Feature-Based Recognition of Control Chart Patterns: A Generalized Approach

Susanta Kumar Gauri [1] and Shankar Chakraborty [2]

[1] SQC & OR Unit, Indian Statistical Institute, India
[2] Department of Production Engineering, Jadavpur University, India
(*Received* August 2005, *accepted* February 2007)

Abstract: Control chart patterns (CCPs) can be associated with certain assignable causes creating problems in the manufacturing processes and thus, the recognition of CCPs can accelerate the diagnostic search, for those causes. Researches in developing CCP recognition systems have traditionally been carried out using standardized or scaled process data to ensure the generalized applicability of such systems. Whereas standardization of data requires additional efforts for estimation of the mean and standard deviation of the underlying process, scaling of process data leads to loss of distinction between the normal and stratification patterns because a stratification pattern is essentially a normal pattern with unexpected low variability. In this paper, a new approach for generalization of feature-based CCP recognition system is proposed, in which the values of extracted shape features from the control chart plot of actual data become independent of the process mean and standard deviation. Based on a set of six shape features, eight most commonly observed CCPs including stratification pattern are recognized using heuristic and artificial neural network techniques and the relative performance of those approaches is extensively studied using synthetic pattern data.

Keywords: Artificial neural network, control chart patterns, generalization of features, heuristics, pattern recognition.

1. Introduction

Statistical process control (SPC) is one of the most effective approaches in total quality management (TQM) and control charts, primarily in the form of $\bar{X}$ chart, remain among the most important tools of SPC. A process is considered to be out-of-control when any point falls beyond the control limits or the control charts display unnatural (non-random) patterns [13]. While the former condition can easily be identified, the latter is difficult to recognize precisely and effectively because the control chart patterns (CCPs) are usually distorted by random noise that occurs naturally in a manufacturing process. *The Statistical Quality Control Handbook* [22] includes various types of control chart patterns (CCPs). Among these, 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), as shown in Figure 1. All other patterns are either special forms of basic CCPs or mixed forms of two or more basic CCPs. Only the NOR pattern is indicative of a process continuing to operate under controlled condition. All other CCPs are unnatural and associated with impending problems requiring pre-emptive actions. The Shewhart control charts consider only the current sample data point to determine the status of a process. Hence, they do not provide any pattern-related information when the process is out-of-control. Many supplementary rules like zone tests or run rules have been developed to assist the users in detecting the CCPs [15, 16]. However, the application of all

these rules can result in excessive number of false alarms.

Detection of an unnatural pattern allows the users to have more insight about the process in addition to the in-control/out-of-control decision. The classified patterns can help the users to identify the potential problems in the process and provide clues or guidelines for effective trouble-shooting. Hence, CCP recognition is the most important supplement and enhancement to the conventional control charting technique. Recently, a bulk of research initiatives have been directed towards developing computer-based algorithms for automated recognition of various control chart patterns aiming to monitor the manufacturing processes in real time environment. The motivation for those research works is to take advantage of the widespread use of automated data acquisition systems for computer charting and analysis of the manufacturing process data. The basic approach remains as defining a moving observation or SPC monitoring window consisting of the most recent N data points and then detecting the control chart pattern that appears in the window. The data points are the sample average values when the process mean is monitored using $\bar{X}$ chart.

The approaches adopted by the researchers for developing automated CCP recognition systems range from the application of expert systems [4, 18, 21] to artificial neural networks (ANN) [2, 6, 8, 10, 17, 19, 20]. The advantage of an expert system or rule-based system is that it contains the information explicitly. This explicit nature of the stored information facilitates the provision of explanations to the operators as and when required. However, the use of rules based on statistical properties has the difficulty that similar statistical properties may be derived for some patterns of different classes, which may create problems of incorrect recognition. The use of ANN techniques has the advantage that it does not require the provision of explicit rules or templates. Rather, it learns to recognize patterns directly through typical example/sample patterns during a training phase. One demerit of ANN is that the information it contains, are implicit and virtually inaccessible to the users. This creates difficulties 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.

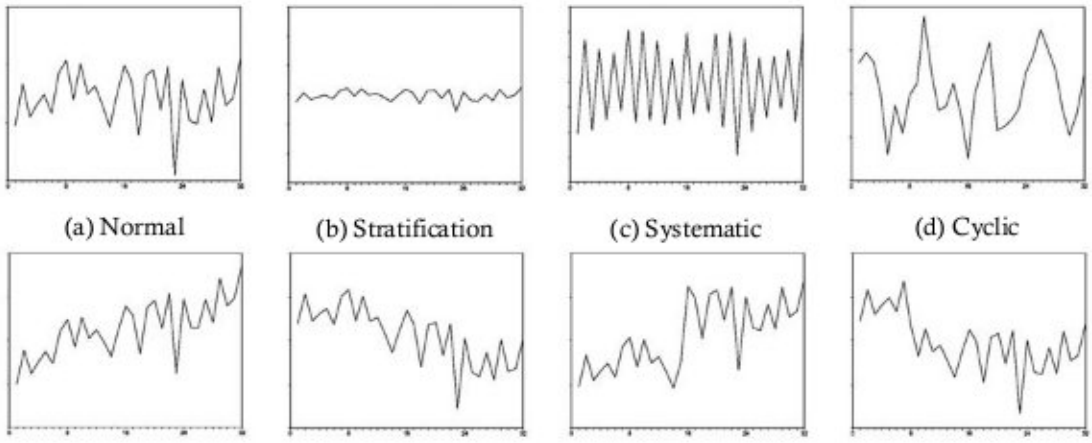
Most of the reported works used raw process data as the input vector for control chart pattern recognition. Only a few researchers [5, 8, 20] have attempted to recognize various CCPs using extracted features from raw process data as the input vector. Whereas Pham and Wani [20] and Gauri and Chakraborty [5] have used extracted shape features from the control chart plot, Hassan *et al.* [8] have utilized extracted statistical features from the input data set. The use of raw process data is computationally inefficient. It requires large neural network size due to which training of the ANN becomes difficult and time consuming. Use of extracted shape features from the control chart plot or statistical features from the input process data can reduce the network size and learning time. However, extraction of statistical features requires considerably large number of observations and their interpretation is also not easily comprehendible to the practitioners. Moreover, the statistical features lose information on the order of the data. The shape features can be extracted from smaller number of observations without losing order of the data. The feature-based approach also provides a greater choice of recognition techniques. Pham and Wani [20] and Gauri and Chakraborty [5] have demonstrated that properly developed heuristics based on the extracted shape features from control chart plot can efficiently differentiate various CCPs. 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. Since extracted shape features represent the main characteristics of the original data in a condensed form, both the feature-based heuristic and ANN approaches can facilitate efficient pattern recognition.

