ORIGINAL RESEARCH

A New Approach for Control Chart Pattern Recognition Using Nonlinear Correlation Measure

Farhatul Janan[1] · Naimur Rahman Chowdhury[1] · Kais Zaman[1]

Abstract

Control charts themselves are unable to predict and recognize unnatural patterns in datasets. Over the years, different methods have been developed to recognize unnatural control chart patterns (CCPs) by quantifying the relationship between a test dataset and a reference dataset. Statistical correlation measure (SCM) and support vector machine (SVM) are two existing methods for recognizing CCPs. While SCM is faulty and cannot accurately recognize patterns in most cases, computational effort in the SVM model is high, and the training process of the model is complex. In this paper, a new methodology has been proposed that can measure the functional relationship between two variables. The model measures the maximal information coefficient (MIC) between a reference dataset and a test dataset to recognize CCPs. To measure the performance, the proposed model has been illustrated with datasets obtained from a major consumer product manufacturing company. The results obtained from the MIC model have been compared with those obtained from SCM and SVM models. The comparison shows that the MIC model is better than the SCM in terms of accuracy in recognizing CCPs. Moreover, the computational time required for the MIC model is less than the SVM model. Furthermore, the MIC model is less expensive and is associated with a simpler algorithm than the SVM model.

Keywords Maximal information coefficient (MIC) · Support vector machine (SVM) · Control chart pattern (CCP)

Introduction

Quality has become one of the most vital market differentiators when it comes to a product or service. With the increasing number of producers for the same consumer product, the market has become highly competitive and manufacturers coming with better quality in their products or service get a competitive advantage. Therefore, manufacturers are focusing on enhancing their product quality to capture the major portion of the demand in the market and give constant satisfaction to their customers. To increase or maintain the quality of their offerings, manufacturers use different statistical process control tools [11]. Control chart, also known as process behavior chart, was first developed by Walter A. Shewhart in the 1920s [26]. The main reason for using control charts is to identify a manufacturing process being in a state of control [33]. However, control charts do not provide any pattern related information if a process deviates from its inherent pattern and follows a particular abnormal pattern [9]. CCPs are important because when any unnatural patterns exist in a process for a long time, it is evident that the unnatural pattern comes from a common source of error or defect. This error can result from machine damage, a problem with raw material, error from workers, etc.

Many authors have already developed different techniques for CCPR over the years. Some of them involve complex algorithms, whereas some have a simple mechanism to recognize CCPs. Cheng [3] developed a method for control chart pattern recognition using a neural network approach. Guh and Hsieh [8] proposed a neural network-based model and in their paper which was used to recognize abnormal patterns. Yang and Yang [31] proposed a fuzzy-soft learning vector quantization to recognize control chart patterns. In contrast with the above-mentioned models that require a complex learning algorithm and large computational effort, Yang and Yang [30] developed a simple mechanism to recognize CCPs using statistical correlation coefficient. In their paper, CCPs were recognized by measuring the statistical correlation between the test data and reference data with a

✉ Kais Zaman
   kaiszaman@yahoo.com

1   Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

set threshold value. The learning algorithm in this model is simple and requires very low computational effort. However, selecting the threshold value for classifying patterns into a certain class and pattern length of data is a big issue in this model, for which the accuracy to detect correct CCPs is lower than the other complex learning models.

Over the years, SVM has become the most popular model for CCPR because of its ability to control data more effectively and perform a large number of operations such as classification, regression, and distribution. SVM is a powerful machine learning tool that has the capability to represent the non-linear relationship and produce a model that generalizes well to unseen data [4]. Lin et al. [13] proposed an SVM-based approach for CCPR in autocorrelated data. Xanthopoulos and Razzaghi [28] in their research have defined a CCPR as an imbalanced supervised learning problem and used a weighted SVM model to show its benefits over traditional SVMs. Later, SVM has been incorporated with other approaches to develop more efficient models for CCPR. To outperform the traditional SVM model, Zhou et al. [37] used fuzzy SVM incorporated with a hybrid kernel function. To optimize the feature selection, they also used a genetic algorithm (GA). A supervised locally linear embedding algorithm was incorporated with SVM to reduce features of high dimension [35]. Zhang et al. [34] proposed a CCPR method by using fusion feature reduction and fireworks algorithm-based multiclass support vector machine (MSVM). Cuentas et al. [5] proposed an SVM-GA based monitoring system for CCPR of correlated processes.

In this paper, a new model is proposed that uses the MIC value to recognize control chart patterns. MIC is based on mutual information analysis (MIA) [7, 27]. Mutual information (MI) is a quantity to estimate a relationship between two random variables, where the relationship between them is either linear or nonlinear [17]. However, using MIA directly to measure non-linear relationships between two datasets is difficult because mutual information computation needs to choose an estimator from a set of existing methods [23]. Moreover, it is difficult to interpret the computed value for MI and compare different obtained values [24]. For these reasons, a statistical tool named MIC was developed by Reshef et al. [20]. It has the same advantages as the MI, since it considers any kind of relationship including non-linear relationships, but has an exact definition, a soft interpretation, and better stability [21]. The main idea of the MIC is that if a relationship exists between a test dataset and a reference dataset, there exists a partition of the data that will allow to include this relationship [36]. Using MIC, a large number of non-linear relationships between the test set and the reference set can be measured [16].

There has been a wide range of applications of MIC in measuring the relationship between datasets. Caban et al. [2] characterized non-linear dependencies among pairs of clinical variables and imaging data using MIC. Tang et al. [25] proposed a cross-platform tool that was used for the rapid computation of MIC to effectively analyze large-scale biological datasets with a substantially reduced computing time through parallel processing. Xu et al. [29] predicted defects from historical data based on MIC with hierarchical agglomerative clustering. Shao et al. [22] used an improved MIC model to analyze railway accidents. MIC has also been used in railway accident analysis. MIC was also used to identify differentially expressed genes by Liu et al. [14].

However, there has not been any application of MIC in CCPR in the existing literature. Being a novel statistical method, MIC has already been used as a robust model to identify different classes with better overall performance, adaptability, and noise immunity from the traditional statistical methods such as SCM [14]. Hence, this paper aims to develop a model to identify the CCPs using MIC, as MIC can capture linear and non-linear relationships, such as cubic, exponential, sinusoidal, etc., and also robust to outliers due to its MI foundation. This makes MIC a feasible approach to identify the relationship between test data and reference data and thus pinpoint the exact pattern of control charts.

