Diagnosing Out-of-Control Signals of Multivariate Control Chart based on Variable Length PSO-SVM

Duo XU[1], Zeshui XU[2]*, Shuixia CHEN[2]

[1] Business School, Stevens Institute of Technology, Hoboken, NJ 07030, USA
nicolexduo@163.com

[2] Business School, Sichuan University, No. 24 South Section 1 Yihuan Road, Chengdu 610064, China
xuzeshui@263.net (**Corresponding author*), chen_shui_xia@163.com

Abstract: Multivariate statistical process control is an essential procedure employed to deliver quality products in modern manufacturing and service industries. Multivariate control charts are an extensively used tool to determine whether a process is performing as intended. Once the control chart detects an abnormal process variable, one difficulty encountered is to interpret the source(s) of the out-of-control signal. Therefore, a novel approach for diagnosing the out-of-control signals in the multivariate process is developed in this paper. The proposed methodology uses the optimized support vector machines (SVM) based on variable-length particle swarm optimization to recognize subclasses of multivariate abnormal process patterns and to identify the responsible variable(s) for their occurrence. Based on a simulation experiment, the proposed approach verifies its capability in accurate classification of the source(s) of out-of-control signal and outperforms the conventional multivariate control scheme based on SVM.

Keywords: Hotelling's T^2 control chart, Support vector machine, Variable-length particle swarm optimization, Parameter optimization.

1. Introduction

Planning for quality, one of the core management systems, can aid manufacturers to gain a competitive edge in the market (Kim et al., 2012). A computer-aided industrial setting contributes to a growing interest in simultaneous monitoring of the correlated process features. As a result, multivariate statistical process control (MSPC) methodologies have been broadly applied to secure safe plants and to improve product quality. The multivariate control chart is one of the most used MSPC techniques to assure the industrial process in a "state of statistical control" and to detect any change that may affect product quality. The quality characteristic measured on a numerical scale is referred to as a process variable. Generally, the goal of control chart analysis lies in detecting shifts in values of process variables and recognizing abnormal process patterns. For the purpose of fault detection, control limits are determined or estimated based on historical reference samples, and the observed values of process variables are then monitored against control limits (Adegoke et al., 2019). A number of studies also aim to recognize abnormal control chart patterns (CCPs) to match these patterns with their assignable causes (Addeh et al., 2018) through feature extraction, feature selection and classification. As a consequence, practitioners are able to distinguish the unpredictable, randomly fluctuated patterns from the inherent, recognizable ones, and to trigger corresponding remedial actions.

Based on solid statistical principles, traditional process control charts are classified into three broad categories, namely: 1) Shewhart chart, 2) Cumulative Sum (CUSUM) chart, and 3) Exponentially Weighted Moving Average (EWMA) chart. The multivariate extension was pioneered by Hotelling (1947), namely T^2. A multivariate T^2 control chart plots the process parameter along with the upper control limit (UCL) and the lower control limit (LCL). Therefore, exceeding any threshold will detect a fault. Consecutive abnormal events may be required to trigger the alarm to avoid the false interpretation. This Shewhart-like chart is memoryless and neglects previous process information, thus it is efficient for detecting large and general shifts in the process parameter of interest (Riaz et al., 2020). Inversely, the multivariate CUSUM (MCUSUM) (Crosier, 1988) and the multivariate EWMA (MEWMA) (Lowry et al., 1992) were developed using both past and current information. They are more sensitive for small-to-moderate process shifts and can function better for individual-observation monitoring (Adegoke et al., 2019). However, these conventional charting techniques do not have competence to identify the sources of out-of-control signals in the manufacturing process. To compensate for this limitation, a substantial body of researches has integrated data driven approaches into control charts to monitor variations. Neural networks (NNs) and support vector machines (SVMs) are the mainstream machine learning algorithms used to supplement an automatic detection for process abnormalities (Weese et al., 2016).

https://doi.org/10.24846/v30i3y202101

NNs have great in-time noise tolerance and often generate promising results due to better estimation capacity and adaptive learning property. Some earlier studies (Chen & Wang, 2004; Niaki & Abbasi, 2005) developed artificial neural network (ANN)-based models for mean or variance shift detection. Ahmadzadeh (2018) used ANN to detect the change point in the process mean vector, and a simulation experiment showed a better performance of ANN over conventional MEWMA on average with a reduced size of searching window. To detect variance changes in the multivariate process, Shao & Lin (2019) proposed a time-delay neural network (TDNN)-based model that accurately classified the contributors of process fault. However, an abnormal variance shift may cause out-of-control signals in both mean and variance charts. Thus, it is reasonable to integrate these process variables into one schema and to simultaneously monitor the variations. Moreover, different NNs structures have broadly been used in recent studies for the diagnostic purpose, e.g. convolutional NN (CNN) (Zan et al., 2020) and evolutionary ANN (Yeganeh et al., 2021). But nevertheless, ANN-based models may suffer from a few drawbacks such as the overfitting problem, unstable results and difficulties in control parameters decisions (e.g., the number of hidden layers/nodes, the momentum term, the learning rate) (Lu et al., 2011). In addition, a classic ANN classifier is trained based on the empirical risk minimization (ERM) principle which may result in long training time, poor generalization capability and the tendency to fall into a local minimum. From these perspectives, ANNs may be less appropriate for quality diagnostics in dynamic industrial processes.

Compared to NNs, SVM achieves greater generalization capability (Cuentas et al., 2017; Yu et al., 2020). Introduced by Cortes & Vapnik (1995), SVM is a versatile classification technique especially when it is used for small sized, nonlinear and high-dimensional data. Various application scenarios have verified its effectiveness for classification and recognition tasks such as industrial fault diagnosis (Gani et al., 2011), system risk identification (Huang et al., 2021), ecological monitoring (Huang et al., 2020) and disease detection (Tuba & Tuba, 2019; Yang & Xu, 2019). Basically, SVM constructs a hyperplane in a high-dimensional space based on the structural risk minimization (SRM) principle. Such construction procures a well trade-off between quality and complexity of the approximation (Zhang, 2005). As a result, SVM can avoid the overfitting problem and reduce the difficulty in model design. Unlike the conventional classification problem, one class SVM (OC-SVM), also referred to as support vector data description (SVDD), specifies only one class of the training data to reduce biases. It was initially applied in the kernel distance-based control chart (*k*-chart) proposed by Sun & Tsung (2003) requiring only in-control preliminary samples. Then, Camci et al. (2008) developed a robust *k*-chart considering limited out-of-control data and the approach obtained very reasonable results for all types of abnormal process states. However, a failure to manage the false alarm rate (i.e., Type I error rate) when deriving the control limit restricts the use of control chart in process monitoring. Fortunately, by changing the width of control limits can effectively deal with this problem. Kim & Kim (2018) proposed a novel chart based on optimal false alarm controlled SVDD. One can precisely control the proportion of out-of-control observations, thus Type I error rate, by adjusting an assigned constant value.

