moving average
FAR	False alarm rate
i.i.d.	Identically and independently distributed
LEWMA	Last value of exponentially weighted moving average
LVQ	Learning vector quantization
MAT	(Mean) × (autocorrelation)
MCUSUM	Multivariate cumulative sum
MEWMA	Multivariate exponentially weighted moving average
MLP	Multilayer perceptrons
MMSV	(Mean) × (mean square value)
MQC	Multivariate quality control
MRWA	Multi-resolution wavelet analysis
MSD	(Mean) × (standard deviation)
MSE	Mean square error
MSPC	Multivariate statistical process control
PCA	Principle component analysis
RA	Recognition accuracy percentage
RBF	Radial basis function
SOM	Kohonen self-organizing mapping
SPC	Statistical process control
traingdx	Gradient decent with momentum and adaptive learning rate

1 Introduction

It is well known that variation in manufacturing processes has become a major source of poor quality. Wear and tear, vibration, machine breakdown, inconsistent material, and lack of human operators are typical sources of process variation.

In monitoring univariate process variation, control charts are widely used for "listening to the voice of the process" [1]. The traditional control charting schemes such as Shewhart [2, 3], cumulative sum (CUSUM) [4], and exponentially weighted moving average (EWMA) [5] control charts are among the most important statistical process control (SPC) tools, which are useful in maintaining univariate process stability. The key feature of control charts, the provision of the method to differentiate between particular processes, is operating within a statistically stable and an unstable state.

In the related study, many manufacturing processes may involve two or more correlated variables, and an appropriate procedure is required to monitor these variables simultaneously. This issue is sometimes called multivariate quality control (MQC) and it has opened the basis for extensive research in the field of multivariate statistical process control (MSPC). The original work in MSPC is T^2 or χ^2 control chart [6], which was developed based on logical extension of Shewhart control chart. Initially, it was applied for monitoring multivariate process of bombsight data during World War II. However, it was found only effective for detecting large deviations in process mean shifts. To improve performance for detecting small deviations in process mean shifts, multivariate cumulative sum (MCUSUM) [7, 8] and multivariate exponentially weighted moving average (MEWMA) [9, 10] control charts were then developed based on logical extension of CUSUM and EWMA control charts, respectively.

The rapid growth in manufacturing technology such as processing methods, precision machines, automatic data acquisition system, and online computerized process monitoring system have led to increase interest in MSPC. The main issue of the most MSPC charting schemes is that they are capable to detect an out-of-control signal (statistically unstable state) in multivariate processes, but they do not directly provide diagnosis information to determine the source variable(s) that is responsible for the out-of-control signal. In practice, this diagnosis information would be greatly useful for a practitioner to find the root cause of error and solution for corrective action. Since the T^2 control chart [6], major research attentions have been given for interpreting out-of-control signals in relation to MSPC charting. Shewhart control charts with Bonferroni-type control limits [11], PCA [12], multivariate profile charts [13], T^2 decomposition [14, 15], and Minimax control chart [16], among others, have been developed for that purpose. Further discussions on this issue can be found in [17–20].

Unstable processes may produce distinct time series of SPC chart patterns. Identification of these patterns coupled with engineering knowledge of the process would lead to more specific diagnosis and troubleshooting. Development in soft computing technology have motivated researchers to explore the use of expert systems, fuzzy inference system, and artificial neural networks (ANN), among others, for automatically and intelligently recognizing patterns in relation to SPC/MSPC charting. Broadly, these artificial intelligence techniques especially ANN have been implemented in various pattern recognition areas such as image processing [21–23], medical diagnosis [24–27], structural control [28–30], handwritten and printed characters [31, 32], and hydraulics and pneumatics [33–35], among others. Discussion for this study is focused on ANN, which is also known as a massively parallel-distributed processor, connectionism, machine learning algorithm, or natural intelligent systems. It is inspired by the structure of the human brain and has the ability to learn highly complex and nonlinear mapping, recall, and generalize knowledge [36]. It is a common neurocomputing technique, recognized as important and emerging methodologies in the area of classification. The advantage with ANN is that it is capable of handling noisy measurements requiring no assumption about the statistical distribution of the monitored data. It learns to recognize patterns directly through typical example patterns during a training phase. Since early 1990s, this modern technique has opened extensive research in the field of control chart pattern recognition (CCPR). The earliest reported works have focused on feasibility study [37–40]. Since then, much progress has been made, whereby the performance of ANN-based CCPR schemes have been enhanced through feature-based and wavelet-denoise input representation techniques, modular and integrated recognizer designs, and multivariate process monitoring and diagnosis. Further discussions on issues in the development of ANN-based CCPR schemes can be found in [41, 42], which are useful for researchers towards facilitating further improvement in this area.

In monitoring and diagnosis of multivariate process mean/variance shifts, researchers have proposed various ANN-based pattern recognition schemes such as ANN-MSPC [43–46], novelty detector [47], modular-ANN [48], ensemble-ANN [49], and multi-module-structure-ANN [50]. The ANN-MSPC schemes combined the MSPC charts (i.e., for monitoring any mean/variance shifts in multivariate processes) with ANN recognizer (i.e., for diagnosis of the source variable(s) that is responsible for mean/variance shifts). The other schemes such as novelty detector, modular-ANN, ensemble-ANN, and multi-module-structure-ANN were designed to perform continuous monitoring and diagnosis simultaneously. In this study, these schemes are referred as bivariate pattern recognition (BPR) since their investigations are mainly focused on two variables.

Based on the above literatures, the existing ANN-based BPR schemes revealed disadvantages in terms of reference bivariate patterns and excess false alarms:

1. Lack of reference bivariate patterns

   In pattern recognition of bivariate process mean shifts, the joint effect in mean shift and cross correlation between two correlated variables should be taken into account. One approach is using several Shewhart control charts. Unique structures of Shewhart control chart patterns may provide useful meaning about univariate process mean shifts but they are unable to indicate cross correlation between the two dependent variables, which often lead to larger false alarm rate (FAR) than assumed. The other approach is using T^2/χ^2 control chart. T^2 statistics patterns may consider the joint effects but they are unable to diagnose the source variable(s) of the process mean shifts. There are limited works reported on modeling of bivariate correlated processes [47, 48] and patterns [45] that can clearly identify the joint effect.

2. Excess false alarms

   The existing MSPC/BPR schemes mainly reported average run length, $ARL_0 \approx 200$ in monitoring bivariate stable processes. This monitoring level, however, produce rather an excess false alarm rate (FAR≈0.5 %) compared to the univariate SPC charts such as Shewhart ($ARL_0 \approx 370$, FAR≈ 0.27 %), CUSUM ($ARL_0 \approx 400$, FAR≈0.25 %), and EWMA ($ARL_0 \approx 500$, FAR≈0.2 %). It is less economical for a practitioner to frequently conduct troubleshooting due to wrong identification of stable process as unstable. When considering this economical aspect, thus, it is important and useful to develop a BPR scheme that could maintain a minimum FAR.

These disadvantages may cause limited scopes and slow development in this area. To overcome this issue, this study aims to investigate a better pattern recognition approach for bivariate correlated processes using feature-based input representation for ANN recognizer. Effective features could reduce noise in raw data, improve recognizer training, and strengthen discrimination capability between normal and shift patterns. Detailed discussion is organized as follows. Section 2 describes several industrial examples of bivariate process variation, while Section 3 presents the patterns and data generation of the bivariate correlated process mean shifts. Section 4 discusses the feature-based ANN scheme, which included raw data and statistical features input representation approaches, features selection, and recognizer design, training, and testing. Section 5 provides the performance comparison between the feature-based ANN against the raw data-based ANN. Section 6 finally outlines some conclusions.

2 Industrial examples of bivariate process variation

The strong need for MQC can be shown by the applications of MSPC charting schemes in various manufacturing industries such as in semiconductor manufacturing [51], plastic polymer manufacturing [52], drug manufacturing [53], and chemical industries [54, 55], among others. In this study, investigation is focused on manufacturing of moving/rotating components in an audio–video device, namely, roller head. This case study was conducted in a local industry. A set of roller heads function for guiding and controlling the movement path of film tapes. The "groove and flange" and the "inner diameters" as shown in Fig. 1 are two geometrical features of a roller head that need for MQC.

Figure 2 shows the sequence of manufacturing process of a roller head. An aluminum extrusion round bar is initially machined with rough cut turning to rough size. It is followed by finish cut turning to size to obtain the geometrical features of groove and flange and inner diameters. The inner diameters are extended to honing process to achieve tight specification according to the bearing sizes. Then, the workpiece is coated using nickel-base alloy electroplating process to obtain hard surface and ready for bearing assembly.