From implementation viewpoint, a developed feature-based CCP recognizer should be applicable to any general process. This objective is usually fulfilled by extracting features from the plot of standardized or scaled process data, instead of extracting features from the plot of actual process data. The process data are standardized using the transformation $z_t = (y_t - \mu)/\sigma$, where y_t and z_t are the observed and standardized values at t^{th} time point respectively, and μ and σ are the process mean and standard deviation respectively. On the other hand, the process data are scaled to (0, 1) or (–1, +1) interval using the linear transformation $z_t = \{(y_t - y_{\min})/(y_{\max} - y_{\min})\}$ or $z_t = \{2(y_t - y_{\min})/(y_{\max} - y_{\min}) - 1\}$ respectively. The standardization of process data assumes that the mean and standard deviation values are known a priori. In practical applications, these values are estimated with additional efforts. A large estimation error for the process mean and standard deviation can also affect the recognition performance adversely. The advantage of scaling of process data is that it requires no additional effort. However, the distinction between normal and stratification patterns is lost when the process data are scaled to (0, 1) or (–1, 1) interval, since a stratification pattern is essentially a normal pattern with unexpected low variability. However, many real life situations demand for a CCP recognition system that will be capable of detecting stratification pattern too and therefore, scaling of the input process data is not often acceptable.

In this paper, a new generalized approach for control chart pattern recognition is proposed, in which the values of the extracted shape features from the plot of actual data are made independent of the process mean and standard deviation. Based on a set of six shape features, eight most commonly observed CCPs including stratification pattern are recognized using tree-structured classification and ANN techniques and their relative performance is extensively studied.

This paper is organized in different sections. The selected set of useful features is presented in section 2. The proposed approach for generalization of the features is discussed in section 3. Section 4 discusses the design of the pattern recognizers followed by the experimental procedure in section 5. Section 6 provides the results and discussions and section 7 concludes the paper.



(a) Normal (b) Stratification (c) Systematic (d) Cyclic

(e) Increasing trend (f) Decreasing trend (g) Upward shift (h) Downward shift

Figure 1. Various control chart patterns.

2. Selected Features

Shape features of control chart patterns can be extracted from different considerations [5] and many of them can be highly correlated. However, a good CCP recognizer should be capable to differentiate different patterns with high accuracy using a minimum number of features and the correlation among those features should be as small as possible. Lower is the association among the features, higher will be the prediction stability [14]. In this paper, therefore, a set of six features, which are having fairly low correlation among themselves, are chosen. These features along with the mathematical expressions for their extractions are described as below:

(i). Ratio between variance of the data points (SD^2) and mean sum of squares of errors of the least square (LS) line representing the overall pattern (*RVE*):

$$RVE = \left[\sum_{i=1}^{N}(y_i-\bar{y})^2 \Big/ (N-1)\right] \Big/ \left[\left\{\sum_{i=1}^{N}(y_i-\bar{y})^2 - \left(\sum_{i=1}^{N} y_i(t_i-\bar{t})\right)^2 \Big/ \sum_{i=1}^{N}(t_i-\bar{t})^2\right\} \Big/ (N-2)\right]. \quad (1)$$

where $t_i = ic\,(i = 1,2,..,N)$ is the distance of i^{th} time point of observation from the origin in the control chart plot, c is the constant linear distance used to represent a sampling interval in the control chart plot, y_i is the observed value of a quality characteristic at i^{th} time point, N is the size of the observation window, $\bar{t} = \sum_{i=1}^{N} t_i / N$ and $\bar{y} = \sum_{i=1}^{N} y_i / N$. The magnitude of *RVE* for normal, stratification, systematic and cyclic patterns is approximately one, while for trend and shift patterns, it is greater than one.

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

The slope (B) of the LS line fitted to the data points in an observation window is given by the following equation:

$$B = \sum_{i=1}^{N} y_i(t_i-\bar{t}) \Big/ \sum_{i=1}^{N}(t_i-\bar{t})^2. \quad (2)$$

It may be noted that the feature B has two characteristics, i.e. (a) its absolute value (*AB*), and (b) its sign (*SB*). Likewise *RVE*, the magnitude of *AB* can differentiate the four patterns that hang around the centerline, i.e. normal, stratification, systematic and cyclic patterns from trend and shift patterns. It has been observed that *RVE* is more powerful than *AB* in discriminating these two groups of patterns. Here, therefore, only the sign characteristic of the slope of the LS line representing the overall pattern has been taken into consideration as a useful feature.

The sign of the slope can be negative or positive. Thus, the feature *SB* can be viewed as a categorical variable, which is '0' if the sign is negative, and '1' otherwise. It can discriminate decreasing versus increasing trend and downward versus upward shift patterns.

(iii). Area between the overall pattern and the LS line per interval in terms of SD^2 (*ALSPI*):

$$ALSPI = \left[ALS/(N-1)\right] \Big/ SD^2, \quad (3)$$

where *ALS* is the area between the pattern and the fitted LS line representing the overall pattern. The value of *ALS* can be easily computed by summing the areas of the triangles and trapeziums that are formed by the LS line and overall pattern. The computation

method is described in details in the next section. The magnitude of *ALSPI* is the highest for STA pattern, lowest for the SYS pattern and intermediate for all other patterns.

