Computers & Industrial Engineering 56 (2009) 1577–1588

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Recognition of control chart patterns using improved selection of features ☆

Susanta Kumar Gauri [a,1], Shankar Chakraborty [b,*]

[a] SQC & OR Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India
[b] Dept. of Production Engineering, Jadavpur University, S.C. Mullick Road, Kolkata 700032, India

A R T I C L E   I N F O

Article history:
Received 4 July 2007
Received in revised form 22 May 2008
Accepted 10 October 2008
Available online 17 October 2008

Keywords:
Control chart patterns
Shape features
Heuristics
Artificial neural network
Pattern recognition

A B S T R A C T

Recognition of various control chart patterns (CCPs) can significantly reduce the diagnostic search process. Feature-based approaches can facilitate efficient pattern recognition. The full potentiality of feature-based approaches can be achieved by using the optimal set of features. In this paper, a set of seven most useful features is selected using a classification and regression tree (CART)-based systematic approach for feature selection. Based on these features, eight most commonly observed CCPs are recognized using heuristic and artificial neural network (ANN) techniques. Extensive performance evaluation of the two types of recognizers reveals that both these recognizers result in higher recognition accuracy than the earlier reported feature-based recognizers. In this work, various features are extracted from the control chart plot of actual process data in such a way that their values become independent of the process mean and standard deviation. Thus, the developed feature-based CCP recognizers can be applicable to any general process.

1. Introduction

Control charts, predominantly in the form of $\bar{X}$ chart, are recognized as one of the most important tools for statistical process control (SPC). However, the effectiveness of the use of control charts largely depends on the ability to detect various types of patterns that are commonly observed on these charts. The control chart patterns (CCPs) can be classified as natural/normal and unnatural/abnormal (Montgomery, 2001). The basic significance of a natural pattern is that it indicates a process under control, whereas an unnatural pattern identifies a process when it is out-of-control. The unnatural patterns can be of various types and each type of unnatural pattern can be associated with specific assignable causes (Western Electric, 1958). Hence, effective identification of the unnatural patterns can greatly narrow down the set of possible assignable causes to be investigated, while significantly reducing the diagnostic search process.

The statistical quality control handbook (Western Electric, 1958) includes 15 common types of CCPs. Amongst these CCPs, there are eight basic CCPs, e.g. normal (NOR), stratification (STA), systematic (SYS), cyclic (CYC), increasing trend (UT), decreasing trend (DT), upward shift (US) and downward shift (DS) patterns, as shown in Fig. 1. All other CCPs, e.g. freaks, grouping or bunching, instability, interaction, mixture, etc. are either special forms of basic CCPs or mixed forms of two or more basic CCPs. Typically, pattern recognition has been based on the visual judgment of the users. Over the years, numerous supplementary rules like zone tests or run rules have been developed to assist the users for detection of unnatural patterns (Nelson, 1984, 1985). Run rules are based on the concept that a run has a low probability of occurrence on a completely random scattering of points around a mean. If a run is detected, then this will indicate that some special causes are affecting the process. One of the main problems with run rules is that the simultaneous application of all these rules is likely to result in excessive number of false alarms. Furthermore, the pattern characteristics which are being looked for may be common to more than one pattern. Therefore, identification and analysis of the unnatural patterns require considerable experience and skill from the part of the practitioners.

With the development of advanced manufacturing technology, many of the functions in manufacturing, traditionally performed by human beings, have now been replaced by machines. Technology searching on automatic recognition of abnormal CCPs has also come to fore. Under the current scenario of widespread use of automated data acquisition systems for computer charting and analysis of the manufacturing process data, the need for implementing an automated system for analysis of the process data has increased manifolds. A number of works aiming automated recognition of CCPs have been reported in the literature. Some researchers (Evans & Lindsay, 1988; Pham & Oztemel, 1992a; Swift & Mize, 1995) have developed expert systems for CCP recognition. Several techniques have been deployed in the knowledge base design of the expert systems, including template matching, run rules

☆ This manuscript was processed by Area Editor E.A. Elsayed.
* Corresponding author. Tel.: +91 033 2414 6153/2548 2655; fax: +91 033 2414 6414/2413 7121.
E-mail addresses: susantagauri@hotmail.com (S.K. Gauri), s_chakraborty00@yahoo.co.in (S. Chakraborty).
[1] Tel.: +91 033 2575 3350/2668 0473; fax: +91 033 2577 6042.

0360-8352/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2008.10.006

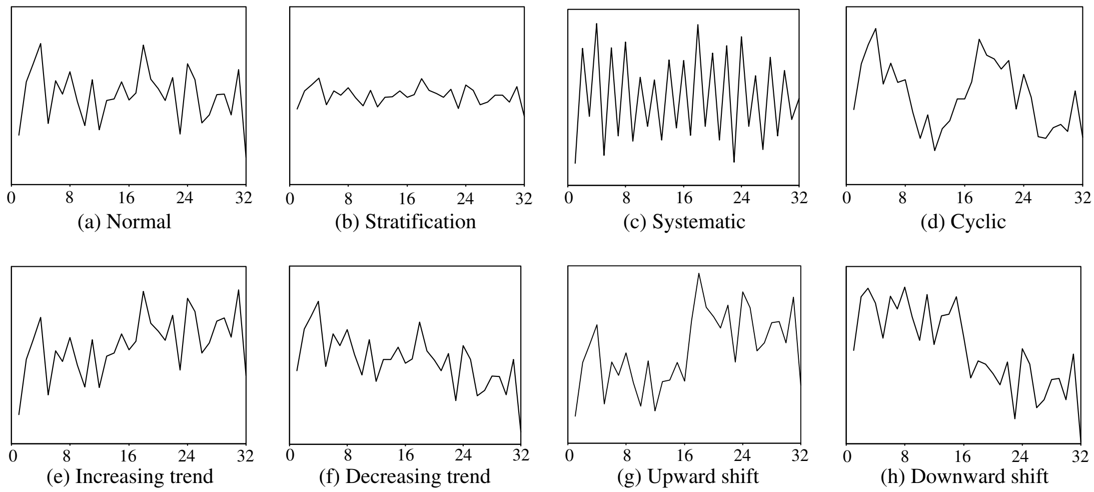

Fig. 1. Various control chart patterns.

and heuristic algorithms. However, the use of rules based on statistical properties in the expert systems has the difficulty that similar statistical properties may be derived for some patterns of different classes, which may create problems of incorrect recognition.

Other researchers (Cheng, 1997; Guh & Shiue, 2005; Guh & Tannock, 1999; Guh, Zorriassatine, Tannock, & O'Brien, 1999; Hwarng & Hubele, 1993; Perry, Spoerre, & Velasco, 2001; Pham & Oztemel, 1992b) have successfully applied artificial neural networks (ANN) for CCP recognition task. The advantage with neural network 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. One disadvantage with neural network is the difficulty in understanding how a particular classification decision has been reached and also in determining the details of how a given pattern resembles with a particular class. In addition, there is no systematic way to select the topology and architecture of a neural network. In general, this has to be found empirically, which can be time consuming.

Pham and Wani (1997) and Gauri and Chakraborty (2006a) have highlighted that each control chart pattern has its own geometric shape and various related features can represent this shape. Different patterns can, therefore, be efficiently discriminated based on these shape features extracted from the control chart plot. Hassan, Nabi Baksh, Shaharoun, and Jamaluddin (2003) have shown that statistical features like mean, skewness etc. can also be useful in differentiating various CCPs. However, one limitation while using the statistical features is that their extraction requires considerably large number of observations and their interpretation is not easily comprehendible to the practitioners. Moreover, the statistical features lose information on the order of the data. The advantage of shape features is that those can be extracted from lesser number of observations without losing order of the data.

The feature-based approaches provide a greater choice of recognition techniques. Pham and Wani (1997) and Gauri and Chakraborty (2006a) have demonstrated that both the properly developed heuristics based on the extracted features and an appropriately designed ANN trained using the extracted features as input vector representation can efficiently differentiate various CCPs. Since extracted features represent the main characteristics of the original data in a condensed form, both the feature-based heuristic and neural network approaches can facilitate efficient pattern recognition. The feature-based heuristic approach has a distinct advantage that the practitioners can clearly understand how a particular pattern has been identified by the use of relevant shape features. On the other hand, feature-based neural network approach results in higher recognition accuracy (Gauri & Chakraborty, 2006a; Pham & Wani, 1997). Extraction of various shape features from the raw pattern data and using those as input vector also reduce the network size and learning time.

Pham and Wani (1997) and Gauri and Chakraborty (2006a) have selected the features for CCP recognition on adhoc basis. The full potential of feature-based approaches is not exploited yet because the set of features that may be the most appropriate for CCP recognition is not known. Gauri and Chakraborty (2006b) have carried out an in-depth study on various potentially useful features of CCPs. They have shown that a large number of features can be extracted from CCPs using various considerations, e.g. fitting least square (LS) line to the data points, computing the area between the overall pattern and mean/LS line, number of crossovers of data points to the mean/LS line, distance between the consecutive data points etc. Further, discrimination between two or more specific patterns can be possible based on more than one shape features. Therefore, selection of the set of most useful features is an important issue for improving the recognition accuracy.