The objective of this paper is to illustrate the proposed MIC model along with two existing CCPR methods, namely SCM and SVM with production data of 1.5 gram (g) coffee minipack obtained from a major consumer product manufacturing company. The results obtained from the MIC model are then compared with results obtained from the SCM as well as the SVM model in terms of both computational accuracy and efficiency.

The remainder of this paper is organized as follows. In the following section, we discuss the existing approaches for CCPR. The subsequent section describes the proposed CCPR approach based on MIC and in the section following that we illustrate the proposed methods with numerical examples. The next section provides conclusions and suggestions for future work.

Existing Methodologies

This section discussed two existing methods for CCPR, the SCM, and SVM models. A brief background on the CCPR problem is provided first.
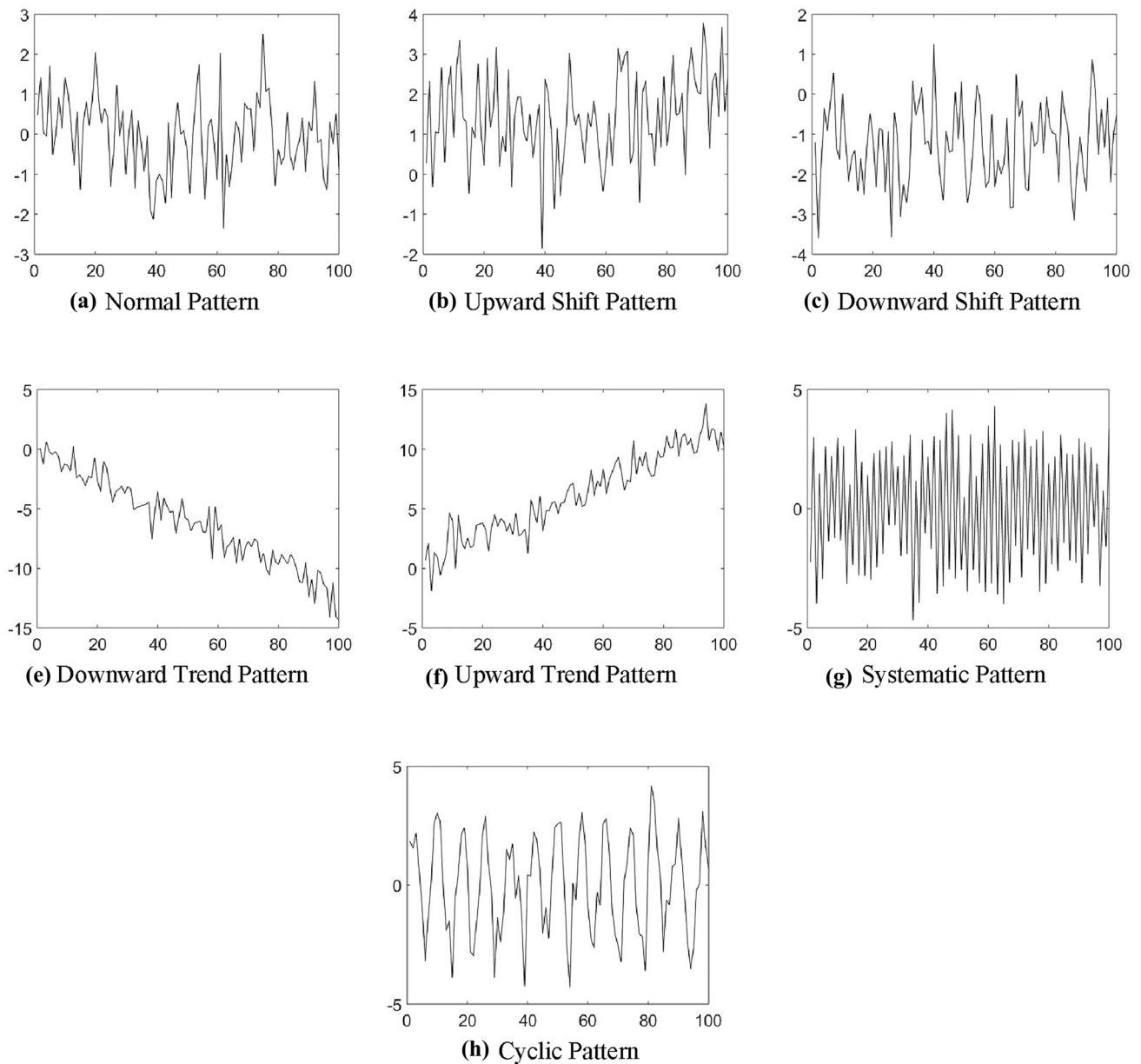
CCPR

CCPR can help to figure out the sources of deviation, if there is information available, and eliminate the error occurring in the future. This facilitates a system to remain in check so that any variations can be easily detected otherwise such alterations will remain hidden in the system

thus affecting the whole lot. There might arise six types of abnormal patterns in a process other than the natural normal pattern. Each of them presents a different form of failure or source of error in a process. In this paper, CCPs have been classified into normal (NOR) and abnormal patterns. Abnormal patterns include upward shift (US), downward shift (DS), upward trend (UT), a downward trend (DT), systematic (SYS), and cyclic (CYC) as shown in Fig. 1.

CCPR Using SCM

The statistical correlation coefficient method, also known as SCM, is the most traditional and simplest method to recognize control chart patterns. In this method, the linear correlation coefficient (r) between a reference dataset and a test dataset is measured to justify if the value of r with a reference set of a certain control chart pattern exceeds a threshold value to match that pattern.

(a) Normal Pattern   **(b)** Upward Shift Pattern   **(c)** Downward Shift Pattern

(e) Downward Trend Pattern   **(f)** Upward Trend Pattern   **(g)** Systematic Pattern

(h) Cyclic Pattern

Fig. 1 Different types of control chart patterns

Framework for SCM

SCM determines the linear correlation between two random datasets. If x and y are two random vectors, then the linear correlation between them can be defined as in Eq. (1):

$$r = \frac{\sum(x_{i\_}\bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_{i\_}\bar{x})^2}\sqrt{\sum(y_{i\_}\bar{y})^2}} \tag{1}$$

where, $\bar{x}$ and $\bar{y}$ are the mean value of x and y, respectively.