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

$$AASBP = \sum_{i=1}^{N-1} \left| \left(y_{i+1} - y_i \right) / \left(t_{i+1} - t_i \right) \right| \Big/ \left(N - 1 \right). \tag{4}$$

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

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 following two features are extracted after segmentation of the observation window so that these two types of deviations can be discriminated, and for their extractions, two types of segmentations approaches, i.e. (a) pre-defined segmentation, where the segment sizes remain fixed, and (b) criterion-based segmentation, where the segment sizes may vary in order to satisfy a desired criterion, have been used.

(v). Range of slopes of straight lines passing through six pair-wise combinations of midpoints of four equal segments (*SRANGE*):

The feature *SRANGE* is extracted based on pre-defined segmentation of the observation window. In this segmentation approach, the total length of the data plot is divided into four equal segments consisting of $(N/4)$ data points, as shown in Figure 2. It is assumed that the total number of observations (N) is so chosen that it can be divisible by 4. The behavior of the process within a segment can be represented by the midpoint of the segment, which is given as

$$\left[\left\{ \sum_{i=k}^{k+(N/4)-1} t_i \Big/ (N/4) \right\}, \left\{ \sum_{i=k}^{k+(N/4)-1} y_i \Big/ (N/4) \right\} \right],$$

where $k = 1,\ (N/4+1),\ (2N/4+1),\ (3N/4+1)$ for the first, second, third and fourth segment respectively. A combination of two midpoints can be obtained in $C_2^4 = 6$ ways implying that six straight lines can be drawn passing through the midpoints of these four segments. So the feature *SRANGE* can be extracted using the following expression:

$$SRANGE = \text{maximum } (s_{jk}) - \text{minimum } (s_{jk}), \tag{5}$$

where $j = 1,2,3;\ k = 2,3,4;\ j < k$. s_{jk} is the slope of the straight line passing through the midpoints of j^{th} and k^{th} segments. The magnitude of *SRANGE* will be higher for shift patterns than trend patterns. The magnitude of *SRANGE* will also be higher for CYC pattern than NOR, STA and SYS patterns, unless each segment of the cyclic pattern consists of a complete cycle.

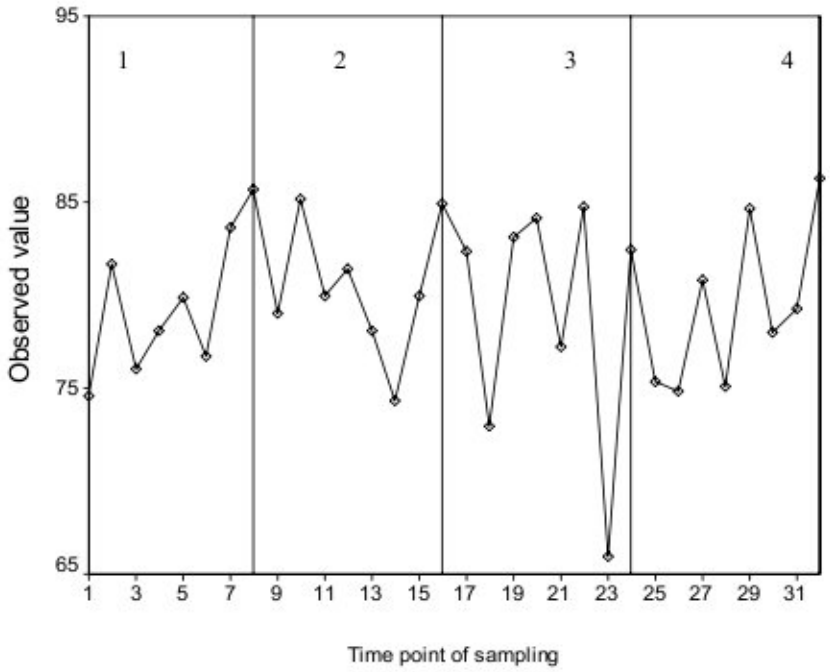


Figure 2. Four segments of equal size in a pattern.

(vi). Sum of absolute slopes of the LS lines fitted to the two segments (*SASPE*):

If the time point of occurrence of a shift can be known and two LS lines are fitted to the observations before and after the shift, each line will be approximately horizontal to the X-axis. However, the time point of occurrence of the shift cannot be known exactly. Therefore, a criterion for segmentation of the observation window into two segments is considered. Here, the defined criterion is minimization of the pooled mean sum of squares of errors (*PMSE*) of the two LS lines fitted to the two segments. Assuming that at least 8 data points are required for fitting a LS line, least square lines are fitted to all possible two segments and the segmentation which leads to the minimum *PMSE* is chosen. Then, the feature *SASPE* is extracted using the following equation:

$$SASPE = \sum_{j=1}^{2} |B_j|, \tag{6}$$

where B_j is the slope of the LS line fitted to j^{th} segment. The magnitude of *SASPE* will be higher for trend patterns than shift patterns.

Table 1 shows the values of pair-wise correlation coefficients among the selected features computed from a set of training samples (see section 5.1). The table reveals that the degrees of association among the selected features are considerably low.

Table 1. Pair-wise correlation coefficients among the selected features.

Selected features	*RVE*	*SB*	*ALSPI*	*AASBP*	*SRANGE*	*SASPE*
RVE	1.00	0.02	–0.29	–0.21	0.07	–0.04
SB	0.02	1.00	0.00	0.14	–0.17	–0.20
ALSPI	–0.29	0.00	1.00	–0.46	–0.34	–0.40
AASBP	–0.21	0.14	–0.46	1.00	–0.02	0.23
SRANGE	0.07	–0.17	–0.34	–0.02	1.00	0.38
SASPE	–0.04	–0.20	–0.40	0.23	0.38	1.00