Gauri and Chakraborty (2006b) have used classification and regression tree (CART) analysis (Breiman, Friedman, Olshen, & Stone, 1984) for selection of the optimal set of features. Given the values of a large number of features, extracted from the learning samples, CART analysis can make automatic selection of the best set of useful features for classification of patterns in the learning samples and develop the rules for pattern classification in the form of decision tree. However, it is found that in this approach of automatic selection, a large number of features including two or more highly correlated features may be selected as useful features. For example, while Gauri and Chakraborty (2006b) have subjected the extracted values of 32 features from the learning samples, a decision tree has been developed based on automatically selected 13 features including *SRANGE* (range of slope of the straight lines passing through six pairwise combination of midpoints of four equal segments) and *BRANGE* (range of slopes of the LS lines fitted to six subsets of *N/2* data points obtained by pairwise combination of four equal segments), which two are highly correlated. On the other hand, if the extracted features from two different sets of learning samples are subjected to CART analysis, it may result in two entirely different sets of useful features. Whereas prediction based on correlated variables can result in prediction instability (Montgomery & Peck, 1982), changes in the set of useful features can lead to confusion among the users. Selecting first the set of most useful features and then subjecting the extracted values of these features to CART analysis can ensure that the automatically generated heuristic rules in the form of classification tree will always be based on the same set of features.

In this paper, almost all the features proposed by Gauri and Chakraborty (2006a, 2006b) are considered with minor modification of some features. All these features are extracted from the plot of actual process data (without scaling or standardization) in such a way that their values will be independent of the process mean and standard deviation. A CART-based systematic approach for feature selection is proposed, using which a set of seven most appropriate features is selected for developing the CCP recognizers. Various CCPs are recognized using heuristic and ANN-based approaches. The relative performance of both these approaches is extensively studied using synthetic pattern data. Since the features are now generalized, the developed CCP recognizers can be applied to any process.

The remainder of this paper is organized in different sections. The list of considered features and the methodologies used for their extraction are presented in Section 2. The procedure for generation of various control chart patterns is described in Section 3. The proposed approach and selection of the set of features are discussed in Section 4. Section 5 discusses the design of the pattern recognizers and the experimental procedure. Section 6 provides the results and discussions. Section 7 concludes the paper.

2. List of features and their extraction methodologies

In total, 30 shape features are taken into consideration from which a set of the most appropriate features will be selected. Most of these features remain the same as proposed by Gauri and Chakraborty (2006a, 2006b) and a few are the modifications of their proposed forms. These features can be broadly classified into three groups, e.g. (a) features extractable without segmentation of the control chart plot, (b) features extractable based on pre-defined segmentation of the control chart plot and (c) features extractable based on criterion-based segmentation of the control chart plot. The mathematical expressions of these features along with their usefulness in pattern discrimination are discussed, in short, as below.

2.1. Features extractable without segmentation of the control chart plot

(i) Absolute slope of the least square (LS) line representing the overall pattern (*AB*):

$$AB = \left| \sum_{i=1}^{N} y_i(t_i - \bar{t}) \Big/ \sum_{i=1}^{N} (t_i - \bar{t})^2 \right| \tag{1}$$

where, $t_i = ic$ (i=1, 2,…,N) is the distance of ith time point of observation from the origin, c is a constant linear distance used to represent a given sampling interval on the control chart plot, y_i is the observed value of a quality characteristic at ith time point, N is the size of the observation window and $\bar{t} = \sum_{i=1}^{N} t_i/N$. The magnitude of *AB* for trend (UT and DT) and shift (US and DS) patterns is greater than zero. For all other patterns, the slope is about zero.

(ii) Sign of slope of the LS line representing the overall pattern (*SB*):

The *SB* can be viewed as a categorical variable, which is '0' if the value of the expression, $\sum_{i=1}^{N} y_i(t_i - \bar{t}) / \sum_{i=1}^{N} (t_i - \bar{t})^2$ is negative and '1' otherwise. It can discriminate DT versus UT and DS versus US patterns.

(iii) Ratio between variance of the data points in the observation window and mean sum of squares of errors of the LS line representing the overall pattern (*RVE*):

$$RVE = [S_{yy}/(N-1)] \Big/ \left[\left(S_{yy} - S_{yt}^2/S_{tt} \right) \Big/ (N-2) \right]. \tag{2}$$

where, S_{yy} is the corrected sum of squares of the observed values, S_{tt} is the corrected sum of squares of the distances of time points from the origin on control chart plot and S_{yt} is the corrected sum of cross-products of the observed values and distances from origin. The magnitude of *RVE* for NOR, STA, SYS and CYC patterns is approximately one, while for trend and shift patterns, it is greater than one.

(iv) Area between the overall pattern and mean line per interval in terms of *SD²* (*ACLPI*):

$$ACLPI = [ACL/(N-1)]/SD^2; \quad SD = \left[\sum_{i=1}^{N} (y_i - \bar{y})^2 \Big/ (N-1) \right]^{1/2} \tag{3}$$

where *ACL* is the area between the pattern and mean line. The value of *ACL* can be easily computed by summing the areas of the triangles and trapeziums that are formed by the mean line and overall pattern using simple algorithm. The magnitude of *ACLPI* is the highest for STA pattern, lowest for SYS pattern and intermediate for all other patterns.

(v) Area between the overall pattern and mean line per mean line crossover in terms of *SD²* (*ACLMLC*):

$$ACLMLC = [ACL/MLC]/SD^2; \quad (MLC \text{ is the number of mean line crossovers}) \tag{4}$$

The relative values of *ACLMLC* are higher for STA and CYC patterns than NOR and SYS patterns, and shift patterns than trend patterns.

(vi) Area between the overall pattern and LS line per interval in terms of *SD²* (*ALSPI*):

$$ALSPI = [ALS/(N-1)]/SD^2 \tag{5}$$

where *ALS* is the area between the pattern and fitted LS line. Likewise *ACL*, the value of *ALS* can also be computed by summing the areas of the triangles and trapeziums that are formed by the LS line and overall pattern. The relative values of *ALSPI* with respect to different patterns vary almost similarly as the feature *ACLPI*.

(vii) Area between the pattern and least square line per LS line crossover in terms of *SD²* (*ALSLSC*):

$$ALSLSC = [ALS/LSC]/SD^2; \quad (LSC \text{ is the number of LS line crossovers}) \tag{6}$$

The relative values of *ALSLSC* with respect to different patterns are almost similar to the feature *ACLMLC*.

(viii) Ratio of area between the pattern and mean line and area between the pattern and LS line:

$$RACLALS = ACL/ALS \tag{7}$$

The value of *RACLALS* for NOR, STA, SYS and CYC patterns will be approximately one, while for trend and shift patterns, it will be greater than one.

(ix) Proportion of the number of crossovers to mean line (*PMLC*):

$$PMLC = \sum_{i=1}^{N-1} o_i \Big/ N \tag{8}$$

where $o_i = 1$ if $(y_i - \bar{y})(y_{i+1} - \bar{y}) < 0$; otherwise $o_i = 0$ and $\bar{y}$ is the mean value of N data points. The magnitude of *PMLC* is the highest for SYS pattern, intermediate for NOR and STA patterns, and least for CYC, shift and trend patterns.

(x) Proportion of the number of crossovers to least square line (*PLSC*):

$$PLSC = \sum_{i=1}^{N-1} o'_i \Big/ N \tag{9}$$

where, $o'_i = 1$ if $(y_i - y'_i)(y_{i+1} - y'_{i+1}) < 0$; otherwise $o'_i = 0$ and y'_i is the least square estimate of ith data point. The magnitude of *PLSC* is the highest for SYS pattern, intermediate for NOR, STA and trend patterns, and least for CYC and shift patterns.

(xi) Proportion of the sum of number of crossovers to mean line and LS line (*PSMLSC*):

$$PSMLSC = \sum_{i=1}^{N-1} (o_i + o'_i) \Big/ 2N \tag{10}$$

The SYS pattern as well as CYC and shift patterns can be differentiated well from other patterns by the magnitude of *PSMLSC*.

(xii) Average distance between the consecutive points in terms of *SD* (*ADIST*):

$$ADIST = \left\{ \sum_{i=1}^{N} \left[(t_{i+1} - t_i)^2 + (y_{i+1} - y_i)^2 \right]^{1/2} \Big/ (N-1) \right\} \Big/ SD \tag{11}$$

The magnitude of *ADIST* is the highest for STA pattern, intermediate for NOR and SYS patterns, and on the lower side for all other patterns.

(xiii) Average absolute slope of straight lines passing through the consecutive points *(AASBP)*:

$$AASBP = \sum_{i=1}^{N-1} \left| (y_{i+1} - y_i)/(t_{i+1} - t_i) \right| \Big/ (N-1) \tag{12}$$

The magnitude of *AASBP* is the highest for SYS pattern, lowest for STA pattern and intermediate for all other patterns.