2.1 Process variation in machining groove and flange

The groove and flange is machined to size for eliminating the film tapes from scratches during movement. Groove length (L) and flange thickness (T) are two dependent process variables being monitored using MSPC scheme. Changes in these variables can be

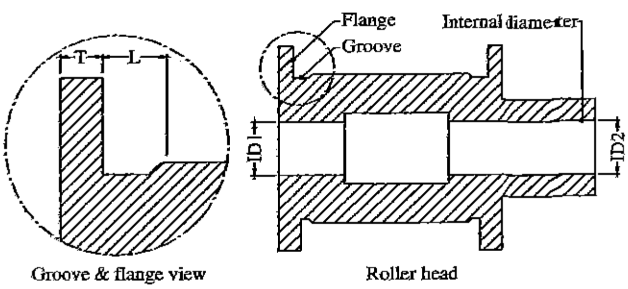


Fig. 1 Groove and flange and inner diameters of a roller head

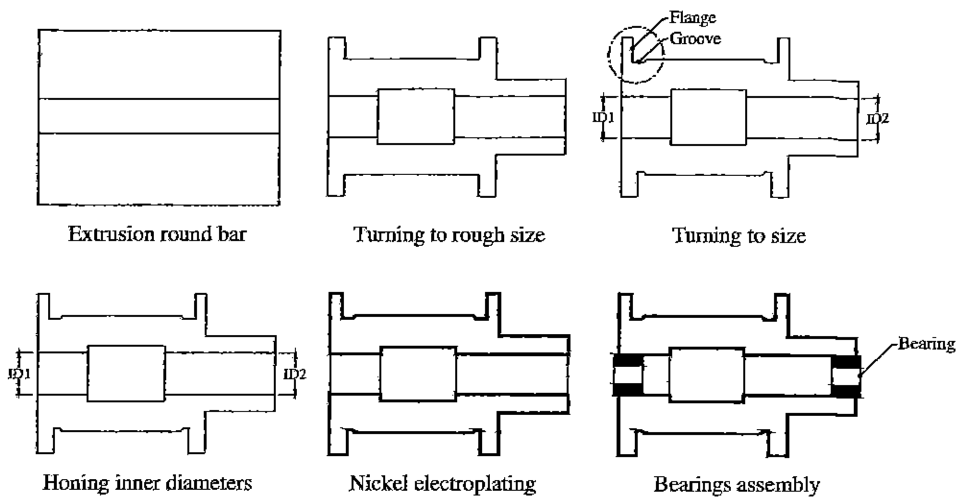


Fig. 2 Sequence of manufacturing process of a roller head

illustrated in Fig. 3 and Table 1. The workpiece is automatically loaded to a pneumatic chuck. In case of tool blunt, L will gradually decrease to $L1$, while T will gradually increase to $T1$. This change yields negative correlation between L and T. In case of loading error, T will suddenly decrease to $T2$, while L is remains stable.

2.2 Process variation in machining inner diameters

Inner diameters are machined in two sequential processes, i.e., turning to size and honing. Inner diameter 1 (ID1) and inner diameter 2 (ID2) are two dependent process variables being monitored using MSPC scheme. Process variation in turning to size process can be illustrated in Fig. 4 and Table 2. The workpiece is automatically loaded to a pneumatic chuck. In case of tool blunt, both inner diameters will gradually decrease and indicate positive correlation. In case of loading error, both inner diameters will suddenly increase and also indicate positive correlation.

Then, process variation in honing process (human-operated) can be illustrated in Fig. 5 and Table 3. Changes in material or human operator can cause sudden changes in both inner diameters. Hard materials generally require a longer time, while soft materials require shorter time for honing inner diameters. In a fixed processing time, changes in hard materials will result in smaller inner diameters, and vice versa.

3 Modeling of bivariate patterns

A large amount of bivariate samples is required to perform a thorough study, i.e., for training and testing the recognizers. Ideally, such samples should be tapped from a real process environment. Since they are not economically available, there is a need for modeling of bivariate patterns and synthetically generating analysis data.

Bivariate process is the simplest case in MQC when only two variables are being monitored dependently. Let

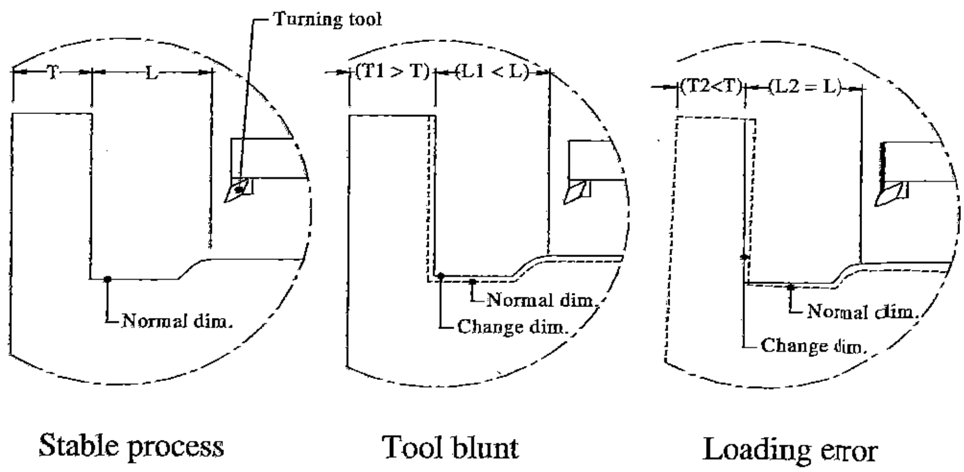
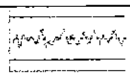
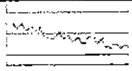
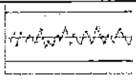
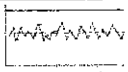
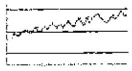
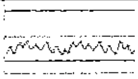
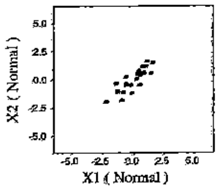
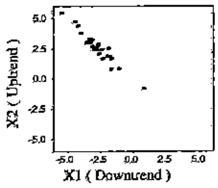
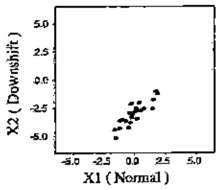


Fig. 3 Process variation in machining groove and flange

Table 1 Process variation in machining groove and flange

	Stable process	Tool blunt	Loading error
X_1 (L)	 Normal	 Downtrend	 Normal
X_2 (T)	 Normal	 Uptrend	 Downshift
Scatter Diagram			

$X_{1i}=(X_{1-1},\ldots, X_{1-24})$ and $X_{2i}=(X_{2-1},\ldots, X_{2-24})$ represent 24 observation samples for a bivariate process. Observation window for both variables start with samples ith$=(1,\ldots,24)$. It is followed with (ith +1), (ith+2), ..., and so on.

In a statistically stable state, samples for both variables are identically and independently distributed (i.i.d.) with zero mean ($\mu_0=0$), unity standard deviation ($\sigma_0=1$), and zero cross correlation ($\rho=0$). Cross correlation shows a measure of degree of linear relationship between the two variables. They yield random patterns when plotted separately on two Shewhart control charts and yield a circle pattern when plotted on a scatter diagram. Scatter diagram yields ellipse patterns when $\rho>0$ as shown in Fig. 6.

Disturbance from assignable causes may suddenly or gradually deteriorate data streams X_{1i} and/or X_{2i} into a statistically unstable state. The structure of an unstable pattern is initially vague, namely, "partially developed pattern". The pattern structure becomes more obvious as it proceeds into "fully developed pattern". This can be illustrated in Fig. 7.

Generally, the occurrence of assignable causes over X_{1i} and/or X_{2i} can be identified by Shewhart control chart patterns such as sudden shifts, trends, cyclic, systematic, and mixture. Investigation for this study was concerned on sudden shift patterns with positive cross correlation ($\rho\geq 0$). This involves seven conditions/categories of bivariate correlated patterns:

Condition 1 Normal (0, 0): both X_{1i} and X_{2i} are stable
Condition 2 Upshift (1, 0): X_{1i} in upward shift, while X_{2i} is stable
Condition 3 Upshift (0, 1): X_{2i} in upward shift, while X_{1i} is stable
Condition 4 Upshift (1, 1): both X_{1i} and X_{2i} are in upward shifts
Condition 5 Downshift (1, 0): X_{1i} in downward shift, while X_{2i} is stable
Condition 6 Downshift (0, 1): X_{2i} in downward shift, while X_{1i} is stable
Condition 7 Downshift (1, 1): both X_{1i} and X_{2i} are in downward shifts

Fig. 4 Process variation in machining turning inner diameters

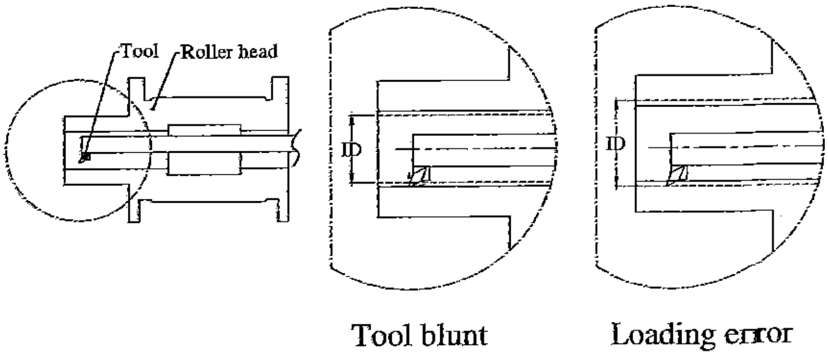
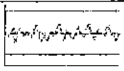
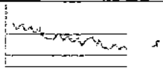
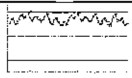
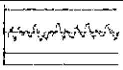
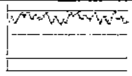
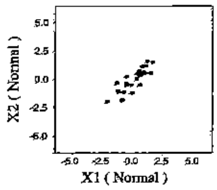
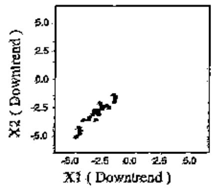
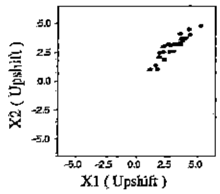


Table 2 Process variation in machining turning inner diameters

	Stable process	Tool blunt	Loading error
X_1 (ID1)	 Normal	 Downtrend	 Upshift
X_2 (ID2)	 Normal	 Downtrend	 Upshift
Scatter Diagram			

3.1 Data generator for bivariate process mean shifts

Synthetic data were simulated based on the following steps [47, 48, 56]:

Step 1: Generate random normal variates for process variable 1 (n_1) and process variable 2 (n_2), which are i.i.d within [−3, +3].

$$n_1 = b \cdot r_1 \tag{1}$$

$$n_2 = b \cdot r_2 \tag{2}$$

where r_1 and r_2 are random normal variates and b is the baseline noise level. In this study, b=1/3 was used.