3. Generalization of Features

The necessary conditions for generalization of the extracted features from the data plot of a normal process are determined first. The i^{th} 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 equations (1), (2) and (4), the values of the corresponding features are obtained as follows:

$$RVE = \left[\sum_{i=1}^{N}(r_i - \bar{r})^2 \Big/ (N-1)\right] \Big/ \left[\left\{\sum_{i=1}^{N}(r_i - \bar{r})^2 - \left(\sum_{i=1}^{N} r_i(i-\bar{i})\right)^2 \Big/ \sum_{i=1}^{N}(i-\bar{i})^2\right\} \Big/ (N-2)\right], \quad (7)$$

$$B = \left[c\sigma \sum_{i=1}^{N} r_i(i-\bar{i})\right] \Big/ \left[c^2 \sum_{i=1}^{N}(i-\bar{i})^2\right], \quad (8)$$

$$AASBP = \frac{\sigma}{c} \sum_{i=1}^{N-1} |r_{i+1} - r_i| \Big/ (N-1). \quad (9)$$

It is noted from equation (7) that the value of *RVE* is given by a function of r_i only for a given *N*. This implies that the feature *RVE* is always independent of μ and σ. The equations (8) and (9) indicate that for given *N*, the values of *B* and *AASBP* will also be expressed as a function of r_i only, if $c = 1\sigma$. It can be shown that the value of slope of the LS line fitted to any number of observations will be independent of μ and σ, if $c = 1\sigma$. This is indicative that the feature *SASPE* will also be independent of the two process parameters under the condition that each sampling interval is represented by a linear distance equal to one standard deviation.

On the other hand, the condition for negative or positive sign of the slope (*B*) will depend on the value of the numerator in equation (8). The sign will be negative, i.e. *SB* = 0 if

$$c\sigma \sum_{i=1}^{N} r_i(i-\bar{i}) < 0 \quad \text{or,} \quad \sum_{i=1}^{N} r_i(i-\bar{i}) < 0 .$$

So, the condition for negative (or similarly positive) sign of the slope does not depend on the values of μ, σ and *c*. This implies that the value of *SB* will be independent of the

process mean and standard deviation without requiring any condition on representation of the sampling intervals in the control chart plot.

As it is noted from equation (3), it is necessary to derive the values of *ALS* and SD^2 for computation of the value of *ALSPI*. For simplicity, let us consider a possible appearance of the first three data points on a control chart plot, as shown in Figure 3 and calculate the respective *ALS* and SD^2 values.

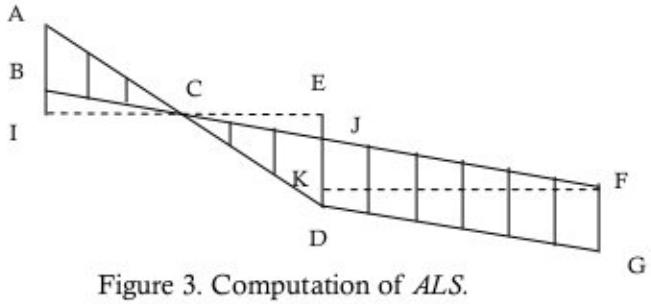


Figure 3. Computation of *ALS*.

In Figure 3, BF represents the LS line, $\hat{y}_i = \hat{b} + \hat{b}_1 t_i$, where, $\hat{y}_i$ is the predicted observation at i^{th} time point, and $\hat{b}$ and $\hat{b}_1$ are the two constant terms. The constants $\hat{b}$ and $\hat{b}_1$ represent the intercept and slope of the LS line respectively. The values of $\hat{b}$ and $\hat{b}_1$ can be estimated as below:

$$\hat{b} = \bar{y} - \hat{b}_1 \bar{t} = \mu + \left\{ \frac{\bar{r} \sum_{i=1}^{3} (i - \bar{i})^2 - \bar{i} \sum_{i=1}^{3} r_i (i - \bar{i})}{\sum_{i=1}^{3} (i - \bar{i})^2} \right\} \sigma = \mu + \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} \sigma , \tag{10}$$

where, $P = \sum_{i=1}^{3} (i - \bar{i})^2$, a function of i only for given N, and

$Q = \sum_{i=1}^{3} r_i (i - \bar{i})$, a function of r_i and i only for given N.

$$\hat{b}_1 = \frac{\sum_{i=1}^{3} y_i (t_i - \bar{t})}{\sum_{i=1}^{3} (t_i - \bar{t})^2} = \frac{\sigma \sum_{i=1}^{3} r_i (i - \bar{i})}{c \sum_{i=1}^{3} (i - \bar{i})^2} = \frac{\sigma}{c} \times \frac{Q}{P}. \tag{11}$$

The points A, D and G, in Figure 3, are the plots of the first three observations on the control chart. Two straight lines IE and KF are drawn parallel to the mean line and the lengths of these lines represent the sampling interval in the control chart plot. Therefore, it can be shown that

IE = KF = c.

$$\text{AB} = |y_1 - \hat{y}_1| = \left| \left[r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right] \sigma \right|. \tag{12}$$

$$\text{DJ} = |y_2 - \hat{y}_2| = \left| \left[r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right] \sigma \right|. \tag{13}$$

$$\text{GF} = |y_3 - \hat{y}_3| = \left| \left[r_3 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{3Q}{P} \right\} \right] \sigma \right|. \tag{14}$$

It may be noted that in Figure 3, ΔABC and ΔCDJ are similar triangles. The triangles ΔBCI and ΔCEJ are also similar. Thus, using the properties of similar triangles, it can be found that

$$\frac{\text{IC}}{\text{CE}} = \frac{\text{BC}}{\text{CJ}} = \frac{\text{AB}}{\text{DJ}}$$

$$\text{or,} \quad \frac{c - \text{CE}}{\text{CE}} = \frac{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{Q} \right\} - \left\{ \frac{Q}{P} \right\} \right|}{\left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|}$$

$$\text{or,} \quad \text{CE} = \frac{\left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right| \times c}{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right| + \left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|} \tag{15}$$

$$\text{IC} = c - CE = \frac{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right| \times c}{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right| + \left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|}. \tag{16}$$

Here, the straight lines IC and CE represent the heights of ΔABC and ΔCDJ respectively. Therefore, the value of *ALS* is obtained as