2.2. Features extractable based on pre-defined segmentation of the control chart plot

The basic difference between trend and shift patterns is that in case of trend patterns, the departure of observations from the target value occurs gradually and continuously, whereas in case of shift patterns, the departure occurs suddenly and then the observations hang around the departed value. The aim of extracting features after segmentation of the observation window is to differentiate these two types of departures. Under the pre-determined segmentation approach, the total length of the data plot is divided into four equal segments, as shown in Fig. 2.

The four segments indicate behaviors of the process in four different quarters of the total time span in an observation window. The behavior of the process in a segment, can be represented by a point,

$$\left\{ \left[\sum_{i=n_1}^{n_1+7} t_i \Big/ 8 \right], \left[\sum_{i=n_1}^{n_1+7} y_i \Big/ 8 \right] \right\}$$

where, $n_1 = 1$, 9, 17 and 25 for the first, second, third and fourth segment, respectively. This is the midpoint of a segment. A combination of two midpoints can be obtained in $C_2^4 = 6$ ways. On the other hand, observations in the four segments can be recombined into subsets of $N/2$ data points in $C_2^4 = 6$ ways. The following eleven features are extracted based on the properties of straight lines drawn through different combinations of the midpoints and fitted LS lines to different subsets of $N/2$ data points.

(i) Absolute average slope of the straight lines passing through six pairwise combinations of midpoints of four equal segments (*AASL*):

$$AASL = \left| \sum_{j,k} s_{jk} \Big/ 6 \right|; \quad (j = 1, 2, 3; \quad k = 2, 3, 4; \quad j < k) \tag{13}$$

where s_{jk} is the slope of the straight line passing through jth and kth midpoints. The magnitude of *AASL* for NOR, STA and SYS patterns will be almost zero, but for trend and shift patterns, its value will be greater than zero. In case of CYC pattern, *AASL* will be around zero if each segment contains a complete cycle; otherwise, it will be greater than zero.

(ii) Sign of average slope of the straight lines passing through six pairwise combinations of midpoints (*SASL*):

Likewise *SB*, *SASL* can be viewed as a categorical variable, which is '0' if the value of the expression, $\sum_{j,k} s_{jk}/6$ is negative and '1' otherwise. It can segregate DT and DS patterns from UT and US patterns, respectively.

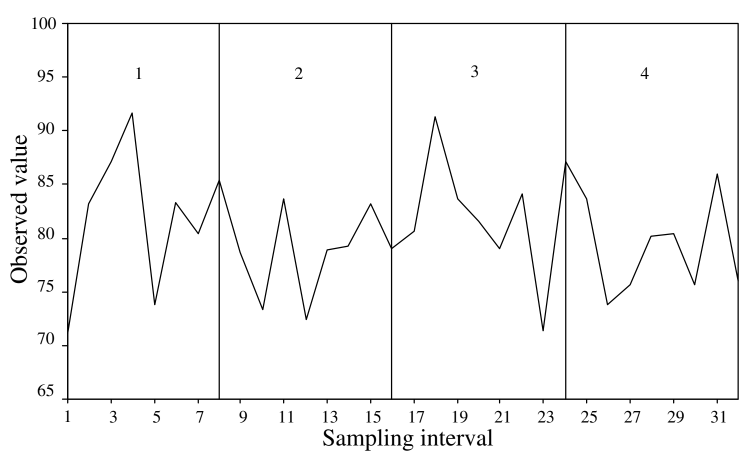

Fig. 2. Four equal segments in a pattern.

(iii) Range of slopes of straight lines passing through six pairwise combinations of midpoints of four equal segments (*SRANGE*):

$$SRANGE = \text{maximum}(s_{jk}) - \text{minimum}(s_{jk}); \quad (j = 1,2,3; \quad k = 2,3,4; \quad j < k) \tag{14}$$

The magnitude of *SRANGE* will be higher for shift patterns than trend patterns. The value of *SRANGE* will also be higher for CYC pattern than NOR, STA and SYS patterns, unless each segment of CYC pattern consists of a complete cycle.

(iv) Absolute average slope of the LS lines fitted to six subsets of $N/2$ data points (*AABL*):

$$AABL = \left| \sum_{j,k} b_{jk} \middle/ 6 \right|; \quad (j = 1,2,3; \quad k = 2,3,4; \quad j < k) \tag{15}$$

where b_{jk} is the slope of the LS line fitted to the observations in jth and kth segments. The magnitudes of *AABL* for different patterns will also vary in the same fashion as *AASL*.

(v) Sign of average slope of the LS lines fitted to six subsets of $N/2$ data points (*SABL*):

Likewise *SB* or *SASL*, *SABL* can be represented as a categorical variable, which is '0' if the value of the expression, $\sum_{j,k} b_{jk}/6$ is negative and '1' otherwise. It can segregate UT and US patterns from DT and DS patterns, respectively.

(vi) Range of slopes of the LS lines fitted to six subsets of $N/2$ data points (*BRANGE*):

$$BRANGE = \text{Maximum}(b_{jk}) - \text{Minimum}(b_{jk}); \quad (j = 1,2,3; \quad k = 2,3,4; \quad j < k) \tag{16}$$

The magnitudes of *BRANGE* for different patterns will vary in the same way as *SRANGE*.

(vii) Ratio of mean sum of squares of errors (*MSE*) of the LS line fitted to overall data and average *MSE* of the LS lines fitted to six subsets of $N/2$ data points (*REAE*):

$$REAE = MSE \middle/ \left[\sum_{j,k} MSE_{jk} \middle/ 6 \right]; \quad (j = 1,2,3; \quad k = 2,3,4; \quad j < k) \tag{17}$$

where MSE_{jk} is the mean sum of squares of errors of the LS line fitted to the observations in jth and kth segments. The magnitude of *REAE* is greater than one for CYC and shift patterns, and about one for NOR, STA, SYS and trend patterns.

(viii) Ratio of variance of the observations (SD^2) and average *MSE* of the LS lines fitted to six subsets of $N/2$ data points (*RVAE*):

$$RVAE = SD^2 \middle/ \left[\sum_{j,k} MSE_{jk} \middle/ 6 \right]; \quad (j = 1,2,3; \quad k = 2,3,4; \quad j < k) \tag{18}$$

The magnitude of *RVAE* is about one for NOR, STA and SYS patterns, and greater than one for CYC, shift and trend patterns.

(ix) Absolute average slope of the LS lines fitted to four subsets of observations (*ADABL*):

$$ADABL = \left| \sum_{j=1}^{4} db_j \middle/ 4 \right| \tag{19}$$

where db_1, db_2, db_3 and db_4 are the slopes of the LS lines fitted to the observations in segments 1 and 2, segments 1, 2 and 3, segments 3 and 4, and segments 2, 3 and 4, respectively. The magnitudes of *ADABL* for different patterns will be similar to the feature *AASL* or *AABL*.

(x) Sign of average slope of the LS lines fitted to four subsets of observations (*SDABL*):

SDABL is again a categorical variable, which can take a value of '0' if the value of the expression, $\sum_{j=1}^{4} db_j/4$ is negative and '1' otherwise. Its usefulness is similar to *SB*, *SASL* and *SABL*.

(xi) Range of slopes of the LS lines fitted to four subsets of observations (*DBRANGE*):

$$DBRANGE = \text{Maximum}(db_j) - \text{Minimum}(db_j); \quad (j = 1,2,\ldots,4) \tag{20}$$

The magnitudes of *DBRANGE* for different patterns will vary in the same way as *SRANGE* or *BRANGE*.

2.3. Features extractable based on criterion-based segmentation of the control chart plot

In case of a shift pattern, the total number of observations can be divided into two segments (before and after the occurrence of a shift) and two LS lines (each approximately horizontal to the *X*-axis) can be fitted well in these segments. However, the time point of occurrence of the shift cannot be known exactly. A criterion-based segmentation into two segments, where the segment sizes may vary in order to satisfy the desired criterion, can take care of this problem. Considering the criterion as the minimization of the pooled mean sum of squares of errors (*PMSE*), the following six features are extracted.

(i) Absolute average slope of the LS lines fitted to two segments (*AABPE*):

$$AABPE = \left| \sum_{j=1}^{2} B_j \right| \middle/ 2 \tag{21}$$

where B_j is the slope of the LS line fitted to jth segment. The value of *AABPE* for trend patterns is higher than shift patterns. On the other hand, among the four patterns that are built surrounding the center line, *AABPE* is the highest for CYC pattern and least for STA pattern.

(ii) Absolute slope difference between the LS line representing the overall pattern and the line segments representing the patterns within the two segments (*ABDPE*):

$$ABDPE = \left| B - \left(\sum_{j=1}^{2} B_j \middle/ 2 \right) \right| \tag{22}$$

The magnitude of *ABDPE* will be higher for shift patterns than trend patterns. On the other hand, the value of *ABDPE* will be higher for CYC and SYS patterns than NOR and STA patterns.