The first step of this method is to develop a training algorithm where each pattern sample has a length of n. Pattern samples can be generated as follows:

Normal pattern:

If $n(t)$ follows a normal distribution then the normal pattern can be described as in Eq. (2).

$$x(t) = n(t) \tag{2}$$

Upward and downward shift patterns:

Upward and downward patterns can be described as in Eq. (3).

$$x(t) = n(t) \pm ud \tag{3}$$

where d is the shift quantity and the values of u are 0 and 1 before and after shifting, respectively.

Upward and downward trend patterns:

Upward and downward trend patterns are described in Eq. (4).

$$x(t) = n(t) \pm te \tag{4}$$

where, e is the value of the slope of the trend.

Cyclic pattern:

The cyclic pattern is shown in Eq. (5).

$$x(t) = n(t) + f\sin\left(\frac{2\pi t}{8}\right) \tag{5}$$

where, f is the value of amplitude.

Systematic pattern:

The systematic pattern can be defined as in Eq. (6).

$$x(t) = n(t) + (-1)^t f \tag{6}$$

where f is the value of amplitude.

The first step of using the linear correlation method is to develop a training algorithm. The training algorithm can be implemented as follows:

- A pattern length of n has to be determined.
- Using a disturbance level of d, a pattern sample generator is selected. A pattern vector x_1 is generated for a pattern length of n, accordingly $x_2, x_3 \ldots, x_N$ pattern vectors are generated.

- From N samples, an estimated pattern vector is determined by Eq. (7) [30].

$$E(x) = \frac{\sum_{i=1}^{N} x_i}{N} \tag{7}$$

Now, to classify the dataset, a classification algorithm is needed. The algorithm is as follows:

- A processing data sequence containing recent n points is regarded as the pattern size to be recognized. The data sequence is input to the recognizer and calculated its statistical correlation coefficient using six reference vectors to obtain outputs for the upward shift, downward shift, upward trend, downward trend, cyclic pattern, and systematic patterns, respectively.
- The pattern for which the value of r is maximum is considered as the winner pattern for the model.

CCPR Using MSVM

Support vector machines (SVMs) were previously designed only to solve binary classification problems [10]. With the increasing need for multiclass classification, different methods have been studied. Currently, there are two types of approaches for MSVM, namely the One-against-one (OAO), and the One-against-all (OAA) method. OAO method involves constructing and combining several binary classifiers, whereas the OAA method involves directly considering all data in one optimization formulation [32]. The formulation to solve multiclass SVM problems in one step has variables proportional to the number of classes. Thus, in order to solve problems with multiclass SVM methods, either a larger optimization problem has to be constructed or several binary classifiers have to be formulated. As a result, MSVM is computationally more expensive to solve a multiclass problem than a binary problem with the same amount of data.

The core concept of the SVM model is to classify two different classes of data by constructing an optimal separating hyperplane (OHS) that maximizes the margin between the two nearest data points belonging to two separate classes [18].

If a dataset is represented by $(x_i, y_i), i = 1, 2, \ldots, l, (x_i, y_i) \in R^{n+1}$, where l is the number of samples and n is the number of features. Each x_i is a sample with n features and a class label $y \in \{-1, +1\}$. The SVM classifies the data point by identifying a separating hyperplane $w^T x + b = 0$, where w is the weight vector and b is the bias. If this hyperplane maximizes the margin, then the Eq. (8) below is valid for all input data.

$$y_i\left(w^T \emptyset(x_i) + b\right) \geq 1 \quad \forall i = 1, 2, \ldots l \tag{8}$$

where, $\emptyset$ is the kernel function.

Using the Kernel function, it is possible to classify non-linear problems. The concept of kernel trick is to map the original data points into a higher dimensional feature space in which they can be separated by a linear classifier. The projection of a linear classifier on the feature space is a non-linear one in the original space.

The separating hyperplane defined by the parameters w and b is used to maximize the margin between two classes by solving the convex optimization problem as shown in Eq. (9).

$$\min \frac{1}{2}\left\|w^2\right\| + C \sum_{i=1}^{l} \xi_i \tag{9}$$

where, C is the parameter denoting the tradeoff between the margin width and the training error and ξ_i is the slack variable to allow the margin constraint to be violated.

In this paper, MSVM is used to classify different pattern classes. Between OAA and OAO MSVM methods, OAO is more suitable to use for practical uses than OAA [10]. In OAO, for a P class problem, there are $\frac{P(P-1)}{2}$ SVM models. Each of the SVMs is trained for separating a class from another [19]. Additionally, the voting result of these SVMs also directs the decision of a testing sample. Whereas in OAA, P class pattern recognition problem is solved. In the first step of OAA, P independent SVMs are to be constructed and each of them is trained to separate one class of samples from all others. The kth SVM is trained with all of the datasets in the kth class with positive labels, and all other examples with negative labels. To test the model after all the SVMs are trained, a sample dataset is an input to all the SVMs. If that sample belongs to a certain class like P_1, then only the SVM which has been trained to separate that particular class from the others will give a positive response [6]. Because of the better suitability in the classification of practical problems, this paper uses OAO.

Feature Selection

Major decision-making in the SVM classifier is selecting the number of appropriate features. Selecting the right features in MSVMs facilitates the recognition of CCPs quickly and accurately. Different classes can be characterized by several shape features. There are nine shape features for discrimination of the CCPs [19]. Features should be chosen such that the proposed statistical features can significantly recognize the patterns quickly and accurately. Some features can be described as follows:

Slope (S) Slope is a feature that represents the least-square line representing the pattern. The magnitude of the slope is approximately zero for normal and cyclic patterns. On the other hand, for trend and shift patterns, the magnitude of the slope is greater or less than zero. As a result, the slope can be a potential feature to differentiate normal and cyclic patterns from shift patterns.

The Number of Mean Crossings (NC1) It is the number of times a pattern crosses the mean line. The value of NC1 is relatively small for shift and trend patterns. It is the highest for normal patterns. For cyclic patterns, its value is an intermediary between those of normal patterns and shift or trend patterns. Therefore, it can differentiate normal patterns from cyclic patterns as well as normal and cyclic patterns from trend and shift patterns.

The Number of Least-Square Line Crossings (NC2) NC2 is the highest for normal and trend patterns and the lowest for shift and cyclic patterns. Thus, it is a good differentiator for normal and trend patterns from shift and cyclic patterns.