ALS = area of ΔABC + area of ΔCDJ + area of trapezium DGFJ

$$= \frac{c\sigma}{2} \left[\frac{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right|^2}{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right| + \left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|} \right.$$

$$\left. + \frac{\left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|^2}{\left| r_1 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{Q}{P} \right\} \right| + \left| r_2 - \left\{ \frac{\bar{r}P - \bar{i}Q}{P} \right\} - \left\{ \frac{2Q}{P} \right\} \right|} \right]$$

$$+\left\{r_2-\left(\frac{\bar{r}P-\bar{i}Q}{P}\right)-\left(\frac{2Q}{P}\right)\right\}+\left\{r_3-\left(\frac{\bar{r}P-\bar{i}Q}{P}\right)-\left(\frac{3Q}{P}\right)\right\}]. \quad (17)$$

and the value of SD^2 is given by the following expression:

$$SD^2=\frac{\sum_{i=1}^{3}(y_i-\bar{y})^2}{3-1}=\frac{\sigma^2\sum_{i=1}^{3}(r_i-\bar{r})^2}{2}. \quad (18)$$

Thus, the value of *ALSPI*, which can be computed using equation (3), will become a function of r_i, σ and *c*. However, if the constant linear distance, *c* is made equal to one process standard deviation, the value of *ALSPI* will be given by a function of r_i only implying that the magnitude of *ALSPI* will become independent of the values of process mean and standard deviation.

The feature *SRANGE* is defined based on the segmentation of the observation window into four equal segments, as shown in Figure 2. The value of the slope of the straight line drawn through the midpoints of first and second segments i.e. s_{12} will be given by the following expression:

$$s_{12}=\frac{\frac{1}{(N/4)}\sum_{i=(N/4)+1}^{(2N/4)}y_i-\frac{1}{(N/4)}\sum_{i=1}^{(N/4)}y_i}{\frac{1}{(N/4)}\sum_{i=(N/4)+1}^{(2N/4)}ic-\frac{1}{(N/4)}\sum_{i=1}^{(N/4)}ic}=\frac{\sigma\left(\sum_{i=(N/4)+1}^{(2N/4)}r_i-\sum_{i=1}^{(N/4)}r_i\right)}{c\left(\sum_{i=(N/4)+1}^{(2N/4)}i-\sum_{i=1}^{(N/4)}i\right)}.$$

The above expression reveals that the value of s_{12} will be given by a function of r_i only if $c=1\sigma$. Similarly, the values of s_{jk} (for all *j* and *k*) will also be independent of the two process parameters, when $c=1\sigma$. Thus, the value of *SRANGE* will be independent of the values of the two process parameters under the condition that each sampling interval in the control chart plot is represented using a linear distance equal to one standard deviation of the underlying process.

All unnatural patterns are generated 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 a normal pattern, i.e. $y_i=\mu+$ a cofficient $\times\sigma$. For example, the equation for modeling an UT 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.

It can, therefore, be concluded that the magnitudes of all the extracted features from the control chart plot in an observation window will become independent of process mean and standard deviation under the condition that each interval in the plot is represented by a linear distance equal to one standard deviation. All the features, therefore, are extracted here assuming that a sampling interval in the control chart plot is represented by a linear distance, $c=1\sigma$.

4. Control Chart Pattern Recognition

4.1. Feature-Based Heuristics

This technique uses simple IF …(condition)…THEN …(action)…heuristic rules. The conditions for the rules set the threshold values of the features and the actions are the classification decisions. The set of heuristics, arranged as a decision tree, can provide easily understood and interpreted information regarding the predictive structure of the data for various features. Given a set of extracted features from the learning samples, classification tree programs like CART (Classification and regression trees) [1] and QUEST (Quick, unbiased, efficient statistical tree) [12] can generate a set of heuristics based on these features arranged as binary decision tree, which can be used for recognition of control chart patterns (CCPs). Both these tree-structured classification algorithms allow automatic selection of the 'right-sized' tree that has the optimal predictive accuracy. The procedures for the 'right-sized' tree selection are not foolproof, but at least, they take the subjective judgment out of the process of choosing the 'right-sized' tree and thus avoid 'over fitting' and 'under fitting' of the data.

The CART performs an exhaustive grid search of all the possible univariate splits while developing a classification tree. CART searches can be lengthy when there are a large number of predictor variables with many levels and it is biased towards choosing the predictor variables with more levels for splits [12]. On the other hand, QUEST employs modification of the recursive quadratic discriminant analysis and includes a number of innovative features for improving the reliability and efficiency of the classification tree that it computes. QUEST is fast and unbiased. QUEST's lack of bias in variable selection for splits is a distinct advantage when some predictor variables have few levels and other have many. Moreover, QUEST does not sacrifice the predictive accuracy for speed [11]. It is planned, therefore, to adopt QUEST algorithm for determining the tree-structured classification rules for CCP recognition.

4.2. Feature-Based Neural Network

Many researchers [6, 8, 10, 17, 19, 20] have successfully used multilayer perceptron (MLP) with back propagation learning rule to solve the pattern classification problems. The advantage of this type of neural network is that it is simple and ideally suited for pattern recognition tasks [9]. One main disadvantage of the back propagation learning rule is reported to be its longer convergence time. However, the problem may not be significant for application to pattern classification, since training is envisaged as an infrequent and off-line activity. The MLP architecture is, therefore, used here for developing the CCP recognition system. The basic structure of an MLP architecture comprises an input layer, one or more hidden layer(s) and an output layer. The input layer receives numerical values from the outside world and the output layer sends information to the users or external devices. The number of nodes in the input layer is set according to the actual number of features used, i.e. six. The number of output nodes is set corresponding to the number of pattern classes, i.e. eight. The processing elements in the hidden layer are used to create internal representations. Each processing element in a particular layer is fully connected to every processing element in the succeeding layer. There is no feed back to any of the processing elements.