(iii) Sum of absolute slope difference between the LS line representing the overall pattern and the individual line segment (*SASDPE*):

$$SASDPE = \sum_{j=1}^{2} |B - B_j| \tag{23}$$

The relative values of *SASDPE* with respect to different patterns are similar to *ABDPE*.

(iv) Sum of absolute slopes of the two line segments (*SASPE*):

$$SASPE = \sum_{j=1}^{2} |B_j| \tag{24}$$

The relative values of *SASPE* with respect to different patterns are same as *AABPE*.

(v) Ratio of *MSE* of the LS line representing the overall pattern and *PMSE* of the LS lines fitted to two segments (*REPEPE*):

$$REPEPE = MSE/PMSE \tag{25}$$

The value of *REPEPE* is higher for shift and CYC patterns, and lesser for all other patterns.

(vi) Ratio of variance of observations SD^2 and *PMSE* of the LS lines fitted to two segments (*RVPEPE*):

$$RVPEPE = SD^2/PMSE \tag{26}$$

The relative values of *RVPEPE* with respect to different patterns are similar to *RVAE*.

It may be noted that Gauri and Chakraborty (2006b) have expressed the features *ACLPI*, *ACLLSC*, *ALSPI* and *ALSLSC* in terms of standard deviation (*SD*) of the data points in the observation window. Here, those features are expressed in units of variance (SD^2) of the data points to ensure that their values are unaffected by the scale of measurement used.

2.4. Feature extraction

As highlighted by Gauri and Chakraborty (2006a, 2006b), scaling of pattern data leads to loss of distinction between NOR and STA patterns and therefore, they have extracted shape features from the plot of actual process data assuming $c = 1$. The problem in this approach is that a developed CCP recognizer can be applicable to the process from which the sample patterns are generated since, the magnitudes of some of the features are dependent on the values of process standard deviation. This limitation can be overcome, if the features can be extracted from standardized process data. The process data can be standardized using the linear transformation: $z_t = (y_t - \mu)/\sigma$, where y_t and z_t are the observed and standardized values at tth time point, respectively, and μ and σ are the process mean and standard deviation, respectively. However, standardization of the data requires that the two process parameters are known.

Studying various features, it is observed that there are two types of features in the list. These are (a) features whose values are completely independent of process mean and standard deviation, and (b) features whose values are completely independent of process mean, but affected by standard deviation. It is further observed that the features, whose values are dependent on the process standard deviation, are also dependent on the value of the linear distance (c) used to represent a sampling interval in the control chart plot. Interestingly, it is found that the values of all these features become independent of the process standard deviation if their values are extracted assuming that each sampling interval in the control chart plot is represented by a constant linear distance, $c = 1\sigma$. The mathematical proof of the above observation is illustrated below for the two features, *AB* and *SB*.

The ith observation from a normal process can be modeled as $y_i = \mu + r_i\sigma$, where, μ, σ and r_i are the values of process mean, standard deviation and standard normal variate, respectively. Therefore, replacing y_i by $\mu + r_i\sigma$ and t_i by ic in Eq. (1), the value of *AB* is obtained as follows:

$$\begin{aligned} AB &= \left| \frac{\sum_{i=1}^{N}(\mu + r_i\sigma)(ic - \bar{i}c)}{\sum_{i=1}^{N}(ic - \bar{i}c)^2} \right| \quad (\text{since, } \bar{t} = \bar{i}c) \\ &= \left| \frac{\left[c\mu\sum_{i=1}^{N}(i-\bar{i}) + c\sigma\sum_{i=1}^{N} r_i(i-\bar{i})\right]}{\left[c^2\sum_{i=1}^{N}(i-\bar{i})^2\right]} \right| = \left| \frac{\left[c\sigma\sum_{i=1}^{N} r_i(i-\bar{i})\right]}{\left[c^2\sum_{i=1}^{N}(i-\bar{i})^2\right]} \right| \end{aligned} \tag{27}$$

Thus, for a given number of observations (N), the value of *AB* is actually given by a function of r_i, σ and c. However, if $c = 1\sigma$, the value of *AB*, as given by Eq. (27), becomes a function of r_i only, implying that the magnitude of *AB* will be independent of the process mean and standard deviation. Similarly, it can also be proved that the values of the features *ADIST*, *AASBP*, *ACLPI*, *ACLMLC*, *ALSPI*, *ALSLSC*, *AASL*, *SRANGE*, *AABL*, *BRANGE*, *ADABL*, *DBRANGE*, *AABPE*, *ABDPE*, *SASDPE* and *SASPE* will become independent of process mean and standard deviation if $c = 1\sigma$.

As mentioned in the previous section, the feature *SB* can be viewed as a categorical variable, which is '0' if the value of the expression, $\sum_{i=1}^{N} y_i(t_i - \bar{t})/\sum_{i=1}^{N}(t_i - \bar{t})^2$ is negative and '1' otherwise. The value of *SB* will be zero, if

$$\begin{aligned} &\sum_{i=1}^{N} y_i(t_i - \bar{t})0 \quad \text{or} \quad \sum_{i=1}^{N}(\mu + r_i\sigma)(ic - \bar{i}c) < 0 \\ &\text{or} \quad c\mu\sum_{i=1}^{N}(i-\bar{i}) + c\sigma\sum_{i=1}^{N} r_i(i-\bar{i}) < 0 \quad \text{or} \quad c\sigma\sum_{i=1}^{N} r_i(i-\bar{i}) < 0 \\ &\text{or} \quad \sum_{i=1}^{N} r_i(i-\bar{i}) < 0 \end{aligned} \tag{28}$$

So, the condition for negative (or similarly positive) value of the slope does not depend on the values of μ, σ and c. This implies that the value of *SB* is independent of the process mean and standard deviation without requiring any condition on c, i.e. *SB* is a robust feature. Similarly, it can be proved that the features *SASL*, *SABL*, *SDABL*, *RVE*, *RACLALS*, *PMLC*, *PLSC*, *PSMLSC*, *REAE*, *RVAE*, *REPEPE* and *RVPEPE* are also robust features.

All unnatural patterns are created due to addition of some extraneous systematic variations into a normal process. All these extraneous variations can be expressed in terms of process standard deviation and consequently, the equation for modeling the observations from an unnatural pattern can be expressed in the form of the equation as used for modeling the normal pattern, i.e. $y_i = \mu + \text{a coefficient} \times \sigma$. For example, the equation for modeling a DT pattern is $y_i = \mu + r_i\sigma - ig$, where $g = w\sigma$ is the gradient and w is a constant. The same equation can be expressed as $y_i = \mu + (r_i - i \times w)\sigma$. The dependence properties of a feature with respect to process mean and standard deviation for all the unnatural patterns will, therefore, be similar to normal pattern.

Thus, the values of all the 30 features will become generalized, i.e. independent of the process mean and standard deviation, if $c = 1\sigma$. It is important to note that in this approach, the generalization of features can be achieved by knowing or estimating only the process standard deviation value, whereas standardization of process data requires two process parameters, i.e. process mean and standard deviation. On the other hand, in this approach, since the features are extracted from actual process data, distinction between NOR and STA patterns is preserved. Therefore, all the features are extracted here assuming that a sampling interval in the control chart plot is represented by a liner distance equal to one process standard deviation. This can facilitate developing a general purpose feature-based CCP recognizer that can recognize all the eight basic CCPs including STA pattern.

3. Generation of sample patterns

Sample patterns are required for assessing the relative importance of various features in discriminating different patterns as well as developing/validating a CCP recognizer. Gauri and Chakraborty (2006a) have considered an observation window consisting of 32 data points and accordingly, they have simulated various control chart patterns for learning and test samples. In this study, the sample patterns are generated using the same pattern equations and values of different pattern parameters, as used by Gauri and Chakraborty (2006a). The parameters along with the equations used for simulating the CCPs are given in Table 1. The values of different parameters for unnatural patterns are randomly varied in a uniform manner between the limits shown. It may be noted that in the current approach for pattern generation, it is assumed that all the sample patterns in an observation window are complete. A set of 2400 (300 × 8) sample patterns containing the eight CCPs is generated from 300 time series of standard normal data and used for the evaluating relative importance of the features and subsequent selection of a set of the most useful features for pattern discrimination. Then, twenty different sets of sample patterns of size 2400 (300 × 8) are simulated for experimentation.

4. Procedure for feature selection

The basic characteristics of various CCPs are logically analyzed first. Then, the eight CCPs are grouped into different pairwise pattern classes in a hierarchical structure, based on similarities and dissimilarities of properties for different patterns, as shown in Fig. 3. It is planned to identify first the most important features for discriminating two pattern classes within each pairwise hierarchical pattern group, obtain the list of most useful features and then, select an initial set of features from them. A judicious selection of features, in this process, can facilitate a good starting point for selection of useful features and prevent inclusion of highly correlated features.