The Average Slope of the Line Segments (AS) There are two line segments for each pattern to fit the data starting from either end of the pattern in addition to its least-square line. The average slope of the line segments is higher for the patterns than for normal, cyclic, and shift patterns. Therefore, it is a good feature to differentiate the patterns from the other patterns.

Other than the above-mentioned shape features, statistical features also have distinguishable properties to solve a classification problem. Mean, standard deviation, skewness, etc. are some of the important statistical features used in SVM models. They are shown respectively in Eqs. (10)–(12).

$$\text{mean} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{10}$$

$$\text{standard deviation (std)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \text{mean})^2}{n}} \tag{11}$$

$$\text{skewness} = \frac{\sum_{i=1}^{n} (x_i - \text{mean})^3}{n(\text{std})^3} \tag{12}$$

Using more features in the classification algorithm gives more accurate results. Also, selecting the right features for a particular classification problem is important.

Framework for SVM Model

The SVM model involves two algorithms. Firstly, the model is trained with training datasets. Training algorithm also involves providing the datasets with appropriate features. Training algorithm can be implemented as follows:

- A pattern length of n has to be determined for each reference dataset.
- A pattern sample generator is selected from the reference pattern equations in Eqs. (2)–(6). A pattern vector x_1 is generated for a pattern length of n, accordingly x_2, x_3…, x_N pattern vectors are generated. N is the number of samples used for a certain pattern.
- From N samples, an estimated pattern vector is determined by Eq. (7).
- Each reference pattern vector is trained with its shape and statistical features. Features are mean, standard deviation, skewness, and the number of crossings.
- Each reference pattern is labeled with a different number.

After the reference datasets are properly trained, a classification algorithm is implemented as follows:

- Input test dataset with corresponding features.
- MSVM model compares the test datasets to all the SVMs trained.
- The output is the label of the pattern best matched with the test data.

In the following section, we have proposed a new technique to recognize CCPs using the MIC model.

Proposed CCPR Model Using MIC

MIC method uses MI between two random variables to determine non-linear dependence or functional relationship between them. Using MIC, we can identify the functional relationship between a test dataset and a reference dataset. If the MIC value is low, that will signal a weak functional relationship between the two datasets. If the MIC value is close to 1, that will signal a strong relationship between the datasets and identify a dataset following the corresponding pattern.

Framework for MIC Model

MIC is based on MIA. Before describing MIC, some tools of mutual information are discussed first.

Let A be a random variable in space A, B a random variable in space B and $\Pr[E]$ is the probability of the event E. The entropy (H) associated with A is defined in Eq. (13).

$$H(A) = -\sum\nolimits_{a \in A}^{n} \Pr[A = a].\log_2(\Pr[A = a]) \tag{13}$$

The entropy associated with B is defined in Eq. (14).

$$H(B) = -\sum\nolimits_{b \in B}^{n} \Pr[B = b].\log_2(\Pr[B = b]) \tag{14}$$

The conditional entropy of A given B can be defined by Eq. (15).

$$H(A, B) = -\sum\nolimits_{a \in A, b \in B}^{n} \Pr[A = a, B = b].\log_2(\Pr[A = a|B = b]) \tag{15}$$

The MI which defines the mutual dependency between random variables can be defined with the help of Eq. (16).

$$\text{MI}(A, B) = H(A) - H(A|B) = H(A) + H(B) - H(A, B) \tag{16}$$

The maximal value of the MI is $H(A) + H(B)$. This measure can detect if there is any relationship linking the variables regardless it is a linear or a non-linear relationship. If the value of MI is 0, the two variables are independent. On the contrary, if the value of the MI is maximal, it indicates that A and B have the strongest relationship. However, computing the mutual information is tricky, when a continuous variable is involved. The issue with MI is that they overestimate the conditional entropy. This has long been perceived as an issue and huge consideration has been committed to creating alternative methods to nullify this limitation [21]. For these reasons, a new method called maximal information coefficient (MIC) was proposed to estimate the probability density function (PDF) of variables using bins [20].

The objective of MIC is to measure if there exists a relationship between two random variables A and B. A and B are coupled in a set $D(A, B)$. A *p-by-q* grid is formed that can be called the partition of the couple (A, B). Here, variable A is distributed in p, and variable B is distributed in q bins. There are a lot of different grids of size *p-by-q*. A^G and B^G are used to denote the distributions of A and B over the grid G and $G(p, q)$ is the set of all grids of size *p-by-q*.

For a particular value of p and q, Eq. (17) defines the maximal mutual information over grids *p-by-q*:

$$I^*(D, p, q) = \max\nolimits_{G \in (p, q)}(I(A^G, B^G)) \tag{17}$$

MIC value between the variables A and B can be defined with the help of Eq. (18).

$$\text{MIC}(p, q) = \max\left(\frac{I^*(D, p, q)}{\log_2 \min(p, q)} \right) \tag{18}$$

MIC value is defined by calculating the ratio in Eq. (18) using the data-dependent binning scheme. MIC value is also bounded by the denominator of the ratio, that's why the range of MIC values is always 0 to 1. MIC can be stated as the estimate of MI, $I^*(D, p, q)$ utilizing constrained binning scheme and divided by

a normalization factor, $log_2 min(p,q)$. It appears that except for highly structured data, MIC values reduce to the estimates of MI as the denominator of the ratio in Eq. (18) becomes 1.

The value of MIC is highly dependent on choosing the number of p and q bins. As the bin sizes affect the value of mutual information, determining the appropriate values of p and q is important. The generality of MIC shares a direct relationship with maximal grid size $B(n)$. If $B(n)$ is set too low, this will result in searching only for simple patterns and weakening the generality. On the other hand, while setting $B(n)$ too high will result in a nontrivial MIC score for independent paired variables under finite samples. Reshef et al. [20] set the $p \times q \mathrel{<} B(n)$, where $B(n) = n^{0.6}$ is the maximal grid size restriction and n is the sample size. This grid size has been used to develop the MIC model for CCPR in this paper.

Using this concept, the MIC model takes a reference dataset of one of the seven CCPs as variable A that is distributed in p and the test dataset as variable B distributed in q. The reference dataset for which the MIC gives the highest value is considered as the actual pattern of the test dataset.