Preliminary investigations are carried out to choose the number of hidden layers, number of nodes in these layers and select an efficient back propagation training algorithm by conducting experiments coded in MATLAB® using its ANN toolbox [3]. Based on the

results of these initial investigations, it is decided to use one hidden layer for the neural network and the number of nodes in the hidden layer is set to 15. Thus, the ANN structure is 6×15×8. Since the supervised training approach is used here, each pattern presentation is tagged with its respective label. The target values for the recognizer's output nodes are represented by a vector of 8 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 highest output for a pattern considered to be correctly classified. Gradient descent with momentum and adaptive learning rate (*traingdx*) (*traingdx* is the code available in MATLAB® for the training algorithm) is found to provide reasonably good performance and more consistent results. It is also more memory-efficient. The *traingdx* is, therefore, chosen here for training of the neural network. The activation functions used are hyperbolic tangent (*tansig*) for the hidden layer and sigmoid (*logsig*) for the output layer. The hyperbolic tangent function transforms the layer inputs to output range from –1 to +1 and the sigmoid function transforms the layer inputs to output range from 0 to 1.

5. Experimental Procedure

5.1. Sample Patterns

Sample patterns are required for assessing the usefulness of various possible shape features in discriminating different patterns as well as developing/validating a CCP recognizer. Ideally, sample patterns should be collected from a real manufacturing process. Since, a large number of patterns are required for developing and validating a CCP recognizer, and as those are not economically available, simulated data are often used. This is a common approach adopted by the other researchers too.

Table 2. Equations and parameters for control chart pattern simulation.

Control chart patterns	Pattern parameters	Parameter values	Pattern equations
Normal	Mean (μ) Standard deviation (σ)	80 5	$y_i = \mu + r_i\sigma$
Stratification	Random noise (σ')	0.2σ to 0.4σ	$y_i = \mu + r_i\sigma'$
Systematic	Systematic departure (d)	1σ to 3σ	$y_i = \mu + r_i\sigma + d \times (-1)^i$
Cyclic	Amplitude (a) Period (T)	1.5σ to 2.5σ 8 and 16	$y_i = \mu + r_i\sigma + a\sin(2\pi i / T)$
Increasing trend	Gradient (g)	0.05σ to 0.1σ	$y_i = \mu + r_i\sigma + ig$
Decreasing trend	Gradient (g)	-0.1σ to -0.05σ	$y_i = \mu + r_i\sigma + ig$
Upward shift	Shift magnitude (s) Shift position (P)	1.5σ to 2.5σ 9, 17, 25	$y_i = \mu + r_i\sigma + ks$; $k = 1$ if $i \geq P$, else $k = 0$
Downward shift	Shift magnitude (s) Shift position (P)	-2.5σ to -1.5σ 9, 17, 25	$y_i = \mu + r_i\sigma + ks$; $k = 1$ if $i \geq P$, else $k = 0$

Note: i = discrete time point at which the pattern is sampled (i = 1,.., 32),
r_i = random value of a standard normal variate at i^{th} time point, and
y_i = sample value at i^{th} time point.

Since a large window size can decrease the recognition efficiency by increasing the time required to detect the patterns, an observation window with 32 data points is

considered here. The equations along with the corresponding parameters used for simulating the eight basic CCPs are given in Table 2. The values of different parameters for the unnatural patterns are randomly varied in a uniform manner between the limits shown. A set of 4000 (500×8) sample patterns are generated from 500 series of standard normal variates. It may be noted that each set contains equal number of samples for each pattern class. This is so done because if a particular pattern class is trained more number of times, the neural network will become biased towards that particular pattern.

5.2. Experiments

Multiple sets of learning samples as well as test samples are needed to rigorously evaluate the recognition and generalization performances of the heuristic and ANN-based CCP recognizers that may be developed based on the selected set of shape features. Here, ten sets of training and ten sets of test samples of size 4000 each are generated for the purpose of the experimentation. Only difference between these twenty sets of sample patterns is in the random generation of standard normal variate and values of different pattern parameters within their respective limits.

Each set of training samples is subjected to the classification tree analysis using QUEST algorithm with the following parameters:

- Prior probabilities for different patterns: proportional to class size
- Misclassification cost of a pattern: equal for all the patterns
- Stopping rule: Prune on misclassification error
- Stopping parameters: (a) value of 'n' for 'Minimum n' rule = 5, and
  (b) value of δ for 'δ Standard Error' rule = 1.0.

This results in ten different classification trees giving ten heuristic-based CCP recognizers. These recognizers are labeled 1.1 – 1.10 in Table 3. The recognition performance of each heuristic-based recognizer is then evaluated using all the ten sets of test samples.

On the other hand, a trained ANN can only accept a certain range of input data. The extracted shape features are, therefore, scaled (pre-processed) such that their values fall within (–1, +1) before their presentation to ANN for the learning process. The neural network is trained ten times by exposing it separately to the ten sets of training samples with the following training parameters:

- Maximum number of epochs = 2500
- Error goal = 0.01
- 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, ten different ANN-based recognizers are developed. All these ANN-based recognizers have the same architecture and differ only in the training data sets used. These recognizers are labeled 2.1 – 2.10 in Table 4. The recognition performance of each ANN-based recognizer is then evaluated using all the ten sets of test samples. As one of the aims of this experimentation is to compare the relative performance of the two types of recognizers, the corresponding heuristic and ANN-based recognizers are developed using the same set of training samples.

6. Results and Discussions

The results obtained during the training and verification phases of the developed heuristic and ANN-based CCP recognizers are given in Tables 3 and 4 respectively. It is observed that the recognition performance is quite good for both these two types of

recognizers. The overall mean percentage of correct recognition achieved by the heuristic-based recognizer at the training and verification (recall) phases are 96.79% and 95.24% respectively, and the overall mean percentage of correct recognition achieved by the ANN-based recognizer at the training and recall phases are 96.38% and 95.92% respectively. It may be noted that the recognition performance of both these two types of recognizers at the recall phase is inferior to the recognition accuracy that is achieved during the training phase. This decrease in recognition accuracy during the recall phase is more for the heuristic-based recognizer than the ANN-based recognizer. This is indicative that the predictive performance for ANN-based recognizer is better than heuristic-based recognizer.