Step 2: Generate random data series for process variable 1 (Y_1) with mean (μ_1) and standard deviation (σ_1).

$$Y_1 = \mu_1 + n_1\sigma_1 \tag{3}$$

Step 3: Generate random data series for process variable 2 (Y_2) dependent to process variable 1 (Y_1). μ_2 and σ_2 are the mean and standard deviation for Y_2, whereas ρ is the correlation coefficient between the two process variables.

$$Y_2 = \mu_2 + \left[n_1\rho + n_2\sqrt{(1-\rho^2)}\right]\sigma_2 \tag{4}$$

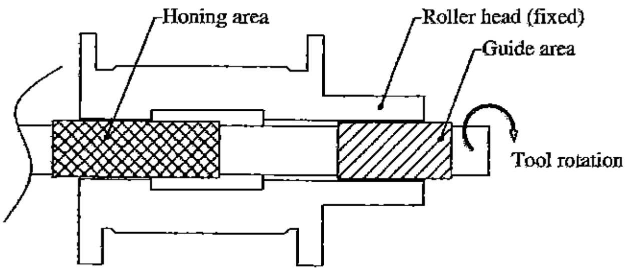


Fig. 5 Schematic diagram of honing tool

Step 4: Compute the means and standard deviations for Y_1 and Y_2 from the generated random data series. These values represent the statistically stable process means and standard deviations for process variable 1 (μ_{01}, σ_{01}) and process variable 2 (μ_{02}, σ_{02}), respectively.

Step 5: Transform random data series into normal or sudden shift patterns to mimic real observation process samples (X_1, X_2). The magnitude of mean shift (h) is expressed in standard deviation of stable process (σ_{01}, σ_{02}):

$$X_1 = h_1(\sigma_{01}/\sigma_1) + Y_1 \tag{5}$$

$$X_2 = h_2(\sigma_{02}/\sigma_2) + Y_2 \tag{6}$$

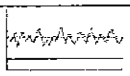
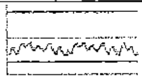
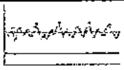
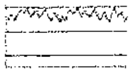
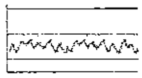
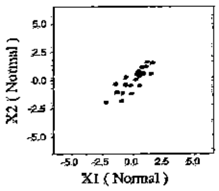
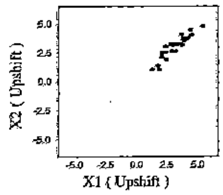
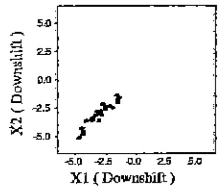
A pair observation sample (X_1, X_2) represents a bivariate vector measured at time t (X_t) that follows the random normal bivariate distribution N (μ_0, Σ_0). μ_0 and $\Sigma_0 = [(\sigma_1{}^2\sigma_{12})\ (\sigma_{12}\sigma_2{}^2)]$ are the mean vector and the covariance matrix for bivariate stable process with variances ($\sigma_1{}^2$, $\sigma_2{}^2$) and covariance (σ_{12}=σ_{21}).

Step 6: Rescale pattern data series into a standardize range, approximately within [−3, +3].

$$Z_1 = (X_1 - \mu_{01})/\sigma_{01} \tag{7}$$

$$Z_2 = (X_2 - \mu_{02})/\sigma_{02} \tag{8}$$

Table 3 Process variation in honing inner diameters

	Stable process	Change to soft material	Change to hard material
X_1 (ID1)	 Normal	 Upshift	 Downshift
X_2 (ID2)	 Normal	 Upshift	 Downshift
Scatter Diagram			

A pair standardized sample (Z_1, Z_2) represents a standardized bivariate vector measured at time t (Z_t) that follows the standardized normal bivariate distribution $N(0, R)$. Zero and R=[(1 ρ) (ρ 1)] are the mean vector and a general correlation matrix for bivariate stable process with unity variances ($\sigma_1^2 = \sigma_2^2 = 1$) and covariance equal to cross correlation ($\sigma_{12} = \sigma_{21} = \rho$).

3.2 Reference patterns for bivariate process mean shifts

As described in an introduction, the joint effect in mean shift and cross correlation between two dependent variables should be considered in pattern recognition of bivariate process mean shifts. Scatter diagram can be used to indicate the distribution of samples in the bivariate processes [45]. The structures of bivariate patterns can be differentiated and recognized effectively in-line with the concurrent changes in mean shifts in the source variable(s) and cross correlation. Pattern properties can be expressed numerically either in the form of standardized samples (raw data) or statistical features. Bivariate stable process (normal (0, 0)) yields a bivariate pattern with an approximately circular shape and zero center position as shown in Fig. 8.

Table 4 summarizes some reference bivariate patterns based on process mean shifts ±3 standard deviations. Structures of bivariate patterns are unique for specific changes in mean shift and cross correlation. The degree of mean shifts can be identified when the center position was shifted away from zero point, whereas the degree of cross correlation can be identified from the ellipse shapes. Thus, recognition of these patterns could lead to a better process monitoring and diagnosis.

4 Pattern recognition using feature-based ANN

Feature-based ANN pattern recognition scheme was investigated for monitoring and diagnosis of mean shifts in bivariate processes. The scheme comprises two main phases, i.e., (1) process patterns and input representation and (2) process monitoring and diagnosis as shown in Fig. 9. In

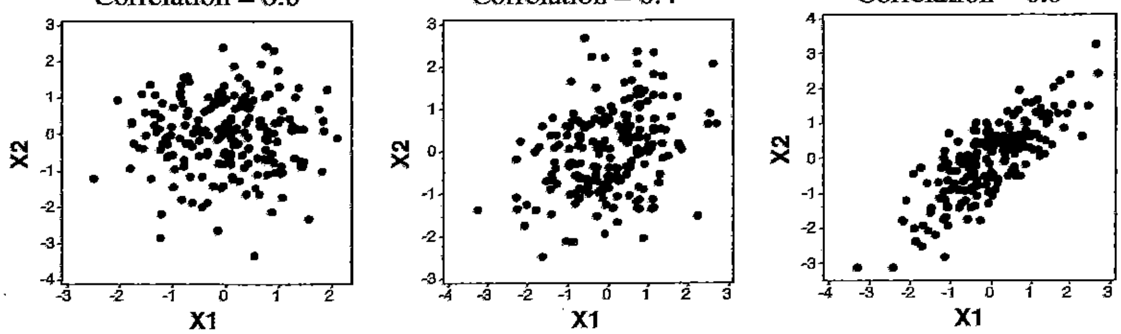


Fig. 6 Changes in cross correlation

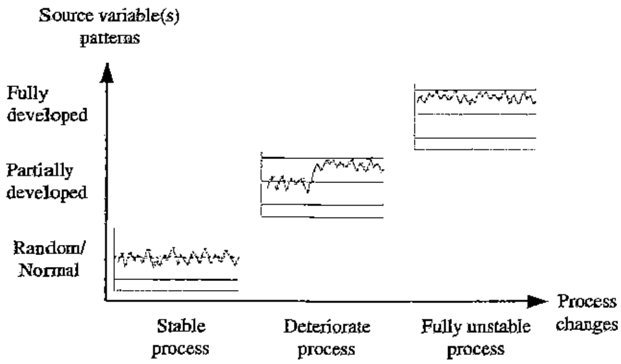


Fig. 7 Process changes in upward shift pattern

phase I, standardized samples (raw data) of two dependent variables were plotted on a scatter diagram to yield bivariate patterns. Statistical features were then extracted from the standardized samples as input representation for an ANN recognizer. In phase II, the trained ANN recognizer was used for monitoring and diagnosis of bivariate process mean shifts. Monitoring refers to the identification of process status either in a statistically stable or unstable state. Diagnosis refers to the identification of the source variable(s) that responsible for a statistically unstable state. In process monitoring, the recognizer should detect the bivariate unstable process as quickly as possible with shorter ARL_1 (minimum type II error). Meanwhile, it should leave the bivariate stable process running as long as possible with longer ARL_0 (minimum false alarms/type I error). In process diagnosis, the recognizer should have the capability to accurately identify the source variable(s) patterns with higher recognition accuracy percentage.

4.1 Input representation

Input representation provides a strong influence on the performance of ANN recognizer. There are various approaches that could be used to represent input data. Raw data-based (standardized samples) is the basic approach [47]. Besides raw data-based approach, feature-based approach that involves extracting useful features from raw data is one of the successful methods for classification in the area of image processing [57, 58]. This approach has also been applied in univariate CCPR, which aims to reduce network size, computational efforts, and training time towards improving the recognition accuracy of ANN recognizers [59–63]. Pham and Wani [59] firstly investigated nine