An algorithm to generate a reference dataset can be implemented as follows:

- A pattern length of n is determined for each reference dataset.
- A pattern sample generator is selected from the reference pattern equations in Eqs. (2)– (6).
- A pattern vector x_1 is generated for a pattern length of n, accordingly $x_2, x_3, \ldots, x_N$ pattern vectors are generated. N is the number of samples used for a certain pattern.
- From N samples, an estimated pattern vector is determined by Eq. (8)

Pattern recognition algorithm by determining MIC value between reference dataset and test dataset is as follows:

- Test dataset of pattern length n is imported.
- MIC value between the test dataset and each reference dataset is determined using Eq. (18).
- Amongst the MIC values, the highest MIC obtained from the reference dataset is considered as the winner pattern for the test dataset.

In the following section, we illustrate the CCPR methodologies with numerical examples.

Numerical Illustrations

Data Collection

To demonstrate control chart pattern recognition, training datasets, and test datasets are to be incorporated in the model to recognize the patterns of the test datasets. Training datasets are generated randomly using MATLAB using Eqs. (2)–(6). For a better understanding of the classification problem, production data of a consumer product is used as the test dataset. In this paper, production datasets of 1.5 g coffee minipack manufactured in a major manufacturing company 'XYZ' have been used. Data from two different coffee powder filling machines are taken from two production lines. For numerical illustration, a portion of a test dataset from Machine-1 and Machine-2 are presented in Tables 10 and 11 respectively. For SCM, each of the datasets has been normalized and used to measure the correlation coefficient with all 7 training datasets. For SVM, the test datasets have been first normalized and then used to determine the SVM features described in "Feature selection". For MIC, the same procedure has been followed as in the SCM method. Additionally, for Machine-1, filling capacity has been gradually increased in each batch and shown in Table 10 to validate the credibility of the three models in recognizing an abnormal CCP. Similarly, other test datasets obtained from company '*XYZ*' have also been used for CCPR.

The selection of pattern length is very important for the datasets because the accuracy of the models depends on it. 1000 randomly generated samples of different patterns where pattern length, $n = 70, 80, 100, 150, 200, 300, 400, 500$ were taken and these three models were used for pattern recognition. The simulation was repeated ten times for each type of pattern and it was seen that for a pattern length of 400, the accuracy was higher. Therefore, for both training and testing datasets, a pattern length of 400 is selected.

SCM

In SCM, firstly, reference data are generated for all six unnatural control chart patterns as well as for the natural normal pattern using Eqs. (2)–(6). Using Eq. (1), correlation coefficient r_{norm}, $r_{u.s.}$, $r_{d.s.}$, $r_{u.t}$, $r_{d.t}$, r_{cyc}, r_{sys} between test dataset and reference dataset is measured. r_{norm}, $r_{u.s.}$, $r_{d.s.}$, $r_{u.t}$, $r_{d.t}$, r_{cyc}, r_{sys} are correlation coefficients, respectively, for NOR, US, DS, UT, DT, CYC, and SYS patterns. Correlation coefficient values for sample test datasets as given in Tables 1 and 11 ("Appendix") are shown in Tables 1 and 2, respectively.

From Table 1, it is seen that in every trial, correlation coefficients' values of upward and downward trends give the highest scores compared to others, thus creating a paradox in recognizing the actual pattern of the datasets. From Table 2, it can be said that correlation coefficients' values are highest for normal patterns in each trial. The highest value for a pattern in each trial is considered as the winner pattern of the dataset for that trial and the pattern that wins the highest number of trials is considered as the actual pattern of the test

Table 1 Correlation coefficient values for Machine-1 dataset

Correlation coefficient	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
r_{norm}	0.0419	0.0431	0.0434	0.0253	0.1013
$r_{u.s}$	0.0024	0.0334	0.0058	0.0027	0.0042
$r_{d.s}$	0.0095	0.0152	0.0152	0.0133	0.0163
$r_{u.t}$	0.9974	0.9975	0.9975	0.9976	0.9975
$r_{d.t}$	0.9974	0.9970	0.9976	0.9975	0.9975
r_{cyc}	0.0184	0.0095	0.0255	0.0165	0.0431
r_{sys}	0.0147	0.0334	0.0407	0.0925	0.0107

Table 2 Correlation coefficient values for Machine-2 dataset

Correlation coefficient	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
r_{norm}	0.0044	0.0399	0.0652	0.0445	0.0052
$r_{u.s}$	0.0021	0.0048	0.0002	0.0011	0.0002
$r_{d.s}$	0.0154	0.0135	0.0187	0.0154	0.0151
$r_{u.t}$	0.0077	0.0084	0.0074	0.0074	0.0073
$r_{d.t}$	0.0074	0.0075	0.0070	0.0076	0.0070
r_{cyc}	0.0400	0.0048	0.0083	0.0441	0.0137
r_{sys}	0.0584	0.0252	0.1217	0.0231	0.0520

Table 3 Class labels

Class label	Control chart pattern
1	Normal
2	Upward shift
3	Downward shift
4	Upward trend
5	Downward trend
6	Cyclic
7	Systematic

Table 4 Feature values of test datasets

Test datasets	Mean	Number of crossings (NC)
Machine-1	1.886213	6
Machine-2	1.55041	50

Table 5 Classifications of test datasets

Test datasets	Output				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Machine 1	4	4	4	4	4
Machine 2	1	1	1	1	1

dataset. That is to say, SCM cannot confirm the pattern for the dataset from Machine-1, but identity Machine-2 dataset as a normal pattern.

Error Analysis for SCM

50 datasets were taken from two coffee machines from the production floor of company 'XYZ' at different times for control chart pattern recognition using all three approaches. Using the SCM model, these 50 datasets have been tested and most of the results were confusing as different trials give different results. By considering the maximum coefficient value, 30% of datasets were correctly recognized and around 70% of results could not accurately recognize the pattern. Hence, the accuracy of this model in recognizing patterns is low as it cannot recognize the nonlinear relations between the datasets. Other factors like randomness in the dataset, data collection methods (timings and frequency), etc., can also affect the accuracy of the SCM model.

SVM Model

In the SVM model, the reference dataset is firstly generated and their appropriate features are determined. In this paper,

the mean value is considered as the statistical, feature and the number of crossings is considered as the shape feature. Using these two features, reference datasets are classified into different classes and classes are labeled as shown in Table 3.

After the reference datasets are labeled and classified in different classes, test datasets are used as inputs to the model and their features are calculated. Table 4 shows samples' feature values for two test datasets.

Once the features are determined, datasets are tested against each trained class and the classifications obtained from the model are given in Table 5.