The percentage of correct recognition for heuristic and ANN-based recognizers at recall phase ranges from 94.55% to 95.78% and from 95.50% to 96.26% respectively. Paired *t*-tests ($\alpha = 0.01$) are conducted for 10 pairs of heuristic and ANN-based recognizers for their performance in terms of percentage of correct classification. The results of statistical significance tests are summarized in Table 5. These results suggest that the difference in recognition accuracy between these two types of recognizers is significant. This confirms that ANN-based recognizers give better recognition performance compared to heuristic-based recognizers.

Table 3. Training and verification (recall) performance of heuristic-based recognizers.

Recognizer number	Training phase		Verification phase	
	Number of splits in the tree	Correct classification (%)	Correct classification (%)	
			Mean	Standard deviation
1.1	21	96.95	95.40	1.29
1.2	21	96.48	95.28	1.52
1.3	24	96.93	95.65	1.39
1.4	23	97.56	95.08	1.63
1.5	20	97.18	95.78	1.13
1.6	24	95.88	94.98	1.46
1.7	23	96.79	95.35	1.39
1.8	22	96.98	95.23	1.27
1.9	24	96.39	94.55	1.44
1.10	25	96.76	95.13	1.66
Overall mean	22.70	96.79	95.24	-

Table 4. Training and verification performance of ANN-based recognizers.

Recognizer number	Training phase		Verification phase	
	Number of epochs	Correct classification (%)	Correct classification (%)	
			Mean	Standard deviation
2.1	2203	96.05	96.28	0.55
2.2	2500	95.88	95.50	0.63
2.3	2377	96.33	95.95	0.59
2.4	2296	96.25	96.05	0.68
2.5	2054	96.35	96.25	0.46
2.6	2322	96.68	95.55	0.58
2.7	1998	96.18	95.73	0.48
2.8	2488	96.63	95.60	0.52
2.9	2208	96.55	96.05	0.61
2.10	2139	96.90	96.23	0.53
Overall mean	2258.5	96.38	95.92	-

Table 5. Statistical significance tests for difference in recall performance.

Hypothesis	$t_{\text{statistics}}$	t_{critical}	Decision
$H_0: \mu_{\text{recall (ANN-Heuristic)}} = 0$ $H_1: \mu_{\text{recall (ANN-Heuristic)}} > 0$	5.12	2.82	Reject H_0

On the other hand, the standard deviation (*sd*) of correct recognition percentage during the recall phase is observed to be consistently higher for heuristic-based recognizers than ANN-based recognizers. The statistical significance for this difference is carried out using *F*-test. The critical value of $F_{9,9}(\alpha = 0.01)$ is 5.35, whereas the computed *F*-statistics $(sd^2_{\text{Heuristic}} / sd^2_{\text{ANN}})$ for all the ten pairs of heuristic and ANN-based recognizers is found to be greater than 5.5. This implies that ANN-based recognizer has better predictive consistency than heuristic-based recognizer. However, heuristic-based recognizer has the distinct advantage that the practitioners can clearly understand how a particular pattern class is identified by the use of relevant shape features. This advantage can compensate the poorer recognition and predictive consistency of heuristic-based recognizer. The best heuristic-based recognizer in terms of consistency of recognition performance is recognizer no. 1.5, and its heuristic rules in the form of classification tree are shown in Figure 4.

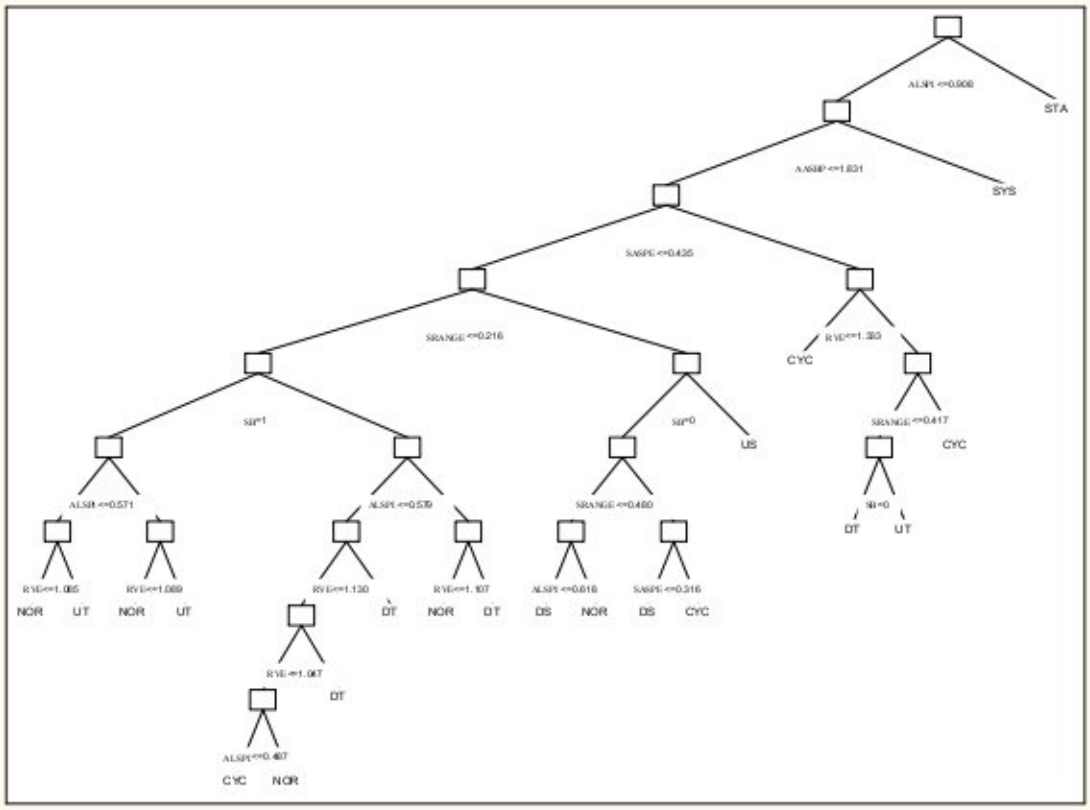


Figure 4. Classification tree for recognition of control chart patterns.