Fig. 8 Pattern for bivariate stable process

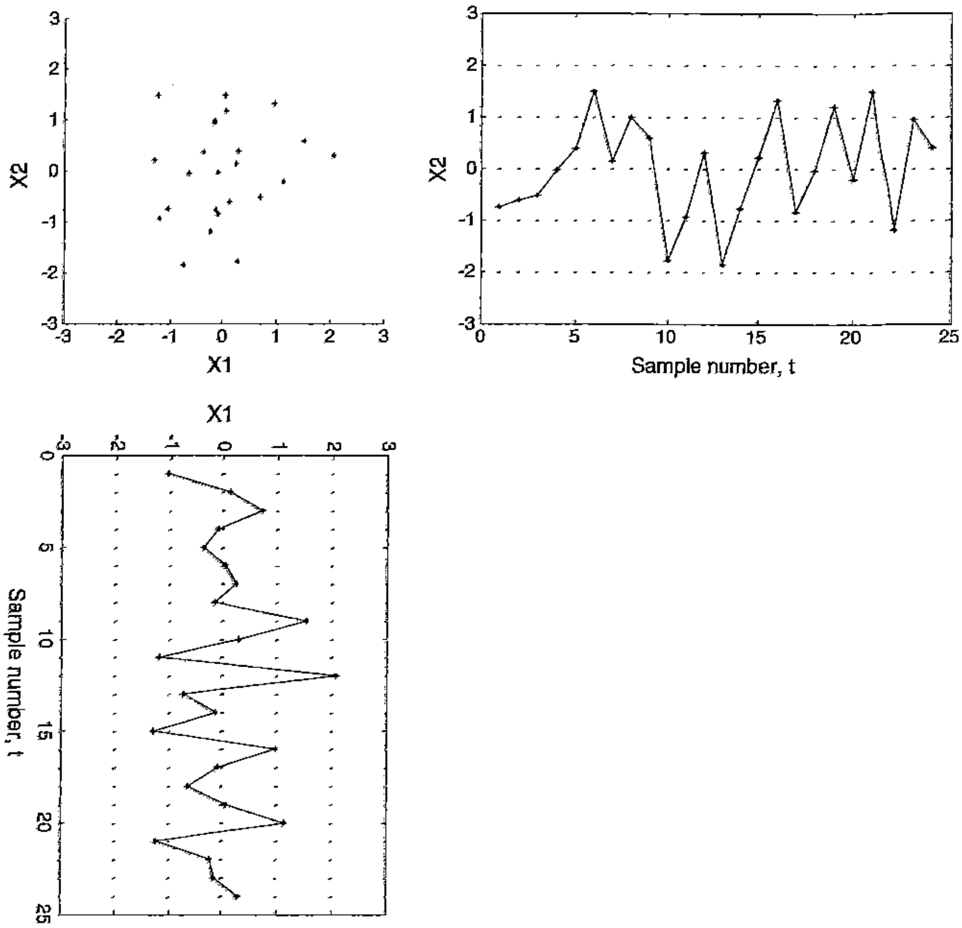
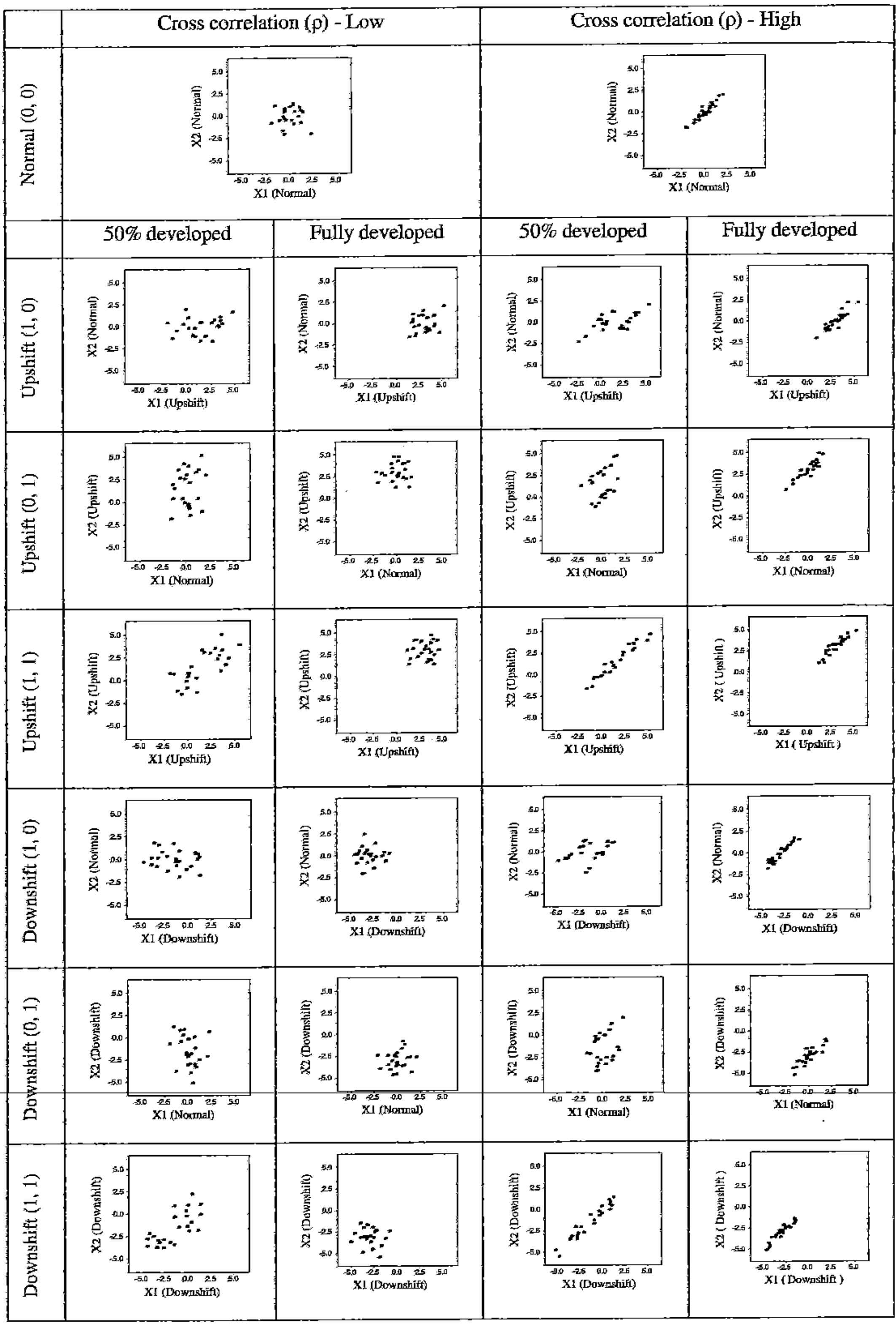


Table 4 Summary of reference bivariate patterns



shape features. Later, Gauri and Chakraborty [60, 61] proposed improved shape features. In the related study, Hassan et al. [62] proposed six statistical features comprising of mean, standard deviation, skewness, mean square value, autocorrelation, and last value of CUSUM. More recently, Guh [63] proposed another set of statistical features

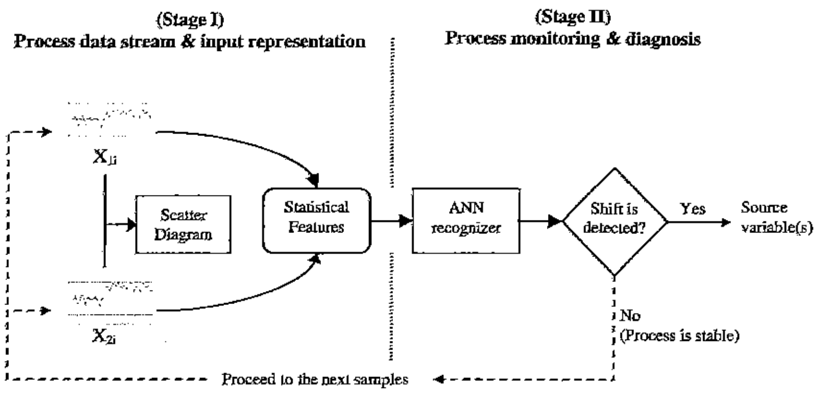


Fig. 9 Conceptual diagram of feature-based ANN scheme

comprising of mean, standard deviation, skewness, kurtosis, slope, and Pearson correlation coefficient.

Several researchers have also applied features with raw data simultaneously such as T^2 statistics with raw data [48] and statistical features with raw data [46, 49] for strengthening pattern properties in BPR. This approach, however, has increased the network size, computational effort, and training time and could be difficult to implement for more complex classification tasks.

The other approach is using multi-resolution wavelet analysis (MRWA) for denoising or filtering raw data through several decomposition levels without changing the network size and training time [64–67]. This approach has been used in univariate CCPR [64, 65] and concurrent pattern recognition [66], and very appealing in other application areas such as for detecting discontinuity and/or abrupt change in signal processing and image processing [67]. The MRWA denoising has played a crucial role for process control or monitoring. Insufficient denoising will distort waveforms and introduce errors. Inversely, excessive denoising will over-smooth the sharp features of underlying signals by recognizing them as noise or outliers [67].

This study utilizes an improved set of statistical features input representation. This improved feature-based approach was then compared to the raw data input representation.

4.1.1 Raw data input representation

Raw data input representation contained 24 consecutive standardized samples for variable 1 and variable 2. Therefore, each pattern was represented by 48 input data (i.e., $Z_{1\_P1}, Z_{1\_P2}, \ldots, Z_{24\_P1}, Z_{24\_P2}$), where $P1$ and $P2$ denote variable 1 and variable 2, respectively.

4.1.2 Statistical features input representation

Statistical features input representation, which were extracted from the raw data comprising of last value of exponentially weighted moving average (LEWMA$_\lambda$) with λ=[0.25, 0.20, 0.15, 0.10], mean (μ), multiplication of mean with standard deviation (MSD), and multiplication of mean with mean square value (MMSV). Each bivariate pattern was represented by 14 input data (i.e., LEWMA$_{0.25_1}$, LEWMA$_{0.20_1}$, LEWMA$_{0.15_1}$, LEWMA$_{0.10_1}$, μ_1, MSD$_1$, MMSV$_1$, LEWMA$_{0.25_2}$, LEWMA$_{0.20_2}$, LEWMA$_{0.15_2}$, LEWMA$_{0.10_2}$, μ_2, MSD$_2$, and MMSV$_2$).