When employing SVM for classification, decisions of both the kernel function type (or the kernel parameter) and the penalty parameter of SVM have significant influences on the generalization ability. These parameters directly affect correct classification rate for the labelled output, thus SVM-based applications always require a search for optimal settings. Table 1 lists some recent research that adopted different optimization algorithms for SVM parameter selection in the field of industrial fault diagnosis. Studies have indicated that genetic algorithm (GA) and particle swarm optimization (PSO) are the most extensively used algorithms for SVM parameter tuning. Moreover, a range of variants based on GA and PSO have been developed attempting to overcome common limitations of the standard version. In terms of kernel function type, radial basis function (RBF) is the most popular SVM kernel in academic research.

In this paper, a framework of mean shift detection is proposed for identifying the source(s) of out-of-control signals in multivariate manufacturing processes. Due to the robust classification power of SVM and the global search capability of PSO, an optimized SVM-based approach is proposed. A multivariate Hotelling's T^2 control chart is

Table 1. Several researches for industrial fault diagnosis based on optimized SVMs

Literature	Proposed model	Optimization algorithm	Kernel function
Chou et al. (2014)	SVM	fast messy genetic algorithm (fmGA)	RBF
Yang et al. (2015)	SVM	artificial colony bee algorithm (BA)	RBF
Zhang et al. (2015)	SVM	barebones particle swarm optimization & differential evolution (BBDE)	RBF
Zhou et al. (2018)	Fuzzy SVM	genetic algorithm (GA)	RBF & Polynomial
Wang et al. (2019)	SVM	saturated and mix-delayed particle swarm optimization (SMDPSO)	RBF & Polynomial
Zeng et al. (2019)	least square SVM	grey wolf optimizer (GWO)	RBF
Zhang et al. (2020)	SVM	fireworks algorithm (FWA)	RBF

utilized to detect any abnormal process variable, and SVM principles are then adopted to recognize the subclasses of abnormal CCPs and to identify the contributor(s) for the occurrence of the process fault. More importantly, an advanced version of PSO, namely variable-length PSO (VLPSO) proposed by Tran et al. (2018), is employed for parameter selection. There are several reasons for basing the proposed model on VLPSOSVM:

1. To overcome the limitation of conventional control charts for MSPC, the SVM principle with parameter optimization has been broadly integrated due to its great generalization capability and recognition performance;

2. Various optimization algorithms are available and comparison analyses have been conducted in a large number of research;

3. Most current PSO-based approaches adopt an inflexible fix-length representation, limiting the performance of PSO for SVM parameter selection. To deal with high-dimensional data in manufacturing systems, a variable-length representation is employed. It narrows the search space and is able to locate the optimal output more efficiently;

4. By applying a length changing mechanism, the PSO searching procedure can avoid being trapped at the local optima.

The remainder of this paper is organized as follows: Section 2 provides the methodologies for the control chart, SVM and VLPSO algorithms used in this paper. Section 3 introduces the proposed VLPSOSVM model. A simulation experiment is described and a comparison analysis is implemented in Section 4, and Section 5 concludes the paper and presents some ideas on future opportunities.

2. Methodology

This section provides a brief review on the conventional Hotelling's T^2 control chart, SVM learning algorithm used for abnormal detection in the multivariate industrial process and VLPSO optimization algorithm used for SVM parameter selection.

2.1 Hotelling's T^2 Control Chart

The control chart plots process means or variance over time and prompts an out-of-control signal if they fall outside the control limits. The most frequently used multivariate control chart for monitoring process mean is the multivariate Hotelling's T^2 chart, which measures the Mahalanobis distance between the center of reference and the moving window dataset. Suppose that p related quality characteristics are observed in the process, denoted by a random vector $X=\left[X_1,X_2,...,X_p\right]^T$. It is assumed that X follows a joint distribution $X\sim N_p(\mu,\Sigma)$, and there are m independent subgroups each with a sample size of n (Nidsunkid et al., 2016). When in-control mean vector μ and covariance matrix Σ are specified to be known, the test statistic using first m preliminary samples can be computed by equation (1)

$$\chi_i^2 = n\left(\overline{X_i}-\mu\right)^T \Sigma^{-1}\left(\overline{X_i}-\mu\right) \quad (1)$$

for $i=1, 2, \ldots, m$, where $\overline{X_i}$ denotes sample mean vector for i^{th} rational subgroup. The χ_i^2 statistic follows a chi-squared distribution with p degrees of freedom when the process is in control. Therefore, the process mean has a UCL given by $\chi_{\alpha,p}^2$, representing the upper α percentile of the chi-squared distribution with p degrees of freedom. Generally, the chi-square limit is employed as an approximation, especially for $m > 100$ (Montgomery, 2020).

Conversely, in the case of unknown μ and Σ, m preliminary samples are used to estimate the pooled vector of sample means $\overline{\overline{X_l}}$ and the pooled sample covariance matrix S. The T^2 statistic plotted on the control chart can be computed by equation (2):

$$T_i^2 = n\left(\overline{X_i} - \overline{\overline{X_l}}\right)^T S^{-1}\left(\overline{X_i} - \overline{\overline{X_l}}\right) \quad (2)$$

for $i = 1, 2, \ldots, m$. To detect any mean shift associated with future observations with unknown μ and Σ, UCL is established by equation (3):

$$L_u = p(m+1)(n-1)/(mn-m-p+1)^* F_{\alpha,p,mn-m-p+1} \quad (3)$$

for m subgroups each with a sample size of n. $F_{\alpha,p,mn-m-p+1}$ represents the upper α percentile of an F-distribution with p and $(mn - m - p + 1)$ degrees of freedom. The LCL is generally set to 0. Since $\chi_i^2 \geq 0$ and $T_i^2 \geq 0$, only UCL is used to test whether the process remains in a state of statistical control in both two cases, i.e., the process is detected out-of-control if $\chi_i^2 \geq \chi_{\alpha,p}^2$ or $T_i^2 \geq L_u$.