6.1. Sensitivity Studies for Pattern Recognition

In this paper, all the shape features are extracted considering that the sampling interval in the control chart plot is represented by a constant distance, $c = 1\sigma$, 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-based recognizers with respect to the estimation error of σ.

To study the sensitivities of pattern recognizers, it is necessary to generate additional test patterns, considering the values of σ different from its known value and then subjecting the patterns to classification by the developed recognizers. All these test data 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. A set of 500 time series with 32 standard normal data, from which the sample patterns leading to the heuristics shown in Figure 4 are developed is taken for generation of the additional test patterns. On the other hand, care is also taken so that the effect of random changes in the values of pattern parameters is not confounded with the effect of deviation of σ from its known value. For the purpose of generation of additional test patterns, 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 2.

Using the above mentioned time series of standard normal data and pattern parameters, five additional test pattern sets with size 4000 each are generated considering the values of σ as 4.50, 4.75, 5.00, 5.25 and 5.50 (implying $\pm 10\%$ deviation of σ from its known value) and they are labeled 1 to 5. These additional test patterns are then subjected to classification using the heuristic and ANN-based recognizers (recognizer no. 1.5 and 2.5 respectively). The achieved recognition performance is shown in Table 6. It is observed 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 $\pm 10\%$ deviation of σ for both the recognizers. The ranges of variation in recognition performance for the heuristic and ANN-based recognizers are 2.45 and 1.45 respectively. This implies that the performance of ANN-based recognizer is relatively less sensitive to estimation error of the process standard deviation than heuristic-based recognizer.

Table 6. Recognition accuracy (%) for additional test patterns.

Type of recognizer	Additional sets of test patterns					Range
	1	2	3	4	5	
Heuristic recognizer	96.05	97.53	98.08	97.08	95.63	2.45
ANN recognizer	97.10	98.18	98.40	98.10	96.95	1.45

7. Conclusions

A new approach for generalization of feature-based CCP recognition system is presented. In this approach, values of the extracted shape features from the control chart plot of actual process data become independent of mean and standard deviation of the underlying process. The advantage of the proposed approach is that one has to know or

estimate the process standard deviation only, whereas in other traditional approaches like standardization of process data, both the process mean and standard deviation values need to be known or estimated. The proposed approach is also advantageous over scaling of the process data into (0, 1) or (–1, +1) interval since in this approach, the distinction between normal and stratification patterns is preserved.

Two CCP recognizers are developed based on a set of six shape features using heuristic and ANN techniques, which can recognize all the eight basic CCPs including stratification pattern. The relative performance of the two feature-based approaches for CCP recognition is extensively examined and their sensitivities with respect to estimation error of process standard deviation are also studied. It is observed that the ANN-based recognizer achieves better recognition performance than the heuristic-based recognizer. The results also indicate that the ANN-based recognizer gives more consistent performance and is less sensitive to estimation error of process standard deviation than the heuristic-based recognizer. However, the heuristic-based recognizer has the distinct advantage that the practitioners can clearly understand how a particular pattern is identified with the help of relevant shape features. This advantage can compensate the poorer recognition performance and consistency of the heuristic-based recognizers.

References

1. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
2. Cheng, C. S. (1997). A neural network approach for the analysis of control chart patterns. *International Journal of Production Research*, 35(3), 667-697.
3. Demuth, H. and Beale, M. (1998). *Neural Network Toolbox User's Guide*. Math Works, Natick, MA.
4. Evans, J. R. and Lindsay, W. M. (1988). A framework for expert system development in statistical quality control. *Computers and Industrial Engineering*, 14(3), 335-343.
5. Gauri, S. K. and Chakraborty, S. (2007). A study on the various features for effective control chart pattern recognition, *International Journal of Advanced Manufacturing Technology*, 34(3-4), 385-398..
6. Guh, R. S., Zorriassatine, F., Tannock, J. D. T. and O'Brien, C. (1999). On-line control chart pattern detection and discrimination – a neural network approach. *Artificial Intelligence in Engineering*, 13(4), 413-425.
7. Guh, R. S. and Shiue, Y. R. (2005). On-line identification of control chart patterns using self-organizing approaches. *International Journal of Production Research*, 43(6), 1225-1254.
8. Hassan, A., Nabi Baksh, M. S., Shaharoun, A. M. and Jamaluddin, H. (2003). Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research*, 41(7), 1587-1603.
9. Haykin, S. (1999). *Neural Network: A Comprehensive Foundation.* Prentice-Hall, Englewood Cliffs, NJ.
10. Hwarng, H. B. and Hubele, N. F. (1993). Back-propagation pattern recognisers for $\overline{X}$ control charts: methodology and performance. *Computers and Industrial Engineering*, 24(2), 219-235.
11. Lim, T.-S., Loh, W.-Y. and Shih, Y.-S. (1997). An empirical comparison of decision trees and other classification methods. *Technical Report 979,* Department of Statistics, University of Wisconsin, Madison.

12. Loh, W.-Y., and Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815-840.

13. Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*. John Wiley and Sons, New York.

14. Montgomery, D. C. and Peck, E. A. (1982). *Introduction to Linear Regression Analysis*. John Wiley and Sons, New York.

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

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

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

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

19. Pham, D. T. and Oztemel, E. (1992[b]). Control chart pattern recognition using neural networks. *Journal of System Engineering*, 2(4), 256-262.

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

21. Swift, J. A. and 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.

22. Western Electric Company. (1958). *Statistical Quality Control Handbook*, Western Electric, Indianapolis.

Authors' Biographies:

Susanta Kumar Gauri is a faculty member in the Statistical Quality Control & Operations Research Division of Indian Statistical Institute (ISI) at Kolkata, India. Besides teaching, he is highly involved in consultancy works in various industries and applied research in the field of quality control, quality assurance and quality management. He has published several papers in various national and international journals during past 8 years.

Shankar Chakraborty is Reader in the Production Engineering Department of Jadavpur University at Kolkata, India. He is highly engaged in teaching and research. His research interests include the field of intelligent manufacturing, quality function deployment, application of analytical hierarchy process and artificial neural networks. He has published many papers in various national and international journals during last 12 years.