(a) Features extraction for LEWMA

The EWMA statistics as can be computed using Eq. (9) incorporates historical data in a form of weighted average of all past and current observation samples [5, 68]:

$$\mathrm{EWMA}_i = \lambda\, X_i + (1-\lambda)\, \mathrm{EWMA}_{i-1} \tag{9}$$

X_i represents the original samples. In this study, the standardized samples (Z_i) were used instead of X_i so that Eq. (9) becomes:

$$\mathrm{EWMA}_i = \lambda\, Z_i + (1-\lambda)\, \mathrm{EWMA}_{i-1} \tag{10}$$

where $0<\lambda\leq 1$ is a constant parameter and i=[1, 2,..., 24] are the number of samples. The starting value of EWMA (EWMA$_0$) was set as zero to represent the process target (μ_0). The last value of EWMA (LEWMA) with λ=[0.25, 0.20, 0.15, and 0.10] was selected empirically for strengthening pattern properties towards improving discrimination capability between normal and shift patterns.

Constant parameter (λ) generally influences the performance of EWMA scheme in monitoring process mean shifts. Several values of λ between [0.05 and 0.40] and width of control limit (L) between [2.615 and 3.054] have been recommended in [5] to have smaller false alarms compared to the Shewhart control chart. Preliminary experiments revealed that the EWMA with small λ (0.05) were more sensitive in identifying small shifts ($\leq$0.75 standard deviations) compared to the large shifts ($\geq$1.25 standard deviations). Inversely, the EWMA with large λ (0.40) were more sensitive in identifying large shifts (≥2.00 standard deviations) compared to small shifts (<1.25 standard deviations).

(b) Features extraction for MSD and MMSV

The MSD and MMSV features were used to magnify the magnitude of mean shift (μ_1, μ_2) for improving discrimination capability between normal and shift patterns.

$$\mathrm{MSD}_1 = \mu_1 \times \mathrm{SD}_1 \tag{11}$$

$$\mathrm{MSD}_2 = \mu_2 \times \mathrm{SD}_2 \tag{12}$$

$$\mathrm{MMSV}_1 = \mu_1 \times \mathrm{MSV}_1 \tag{13}$$

$$\mathrm{MMSV}_2 = \mu_2 \times \mathrm{MSV}_2 \tag{14}$$

where (μ_1, μ_2), (SD_1, SD_2), and (MSV_1, MSV_2) are the means, standard deviations, and mean square values for process variable 1 and process variable 2, respectively. The computation of mean square value can be found in [62].

The ANN training will be more efficient when input data fall within a certain range [69]. As such in this study, the input data were normalized to a compact range between [−1 and 1] as has been implemented in [60–62]:

$$Zn_t = [2 \times (Z_t - Z_{\min})/(Z_{\max} - Z_{\min})] - 1 \tag{15}$$

where, Zn_t normalized value from raw data or features,
Z_t raw data or features,
$Z_{\max}$ maximum value of raw data or features,
$Z_{\min}$ minimum value of raw data or features.

The maximum and the minimum values for normalization were taken from the overall training patterns. Normalization is often useful particularly when the values of input data differ significantly [70].

(c) Features selection

The choice of components of statistical features is very important. The presence of too many input features can burden the training process and lead to inefficient recognition [62]. Initially, 12 candidate features were considered based on summary statistics (i.e., univariate and multivariate statistics) as summarized in Table 5, whereby some of them have been used by other researchers. Nine candidate features were then omitted, while six candidate features were short listed for further experiments:

- *To avoid confusion in pattern classification*. Standard deviation, mean square value, autocorrelation, and multivariate statistics (i.e., T^2, MCUSUM, MEWMA, Euclidean distance, and cross correlation) were omitted because these features gave the same vector (positive value) for both upward and downward shift patterns and increase confusion in pattern classification.
- *To improve ARL_0 results and to reduce input data*. Preliminary experiments revealed that besides having comparable ARL_1 results in detecting process mean shifts, the EWMA scheme shows better ARL_0 result compared to the two-sided CUSUM scheme in identifying stable process. Each sample in the EWMA scheme also requires only one data compared to the two-sided CUSUM scheme that requires two data (positive and negative CUSUMs). Thus, last EWMA was selected, while last CUSUM was omitted.

Table 5 Summary of candidate features towards achieving optimal features set

Candidate features	Omitted features	Short listed features	Optimal features
Univariate statistics:	• Last CUSUM	• Last EWMA	• Last EWMA
• Last CUSUM	• Standard deviation	• Mean	• Mean
• Last EWMA	• Mean square value	• Mean × standard deviation	• Mean × standard deviation
• Standard deviation	• Autocorrelation	• Mean × mean square value	• Mean × mean square value
• Mean square value	• MCUSUM	• Mean × autocorrelation	
• Autocorrelation	• MEWMA	• Slope	
• Slope	• Euclidean distance		
	• Cross correlation		
Multivariate statistics:			
• T^2			
• MCUSUM			
• MEWMA			
• Cross correlation			
• Euclidean distance			

- *To enhance the effect of mean and basic statistics.* Mean, standard deviation, mean square value, and autocorrelation are several basic parameters in summary statistics. To strengthen pattern properties, the effects of mean could be enhanced by multiplying the mean with standard deviation, mean square value, and autocorrelation.
- *To investigate useful features as used by other researchers.* Mean and slope features have been applied in univariate CCPR [62, 63]. Since these features produce different vectors for upward and downward shift patterns, it should be useful to investigate their effect when apply into BPR.

Using the short listed features above, 13 candidate ANN architectures were trained using various combinations of statistical features as marked by 'a' in Table 6. The desired training performance was set as the following:

1. *Primary performance.* To reduce false alarms (longer ARL_0) in identifying bivariate stable process, the trained ANN recognizer should achieve an excellent recognition accuracy percentage, i.e., above 99 % in classifying normal patterns.
2. *Secondary performance.* To improve sensitivity in detecting mean shifts (smaller ARL_1) and accurate diagnosis of the source variables (higher recognition accuracy percentage (RA)), the trained ANN recognizer should achieve a high recognition accuracy percentage, i.e., above 95 % in classifying shift patterns.

The selected ANN architecture was 14×22×7 with very good recognition accuracy percentage for normal patterns (99.53 %) and shift patterns (95.72 %). This feature-based ANN recognizer yields a smaller network size, less computational efforts, and shorter training time compared to the raw data-based ANN recognizer. Further comparison on the network architectures is discussed in Section 4.2.

4.2 Design of ANN recognizer

Various ANN-based models such as multilayer perceptrons (MLP), learning vector quantization (LVQ), radial basis function (RBF), adaptive resonance theory (ART), and Kohonen self-organizing mapping (SOM) have been applied in pattern recognition. MLP and LVQ are from supervised training, whereas ART and Kohonen SOM are from unsupervised training. Selection of the model depends on the problem.

MLP model trained with back-propagation algorithm was selected for this study since it has been effectively implemented in univariate CCPR [62, 63, 71–76]. This model basically consisted of an input layer, one or more hidden layer(s), and an output layer. The number of layers and neurons in each layer could influence ANN performance. Thus, it should be properly selected during the design stage. As a general guideline, the neural network size should be as small as possible to allow for efficient computation [72].

Figure 10 compares the network architectures between feature-based ANN and raw data-based ANN as described in Section 4.1.2 based on three-layer MLP model. The size of input representation determined the number of input neurons. Raw data input representation requires 48 neurons,

Table 6 Thirteen candidates ANN architectures

No.	Candidate ANN	Short listed features						% Recognition accuracy		MSE×10^{-3}	Epoch	Time (s)
		LEWMA	Mean	MSD	MMSV	MAT	Slope	Nor	Shift			
1	8×8×7	a						96.39	96.13	7.56	232	55
2	10×16×7	a	a					96.81	97.13	5.23	213	67
3	10×16×7	a		a				97.37	96.83	6.01	176	55
4	10×16×7	a			a			98.39	95.85	10.6	205	64
5	10×16×7	a				a		98.15	94.94	11.6	225	70
6	14×22×7	a		a	a	a		99.14	94.53	10.8	238	91
7	**14×22×7**	**a**	**a**	**a**	**a**			**99.53**	**95.72**	**9.36**	**249**	**100**
8	10×16×7	a					a	95.51	96.75	6.51	255	81
9	16×10×7	a	a	a	a	a		99.46	95.16	16.9	193	49
10	12×16×7	a	a				a	96.85	97.76	5.86	197	61
11	16×10×7	a		a	a	a	a	98.79	96.35	10.1	263	64
12	18×24×7	a	a	a	a	a	a	99.05	96.64	10.2	237	98
13	16×10×7	a	a	a	a		a	97.00	97.31	8.67	312	78
	48×26×7	Raw data						98.70	98.48	3.41	293	136

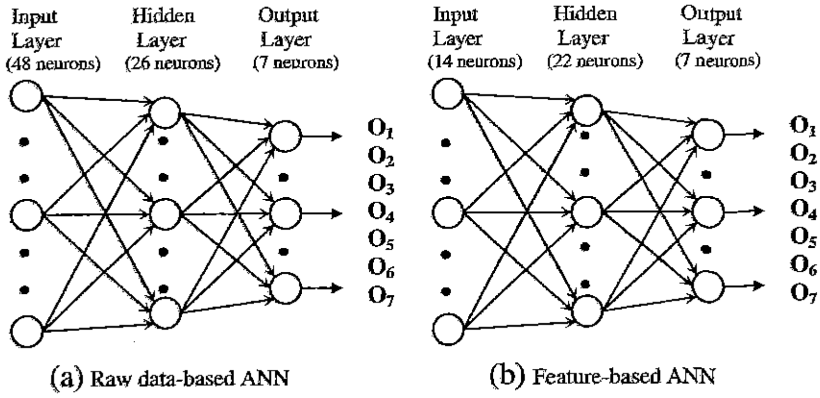


Fig. 10 Network architectures of three-layer MLP models. **a** Raw data-based ANN; **b** feature-based ANN

while statistical features input representation requires only 14 neurons. The output layer contains seven neurons, which was determined according to the number of pattern categories. One hidden layer with 26 neurons and 22 neurons were determined empirically for raw data-based ANN and feature-based ANN, respectively. The experiments showed that initially, the training results improved in-line with the increment in the number of neurons. Once the neurons exceeded the required numbers, further increment of the neurons did not improve the training results but revealed poorer results. These excess neurons could burden the network computationally, reduces the network generalization capability, and increases the total training time.