The logic behind the defined hierarchical structure, as shown in Fig. 3, can be explained as follows. The eight most commonly observed CCPs can be grouped into two broad classes, e.g. UT, DT, US and DS as a single class of patterns where the patterns move away from the center line, and NOR, STA, SYS and CYC are another class where the patterns are built surrounding the center line. UT, DT, US and DS can again be grouped into two sub-classes, i.e. UT and DT (where the departure of observations from the mean value occurs gradually and continuously) vs. US and DS (where the departure occurs suddenly and then the observations hang around the departed value). Increasing trend and decreasing trend patterns can, then, be segregated into two individual classes, i.e. UT (where the departure of observations occurs in the higher side) vs. DT (where the departure of observations occurs in the lower side). Similarly, the other part of the hierarchical structure in Fig. 3 is drawn using the similarities and dissimilarities of characteristics for NOR, STA, SYS and CYC patterns. Thus, the eight most common CCPs can be arranged into the following seven pairwise hierarchical groups of pattern classes,

(a) NOR, STA, SYS and CYC patterns vs. UT, DT, US and DS patterns
(b) NOR and STA patterns vs. SYS and CYC patterns
(c) US and DS patterns vs. UT and DT patterns
(d) NOR pattern vs. STA pattern
(e) SYS pattern vs. CYC pattern
(f) US pattern vs. DS pattern
(g) UT pattern vs. DT pattern

In the process of automatic selection of features using CART analysis, the relative importance of various features in discriminating different pattern classes is computed (Breiman et al., 1984). Analysis of the values of various features extracted from the learning samples of each hierarchical pattern group using CART analysis can reveal the relative importance of different features in discriminating two pattern classes within these pattern groups. Then, the initial set of useful features can be selected easily. However, the selected initial set of features may not be adequate for accurately

Table 1
Parameters for simulating control chart patterns.

Control chart patterns	Pattern parameters	Parameter values	Pattern equations
NOR	Mean (μ)	80	$y_i = \mu + r_i\sigma$
	Standard deviation (σ)	5	
STA	Random noise (σ')	$0.2(\sigma)$ to $0.4(\sigma)$	$y_i = \mu + r_i\sigma'$
SYS	Systematic departure (d)	$1(\sigma)$ to $3(\sigma)$	$y_i = \mu + r_i\sigma + d \times (-1)^i$
CYC	Amplitude (a)	$1.5(\sigma)$ to $2.5(\sigma)$	$y_i = \mu + r_i\sigma + a\sin(2\pi i/T)$
	Period (T)	8 and 16	
UT	Gradient (g)	$0.05(\sigma)$ to $0.1(\sigma)$	$y_i = \mu + r_i\sigma + ig$
DT	Gradient (g)	$-0.1(\sigma)$ to $-0.05(\sigma)$	$y_i = \mu + r_i\sigma - ig$
US	Shift magnitude (s)	$1.5(\sigma)$ to $2.5(\sigma)$	$y_i = \mu + r_i\sigma + ks$;
	Shift position (P)	9, 17, 25	$k = 1$ if $i \geq P$, else $k = 0$
DS	Shift magnitude (s)	$-2.5(\sigma)$ to $-1.5(\sigma)$	$y_i = \mu + r_i\sigma - ks$;
	Shift position (P)	9, 17, 25	$k = 1$ if $i \geq P$, else $k = 0$

Note: i = discrete time point at which the pattern is sampled ($i = 1,\ldots,32$), r_i = random value of a standard normal variate at ith time point, and y_i = sample value at ith time point.

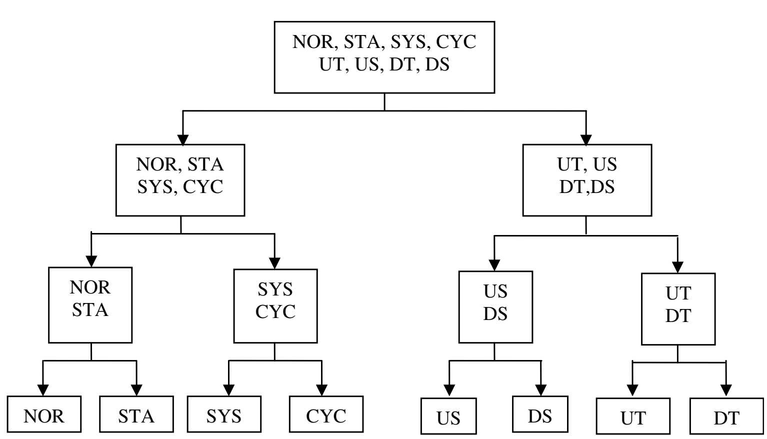

Fig. 3. Pairwise classification of CCPs in a hierarchical structure.

segregating all the eight CCPs. From the practical considerations, a good CCP recognizer should be capable to differentiate all the CCPs with high accuracy using a minimum number of features. Keeping these requirements in mind, the following CART-based four-step systematic approach is adopted for selection of the features.

(a) Take a set of learning samples consisting of all the eight patterns. Then, create learning samples for the seven pairwise pattern groups and determine the relative importance of different features in discriminating pattern classes within these pattern groups using CART analysis. Subsequently, select an initial set of most useful features taking into consideration the relative importance of various features for different pattern groups.
(b) Derive the classification rules by subjecting the extracted values of this initial set of features from the original learning samples to CART analysis and obtain the confusion matrix (a table summarizing the tendency of the recognizer to classify a recognized pattern into a correct class or into any of the other seven possible wrong classes) under the derived classification rules.
(c) Find out the pair of patterns where the confusion is the maximum. Then search out the features that are powerful in discriminating the confusing pair of patterns and among them, choose one that has considerably smaller correlation with the already selected features for inclusion in the selected set of features.
(d) Repeat steps 2 and 3 until no new useful feature is found whose correlation with the other selected features is considerably low.

4.1. Selection of features

A set of learning samples of size 2400 (300×8) is selected. The learning samples for the pairwise pattern group (a) are created by re-coding NOR, STA, SYS and CYC patterns as ‘1’, and UT, US, DT and DS patterns as ‘2’. So, the set of learning samples for pattern group (a) consists of 2400 (1200×2) patterns from two pattern classes. The learning samples for pattern group (b) are created by taking out NOR, STA, SYS and CYC patterns from the original set of learning samples and then, re-coding NOR and STA patterns as ‘1’, and SYS and CYC patterns as ‘2’. Thus, the set of learning samples for pattern group (b) consists of 1200 (600×2) patterns from two classes. Similarly, the learning samples for each pairwise pattern group are obtained.

From each of the seven pattern groups, values of various features are extracted and then, subjected to CART analysis separately using the following specifications: misclassification cost to be equal for all the patterns, measure of goodness of fit for split as Gini measure, $n = 2$ for ‘Minimum n’ rule, pruning method as minimal cost-complexity cross-validation (CV) pruning and $\delta = 0.5$ for ‘δStandard Error’ rule. The relative importance (RI) of various features for segregating pattern classes within the seven pairwise pattern groups are obtained in terms of bar diagrams. In these figures, the most important feature has a measure of 100 and others have measures in the range of 0 to 100. The bar diagrams indicating RI of various features for pattern groups (a) to (f) are shown in Fig. 4(i)–(vi). The RI of different features for pattern group (g) is similar to pattern group (f) and therefore, is not shown in this figure.

Examination of the bar diagrams in Fig. 4(i)–(vi) reveals that the most important features for discriminating the pattern classes can be uniquely identified for the pattern groups (a), (b) and (c), and these are *AASL*, *ACLPI* and *SRANGE*, respectively. So these three features are selected straightway. Four features (categorical variables) *SB*, *SASL*, *SABL* and *SDABL* are equally useful for distinguishing pattern classes within pattern groups (f) and (g), and so any one of these features may be chosen for segregation of US pattern from DS pattern or UT pattern from DT pattern. Here, *SB* is chosen arbitrarily. On the other hand, there are more than one features that are equally powerful in discriminating pattern classes within pattern groups (d) and (e). The features *AASBP*, *ADIST*, *ACLPI* and *ALSPI* are equally important for differentiating pattern classes within pattern group (d) and among these, *ACLPI* has already been selected. So no more features are selected in consideration to discrimination of pattern classes within pattern group (d). Fig. 4(v) reveals that the pattern classes within pattern group (e) can be differentiated equally by any of the features *ADIST*, *PMLC*, *PLSC*, *PSMLSC*, *ACLMLC* and *ALSLSC*. The degree of association of these features with the already selected four features, i.e. *AASL*, *ACLPI*, *SRANGE* and *SB* are examined by estimating the pairwise correlation coefficient values between the features. The computed correlation coefficient values reveal that except for the features *PSMLSC*, *PLSC* and *PMLC*, all other features have very high correlation with at least one of the already selected features. With respect to discrimination of patterns within pattern class (e), therefore, *PSMLSC*, *PLSC* or *PMLC* only may be selected as an additional useful feature. Here, *PSMLSC* is selected since its correlation properties are found to be most desirable.

The initial set of features, therefore, consists of *AASL*, *ACLPI*, *SRANGE*, *SB* and *PSMLSC*. The extracted values of these features from the original learning samples containing eight CCPs are, then, subjected to CART analysis keeping all the specifications same as mentioned earlier. This results in a set of heuristic rules in the form of classification tree, which is, then, used to recognize various CCPs in the learning samples and the corresponding confusion matrix is obtained, as shown in Table 2.