2.2 Support Vector Machines

SVM is a powerful classification technique particularly for small- and medium-sized datasets. The basic principle of SVM is to transform the input data into a high-dimensional feature space using a mapping function φ and to find the optimal hyperplane that separates the labelled data. In a p-dimensional input space, a hyperplane is a flat $(p - 1)$ dimensional subspace that separates different classes of the observations. When the distance between the hyperplane and the nearest data points (referred to as support vectors) of different classes is maximized, SVM achieves higher classification accuracy. To clarify the problem, let $\{x_i, y_i\}_{i=1}^n$, $x_i \in \mathbb{R}^p$, $y_i \in \{-1, 1\}$ be the sample training set D with n sample observations. The quadratic optimization problem can then be formulated as in equation (4):

$$\begin{aligned} &\min_{w,b,\xi} 1/2\, w^T w + C\sum_{i=1}^{n} \xi_i \\ &s.t. y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i \end{aligned} \quad (4)$$

where ξ_i is non-negative for $i = 1, 2, \ldots, n$. In (4), w is the hyperplane coefficient vector, C is the penalty parameter and b is the bias term. C controls the trade-off between the confidence interval and classification errors. Therefore, it is used to reduce the misclassification error and to avoid the occurrence of over-fitting. A non-negative slack variable ξ_i is introduced in the soft margin objective to obtain a more flexible buffer of boundary.

Such convex quadratic optimization with linear constraints is known as a quadratic programming (QP) problem. However, it can be confusing when there are more features than the training instances. In this case, fortunately, the QP problem can be transformed into an unconstrained dual problem using the Lagrange multiplier given by equation (5):

$$\begin{aligned} L(w,b,\alpha) &= 1/2\, w^T w + \sum_{i=1}^{n} \alpha_i \\ &- \sum_{i=1}^{n} \alpha_i y_i \left(w^T \varphi(x_i) + b\right) \end{aligned} \quad (5)$$

where $\alpha_i \geq 0$ for $i = 1, 2, \ldots, n$. An optimal solution can be guaranteed at the stationary point $(\hat{w}, \hat{b}, \hat{\alpha})$ when $\frac{\partial}{\partial b} L(w,b,\alpha)$ equals 0. Therefore, the dual form of the SVM problem can be obtained using equation (6):

$$\min_{\alpha} 1/2 \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \varphi(x_i)^T \varphi(x_j) - \sum_{i=1}^{n} \alpha_i \quad (6)$$

where $\alpha_i \geq 0$ for $i = 1, 2, \ldots, n$. The dual problem requires the transformation φ applied to all the training instances. Introducing a kernel function into the transformation process can be more computationally efficient, by computing $\varphi(x_i)^T \varphi(x_j)$ based only on the original vectors xi and xj. Additionally, for non-linearly separable problem, the kernel function is applied to map the training instance to a high-dimensional feature space, where a linear separating hyperplane can then be obtained with maximal margin. The most commonly used kernel functions are linear, polynomial, RBF and sigmoid, and they are comparable under low dimension. In this paper, the Gaussian RBF kernel is used, given by equation (7):

$$K(x_i, x_j) = exp\{-\gamma \| x_i - x_j \|^2\} \quad (7)$$

where γ is the kernel parameter. RBF is appropriate and sufficient in many complex cases, and it has fewer parameters that need to be determined.

2.3 Variable-length Particle Swarm Optimization

PSO is a population-based evolutionary optimization algorithm developed by Kennedy & Eberhart (1995), well-known for its global

search ability. Each particle in the swarm represents a candidate solution with two attribute vectors, position and velocity, and it is randomly distributed in the solution search space. However, the standard PSO representations are of a fixed length, i.e. the length of each particle in the swarm equals the initial number of features (dimensionality). As a result, the fix-length representation requires much computer memory and long computing time when applied on high-dimensional data. To overcome this limitation, Tran et al. (2018) developed a more effective and flexible PSO-based algorithm with a variable-length representation that reduces the search space. By allowing particles with different lengths to learn from each other, the swarm can maintain an appropriate level of diversity, thus alleviating the premature convergence issue. The following content introduces the representation of VLPSO and the length changing procedure adopted to obtain the optimal output more efficiently.

2.3.1 Variable-length Representation

A VLPSO representation is still vector-based as standard PSO and each particle with a different length L has three attributes including position, velocity and exemplar. An exemplar of the particle for a dimension d is assigned by VLPSO, and particles in each dimension then move towards the best position that the exemplar has explored so far (*pbest*). The position and velocity of the i^{th} particle are updated through iteration steps using equations (8) and (9), respectively,

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{8}$$

$$v_{id}^{t+1} = \omega * v_{id}^{t} + c * r_{id} * \left(p_{exmplr(id)d}^{t} - x_{id}^{t} \right) \tag{9}$$

where x_{id}^{t} and v_{id}^{t} denote the position and the moving speed of the i^{th} particle in dimension d at time t. In (9), $exmplr(id)$ returns the exemplar of the i^{th} particle in dimension d, and $p_{exmplr(id)d}^{t}$ represents its best position in the swarm at time t. ω is the inertia weight and a reduced ω will result in stronger convergence and greater ability to locate the global optima. c is an acceleration constant and $r_{id} \in [0,1]$ is a random value assigned with a uniform distribution.

In addition, the learning probability *Pc* and the renew exemplar count are recorded in the variable-length representation. VLPSO enables particles with different L to communicate with each other by allowing each particle to learn either from *pbest* of its own or from *pbest* of the assigned exemplar. This decision is based on an adaptive *Pc* given by equation (10):

$$P_c = 0.05 + 0.45 \frac{exp^{\frac{10(rank(i)-1)}{S-1}}}{exp^{10} - 1} \tag{10}$$

where S denotes the population size and rank(i) denotes the fitness rank of i^{th} particle. Tran et al. (2018) suggested that the particles learn from the one with better fitness, i.e., smaller *Pc.* Therefore, if a random number drawn from a uniform distribution is greater than *Pc* of the i^{th} particle, the i^{th} particle will be assigned as an exemplar for d. Otherwise, two different particles both with $L > d$ randomly picked and the one scoring better fitness is assigned as the exemplar for d. If none of the particles in d fails to meet the condition $L > d$, d learns from its own *pbest* instead. Moreover, when *pbest* stops improving for α iterations, exemplars are renewed and the number of iterations *pbest* remains unchanged is recorded as the renew exemplar count.