4.3 Training and testing of ANN recognizer

Partially developed patterns were applied for training the ANN recognizers as described in Section 4.1.2. Detail parameters of the training patterns are summarized in Table 7. It should be noted that total training pattern for bivariate process mean shifts=[100×(total combination of shift)×(total combination of cross correlation)], whereas total training pattern for bivariate normal process=[1,500×(total combination of cross correlation)]. On the other hand, dynamic patterns were used for testing, which is suited for online process monitoring as addressed in [48, 74, 75].

The above design was implemented in MATLAB 2009a. MLP model was trained using "gradient decent with momentum and adaptive learning rate" (traingdx). The other training parameters setting were learning rate (0.05), learning rate increment (1.05), maximum number of epochs (1,500), and error goal (0.001), whereas the network performance was based on mean square error (MSE). Hyperbolic tangent function was used for hidden layer, while sigmoid function was used for an output layer. The training session was stopped either when the number of training epochs was met or the required MSE has been reached. Table 8 summarizes the target outputs, whereby the maximum values (0.9)

Table 7 Parameters for the training patterns

Pattern category	Mean shift (in std dev)	Starting point of sudden shift (SP)	Cross correlation (ρ)	Number of training pattern
N (0, 0)	$X1$: 0.00 $X2$: 0.00	N/A	0.1, 0.3, 0.5, 0.7, 0.9	1,500×5=7,500
US (1, 0)	$X1$: 1.00, 1.25, ..., 3.00 $X2$: 0.00, 0.00, ..., 0.00	For shift range ± [1.00, 1.50], SP = sample 9th		100×9×5=4,500
US (0, 1)	$X2$: 0.00, 0.00, ..., 0.00 $X1$: 1.00, 1.25, ..., 3.00	For shift range ± [1.75, 2.25], SP = sample 13th		100×9×5=4,500
US (1, 1)	$X1$: 1.00, 1.00, 1.25, 1.25, ..., 3.00 $X2$: 1.00, 1.25, 1.00, 1.25, ..., 3.00	For shift range ± [2.50, 3.00], SP = sample 17th		100×25×5=12,500
DS (1, 0)	$X1$: −1.00, −1.25, ..., −3.00 $X2$: 0.00, 0.00, ..., 0.00			100×9×5=4,500
DS (0, 1)	$X2$: 0.00, 0.00, ..., 0.00 $X1$: −1.00, −1.25, ..., −3.00			100×9×5=4,500
DS (1, 1)	$X1$: −1.00, −1.00, −1.25, −1.25, ..., −3.00 $X2$: −1.00, −1.25, −1.00, −1.25, ..., −3.00			100×25×5=12,500

Table 8 Target outputs for an ANN

Pattern	O_1	O_2	O_3	O_4	O_5	O_6	O_7
N (0, 0)	0.9	0.1	0.1	0.1	0.1	0.1	0.1
US (1, 0)	0.1	0.9	0.1	0.1	0.1	0.1	0.1
US (0, 1)	0.1	0.1	0.9	0.1	0.1	0.1	0.1
US (1, 1)	0.1	0.1	0.1	0.9	0.1	0.1	0.1
DS (1, 0)	0.1	0.1	0.1	0.1	0.9	0.1	0.1
DS (0, 1)	0.1	0.1	0.1	0.1	0.1	0.9	0.1
DS (1, 1)	0.1	0.1	0.1	0.1	0.1	0.1	0.9

in each row identify the corresponding neuron expected to secure the highest output for patterns correctly classified.

The coded "difference" (D) as in Eq. 16 was used to differentiate between the correctly classified and wrongly classified patterns:

$$D = (\text{target output neuron}) - 7 \times (\text{actual output neuron}) \quad (16)$$

The constant value (7) was set according to the number of pattern categories. For example, normal (0, 0) pattern that is correctly classified was determined from value [1–7(1)= −6], while normal (0, 0) pattern that is wrongly classified as upshift (1, 0) was determined from value [1–7(2)=−13]. Table 9 shows a matrix of coded difference for the correctly classified and wrongly classified patterns.

5 Results and discussion

The monitoring and diagnosis performances of feature-based ANN scheme against raw data-based ANN scheme were evaluated based on ARL and RA. Three thousand testing patterns for specific magnitudes of mean shift and cross correlation were used. The ARL results were also compared to the traditional MSPC charts such as χ^2 [6], MC1 [8], and MEWMA [9] as reported in the literature.

5.1 Monitoring performance

The ARL were simulated based on the correctly classified patterns. However, the ARL_1 results cannot be compared directly and precisely with the traditional MSPC charts due to the difference in terms of pattern categories. To overcome this problem, the ARL_1 results for the specific magnitudes of mean shift and cross correlation were estimated by averaging the ARL_1 on six shift pattern categories as summarized in Table 10. Type I error (false alarm rate) and type II error were then computed based on the ARL results [68]. The results support the conclusion that the mean shift with larger magnitudes can be detected more quickly with shorter ARL_1 (smaller type II error).

In identifying bivariate stable process, the feature-based ANN scheme gave a longer ARL_0 (smaller type I error) compared to the raw data-based ANN scheme and the traditional MSPC charts. This result is slightly better than the Shewhart control chart ($ARL_0 \approx 370$). On the other hand, the feature-based ANN gave shorter ARL_1 (smaller type II error) for mean shifts ≥1.5 standard deviations, while the raw data-based ANN gave shorter ARL_1 for mean shifts ≤1.0 standard deviations in detecting bivariate process mean shifts. These results indicated that the feature-based ANN is more sensitive for detecting moderate and large mean shifts, while the raw data-based ANN is more sensitive for small mean shifts. Detection capability as shown by the feature-based ANN is comparable to the MC1 [8] and the MEWMA [9] control charts. Overall, it is clear that the feature-based ANN scheme exhibits a strong capability for online monitoring mean shifts in bivariate processes.

5.2 Diagnosis performance

The RA measure how accurate an ANN scheme could perform classification for the source variable(s) patterns towards diagnosing who cause the problem. Besides the overall RA results, the average RA for the specific magnitudes of mean shift and cross correlation as summarized in Table 11 could be used to compare diagnosis performances between the two ANN schemes. The results support the conclusion that the source variable(s) patterns with larger magnitudes of mean shifts can be more accurately classified.

Besides a good overall RA (>93 %), both ANN schemes provided excellent average RA results (≥95 %) for mean

Table 9 Matrix of coded difference for correctly classified and wrongly classified patterns

Pattern	N (0, 0)	US (1, 0)	US (0, 1)	US (1, 1)	DS (1, 0)	DS (0, 1)	DS (1, 1)
N (0, 0)	T (−6)	F (−13)	F (−20)	F (−27)	F (−34)	F (−41)	F (−48)
US (1, 0)	F (−5)	T (−12)	F (−19)	F (−26)	F (−33)	F (−40)	F (−47)
US (0, 1)	F (−4)	F (−11)	T (−18)	F (−25)	F (−32)	F (−39)	F (−46)
US (1, 1)	F (−3)	F (−10)	F (−17)	T (−24)	F (−31)	F (−38)	F (−45)
DS (1, 0)	F (−2)	F (−9)	F (−16)	F (−23)	T (−30)	F (−37)	F (−44)
DS (0, 1)	F (−1)	F (−8)	F (−15)	F (−22)	F (−29)	T (−36)	F (−43)
DS (1, 1)	F (0)	F (−7)	F (−14)	F (−21)	F (−28)	F (−35)	T (−42)