Examination of the confusion matrix reveals that confusing shift patterns as the trend patterns is the maximum (5.67% for US patterns, and 5.00% for DS patterns). The CYC patterns are also highly confused (3.33%) with the DS patterns. These results suggest that inclusion of an additional feature that can distinguish between shift and trend patterns can be quite useful. It is noted from Fig. 4(iii) that the candidate features for further selection with respect to possible reduction in confusion between shift and trend patterns are *BRANGE*, *DBRANGE*, *REAE* and *REPEPE*. However, *BRANGE* and *DBRANGE* are found to be quite highly correlated with *SRANGE* (already selected). The pairwise correlation coefficients of these two features with *SRANGE* are noted to be more than 0.89. Therefore, *BRANGE* and *DBRANGE* are not considered for inclusion. Among the remaining features, *REAE* is included as an additional predictor variable because it leads to higher reduction in confusion between trend and shift patterns.

On the other hand, with the aim to identify additional useful features for discriminating CYC and DS patterns, if any, the learning samples consisting of CYC and DS patterns are segregated. Then, all the features extracted from these samples are subjected to CART analysis and the relative importance of various features in differentiating CYC and DS patterns are obtained. It indicates that *ABDPE* is the most powerful feature for discriminating CYC and DS patterns. The computed pairwise correlation coefficient values with each of the already selected six features reveal that its associations with those features are considerably low. So, *ABDPE* is also selected as a useful feature. Finally, seven features are selected as the most important features for recognition of various CCPs and these are *AASL*, *ACLPI*, *SRANGE*, *SB*, *PSMLSC*, *REAE* and *ABDPE*. The pairwise correlation coefficient values between these seven features, as shown in Table 3, indicate that the associations among them are fairly low. This implies that a CCP recognizer developed based on these features may result in quite good prediction stability.

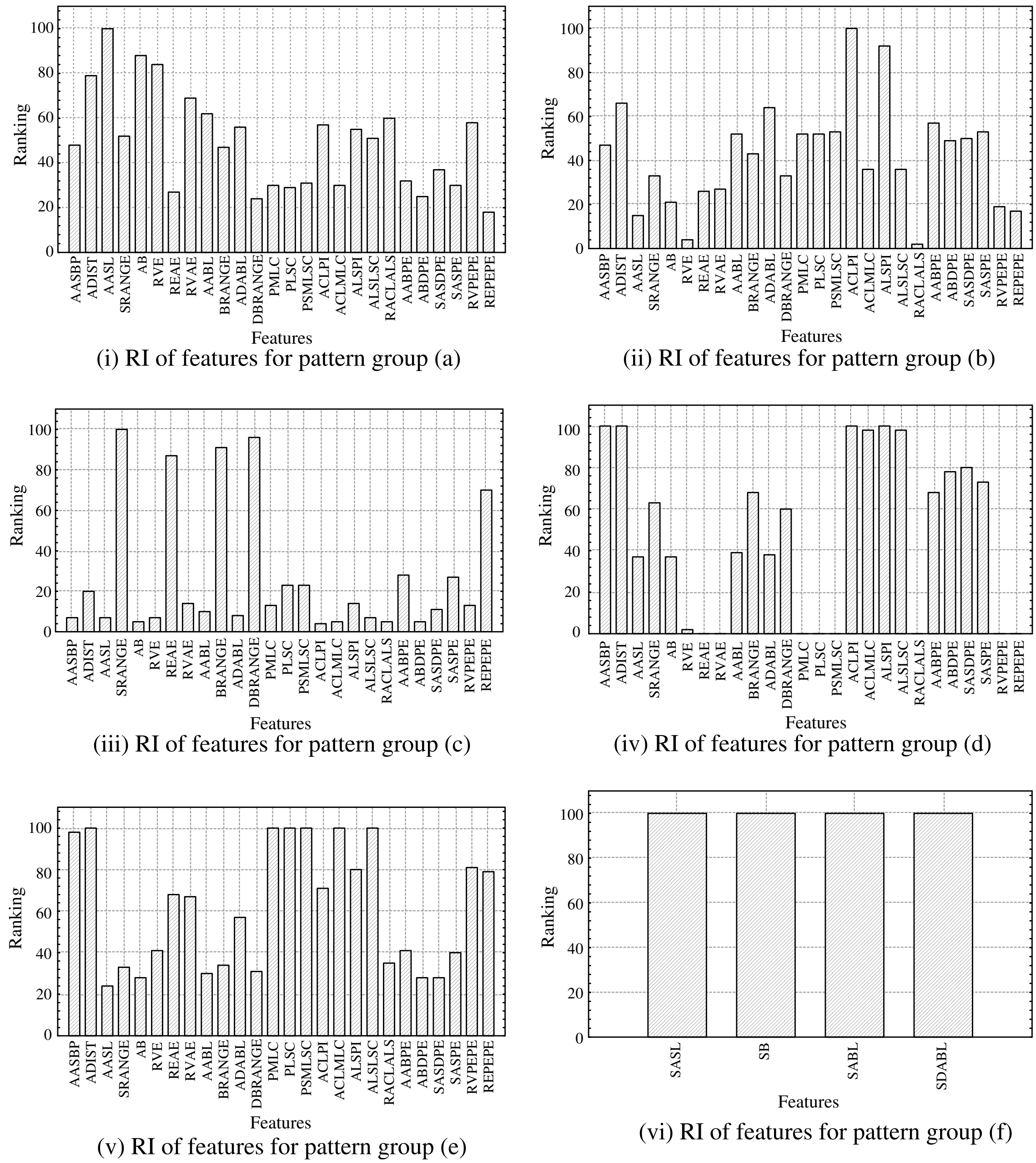

Fig. 4. Relative importance of various features for different pairwise hierarchical pattern groups.

5. Pattern recognizer design

Using extracted features from the control chart plots, Pham and Wani (1997) and Gauri and Chakraborty (2006a) have designed two types of CCP recognizers, e.g. heuristic rule-based and ANN-based recognizers. While Pham and Wani (1997) have developed the heuristic rules by manually examining the values of the extracted features, Gauri and Chakraborty (2006a) have derived the heuristic rules for CCP recognition using CART analysis. The advantage of using CART algorithm is that it automatically selects the 'right-sized' tree that has the optimal predictive accuracy. The procedure for the 'right-sized' tree selection is not foolproof, but at least, it takes the subjective judgment out of the process of choosing the 'right-sized' tree and thus avoids 'over fitting' and 'under fitting' of the data (Breiman et al., 1984). The CART-based approach is adopted here for deriving the feature-based heuristic rules for CCP recognition.

On the other hand, multilayer perceptrons (MLP) neural network is simple and ideally suited for pattern recognition tasks (Haykin, 1999), and therefore, an MLP is used here as pattern recognizer. Its basic structure comprises an input layer, one or more hidden layer(s) and an output layer. The number of nodes in the input layer is set according to the actual number of features used, i.e. seven. The number of output nodes is also set corresponding to the number of pattern classes, i.e. eight. This neural network has only one hidden layer and the number of nodes in the hidden layer is set to 12 (chosen empirically). Thus the ANN structure is $7 \times 12 \times 8$. Since it is planned to use the supervised training approach, each pattern presentation is tagged with its respective label. The target values for the recognizer's output nodes are represented by a vector of eight elements, e.g. the desired output vector for NOR and STA patterns are [0.9, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1] and [0.1, 0.9, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1], respectively. The maximum value (0.9) identifies the corresponding node that is expected to secure the

Table 2
Confusion (%) in pattern recognition under the rules with the initial set of features.

True pattern class	Identified pattern class							
	NOR	STA	SYS	CYC	UT	US	DT	DS
NOR	98.00	0.00	0.00	1.34	0.33	0.33	0.00	0.00
STA	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00
SYS	0.33	0.00	99.67	0.00	0.00	0.00	0.00	0.00
CYC	0.67	0.00	0.67	95.00	0.00	0.33	0.00	3.33
UT	1.00	0.00	0.00	0.33	98.34	0.33	0.00	0.00
US	0.00	0.00	0.00	1.33	5.67	93.00	0.00	0.00
DT	1.00	0.00	0.00	0.00	0.00	0.00	98.67	0.33
DS	0.33	0.00	0.67	2.00	0.00	0.00	5.00	92.00

Table 3
Pairwise correlation coefficients between the selected features.

Selected features	*SB*	*ACLPI*	*PSMLSC*	*AASL*	*SRANGE*	*REAE*	*ABDPE*
SB	1.00	−0.01	0.26	−0.03	−0.17	−0.17	−0.19
ACLPI	−0.01	1.00	−0.01	−0.46	−0.32	−0.16	−0.46
PSMLSC	0.26	−0.01	1.00	−0.33	−0.45	−0.49	−0.17
AASL	−0.03	−0.46	−0.33	1.00	0.47	0.41	0.08
SRANGE	−0.17	−0.32	−0.45	0.47	1.00	0.61	0.38
REAE	−0.17	−0.16	−0.49	0.41	0.61	1.00	0.26
ABDPE	−0.19	−0.46	−0.17	0.08	0.38	0.26	1.00

highest output for a pattern considered to be correctly classified. Gradient descent with momentum and adaptive learning rate (*traingdx*) is used as back propagation training algorithm, since it provides reasonably good performance and more consistent results (Gauri & Chakraborty, 2006a; Hassan et al., 2003). The network performance is measured using the mean squared error value.