2.3.2 Length Changing Procedure

In addition, a length changing mechanism is employed to help PSO jump out of the local optima when the best position of the population (*gbest*) does not change for β iterations. To reduce the search space, the entire swarm is divided into a predefined number of divisions. Particles of the same length are in the same division and they can still represent different feature sets. The length of particle can be obtained by equation (11):

$$ParLen_V = MaxLen * V / (number\ of\ divisions) \tag{11}$$

where *MaxLen* represents the dimensionality of the current swarm. V refers to the division where the particle is in, i.e. V = 3 when the particle is contained in division 3. The mechanism aims to scale the search process into the best division with the highest average fitness by resizing particles in other divisions. The particle length of best division (*BestLen*) is kept unchanged. If the *ParLenV* > *BestLen*, the *ParLenV* will be automatically shortened using *BestLen* as the new *MaxLen* of the swarm. Otherwise, more dimensions will be appended at the end of the representation and then randomly initialized.

3. Proposed Model

In this paper, it is assumed that the mean vector and covariance matrix of the process data can be accurately estimated given enough data. Since the proposed model only serves for mean shift detection and shifted variable identification, it is also assumed that abnormal process variables lead to changes only in the mean vector. In addition, only abrupt conditions are considered here, indicating that a process remains in-control until a mean shift occurs and the process data before and after the change point are all independently and identically distributed. An out-of-control variable will not become in-control even if faults are detected and removed.

In terms of the model design, the schema is divided into two parts. The first comprises a multivariate T^2 control chart to detect an abnormal situation. When the process observation continuously falls outside the control limit, i.e., $\chi_i^2 > \chi_{\alpha,p}^2$ or $T_i^2 > L_u$, the control chart generates an out-of-control alarm. The group of observations that triggers the alarm is represented with a binary code and then collected to be diagnosed. In a p-dimensional manufacturing process, the process mean vector has 2^p states, either in-control or out-of-control. In other words, once the test statistic exceeds the UCL, there exist $2^p - 1$ abnormal states that need to be distinguished from the total possible states.

Subsequently, a pattern recognition problem is constructed based on the VLPSOSVM classifier for the identification of shifted variables. The SVM algorithm adopts Gaussian RBF kernel to distinguish types of abnormal process patterns. The parameters that need to be selected with caution are the penalty parameter C of SVM and the kernel parameter γ. An inappropriate C may lead to the overfitting problem and γ determines the structure of the high-dimensional feature space. In this study, VLPSO is employed to tune these parameters of SVM following the process shown in Figure 1, in order to achieve a better recognition performance.

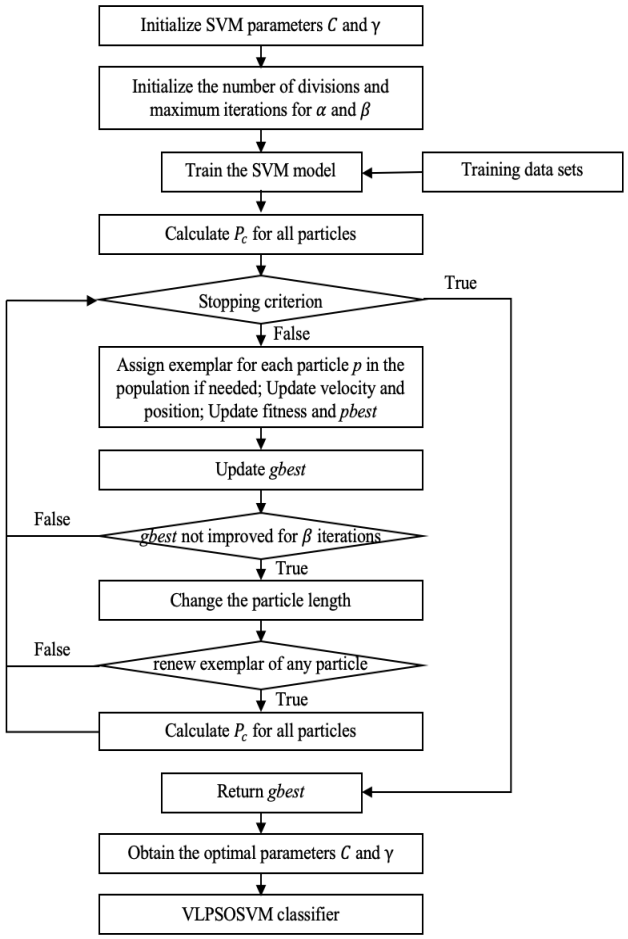

Figure 1. The architecture of SVM parameter optimization using VLPSO

A particle in the swarm consists of C and γ. Initially, velocity and position of a group of particles are given randomly. The ranking of the i^{th} particle can be quantified by feeding its position to the fitness function. The best searching position for the i^{th} particle is denoted as $P_i = \left(p_{iC}, p_{i\gamma}\right)$ (*pbest*), and that for the entire swarm as $P_g = \left(p_{gC}, p_{g\gamma}\right)$ (*gbest*). The main steps of VLPSOSVM are given as follows:

Step 1. Generate the training and the testing data sets to prepare for SVM modeling;

Step 2. Generate initial particles $\left(C, \gamma\right)$ and initial VLPSO parameters including population size, number of divisions, constant *c*, maximum iteration, threshold for selected feature, maximum iteration for renew exemplars (α) and for length changing (β);

Step 3. Train the SVM model with 500 training data sets generated. The parameter pairs $\left(C, \gamma\right)$ change along with the flying of variable-length particles;

Step 4. Calculate the learning probability *Pc* using equation (10) for all the particles;

Step 5. Check against stopping criterion. Repeat Steps 6 to 9 until the maximum iteration is satisfied;

Step 6. Each particle *p* in the population, if *p* has a lower *Pc* than a random value drawn from a uniform distribution, is assigned as the exemplar for *d*. Otherwise, the one scoring better fitness value of two random particles is selected as the exemplar. Once the exemplar for *d* is determined, update the velocity and position for *p* using equations (9) and (8). Update fitness value and *pbest*;

Step 7. If *pbest* has not been improved for α iterations, the exemplar for the particle *p* is renewed and the renew exemplar count (the number of iterations until renewed) is recorded. Update *gbest* for the population;

Step 8. If *gbest* has not been improved for β iterations, the length of particles is changed using equation (11) with new *MaxLen* of the swarm. Otherwise, the stopping criterion is checked against. Generally, β needs to be large enough for VLPSO to converge;

Step 9. If the exemplar is renewed for any particle in the swarm, then *Pc* is assessed for all the particles using equation (10). Otherwise, the stopping criterion is checked against;

Step 10. Once maximum iteration is satisfied, the procedure is finished and *gbest* is returned;

Step 11. Obtain the optimal parameter setting $\left(C, \gamma\right)$ and finalize the VLPSOSVM classifier for diagnosing abnormal patterns.