T correctly classified, *F* wrongly classified, *( )* difference value

Table 10 Comparison of the ARLs results

Pattern category	Mean shift (SD)		Feature-based ANN		Raw data-based ANN		χ^2 UCL=10.6 ρ=0.0	MC1 k=0.50 h=4.75 ρ=0.0	MEWMA r=0.10 h=8.66 ρ=0.0
	$X1$	$X2$	ρ=0.3/0.8		ρ=0.3/0.8				
N (0, 0)	0.00	0.00	397.6/NA (0.0025/NA)	Average	240.7/NA (0.0042/NA)	Average	200 (0.005)	203 (0.0049)	200 (0.005)
US (1, 0)	0.75	0.00	19.16/21.69	16.40 / 17.45 (0.939 / 0.943)	16.57/17.59	14.60 / 14.83 (0.932 / 0.933)			
US (0, 1)	0.00	0.75	16.78/17.96		14.39/14.72				
US (1, 1)	0.75	0.75	12.46/11.67		13.16/12.61				
DS (1, 0)	−0.75	0.00	19.32/21.04		14.21/14.79				
DS (0, 1)	0.00	−0.75	18.46/20.94		15.98/16.46				
DS (1, 1)	−0.75	−0.75	12.29/11.41		13.30/12.83				
US (1, 0)	1.00	0.00	11.76/ 12.65	10.49/10.84 (0.905 / 0.908)	10.84/11.14	10.02/10.06 (0.900 / 0.901)	42.00 (0.976)	9.28 (0.892)	10.20 (0.902)
US (0, 1)	0.00	1.00	10.80/ 11.06		9.84/9.80				
US (1, 1)	1.00	1.00	8.30/ 7.99		9.29/9.18				
DS (1, 0)	−1.00	0.00	12.65/ 13.22		9.90/10.06				
DS (0, 1)	0.00	−1.00	11.56/ 12.44		10.76/10.84				
DS (1, 1)	−1.00	−1.00	7.84/7.65		9.50/9.32				
US (1, 0)	1.50	0.00	6.55/6.61	5.95/5.97 (0.832 / 0.832)	7.20/7.16	6.69/6.68 (0.851 / 0.850)	15.80 (0.937)	5.23 (0.809)	6.12 (0.837)
US (0, 1)	0.00	1.50	6.11/6.11		6.49/6.47				
US (1, 1)	1.50	1.50	4.86/4.79		6.37/6.24				
DS (1, 0)	−1.50	0.00	7.17/7.29		6.57/6.71				
DS (0, 1)	0.00	−1.50	6.36/6.50		6.96/7.04				
DS (1, 1)	−1.50	−1.50	4.65/4.53		6.54/6.45				
US (1, 0)	2.00	0.00	4.49/4.47	4.15/4.17 (0.759 / 0.760)	5.54/5.54	5.17/5.15 (0.807 / 0.806)	6.90 (0.855)	3.69 (0.729)	4.41 (0.773)
US (0, 1)	0.00	2.00	4.30/4.32		5.03/5.04				
US (1, 1)	2.00	2.00	3.48/3.47		4.99/4.91				
DS (1, 0)	−2.00	0.00	5.01/5.03		4.95/4.95				
DS (0, 1)	0.00	−2.00	4.38/4.41		5.42/5.43				
DS (1, 1)	−2.00	−2.00	3.25/3.30		5.11/5.02				
US (1, 0)	2.50	0.00	3.38/3.37	3.23/3.22 (0.690 / 0.689)	4.51/4.56	4.25/4.23 (0.765 / 0.764)	3.50 (0.714)	2.91 (0.656)	3.51 (0.715)
US (0, 1)	0.00	2.50	3.35/3.33		4.12/4.14				
US (1, 1)	2.50	2.50	2.77/2.73		4.16/4.12				
DS (1, 0)	−2.50	0.00	3.87/3.80		4.07/4.04				
DS (0, 1)	0.00	−2.50	3.41/3.43		4.39/4.41				
DS (1, 1)	−2.50	−2.50	2.61/2.63		4.23/4.13				
US (1, 0)	3.00	0.00	2.80/2.79	2.67/2.66 (0.625 / 0.624)	3.90/3.90	3.65/3.62 (0.726 / 0.724)	2.20 (0.545)	2.40 (0.583)	2.92 (0.658)
US (0, 1)	0.00	3.00	2.74/2.73		3.57/3.57				
US (1, 1)	3.00	3.00	2.31/2.28		3.62/3.54				
DS (1, 0)	−3.00	0.00	3.15/3.15		3.37/3.41				
DS (0, 1)	0.00	−3.00	2.83/2.82		3.77/3.75				
DS (1, 1)	−3.00	−3.00	2.17/2.20		3.64/3.53				

Items in the parentheses are values of type I error (FAR) and type II error

shifts ≥1.5 standard deviations. This revealed a strong capability for classifying the source variable(s) patterns particularly in dealing with moderate and large mean shifts. The average RA is still high (≥90 %) when dealing with small shift, that is, 1.00 standard deviations. Overall, the feature-based ANN scheme is able to identify the source variable(s) patterns as accurate as the raw data-based ANN scheme when dealing with low cross correlated processes. However, the raw data-based ANN scheme performed slightly better diagnosis in dealing with high cross correlated processes.

6 Conclusions

This paper proposes a feature-based input representation applied into ANN for monitoring and diagnosis of mean

Table 11 Comparison of the RA results

Pattern category	Mean shift (σ)		Feature-based ANN			Raw data-based ANN		
	X1	*X2*	ρ=0.3	ρ=0.8	Average ρ=0.3/0.8	ρ=0.3	ρ=0.8	Average ρ=0.3/0.8
US (1, 0)	0.75	0.00	87.6	85.4	87.1 / 85.5 %	87.0	87.0	87.2/91.5
US (0, 1)	0.00	0.75	85.2	80.2		87.8	87.4	
US (1, 1)	0.75	0.75	89.6	99.8		87.1	99.6	
DS (1, 0)	−0.75	0.00	82.6	72.4		89.4	89.1	
DS (0, 1)	0.00	−0.75	83.7	75.6		87.2	86.3	
DS (1, 1)	−0.75	−0.75	94.0	99.7		84.4	99.3	
US (1, 0)	1.00	0.00	92.6	88.7	91.2 / 90.0 %	91.1	88.3	91.5/93.9
US (0, 1)	0.00	1.00	89.7	87.8		90.9	90.5	
US (1, 1)	1.00	1.00	94.0	99.9		92.6	99.9	
DS (1, 0)	−1.00	0.00	85.4	81.3		94.4	94.0	
DS (0, 1)	0.00	−1.00	87.1	82.4		91.1	91.2	
DS (1, 1)	−1.00	−1.00	98.2	100		88.9	99.4	
US (1, 0)	1.50	0.00	96.7	96.5	94.8 / 95.0 %	95.8	95.7	94.7/97.0
US (0, 1)	0.00	1.50	96.6	95.9		95.4	95.4	
US (1, 1)	1.50	1.50	94.5	100		93.6	99.9	
DS (1, 0)	−1.50	0.00	90.9	88.1		95.9	95.9	
DS (0, 1)	0.00	−1.50	92.1	89.3		95.5	95.0	
DS (1, 1)	−1.50	−1.50	98.1	100		92.0	99.8	
US (1, 0)	2.00	0.00	97.6	97.4	95.8 / 96.0 %	96.5	96.7	95.0/97.4
US (0, 1)	0.00	2.00	97.8	96.8		95.0	95.0	
US (1, 1)	2.00	2.00	95.4	99.9		93.8	99.9	
DS (1, 0)	−2.00	0.00	92.5	91.0		96.2	96.6	
DS (0, 1)	0.00	−2.00	92.3	91.0		95.6	96.4	
DS (1, 1)	−2.00	−2.00	99.0	100		92.7	99.8	
US (1, 0)	2.50	0.00	98.0	97.7	96.3 / 96.9 %	96.9	97.3	95.9/97.7
US (0, 1)	0.00	2.50	98.2	97.7		95.7	95.5	
US (1, 1)	2.50	2.50	96.8	99.8		94.5	99.8	
DS (1, 0)	−2.50	0.00	93.2	93.2		97.9	97.4	
DS (0, 1)	0.00	−2.50	92.1	92.8		97.0	96.1	
DS (1, 1)	−2.50	−2.50	99.2	100		93.1	99.8	
US (1, 0)	3.00	0.00	97.9	97.9	96.4 / 97.2 %	97.7	97.4	96.0/97.7
US (0, 1)	0.00	3.00	98.7	98.4		95.8	95.5	
US (1, 1)	3.00	3.00	96.5	99.8		95.1	99.8	
DS (1, 0)	−3.00	0.00	93.3	93.3		98.3	97.6	
DS (0, 1)	0.00	−3.00	92.4	93.6		96.3	96.1	
DS (1, 1)	−3.00	−3.00	99.4	99.9		92.5	99.5	

shifts in bivariate processes. The univariate SPC/CCPR techniques are no longer sufficient to monitor dependent/correlated processes in today's manufacturing industries such as in manufacturing of audio-video device components, among other products. While there have been significant advances in univariate CCPR, very limited works have been reported in BPR.

Lack of reference multivariate/bivariate patterns and excess false alarms seems to have hindered advances in this area. This paper provides some reference patterns of bivariate process mean shifts plotted on scatter diagrams. An improved set of statistical features input representation was utilized into an ANN training and testing for improving discrimination capability between bivariate normal and bivariate mean shift patterns. The optimal set of statistical features was determined through a proper features selection and analysis. This approach gave a smaller network size and strong monitoring and diagnosis capabilities. The proposed scheme generally reveals a better monitoring and a comparable diagnosis performances compared to the raw data-based ANN. Nevertheless, the raw data-based ANN could be an alternative scheme when the recognition accuracy is a main concern.

For the future work, we plan to conduct a more extensive investigation towards heightening the performance of the feature-based ANN for identifying smaller mean shifts. Investigation also will be extended for other causable patterns such as trend and cyclic.

Acknowledgments The authors would like to thank Universiti Tun Hussein Onn Malaysia, Universiti Teknologi Malaysia, and Ministry of Higher Education of Malaysia who are sponsoring this work. The constructive suggestions from the Editor in Chief of IJAMT and the anonymous reviewers have greatly aided us to improve the contents as well as the presentation of this paper.