From Table 5, it is clear that the SVM model identifies the Machine-1 dataset as upward trend pattern and the Machine-2 dataset as normal pattern, since it gives the same output in all the trials.

Error Analysis for SVM

From the analysis of the test dataset of pattern length n of 400, it is evident that SVM, for its highly developed training algorithm, gives the same classification result in all the trials, thus considered as 100% accurate in recognizing CCPs using appropriate features. The result shows that SVM

performs well when there is a distinct margin of class separation. It also uses a subset of the training set in the decision functions called support vectors, so it is also memory efficient. All the 50 test datasets mentioned in "Error analysis for SCM" are checked with the SVM model and the accuracies of the results obtained from other models have been identified using the results obtained from SVM model. However, one of the significant problems of SVM is the trouble to interpret except if the features are interpretable. Values of features in different classes must not intersect with each other. For that reason, the selection of an optimum number of features and their corresponding ranges are necessary and difficult tasks for the proper separation and identification of classes based on those features. In other words, the selection of features for the classification of data is a vital task in SVM. It also tends to be computationally costly and it highly depends on kernel function [1, 15]. All these negative traits of the SVM model in pattern recognition create the need to develop or identify another model to overcome the disadvantages of SVM while maintaining accuracy. Hence, results from the other two methods are compared with the SVM model to search for a method that can replace the SVM model by providing substantial accuracy and overcoming the computational complexity of the SVM model.

MIC Model

Using the test datasets as given in Tables 10 and 11, the MIC model compares the MIC values between the test datasets and training datasets. Using Eq. (18), MIC values are measured for the sample datasets of Machine-1 and Machine-2 as shown in Tables 6 and 7, respectively.

As discussed earlier in "Framework for MIC model", the highest MIC obtained from the reference dataset is considered as the winner pattern for the test dataset. It is seen in Table 6 that, in every trial, the MIC value is the maximum for the upward trend. Therefore, we identify the Machine-1 dataset as upward trend pattern. Similarly, it can be observed in Table 7 that, in every trial, the MIC value is the maximum for the normal pattern. Therefore, the proposed framework identifies the Machine-2 dataset as normal pattern.

Error Analysis for MIC

MIC is an equitable measure of dependence as it has the ability to provide similar scores to relationships with similar noise levels [20]. Because of this property, MIC is more useful in identifying the strongest association among many significant associations in a dataset, compared to other alternatives such as MI estimation, distance correlation, Spearman correlation coefficient, etc. However, all statistical problems do not call for a self-equitable measure of dependence. For example, if data which are to be measured are of a limited type and noise in that data are approximated as Gaussian,

Table 6 MIC values for Machine-1 dataset

Control Chart Patterns	MIC value				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Normal	0.1966	0.1667	0.1850	0.1674	0.1811
Up shift	0.2087	0.2061	0.1791	0.1937	0.1889
Down shift	0.1902	0.1755	0.1838	0.1599	0.1696
Up trend	1.0000	1.0000	1.0000	1.0000	1.0000
Down trend	0.1966	0.1616	0.1679	0.1617	0.1899
Cyclic	0.1634	0.1532	0.1656	0.1498	0.1665
Systematic	0.1898	0.1804	0.1717	0.1905	0.1765

Table 7 MIC values for Machine-2 dataset

Control Chart Patterns	MIC value				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Normal	0.4714	0.3176	0.3528	0.4939	0.4121
Up shift	0.1573	0.1663	0.1780	0.1321	0.1715
Down shift	0.1730	0.1664	0.1747	0.1733	0.1547
Up trend	0.1842	0.1862	0.1903	0.1775	0.1839
Down trend	0.1639	0.1880	0.1666	0.1566	0.1667
Cyclic	0.1660	0.1559	0.1830	0.1450	0.1404
Systematic	0.1345	0.1970	0.1447	0.1688	0.1567

squared Pearson correlation R^2 gives more accurate results than estimates of mutual information [12]. In the case of larger data with unknown noise, MIC gives more accurate results. After testing 50 test datasets as mentioned in "Error analysis for SCM" with the MIC model, it is seen that it can accurately recognize patterns 85% of the time. MIC model can estimate both linear and nonlinear correlations between datasets considering the noise levels in datasets. Therefore, the accuracy of the MIC model is better than the SCM model. Though its accuracy is slightly less than the SVM model, it still can replace SVM in places where a few errors in the system are acceptable.

Comparison of Results

After recognizing CCPs using the traditional SCM, SVM and MIC, it can be stated that traditional SCM results in 70% error which is the highest among the three models. As it can only measure the linear relationship between two datasets, it gives inaccurate results for most of the cases where linearity cannot properly define the relation between a test and a training dataset. The output obtained from the MIC model clarifies that it performs better in terms of accuracy than the SCM method as it is able to identify the non-linear relationship between any two datasets. MIC model can recognize CCPs having pattern length n of 400 with 85% accuracy. Among the three models, the SVM model is the most

Table 8 Accuracies of the three models in CCPR

Model	Accuracy (%)
SCM	30
MIC	85
SVM	100

Table 9 Computational time for three models

Model	Computational time (s)
SCM	0.23
MIC	0.31
SVM	0.39

advanced and accurate model which recognizes CCPs for this study with a 100% accuracy. The training algorithm and feature selection in the SVM model allow it to perform better than both SCM and MIC models. In Table 8, accuracies for statistical correlation measure and MIC model are shown.

For the pattern length, n of 400, the computational time for three models was calculated as given in Table 9. From Table 9, it is seen that SCM needs the lowest time for computing the result. However, this model has very low accuracy. MIC model needs 0.31 s to calculate output, whereas the SVM model needs 0.39 s to recognize a certain pattern. It is evident that the MIC model is 20.5% faster than the SVM model in CCPR because of its simpler learning algorithm than SVM.

The comparison among the three models suggests that the MIC model is the most feasible model when 10–15% error can be considered allowable in the output, because of its simplicity, accuracy and decreased computational time.

Conclusions

In this highly competitive market, no product can sustain profitability without maintaining consistent quality. Manufacturing process plays a vital role in maintaining the product quality. As a result, maintaining a highly controlled manufacturing process has become mandatory. Control charts are one of the most useful statistical tools used to monitor a process. With the use of a control chart pattern recognizer, any type of CCP can be recognized very fast and a structured decision can be made to resolve a certain anomaly or defect that may arise in the production process.