Once it is finalized, the VLPSOSVM model is then used to recognize the shifted variable(s) . Take a bivariate process (p = 2) for illustration with mean vector μ_0 and covariance matrix Σ. By representing a process variable that is out-of-control as 1 and that is in-control as 0, an in-control normal process state with two variables as (0 0) can be identified. When p = 2, there exist 3 out-of-control states, (1 0), (0 1) and (1 1), representing fault process patterns recognized in μ_1, μ_2 or in both, respectively. In other words, the process mean vectors shift following three patterns: (1) only the first of two variables shifts λ units of standard deviation σ_{11}, $\left(\mu_1 + \lambda\sigma_{11}, \mu_2\right)$; (2) only the second of two variables shifts λ units of σ_{22} and $\left(\mu_1, \mu_2 + \lambda\sigma_{22}\right)$; (3) both two variables shift λ units of σ_{11} and σ_{22}, respectively, $\left(\mu_1 + \lambda\sigma_{11}, \mu_2 + \lambda\sigma_{22}\right)$. When an overall out-of-control signal is prompted by the Hotelling's T^2 control chart in the detection process, the VLPSOSVM is then implemented to identify and locate the sources of mean shifts, i.e., μ_1, μ_2 or both. When designing the process, sufficient datasets are required to train SVM adequately and different shift magnitude λ in the process mean vector are considered. When the output value of the trained SVM is close to either zero or one, they are rounded down or up, respectively.

4. Simulation Experiment

The function of the proposed model lies in identifying the abnormal variable(s) in the multivariate manufacturing process by formulating a pattern recognition problem. The representation of the training sets has a strong influence on the performance of the VLPSOSVM recognizer. However, real-world manufacturing systems usually rely on well-equipped facilities to collect various data from sensors, thus data acquisition costs much if such equipment is absent. In this regard, simulation is a simplified but useful alternative for abnormal CCP investigation in the multivariate industrial process (Zorriassatine et al., 2003).

To illustrate how the proposed VLPSOSVM works for the pattern classification in the diagnostic process, a bivariate industrial case derived from a chemistry process (Montgomery, 2020) is provided. The case was also studied by Cheng et al. (2011) and Li et al. (2013). The only two input variables (X1 and X2) are target mean values of the process from a chemistry plant. Based on the first fifteenth process observation data, the reference mean vector and the reference covariance matrix were given as:

$$\overline{\overline{X}}_I = \begin{bmatrix} 10 \\ 10 \end{bmatrix} \text{ and } S = \begin{bmatrix} 0.7986 & 0.6973 \\ 0.6973 & 0.7343 \end{bmatrix}.$$

When training data, Monte Carlo simulation is used to generate sufficient data sets. In this bivariate example, 22 process patterns can be discovered with mean $\mu_1 = \mu_2 = 10$ and standard deviation $\sigma_{11} = \sqrt{0.7986} = 0.8936$ and $\sigma_{22} = \sqrt{0.7343} = 0.8569$. The four process patterns can be presented as (0 0) (both μ_1 and μ_2 are in-control), (0 1) (μ_1 is in-control while μ_2 is out-of-control), (1 0) (μ_1 is out-of-control while μ_2 is in-control), and (1 1) (both μ_1 and μ_2 are out-of-control). Therefore, a VLPSOSVM that has two variables (X1, X2) in input terminal and four nodes (0 0), (0 1), (1 0) and (1 1) in output terminal is designed. The relevant parameters selected following a VLPSO experiment are illustrated in Table 2. As a result, the optimal combinations of the penalty parameter and the kernel parameter (C, γ) of SVM classifier obtained are C = 1.46 and γ = 6.86. The whole process was realized by Python 3.7.

Table 2. VLPSO parameter setting

Parameters	Value
population size	300
maximum iteration	100
acceleration constant	1.5
inertia weight	$0.9\text{-}0.5 * \frac{current\ iteration}{max\ iteration}$
threshold for selected feature	0.6
max iteration for renew exemplars	7
number of divisions	8
max iteration for length changing	6

Using the generation function for multivariate normal random data sets in MATLAB, the training sets are firstly generated by shifting μ_1 and μ_2 1.0 standard deviation, i.e. λ = 1.0, following three abnormal patterns as $(\mu_1, \mu_2 + 1.0\sigma_{22})$, $(\mu_1 + 1.0\sigma_{11}, \mu_2)$ and $(\mu_1 + 1.0\sigma_{11}, \mu_2 + 1.0\sigma_{22})$ to verify the classification capability of SVM for an individual observation, i.e., n = 1. In addition, to evaluate the model efficiency for various shift magnitudes, λ = 1.5, 2.0, 2.5, 3.0 is also considered for each fault mode. The T^2 statistic is also applied to detect the overall out-of-control signal for each sample vector with type I error α = 0.05. At the end, 500 examples are obtained for each shift size in each out-of-control state and are then used as the testing set for VLPSOSVM.

The following content compares the classification efficiency of the conventional SVM using gradient descent search, an optimized SVM (OSVM) (Li et al., 2013) using GA and the proposed VLPSOSVM using VLPSO for parameter optimization. The optimal parameter settings for OSVM obtained using GA are C = 1.51 and γ = 1.08.

Based on the same simulation experiment, Table 3 compares the correct classification count for four possible process states, (0 0), (0 1), (1 0) and (1 1), in the bivariate example using OSVM and the proposed VLPSOSVM. The shift magnitude represents the case when process mean vector shifts 1.0, 1.5, 2.0, 2.5 and 3.0 standard deviation(s) from the average under different states. The classification results of VLPSOSVM (the first row) and OSVM (the second row) are contrasted for each shift magnitude. In most cases, especially for large mean shifts, VLPSOSVM realizes a better recognition performance (shown as bold italics).