References

1. Oakland JS (1996) Statistical process control. Butterworth-Heinemann, Oxford
2. Shewhart WA (1931) The economic control of quality manufactured products. Van Nostrand, New York
3. Nelson LS (1989) Standardization of Shewhart control charts. J Qual Technol 21(4):287–289
4. Crosier RB (1986) A new two-sided cumulative sum quality control schemes. Technometrics 28(3):187–194
5. Lucas JM, Saccucci MS (1990) Exponentially weighted moving average control schemes: properties and enhancements. Technometrics 32(1):1–12
6. Hotelling HH (1947) Multivariate quality control. In: Eisenhart C, Hastay MW, Wallis WA (eds) Techniques of statistical analysis. McGraw-Hill, New York
7. Crosier RB (1988) Multivariate generalizations of cumulative sum quality control schemes. Technometrics 30(3):291–303
8. Pignatiello JJ, Runger GC (1990) Comparisons of multivariate CUSUM charts. J Qual Technol 22(3):173–186
9. Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) Multivariate exponentially weighted moving average control chart. Technometrics 34(1):46–53
10. Prabhu SS, Runger GC (1997) Designing a multivariate EWMA control chart. J Qual Technol 29(1):8–15
11. Alt FB (1985) Multivariate quality control. In: Johnson NL, Kotz S (eds) Encyclopedia of statistical sciences 6. Wiley, New York
12. Jackson JE (1991) A user's guide to principle component. Wiley, New Jersey
13. Fuchs C, Benjamini Y (1994) Multivariate profile charts for statistical process control. Technometrics 36(2):182–195
14. Mason RL, Tracy ND, Young JC (1995) Decomposition of T^2 for multivariate control chart interpretation. J Qual Technol 27(2):99–108
15. Mason RL, Tracy ND, Young JC (1996) Monitoring a multivariate step process. J Qual Technol 28(1):39–50
16. Sepulveda A, Nachlas JA (1997) A simulation approach to multivariate quality control. Comput Ind Eng 33(1–2):113–116
17. Lowry CA, Montgomery DC (1995) A review of multivariate control charts. IIE Trans 27(6):800–810
18. Kourti T, MacGregor JF (1996) Multivariate SPC methods for process and product monitoring. J Qual Technol 28(4):409–428
19. Mason RL, Tracy ND, Young JC (1997) A practical approach for interpreting multivariate T^2 control chart signals. J Qual Technol 29(4):396–406
20. Bersimis S, Psarakis S, Panaretos J (2007) Multivariate statistical process control charts: an overview. Qual Reliab Eng Int 23:517–543
21. Wöhler C, Anlauf JK (1999) A time delay neural network algorithm for estimating image-pattern shape and motion. Image Vis Comput 17:281–294
22. Singh V, Mishra R (2006) Development a machine vision system for spangle classification using image processing and artificial neural network. Surf Coat Technol 201:2813–2817
23. Liang SF, Lu SM, Chang JY, Lin CT (2008) A novel two-stage impulse noise removal technique based on neural networks and fuzzy decision. IEEE Trans Fuzzy Syst 16(4):863–873
24. Christodoulou CI, Pattichis CS (1999) Unsupervised pattern recognition for the classification of EMG Signals. IEEE Trans Biomed Eng 46(2):169–178
25. Güven A, Kara S (2006) Diagnosis of the macular diseases from pattern electroretinography signals using artificial neural networks. Expert Syst Appl 30:361–366
26. Altunay S, Telatar Z, Erogul O, Aydur E (2009) A new approach to uninary system dynamics problems: evaluation and classification of uroflowmeter signals using artificial neural networks. Expert Syst Appl 36:4891–4895
27. Er O, Yumusak N, Temurtas F (2010) Chest diseases diagnosis using artificial neural networks. Expert Syst Appl 37:7648–7655
28. Chen CW (2009) Modeling and control for nonlinear structural systems via a NN-based approach. Expert Syst Appl 36:4765–4772
29. Chen CW, Yeh K, Liu KFR (2009) Adaptive fuzzy sliding mode control for seismically excited bridges with lead rubber bearing isolation. Int J Uncertain Fuzziness Knowl-Based Syst 17(5):705–727
30. Chen CY, Lin JW, Lee WI, Chen CW (2010) Fuzzy control for an oceanic structure: a case study in time-delay TLP system. J Vib Control 16(1):147–160
31. Amin A (2000) Recognition of printed Arabic text based on global features and decision tree learning techniques. Pattern Recogn 33:1309–1323
32. Bhattacharyya K, Sarma KK (2009) ANN-based innovative segmentation method for handwritten text in assamese. Int J Comput Sci Issues 5:9–16
33. Marcu T, Seliger BK, Stücher R (2008) Design of fault detection for a hydraulic looper using dynamic neural networks. Control Eng Pract 16:192–213
34. Demetgul M, Tansel IM, Taskin S (2009) Fault diagnosis of pneumatic systems with artificial neural network algorithms. Expert Syst Appl 36:10512–10519
35. Jia ZY, Ma JW, Wang FJ, Liu W (2010) Characteristics forecasting of hydraulic valve based on grey correlation and ANFIS. Expert Syst Appl 37:1250–1255
36. Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, New Jersey
37. Hwarng HB, Hubele NF (1993) Back-propagation pattern recognizers for X-bar control charts: methodology and performance. Comput Ind Eng 24:219–235
38. Velasco T (1993) Back propagation artificial neural networks for the analysis of quality control charts. Comput Ind Eng 25(1–4):397–400
39. Pham DT, Oztemel E (1993) Control chart pattern recognition using combinations of multilayer perceptrons and learning vector quantization neural networks. Proc Inst Mech Eng 207:113–118
40. Pham DT, Oztemel E (1994) Control chart pattern recognition using learning vector quantization networks. Int J Prod Res 32:721–729
41. Zorriassatine F, Tannock JDT (1998) A review of neural networks for statistical process control. J Intell Manuf 9:209–224
42. Masood I, Hassan A (2010) Issues in development of artificial neural network-based control chart pattern recognition schemes. Eur J Sci Res 39(3):336–355

43. Chen LH, Wang TY (2004) Artificial neural networks to classify mean shifts from multivariate χ^2 chart signals. Comput Ind Eng 47:195–205
44. Niaki STA, Abbasi B (2005) Fault diagnosis in multivariate control charts using artificial neural networks. Qual Reliab Eng Int 21:825–840
45. Cheng CS, Cheng HP (2008) Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines. Expert Syst Appl 35:198–206
46. Yu JB, Xi LF, Zhou XJ (2009) Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble. Eng Appl Artif Intell 22:141–152
47. Zorriassatine F, Tannock JDT, O'Brien C (2003) Using novelty detection to identify abnormalities caused by mean shifts in bivariate processes. Comput Ind Eng 44:385–408
48. Guh RS (2007) On-line identification and quantification of mean shifts in bivariate processes using a NN-based approach. Qual Reliab Eng Int 23:367–385
49. Yu JB, Xi LF (2009) A neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes. Expert Syst Appl 36:909–921
50. El-Midany TT, El-Baz MA, Abd-Elwahed MS (2010) A proposed framework for control chart pattern recognition in multivariate process using artificial neural networks. Expert Syst Appl 37:1035–1042
51. Park S-J, Lee M-S, Shin S-Y, Cho K-H, Lim J-T, Cho B-S, Y-H Jei, Kim M-K, Park C-H (2005) Run-to-run overlay control of steppers in semiconductor manufacturing systems based on historical data analysis and neural network modeling. IEEE Trans Semicon Manuf 18(4):605–613
52. Mason RL, Chou YM, Young JC (2001) Applying Hotelling T^2 statistic to batch processes. J Qual Technol 33(4):466–479
53. Parra MGL, Loziza PR (2003–2004) Application of multivariate T^2 control chart and Mason-Tracy decomposition procedure to the study of the consistency of impurity profiles of drug substances. Quality Engineering 16(1):127–142
54. Singh R, Gilbreath G (2002) A real-time information system for multivariate statistical process control. Int J Prod Econ 75:161–172
55. Miletic I, Quinn S, Dudzic M, Vaculik V, Champagne M (2004) An industrial perspective on implementation on-line applications of multivariate statistics. J Process Control 14:821–836
56. Lehman RS (1977) Computer simulation and modeling: an introduction. Lawrence Erlbaum, London
57. Brunzell H, Eriksson J (2000) Feature reduction for classification of multidimensional data. Pattern Recog 33(12):1741–1748
58. Klosgen W, Zytkow JM (2002) Handbook of data mining and knowledge discovery. Oxford University Press, London
59. Pham DT, Wani MA (1997) Feature-based control chart pattern recognition. Int J Prod Res 35(7):1875–1890
60. Gauri SK, Chakraborty S (2006) Feature-based recognition of control chart patterns. Comput Ind Eng 51:726–742
61. Gauri SK, Chakraborty S (2008) Improved recognition of control chart patterns using artificial neural networks. Int J Adv Manuf Technol 36:1191–1201
62. Hassan A, Nabi Baksh MS, Shaharoun MA, Jamaludin H (2003) Improved SPC chart pattern recognition using statistical features. Int J Prod Res 41(7):1587–1603
63. Guh RS (2010) Simultaneous process mean and variance monitoring using artificial neural networks. Comput Ind Eng 58:739–753
64. Al-Assaf Y (2004) Recognition of control chart patterns using multi-resolution wavelets analysis and neural networks. Comput Ind Eng 47:17–29
65. Assaleh K, Al-Assaf Y (2005) Features extraction and analysis for classifying causable patterns in control charts. Comput Ind Eng 49:168–181
66. Chen Z, Lu S, Lam S (2007) A hybrid system for SPC concurrent pattern recognition. Adv Eng Inform 21:303–310
67. Wang CH, Kuo W, Qi H (2007) An integrated approach for process monitoring using wavelet analysis and competitive neural network. Int J Prod Res 45:227–244
68. Montgomery DC (2005) Introduction to statistical quality control. Wiley, New Jersey
69. Demuth H, Beale M (1998) Neural network toolbox, user's guide. The MathWorks, Natick
70. Bishop C (1995) Neural network for pattern recognition. Oxford University Press, New York
71. Cheng CS (1995) A multi-layer neural network model for detecting changes in the process mean. Comput Ind Eng 28:51–61
72. Cheng CS (1997) A neural network approach for the analysis of control chart patterns. Int J Prod Res 35(3):667–697
73. Guh RS, Hsieh YC (1999) A neural network based model for abnormal pattern recognition of control charts. Comput Ind Eng 36:97–108
74. Guh RS, Tannock JDT, O'Brien C (1999) Intellispc: a hybrid intelligence tool for on-line economical statistical process control. Expert Syst Appl 17:195–212
75. Guh RS, Zoriassatine F, Tannock JDT, O'Brien C (1999) On-line control chart pattern detection and discrimination—a neural network approach. Artif Intell Eng 13:413–425
76. Perry MB, Spoerre JK, Velasco T (2001) Control chart pattern recognition using back propagation artificial neural networks. Int J Prod Res 39:3399–3418