In this paper, a new methodology has been proposed for CCPR using MIC values between two datasets. The results obtained from the MIC model are then compared with the results obtained from the traditional SCM and SVM models. Results reveal that the MIC model can be a potential alternative to the SVM model as it can certainly reduce the computational effort and cost and yet can provide satisfactory classification accuracy, whereas the SCM model is not suitable for CCPR.

The existing MIC model begins to fluctuate its output when the sample size is large. The future scope of this research involves using a more optimized MIC model that is able to handle a large number of samples each having a large pattern length without deviating from a constant performance. Another scope is optimizing the *p-by-q* grid size of the model that can distribute random variables in a more efficient way to recognize control chart patterns accurately in all cases. Moreover, a hybrid MIC model, such as random forest regression-based MIC can also be used for better accuracy in CCPR. If these improvements could be brought in the MIC model, it would be able to successfully replace the SVM model in industrial applications by achieving higher accuracy. Thus, it would be able to reduce the computational effort and cost for CCPR.

Appendix

Table 10 Test data of 1.5 g coffee minipack from machine-1

	Amount of coffee filled in each minipack (g)						
	Pack-1	Pack-2	Pack-3	Pack-4	Pack-5	Pack-6	Pack-7
Batch-1	1.58	1.54	1.52	1.50	1.59	1.56	1.48
Batch-2	1.53	1.61	1.52	1.52	1.57	1.58	1.56
Batch-3	1.62	1.51	1.59	1.58	1.57	1.61	1.66
Batch-4	1.63	1.61	1.68	1.64	1.72	1.69	1.70
Batch-5	1.74	1.72	1.71	1.68	1.82	1.74	1.73
Batch-6	1.80	1.82	1.80	1.86	1.82	1.82	1.78
Batch-7	1.69	1.85	1.81	1.86	1.70	1.86	1.86
Batch-8	1.78	1.86	1.80	1.89	1.89	1.90	1.89
Batch-9	1.94	1.87	1.96	1.89	1.99	2.00	1.99
Batch-10	2.03	2.01	2.00	1.97	2.02	2.08	1.97
Batch-11	2.01	2.00	2.07	2.02	2.14	2.05	2.05
Batch-12	2.08	2.05	2.12	2.05	2.03	2.09	2.03
Batch-13	2.12	2.11	2.12	2.12	2.08	2.24	2.29
Batch-14	2.24	2.24	2.32	2.23	2.16	2.27	2.25
Batch-15	2.29	2.27	2.26	2.20	2.23	2.24	2.32

Table 11 Test data of 1.5 g coffee minipack from machine-2

	Amount of coffee filled in each minipack (g)						
	Pack-1	Pack-2	Pack-3	Pack-4	Pack-5	Pack-6	Pack-7
Batch-1	1.60	1.53	1.50	1.46	1.58	1.63	1.70
Batch-2	1.63	1.44	1.58	1.57	1.60	1.57	1.58
Batch-3	1.59	1.50	1.50	1.55	1.58	1.50	1.57
Batch-4	1.51	1.53	1.51	1.53	1.57	1.50	1.51
Batch-5	1.54	1.55	1.56	1.58	1.55	1.60	1.54
Batch-6	1.55	1.55	1.58	1.55	1.57	1.55	1.50
Batch-7	1.53	1.54	1.55	1.50	1.54	1.53	1.57
Batch-8	1.60	1.51	1.57	1.58	1.56	1.58	1.55
Batch-9	1.53	1.56	1.57	1.53	1.51	1.60	1.51
Batch-10	1.57	1.56	1.55	1.58	1.50	1.53	1.49
Batch-11	1.55	1.61	1.47	1.54	1.59	1.52	1.59
Batch-12	1.57	1.58	1.51	1.56	1.54	1.57	1.44
Batch-13	1.57	1.50	1.53	1.54	1.50	1.53	1.53
Batch-14	1.61	1.57	1.53	1.59	1.56	1.69	1.53
Batch-15	1.61	1.56	1.51	1.53	1.59	1.55	1.56

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Alyami R, Alhajjaj J, Alnajrani B, Elaalami I, Alqahtani A, Aldhafferi N, Owolabi TO, Olatunji SO. Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines. In: 2017 international conference on informatics, health and technology, ICIHT 2017. 2017. https://doi.org/10.1109/ICIHT.2017.7899011.
2. Caban JJ, Bagci U, Mehari A, Alam S, Fontana JR, Kato GJ, Mollura DJ. Characterizing non-linear dependencies among pairs of clinical variables and imaging data. Proc Annu Int Conf IEEE Eng Med Biol Soc (EMBS). 2012. https://doi.org/10.1109/EMBC.2012.6346521.
3. Cheng C-S. A neural network approach for the analysis of control chart patterns. Int J Prod Res. 1997;35(3):667–97. https://doi.org/10.1080/002075497195650.
4. Cheng CS, Cheng HP, Huang KK. A support vector machine-based pattern recognizer using selected features for control chart patterns analysis. In: IEEM 2009—IEEE international

conference on industrial engineering and engineering management. 2009. p. 419–423. https://doi.org/10.1109/IEEM.2009.5373318.

5. Cuentas S, García E, Peñ Abaena-Niebles R. An SVM-GA based monitoring system for pattern recognition of autocorrelated processes. Soft Comput. 2022;2022:1–20. https://doi.org/10.1007/S00500-022-06955-7.

6. Fu J, Lee S. A multi-class SVM classification system based on learning methods from indistinguishable chinese official documents. Expert Syst Appl. 2012;39(3):3127–34. https://doi.org/10.1016/J.ESWA.2011.08.176.

7. Gierlichs B, Batina L, Tuyls P, Preneel B. Mutual information analysis. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 5154 LNCS. 2008. p. 426–442. https://doi.org/10.1007/978-3-540-85053-3_27

8. Guh RS, Hsieh YC. A neural network based model for abnormal pattern recognition of control charts. Comput Ind Eng. 1999;36(1):97–108. https://doi.org/10.1016/S0360-8352(99)00004-2.

9. la Gutiérrez HD, Pham DT. Identification of patterns in control charts for processes with statistically correlated noise. Int J Prod Res. 2017;56(4):1504–20. https://doi.org/10.1080/00207543.2017.1360530.

10. Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw. 2002;13(2):415–25. https://doi.org/10.1109/72.991427.