5.1. Experimentation

Ten different sets of samples of size 2400 (300×8) each are used as learning/training samples for developing the CCP recognizers and another ten sets are utilized as test samples. Values of the selected seven features are extracted from each set of the training samples and subjected to CART analysis using the same specifications as used for the analysis during feature selection. This results in 10 different classification trees. In this process, 10 different heuristic-based CCP recognizers are developed. These recognizers are labeled as 1.1–1.10 in Table 4.

On the other hand, a trained ANN can only accept a certain range of input data. The extracted shape features, therefore, are scaled such that their values fall within [−1, +1] interval before their presentation to the ANN for the learning process. The neural network is trained 10 times by exposing it separately to the extracted features from the same 10 sets of training samples with the following training parameters:

- Maximum number of epochs = 2500
- Error goal = 0.008
- Learning rate = 0.1
- Momentum constant = 0.5
- Ratio to increase learning rate = 1.05
- Ratio to decrease learning rate = 0.7

The training is stopped whenever either the error goal has been achieved or the maximum allowable number of training epochs has been met. In this process, 10 different ANN-based recognizers are developed. All these ANN recognizers have the same architecture and differ only in the training data sets used. These recognizers are labeled as 2.1–2.10 in Table 5. The corresponding heuristic and ANN-based recognizers, e.g. 1.1 and 2.1 are developed from the same set of training samples.

In addition, with the aim to facilitate comparison of the recognition performance of the currently developed heuristic and ANN recognizers with the recognizers that can be developed using Gauri and Chakraborty (2006a) proposed eight features, 10 different heuristic and 10 different ANN recognizers are developed from the same 10 sets of training samples but using extracted values of the Gauri and Chakraborty (2006a) proposed eight features.

The recognition performance of all the heuristic and ANN-based recognizers are, finally, evaluated using ten sets of test samples. All the procedures for training and verification are coded in MATLAB using its ANN toolbox (Demuth & Beale, 1998).

6. Results and discussion

Training and verification performance of the 10 feature-based heuristic and ANN recognizers are shown in Tables 4 and 5, respectively. The overall mean percentage of correct recognition for the heuristic and ANN recognizers at the training phase are 96.93% and 96.95%, respectively. On the other hand, the overall mean percentage of correct recognition for the heuristic and ANN recognizers at the verification phase are 95.46% and 96.66%, respectively. Paired t-test (Walpole, Myers, & Myers, 1998) is conducted for 10 pairs of heuristic and ANN recognizers for their performance in terms of percentage of correct classification. The computed t-statistic is found to be 5.20, which is greater than 2.82, the critical value of t_9 ($\alpha = 0.01$). This implies that the difference in recognition accuracy between these two types of recognizers is statistically significant. This confirms that the feature-based ANN recognizers

Table 4
Training and verification (recall) performance of heuristic-based recognizers.

Recognizer number	Training phase		Verification phase	
	Number of splits in the tree	Percentage of correct classification	Percentage of correct classification	
			Mean	Standard deviation
1.1	21	97.50	95.79	1.26
1.2	19	96.83	95.54	1.06
1.3	21	96.71	96.13	1.86
1.4	20	97.33	95.68	1.71
1.5	22	95.88	95.67	1.36
1.6	21	96.88	95.21	1.33
1.7	23	97.33	95.63	1.42
1.8	20	96.79	94.83	1.18
1.9	24	97.23	95.75	1.63
1.10	20	96.83	94.33	1.52
Overall mean	21.1	96.93	95.46	–

Table 5
Training and verification (recall) performance of ANN-based recognizers.

Recognizer number	Training phase		Verification phase	
	Number of epochs	Percentage of correct classification	Percentage of correct classification	
			Mean	Standard deviation
2.1	1155	96.73	96.53	0.52
2.2	988	97.67	97.22	0.42
2.3	1206	97.61	97.13	0.81
2.4	1166	96.52	96.36	0.63
2.5	978	96.78	96.38	0.58
2.6	1300	96.33	96.02	0.48
2.7	1323	96.82	96.63	0.56
2.8	1008	97.5	97.22	0.51
2.9	812	96.25	96.08	0.63
2.10	1248	97.25	97.05	0.65
Overall mean	1118.4	96.95	96.66	–

give better recognition accuracy compared to heuristic recognizers. This result is in conformity with Gauri and Chakraborty (2006a).

On the other hand, the standard deviation (*sd*) of correct recognition percentage during the recall phase is observed to be consistently higher for heuristic recognizers than ANN recognizers. The statistical significance for the difference is carried out using *F*-test. The *F*-statistics $sd^2_{\text{Heuristic}}/sd^2_{\text{ANN}}$ for all the ten pairs of heuristic and ANN recognizers are computed and the minimum value is found to be 5.41, whereas the critical value of $F_{9,9}$ ($\alpha = 0.01$) is 5.35. These results confirm that the standard deviation of correct recognition percentage is always higher for heuristic recognizers than ANN recognizers. This implies that ANN recognizers have better predictive consistency than heuristic recognizers. The best heuristic recognizer in terms of consistency of recognition performance is recognizer no. 1.2 and its heuristic rules in the form of classification tree are shown in Fig. 5. On the other hand, ANN recognizer 2.2 is found to be the best with respect to both accuracy and consistency in recognition performance.

Table 6 shows comparison of the recognition accuracy achieved during verification phase by the heuristic and ANN recognizers developed using the currently selected seven features and Gauri and Chakraborty (2006a) proposed eight features. The differences in recognition percentages for the heuristic as well as ANN recognizers are tested using paired *t*-test. The computed values of *t*-statistic are found to be 7.75 and 3.98 for the heuristic and ANN recognizers, respectively, both of which are greater than 2.82, the critical value of t_9 ($\alpha = 0.01$). This implies that the respective differences in recognition accuracies are statistically significant. So it can be concluded that both the heuristic and ANN recognizers developed using currently selected seven features result in higher recognition accuracy than the heuristic and ANN recognizers, respectively, developed using Gauri and Chakraborty (2006a) proposed eight features. In spite of selection of lesser number of features in the current approach, on an average 0.54% and 0.57% higher accuracies are achieved for heuristic and ANN recognizers, respectively.

Table 6
Comparison of recognition accuracies.

Training set	Recognition accuracy (%) of			
	Heuristic recognizers developed using		ANN recognizers developed using	
	Currently selected seven features	Gauri and Chakraborty (2006a) proposed eight features	Currently selected seven features	Gauri and Chakraborty (2006a) proposed eight features
1	95.79	95.08	96.53	95.88
2	95.54	95.17	97.22	96.38
3	96.13	95.63	97.13	95.75
4	95.68	95.17	96.36	96.04
5	95.67	95.46	96.38	95.83
6	95.21	94.75	96.02	95.67
7	95.63	95.17	96.63	96.21
8	94.83	93.79	97.22	96.08
9	95.75	95.21	96.08	96.29
10	94.33	93.75	97.05	96.75
Mean	95.46	94.92	96.66	96.09

6.1. Sensitivity studies for pattern recognition

In this study, all the features based on which the CCP recognizers are developed, are extracted assuming that a sampling interval in the control chart plot is represented by a constant distance,