Table 3. The results of classification for the bivariate process states using OSVM and VLPSOSVM

Shift magnitude	Correct classification count			
	(0 0)	(0 1)	(1 0)	(1 1)
(0, 0)	**500***	0	0	0
	500	0	0	0
(1, 0)	0	**478**	0	32
	0	**480**	1	19
(0, 1)	0	0	**473**	27
	0	3	**484**	13
(1, 1)	0	17	18	**465**
	0	8	10	**482**
(1.5, 0)	0	***492***	0	8
	0	***491***	0	9

(0, 1.5)	0	0	***493***	7
	0	0	***488***	12
(1.5, 1.5)	0	16	14	**470**
	0	5	5	**490**
(2, 0)	0	***497***	0	3
	0	***490***	1	9
(0, 2)	0	0	***491***	9
	0	1	***488***	11
(2, 2)	0	8	8	**484**
	0	7	8	**485**
(2.5, 0)	0	***497***	0	3
	0	***490***	0	10
(0, 2.5)	0	0	***499***	1
	0	0	***492***	8
(2.5, 2.5)	0	1	1	***498***
	0	5	4	***491***
(3, 0)	0	***500***	0	0
	0	***494***	0	6
(0, 3)	0	0	***500***	0
	0	0	***495***	5
(3, 3)	0	0	1	***499***
	0	7	6	***487***

*Bold numbers denote the number of times (out of 500) the correct diagnosis has been observed

To better illustrate the efficiency of the proposed VLPSOSVM schema in recognizing the abnormal variable(s) from all possible process states, the classification accuracy rates of the conventional SVM, OSVM, and VLPSOSVM approaches trained and tested using the same simulation datasets are compared. The results are shown in Table 4. More straightforward, the average classification accuracy rates (correct classification counts out of 500) of SVM, OSVM and VLPSOSVM for total 16 shift patterns are 94.23, 97.96 and 98.10, respectively.

VLPSOSVM outperforms standard SVM in diagnosing variations for nearly all shift magnitudes (shown as bolds) except when both two process mean vectors shift 1.5 standard deviations. Compared to OSVM, VLPSOSVM achieves better average recognition accuracy in most cases (10 in 15, shown as bold italics) especially for moderate to large mean shifts, indicating the importance of appropriate choices of the penalty parameter C and the kernel parameter γ. Concretely, VLPSO can strengthen the pattern feature selection and assist the SVM based approach to diagnose more efficiently the out-of-control signals in manufacturing processes.

Table 4. Correct classification rates of SVM, OSVM and VLPSOSVM

Shift magnitude	Correct classification percentage (%)		
	SVM	OSVM	VLPSOSVM
(0, 0)	96.0	100	**100**
(1, 0)	93.6	96.6	**95.6**
(0, 1)	93.6	96.8	***97.0***
(1, 1)	92.4	96.4	**93.0**
(1.5, 0)	92.2	98.2	***98.4***
(0, 1.5)	94.2	97.6	***98.6***
(1.5, 1.5)	95.0	98.0	94.0
(2, 0)	94.8	98.0	**99.4**
(0, 2)	94.0	97.6	***98.2***
(2, 2)	93.6	97.0	**96.8**
(2.5, 0)	96.0	98.0	***99.4***
(0, 2.5)	94.4	99.4	***99.8***
(2.5, 2.5)	92.0	99.2	***99.6***
(3, 0)	96.0	98.8	***100***
(0, 3)	95.6	99.0	***100***
(3, 3)	94.4	97.4	***99.8***
Averaged	94.23	97.96	98.10

Table 5. Correct classification rates of VLPSOSVM for different sample sizes

Shift magnitude	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
(0, 0)	100	100	100	100	100
(1, 0)	95.6	98.6	99.4	99.2	99.6
(0, 1)	97.0	94.2	97.2	97.6	100
(1, 1)	93.0	88.5	96.4	98.0	99.2
(1.5, 0)	98.4	99.0	100	100	100
(0, 1.5)	98.6	99.2	100	100	100
(1.5, 1.5)	94.0	97	99.4	100	100
(2, 0)	99.4	100	100	100	100
(0, 2)	98.2	99.8	100	100	100
(2, 2)	96.8	99.6	100	100	100
(2.5, 0)	99.4	99.4	100	100	100
(0, 2.5)	99.8	99.8	100	100	100
(2.5, 2.5)	99.6	99.6	100	100	100
(3, 0)	100	100	100	100	100
(0, 3)	100	100	100	100	100
(3, 3)	99.8	99.8	100	100	100
Averaged	98.10	98.40	99.49	99.67	99.92

In addition to individual observation monitoring ($n = 1$), the diagnostic accuracy of VLPSOSVM is investigated for sub-grouped sample data with a sample size $n = 2, 3, 4, 5$, assuming that all the relevant parameters remain the same as in the bivariate case mentioned above. Based on the results in Table 5, VLPSOSVM shows a stronger capability in detecting out-of-control variable(s) for larger sample sizes.

Moreover, the results for $n = 1, 3, 5$ are graphically illustrated in Figure 2. This figure plots the relationship between two target variables X1 and X2 in the bivariate process. Subplots (a) to (e) represent the mean vectors shifted $\lambda\sigma$, where $\lambda = 1.0, 1.5, 2.0, 2.5, 3.0$, with a sample size of 1, 3 or 5, respectively. On the diagonal line of each subplot is the distribution map of these two attributes and on the non-diagonal is the correlation map between them, for three different abnormal patterns (0 1), (1 0) and (1 1). The graphical results show that with an increased λ under the constant n, the classification boundary of three different out-of-control process patterns in the scatter plot will be clearer, the distribution clusters will be more obvious, and the classification accuracy rate will increase. In the case of the constant λ with an increased n, the classification of three different groups can also be discovered and the accuracy increases sharply. When n and λ increase to a certain number at the same time, process patterns of interest are in a completely separated state and the classification accuracy rate is expected to reach 100% in this bivariate case.