11. Huang XW, Emura T. Model diagnostic procedures for copula-based Markov chain models for statistical process control. Commun Stat Stimul Comput. 2019;50(8):2345–67. https://doi.org/10.1080/03610918.2019.1602647.

12. Kinney J, National Academy Sciences, et al. Equitability, mutual information, and the maximal information coefficient. Natl Acad Sci. 2014. https://doi.org/10.1073/pnas.1309933111.

13. Lin SY, Guh RS, Shiue YR. Effective recognition of control chart patterns in autocorrelated data using a support vector machine based approach. Comput Ind Eng. 2011;61(4):1123–34. https://doi.org/10.1016/J.CIE.2011.06.025.

14. Liu H-M, Yang D, Liu Z-F, Hu S-Z, Yan S-H, He X-W. Density distribution of gene expression profiles and evaluation of using maximal information coefficient to identify differentially expressed genes. PLoS ONE. 2019;14(7): e0219551. https://doi.org/10.1371/JOURNAL.PONE.0219551.

15. Mohamed A. Comparative study of four supervised machine learning techniques for classification. Int J Appl Sci Technol. 2017;7(1). https://www.academia.edu/download/54482697/2.pdf.

16. Morelli MS, Greco A, Valenza G, Giannoni A, Emdin M, Scilingo EP, Vanello N. Analysis of generic coupling between EEG activity and PETCO2 in free breathing and breath-hold tasks using maximal information coefficient (MIC). Sci Rep. 2018;8(1):1–9. https://doi.org/10.1038/s41598-018-22573-6.

17. Panos B, Kleint L, Voloshynovskiy S. Exploring mutual information between IRIS spectral lines. I. Correlations between spectral lines during solar flares and within the quiet Sun. Astrophys J. 2021;912(2):121. https://doi.org/10.3847/1538-4357/ABF11B.

18. Pontil M, Verri A. Properties of support vector machines. Neural Comput. 1998;10(4):955–74. https://doi.org/10.1162/089976698300017575.

19. Ranaee V, Ebrahimzadeh A, Ghaderi R. Application of the PSO–SVM model for recognition of control chart patterns. ISA Trans. 2010;49(4):577–86. https://doi.org/10.1016/J.ISATRA.2010.06.005.

20. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. Science. 2011;334(6062):1518–24. https://doi.org/10.1126/SCIENCE.1205438.

21. Reshef D, Reshef Y, Mitzenmacher M, Sabeti P. Equitability analysis of the maximal information coefficient, with comparisons. 2013. https://arxiv.org/abs/1301.6314v2.

22. Shao F, Li K, Xu X. Railway accidents analysis based on the improved algorithm of the maximal information coefficient. Intell Data Anal. 2016;20(3):597–613. https://doi.org/10.3233/IDA-160822.

23. Sharmin S, Shoyaib M, Ali AA, Khan MAH, Chae O. Simultaneous feature selection and discretization based on mutual information. Pattern Recogn. 2019;91:162–74. https://doi.org/10.1016/J.PATCOG.2019.02.016.

24. Suzuki T, Sugiyama M, Kanamori T, Sese J. Mutual information estimation reveals global associations between stimuli and biological processes. BMC Bioinform. 2009;10(1):1–12. https://doi.org/10.1186/1471-2105-10-S1-S52.

25. Tang D, Wang M, Zheng W, Wang H. RapidMic: rapid computation of the maximal information coefficient. Evolut Bioinform. 2014. https://doi.org/10.4137/EBO.S13121.

26. Vance LC. A bibliography of statistical quality control chart techniques. J Qual Technol. 2018;15(2):59–62. https://doi.org/10.1080/00224065.1983.11978845.

27. Wang R, Li H, Chen M, Dai Z, Zhu M. MIC-KMeans: a maximum information coefficient based high-dimensional clustering algorithm. Adv Intell Syst Comput. 2019;764:208–18. https://doi.org/10.1007/978-3-319-91189-2_21/FIGURES/6.

28. Xanthopoulos P, Razzaghi T. A weighted support vector machine method for control chart pattern recognition. Comput Ind Eng. 2014;70(1):134–49. https://doi.org/10.1016/J.CIE.2014.01.014.

29. Xu Z, Xuan J, Liu J, Cui X. MICHAC: defect prediction via feature selection based on maximal information coefficient with hierarchical agglomerative clustering. In: 2016 IEEE 23rd international conference on software analysis, evolution, and reengineering, SANER 2016. 2016. p. 370–381. https://doi.org/10.1109/SANER.2016.34.

30. Yang JH, Yang MS. A control chart pattern recognition system using a statistical correlation coefficient method. Comput Ind Eng. 2005;48(2):205–21. https://doi.org/10.1016/J.CIE.2005.01.008.

31. Yang M-S, Yang J-H. A fuzzy-soft learning vector quantization for control chart pattern recognition. Int J Prod Res. 2002;40(12):2721–31. https://doi.org/10.1080/00207540210137639.

32. Yang X, Yu Q, He L, Guo T. The one-against-all partition based binary tree support vector machine algorithms for multi-class classification. Neurocomputing. 2013;113:1–7. https://doi.org/10.1016/J.NEUCOM.2012.12.048.

33. Yeganeh A, Shadman A, Abbasi SA. Enhancing the detection ability of control charts in profile monitoring by adding RBF ensemble model. Neural Comput Appl. 2022. https://doi.org/10.1007/S00521-022-06962-7/FIGURES/9.

34. Zhang M, Yuan Y, Wang R, Cheng W. Recognition of mixture control chart patterns based on fusion feature reduction and fireworks algorithm-optimized MSVM. Pattern Anal Appl. 2018;23(1):15–26. https://doi.org/10.1007/S10044-018-0748-6.

35. Zhao C, Wang C, Hua L, Liu X, Zhang Y, Hu H. Recognition of control chart pattern using improved supervised locally linear embedding and support vector machine. Proc Eng. 2017;174:281–8. https://doi.org/10.1016/J.PROENG.2017.01.138.

36. Zhao X, Deng W, Shi Y. Feature selection with attributes clustering by maximal information coefficient. Proc Comput Sci. 2013;17:70–9. https://doi.org/10.1016/J.PROCS.2013.05.011.

37. Zhou X, Jiang P, Wang X. Recognition of control chart patterns using fuzzy SVM with a hybrid kernel function. J Intell Manuf. 2015;29(1):51–67. https://doi.org/10.1007/S10845-015-1089-6.