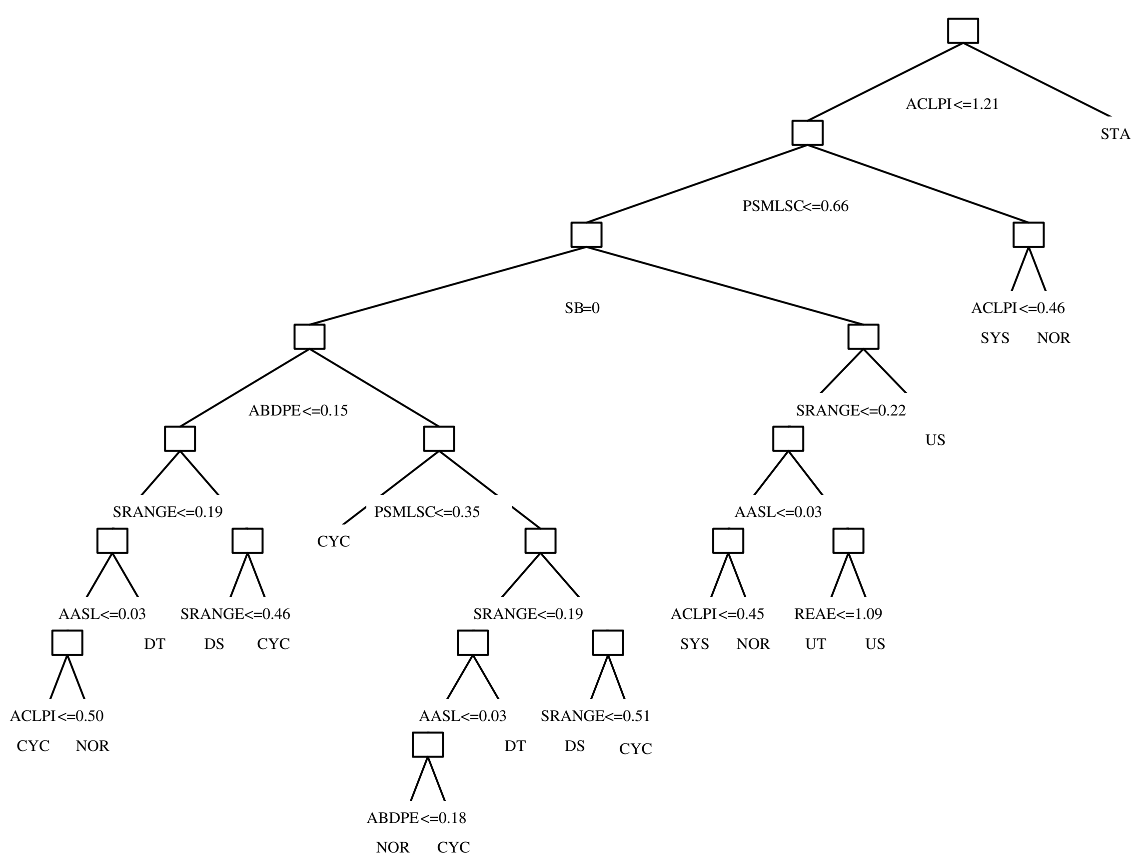

Fig. 5. Classification tree for recognition of control chart patterns.

Table 7
Recognition accuracy (%) for additional test patterns.

Type of recognizer	Additional sets of test patterns					Range
	1	2	3	4	5	
Heuristic recognizer	96.17	97.58	98.21	96.92	95.46	2.75
ANN recognizer	97.21	98.17	98.54	98.13	96.88	1.66

$c = 1\sigma$, where $\sigma = 5$ (known). In reality, the value of σ is to be estimated. Therefore, it is planned to study the sensitivities of both the heuristic and ANN recognizers with respect to the estimation error of σ.

To study the sensitivity of pattern recognizers, it is necessary to generate additional test patterns, considering the values of σ different from its known value and then subjecting them to classification using the developed recognizers. All these test pattern sets should be simulated using the same series of standard normal data to ensure that the sequences of randomness are similar in all these test sets; otherwise, the effect of changes in the sequences of randomness will be confounded with the effect of departure of σ from its known value. The set of 300 time series of standard normal data, from which the sample patterns leading to the heuristics shown in Fig. 5 are developed, is taken for generation of additional test patterns. On the other hand, care is also taken so that the effect of random changes in the values of different pattern parameters is not confounded with the effect of deviation of σ from its known value. For the purpose of additional test patterns generation, therefore, the values of various pattern parameters are fixed as: $\mu = 80$, $\sigma' = 0.3\sigma$, $d = 2\sigma$, $a = 2\sigma$, $T = 16$, $g = \pm 0.075\sigma$, $s = \pm 2\sigma$ and $P = 16$. These values are mostly the midpoints of the ranges of respective pattern parameters, as shown in Table 1.

Using the above-mentioned time series of standard normal data and pattern parameters, five additional test pattern sets with size 2400 (300×8) each are generated considering the values of σ as 4.50, 4.75, 5.00, 5.25 and 5.50 (implying ±10% deviation of σ from its known value) and they are labeled as 1–5. These additional test patterns are then subjected to classification using the heuristic and ANN recognizers (recognizer no. 1.2 and 2.2, respectively). The corresponding recognition accuracies are shown in Table 7. It is noted that for both the recognizers, the recognition performance is the maximum for the additional pattern set number 3, which is generated assuming $\sigma = 5$. More and more is the deviation in the value of σ, poorer becomes the recognition performance. However, the recognition performance remains reasonably well within the ±10% deviation of σ for both the recognizers. The range of variation in recognition performance for heuristic and ANN recognizers is 2.75% and 1.66%, respectively. This implies that the performance of ANN recognizers is relatively less sensitive to estimation error of process standard deviation than heuristic recognizers.

7. Conclusions

Selection of an optimal set of features is an important issue for exploiting the full potentiality of feature-based systems in recognition of various CCPs. In this paper, a set of seven most useful features is selected from a large number of potentially useful features using a CART-based systematic approach. Based on these selected features, eight most commonly observed CCPs are recognized using heuristic and ANN techniques. Extensive evaluation of the two types of recognizers reveals that ANN-based recognizers can achieve better recognition performance as well as consistency than heuristic-based recognizers. It is further observed that both the recognizers result in better recognition performance than the previous feature-based recognizers, in spite of using lesser number of features. In this study, various features are extracted in such a way that their values become independent of the process mean and standard deviation. Thus, these feature-based CCP recognizers can be applicable to any general process. For this generalization, the process standard deviation only needs to be known or estimated. One limitation of the proposed approach for CCP recognition is that in this approach, all the CCPs are assumed to be complete in the recognition window.

Future research works may make improvements in two directions. Firstly, using real time data will provide an ultimate test for verifying their universal applicability in SPC. Secondly, each feature is essentially a function of sample observations, i.e. statistic. Studying the statistical properties of various features may provide important guidelines to develop better SPC pattern recognizers.

References

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks.

Cheng, C. S. (1997). A neural network approach for the analysis of control chart patterns. *International Journal of Production Research, 35*(3), 667–697.

Demuth, H., & Beale, M. (1998). *Neural network toolbox user's guide*. Natick, MA: Math Works.

Evans, J. R., & Lindsay, W. M. (1988). A framework for expert system development in statistical quality control. *Computers and Industrial Engineering, 14*(3), 335–343.

Gauri, S. K., & Chakraborty, S. (2006a). Feature-based recognition of control chart patterns. *Computers and Industrial Engineering, 51*, 726–742.

Gauri, S. K., & Chakraborty, S. (2006b). A study on the various features for effective control chart pattern recognition. *International Journal of Advanced Manufacturing Technology*. 10.1007/s00170-006-0591-6.

Guh, R. S., & Shiue, Y. R. (2005). On-line identification of control chart patterns using self-organizing approaches. *International Journal of Production Research, 43*(6), 1225–1254.

Guh, R. S., & Tannock, J. D. T. (1999). Recognition of control chart concurrent patterns using a neural network approach. *International Journal of Production Research, 37*(8), 1743–1765.

Guh, R. S., Zorriassatine, F., Tannock, J. D. T., & O'Brien, C. (1999). On-line control chart pattern detection and discrimination – a neural network approach. *Artificial Intelligence in Engineering, 13*(4), 413–425.

Hassan, A., Nabi Baksh, M. S., Shaharoun, A. M., & Jamaluddin, H. (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research, 41*(7), 1587–1603.

Haykin, S. (1999). *Neural network: A comprehensive foundation*. Englewood Cliffs, NJ: Prentice-Hall.

Hwarng, H. B., & Hubele, N. F. (1993). Back-propagation pattern recognizers for $\bar{X}$ control charts: Methodology and performance. *Computers and Industrial Engineering, 24*(2), 219–235.

Montgomery, D. C. (2001). *Introduction to statistical quality control*. New York: John Wiley and Sons.

Montgomery, D. C., & Peck, E. A. (1982). *Introduction to linear regression analysis*. New York: John Wiley and Sons.

Nelson, L. S. (1984). The Shewhart control chart–test for special causes. *Journal of Quality Technology, 16*(4), 237–239.

Nelson, L. S. (1985). Interpreting Shewhart $\bar{X}$ control chart. *Journal of Quality Technology, 17*(2), 114–117.

Perry, M. B., Spoerre, J. K., & Velasco, T. (2001). Control chart pattern recognition using back propagation artificial neural networks. *International Journal of Production Research, 39*(15), 3399–3418.

Pham, D. T., & Oztemel, E. (1992a). XPC: An on-line expert system for statistical process control. *International Journal of Production Research, 30*(12), 2857–2872.

Pham, D. T., & Oztemel, E. (1992b). Control chart pattern recognition using neural networks. *Journal of System Engineering, 2*(4), 256–262.

Pham, D. T., & Wani, M. A. (1997). Feature-based control chart pattern recognition. *International Journal of Production Research, 35*(7), 1875–1890.

Swift, J. A., & Mize, J. H. (1995). Out-of-control pattern recognition and analysis for quality control charts using LISP-based systems. *Computers and Industrial Engineering, 28*(1), 81–91.

Walpole, R. E., Myers, R. H., & Myers, S. L. (1998). *Probability and statistics for engineers and scientists* (6th ed.). New York: Macmillan.

Western Electric. (1958). *Statistical quality control handbook*. Indianapolis: Western Electric Company.