5. Conclusion and Future Opportunities

Multivariate control charts are one of the most frequently used MSPC tools for manufacturing process monitoring. They generate an alarm once the process monitored goes out-of-control due to abnormal shifts observed in the process variable (i.e. process mean or variance). Under this circumstance, identifying the contributor(s) of abnormal patterns is an important task with practical significance. In this paper, a novel VLPSOSVM based model has been proposed which comprises two modules. In the detection module, the Hotelling's T^2 statistic is applied to detect an abnormality in the process mean vector. The group of observed process variables that triggers an out-of-control signal is then used as the input of VLPSOSVM model. The recognized shift patterns and the corresponding shift magnitudes are output in the diagnostic module. To obtain an optimal SVM setting based on Gaussian RBF kernel, VLPSO has been adopted to optimize the penalty parameter C and the kernel parameter γ. Based on a simulation experiment from a bivariate industrial example, it can be concluded

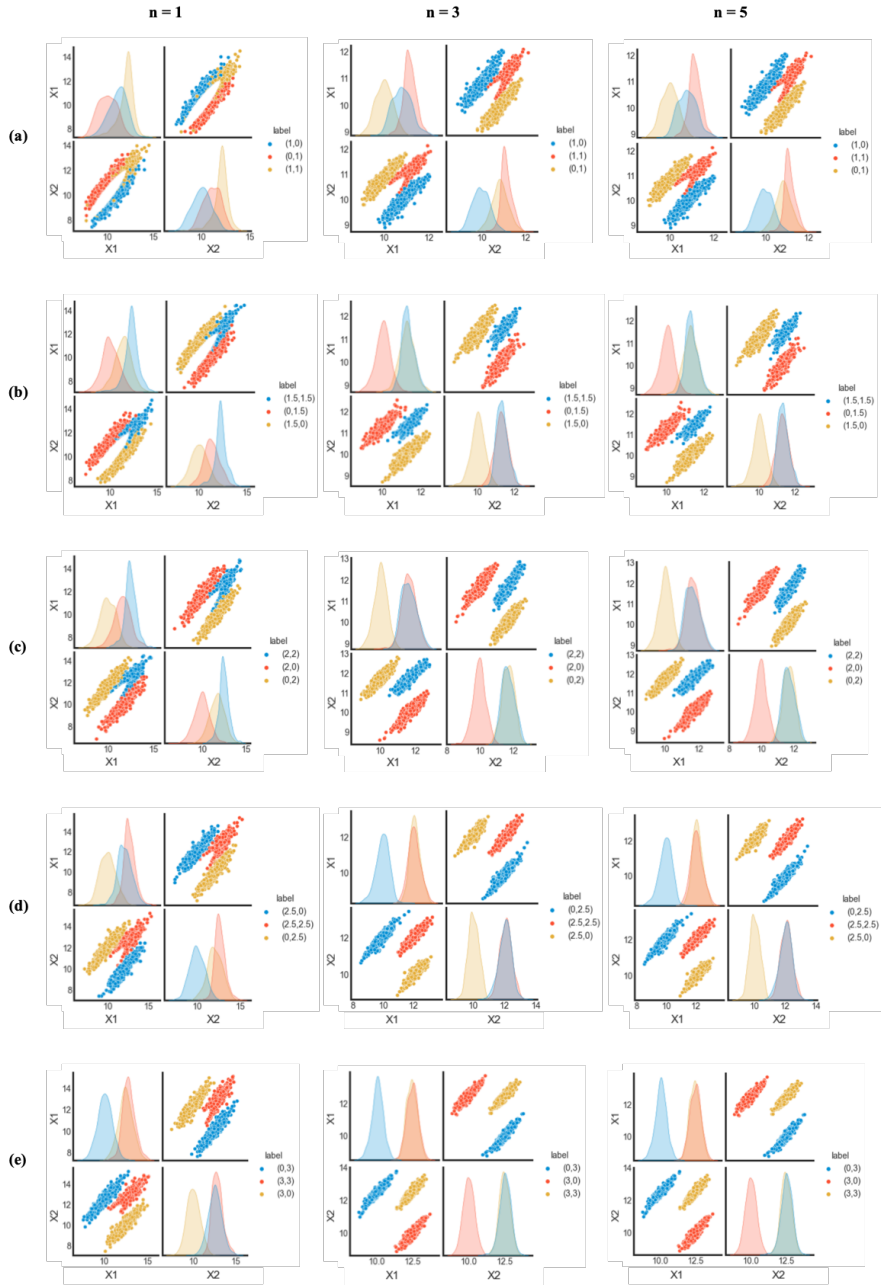

Figure 2. The relationship plots between two target variables in the bivariate process

that VLPSOSVM achieves better average classification accuracy with great generalization ability compared to other SVM models using gradient descent search and GA for parameter optimization. The comparison indicates that VLPSOSVM provides a competitive alternative to the existing MSPC approach.

Some contributions are made in this paper, however, there still exist several limitations. In the proposed VLPSOSVM approach, the focus is on the VLPSO representation and the length changing mechanism. An attempt can be made by combining local search with VLPSO *pbest* to further enhance VLPSO performance. In terms of the simulation experiment, only a bivariate industrial example was used to evaluate the effectiveness of VLPSOSVM. Therefore, a multivariate case can be considered in the future to improve its robustness for high-dimensional classification tasks. In addition to grid search and GA for parameter optimization, a comparison analysis can be implemented between VLPSOSVM and SVM based on other variants of PSO algorithm (e.g., SMDPSO (Wang et al., 2019)), regarding both classification accuracy and computation time.

REFERENCES

Addeh, A., Khormali, A. & Golilarz, N. A. (2018). Control chart pattern recognition using RBF neural network with new training algorithm and practical features, *ISA Transactions*, *79*, 202-216.

Adegoke, N. A., Abbasi, S. A., Smith, A. N., Anderson, M. J. & Pawley, M. D. (2019). A multivariate homogeneously weighted moving average control chart, *IEEE Access*, *7*, 9586-9597.

Ahmadzadeh, F. (2018). Change point detection with multivariate control charts by artificial neural network, *The International Journal of Advanced Manufacturing Technology*, *97*(9), 3179-3190.

Camci, F., Chinnam, R. B. & Ellis, R. D. (2008). Robust kernel distance multivariate control chart using support vector principles, *International Journal of Production Research*, *46*(18), 5075-5095.

Chen, L. H. & Wang, T. Y. (2004). Artificial neural networks to classify mean shifts from multivariate $\chi 2$ chart signals, *Computers & Industrial Engineering*, *47*(2-3), 195-205.

Cheng, Z. Q., Ma, Y. Z. & Bu, J. (2011). Variance shifts identification model of bivariate process based on LS-SVM pattern recognizer, *Communications in Statistics – Simulation and Computation®*, *40*(2), 274-284.

Chou, J. S., Cheng, M. Y., Wu, Y. W. & Pham, A. D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification, *Expert Systems with Applications*, *41*(8), 3955-3964.

Cortes, C. & Vapnik, V. (1995). Support-vector networks, *Machine Learning*, *20*(3), 273-297.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes, *Technometrics*, *30*(3), 291-303.

Cuentas, S., Peñabaena-Niebles, R. & Garcia, E. (2017). Support vector machine in statistical process monitoring: a methodological and analytical review, *The International Journal of Advanced Manufacturing Technology*, *91*(1), 485-500.

Gani, W., Taleb, H. & Limam, M. (2011). An assessment of the kernel-distance-based multivariate control chart through an industrial application, *Quality and Reliability Engineering International*, *27*(4), 391-401.

Hotelling, H. (1947). Multivariate quality control, *Techniques of Statistical Analysis*, 111-184.

Huang, S., Zheng, X., Ma, L., Wang, H., Huang, Q., Leng, G., Meng, E. & Guo, Y. (2020). Quantitative contribution of climate change and human activities to vegetation cover variations based on GA-SVM model, *Journal of Hydrology*, *584*, 124687.

Huang, W., Liu, H., Zhang, Y., Wei, M., Tong, C., Xiao, W. & Shuai, B. (2021). Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM, *Applied Soft Computing*, *109*(5), 107541.

Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 – International Conference on Neural Networks*, *4* (pp. 1942-1948).

Khormali, A. & Addeh, J. (2016). A novel approach for recognition of control chart patterns: Type-2 fuzzy clustering optimized support vector machine, *ISA Transactions*, *63*, 256-264.

Kim, S. B., Jitpitaklert, W., Park, S. K. & Hwang, S. J. (2012). Data mining model-based control charts for multivariate and autocorrelated processes, *Expert Systems with Applications*, *39*(2), 2073-2081.

Kim, Y. & Kim, S. B. (2018). Optimal false alarm controlled support vector data description for multivariate process monitoring, *Journal of Process Control*, *65*, 1-14.

Li, T. F., Hu, S., Wei, Z. Y. & Liao, Z. Q. (2013). A framework for diagnosing the out-of-control signals in multivariate process using optimized support vector machines, *Mathematical Problems in Engineering*, 9 pp. Article ID 494626.

Lowry, C. A., Woodall, W. H., Champ, C. W. & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart, *Technometrics*, *34*(1), 46-53.

Lu, C. J., Shao, Y. E. & Li, P. H. (2011). Mixture control chart patterns recognition using independent component analysis and support vector machine, *Neurocomputing*, *74*(11), 1908-1914.

Montgomery, D. C. (2020). *Introduction to Statistical Quality Control*. John Wiley & Sons.

Niaki, S. T. A. & Abbasi, B. (2005). Fault diagnosis in multivariate control charts using artificial neural networks, *Quality and Reliability Engineering International*, *21*(8), 825-840.

Nidsunkid, S., Borkowski, J. J. & Budsaba, K. (2016). The effects of violations of assumptions in multivariate Shewhart control charts. In *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, (pp. 214-218).

Riaz, M., Saeed, U., Mahmood, T., Abbas, N. & Abbasi, S. A. (2020). An improved control chart for

monitoring linear profiles and its application in thermal conductivity, *IEEE Access*, *8*, 120679-120693.

Shao, Y. E. & Lin, S. C. (2019). Using a time delay neural network approach to diagnose the out-of-control signals for a multivariate normal process with variance shifts, *Mathematics*, *7*(10), 959.

Sun, R. & Tsung, F. (2003). A kernel-distance-based multivariate control chart using support vector methods, *International Journal of Production Research*, *41*(13), 2975-2989.

Tran, B., Xue, B. & Zhang, M. (2018). Variable-length particle swarm optimization for feature selection on high-dimensional classification, *IEEE Transactions on Evolutionary Computation*, *23*(3), 473-487.

Tuba, M. & Tuba, E. (2019). Generative adversarial optimization (GOA) for acute lymphocytic leukemia detection, *Studies in Informatics and Control, 28*(3), 245-254. DOI: 10.24846/v28i3y201901

Wang, C., Zhang, Y., Song, J., Liu, Q. & Dong, H. (2019). A novel optimized SVM algorithm based on PSO with saturation and mixed time-delays for classification of oil pipeline leak detection, *Systems Science & Control Engineering*, *7*(1), 75-88.

Weese, M., Martinez, W., Megahed, F. M. & Jones-Farmer, L. A. (2016). Statistical learning methods applied to process monitoring: An overview and perspective, *Journal of Quality Technology*, *48*(1), 4-24.

Yang, D., Liu, Y., Li, S., Li, X. & Ma, L. (2015). Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm, *Mechanism and Machine Theory*, *90*, 219-229.

Yang, L .Y. & Xu, Z. S. (2019). Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning, *International Journal of Machine Learning and Cybernetics*, *10*(3), 591-601.

Yeganeh, A., Pourpanah, F. & Shadman, A. (2021). An ANN-based ensemble model for change point estimation in control charts, *Applied Soft Computing*, *110*, 107604.

Yu, D. J., Xu, Z. S. & Wang, X. Z. (2020). Bibliometric analysis of support vector machines research trend: A case study in China *International Journal of Machine Learning and Cybernetics*, *11*, 715-728.

Zan, T., Liu, Z., Wang, H., Wang, M., & Gao, X. (2020). Control chart pattern recognition using the convolutional neural network, *Journal of Intelligent Manufacturing*, *31*(3), 703-716.

Zeng, B., Guo, J., Zhu, W., Xiao, Z., Yuan, F. & Huang, S. (2019). A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM, *Energies*, *12*(21), 4170.

Zhang, M. G. (2005). Short-term load forecasting based on support vector machines regression. In *2005 International Conference on Machine Learning and Cybernetics*, *7* (pp. 4310-4314).

Zhang, M., Yuan, Y., Wang, R. & Cheng, W. (2020). Recognition of mixture control chart patterns based on fusion feature reduction and fireworks algorithm-optimized MSVM, *Pattern Analysis and Applications*, *23*(1), 15-26.

Zhang, X., Qiu, D. & Chen, F. (2015). Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis, *Neurocomputing*, *149*, 641-651.

Zhou, X., Jiang, P. & Wang, X. (2018). Recognition of control chart patterns using fuzzy SVM with a hybrid kernel function, *Journal of Intelligent Manufacturing*, *29*(1), 51-67.

Zorriassatine, F., Tannock, J. D. T. & O'Brien, C. (2003). Using novelty detection to identify abnormalities caused by mean shifts in bivariate processes, *Computers & Industrial Engineering*, *44*(3), 385-408.