Decision Science Letters 10 (2021) 39–52

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

Monitoring image-based processes using a PCA-based control chart and a classification technique

Setareh Kazemi[a] and Seyed Taghi Akhavan Niaki[a*]

[a]*Department of Industrial Engineering, Sharif University of Technology, P.O. Box 11155-9414 Azadi Ave., Tehran 1458889694, Iran*

C H R O N I C L E

Article history:
Received August 22, 2020
Received in revised format:
August 31, 2020
Accepted October 11 2020
Available online
October 11, 2020

Keywords:
SPC
PCA
Classification
LDA
QDA
KNN
SVM

A B S T R A C T

Machine vision systems are among the novel tools proven to be useful in different applications, among which monitoring and controlling manufacturing processes is one of the most important ones. However, due to the complexity resulted from high-dimensional image data and their inherent correlations, the acquisition of traditional statistical process control tools seems inapplicable. To overcome the shortcomings of the traditional methods in this regard, a statistical model is proposed in this paper which utilizes the concepts of both the PCA-based T^2 control chart and the classification methods to develop a tool capable of controlling an image-based process. By defining the warning zones, collected data taken from an image-based process are classified into more than the two classes related to in-control and out-of-control processes. This helps practitioners to define rules to make it easier to realize when the process is getting out of control. Through simulation, the accuracy performance and the speed of four different types of classifiers including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), kth nearest neighbors (KNN), and support vector machine (SVM) are assessed in different scenarios, based on which the functionality of the proposed approach is evaluated in in-control and out-of-control conditions.

1. Introduction

It has been many years that statistical monitoring and controlling production processes has become an inseparable part of any manufacturing system with a drive to not only survive in today's competing world but also excel in doing so. To answer this need, there have been many statistical methods proposed by different researchers around the world. Undoubtedly the control charts, firstly introduced by Shewhart (1924), as effective tools to help reduce the variabilities that cause a process to exit from the controlled state, were among the world-changing methods in this regard. Since the advent of control charts, many scientists have developed them in various ways to find more responsive ones to the scope of their work. Nevertheless, to cope with the ever-growing technology better and to respond to the needs of today's businesses more quickly, modern equipment such as machine vision systems (MVS), sensors, and cameras, have come to assist the traditional controlling systems in a faster and more accurate way. To monitor and control a process this way, the production line is equipped with a camera that takes photos of the products first. Then, as opposed to the traditional way in which an operator controls the quality of products on time with a high possibility of errors and sometimes needed destructive practices, the digital images are analyzed by an online monitoring system with more accuracy and a higher pace. However, despite all the advantages, image-based monitoring systems bring, depending on the number of variables that are needed to be monitored, the nature of the manufacturing system, time, and cost limitations, the complexity of this system varies.

* Corresponding author.
E-mail address: Niaki@Sharif.edu (S. T. A. Niaki)

doi: 10.5267/j.dsl.2020.10.005

The use of images to monitor and control a production system takes huge data sets consisting of multi-dimensional variables that are highly correlated. A widely known method that helps reduce the correlation between the variables is the so-called principal components analysis (PCA). In this method, the original variables are projected to a new space, in which they become uncorrelated. Principle components are a linear combination of the previous variables. To monitor the new multivariate principal components, effective multivariate control charts are needed. In this paper the T^2 control chart is used for this purpose. In this method, the T^2-statistic is calculated using the covariance matrix of the principal components. To find whether or not a process is in-control, the T^2-statistic is compared with the control limits. These control limits can be defined using classification methods which are among supervised statistical learning methods. The classifiers assign the obtained samples to in-control or out-of-control classes, hence can be employed as new and effective methods to substitute the primary control charts. Utilizing classifiers also allows one to divide observations into more than two categories. Therefore, by defining warning zones in the in-control region, monitoring and controlling the process will become more accurate as tracking observations which fall into warning classes decreases the probability that the process will go out of control. Besides, the performance accuracy and the speed of four different classifiers, namely linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machines (SVM), and kth nearest neighbor (KNN) are measured in a simulated process evaluated in both in-control and out-of-control state.

The paper is organized as follows. In Section 2, a review of the literature and background is briefed. Section 3 presents the methodology proposed in this research. In Section 4, the final results of applying the methodology on the simulated processes are analyzed. Finally, concluding remarks and future research directions are presented in Section 5.

2. Literature and background

Irresistible random and natural variations occur in every manufacturing system. As long as the process is functioning in the presence of these variations, it is called “in-control”. However, when there are other significant variations (assignable causes) that happen as a result of an error caused by machine, material, or human, the process falls into an “out-of-control” condition (Montgomery 2007). Control charts are one of the online-monitoring tools that by accurate and in-time detection of assignable causes assist in reducing or omitting such variations in the process. Control charts are used in two phases: In Phase 1, a data set is first extracted from the process for retrospective analysis. Then, control limits are designed to track whether or not the process is in-control. In case the chart statistic falls out of the control limits and hence the process is detected to be out-of-control, it is analyzed to find assignable causes. Next, such causes are removed from the process and the control limits are re-estimated accordingly until a set of in-control samples are obtained based on which the parameters of the process are estimated. In Phase 2, it is assumed that the main assignable causes that move the process out of control are identified and removed in Phase 1. Therefore, the main purpose is to monitor the process. When monitoring quality attributes such as color or appearance of a product, common statistical process control (SPC) methods, and control charts can no longer play an effective role. For instance, products such as LCD panels, food, and tiles fall into this group. In the traditional monitoring systems, trained operators are responsible to monitor these kinds of processes. Nevertheless, operators’ judgment could lead to classification variability due to fatigue and lack of concentration (repeatability) or variability in classification could occur by several operators who had a different understanding of the process (reproducibility) (Prats-Montalbán & Ferrer 2014). Thus, due to the drawbacks of traditional monitoring systems and their incompetency to monitor several products at the same time, machine vision systems (MVS) come to the picture to play a more efficient role. A MVS mostly consists of an image-capturing machine (like cameras or digital sensors) and computer systems to analyze the data obtained (Malamas et al., 2003). The pervasive applications of MVS can be tracked from their early emergence for medical cases (Ledley, 1964; Strauss et al., 1971; Ballard & Sklansky, 1973; Bellon, et al., 1995) and satellite data (Bernstein, 1976; Brayer et al., 1977) to their industrial applications in the following decades (Goldstein & Nagler, 1988; Cohen, et al., 1991; Boukouvalas et al., 1997; Malamas et al., 2003; Yu & MacGregor, 2003; Yu & MacGregor, 2004, Szatvanyi et al., 2006; Wójcik & Kotyra, 2009; Castiñeira et al., 2012; Dutta et al. 2016a; Dutta et al., 2016b; Dutta et al., 2016c; Singh et al., 2016).

After the image is taken from the process, the process can be analyzed through different image analysis methods. Various methods have been presented for this sake, among which the multivariate image analysis (MIA) has proven to be quite effective (Duchesne et al., 2012; Prats-Montalbán & Ferrer 2014). Image analysis is part of the vast field of image monitoring with the following three main steps:

1. *Image preprocessing*: This stage consists of improving image quality by noise reduction, edge detection, and validating the image analysis stage.
2. *Image compression*: This will decrease the needed memory by omitting unwanted parts of the image.
3. *Image analysis*: This step aides to obtain various numeric values/graphs related to image specifications that can be employed in classification, defect detection, or prediction of quantitative characteristics of the image (Prats-Montalbán et al., 2011)

The above third step is where statistical process control methods can assist in monitoring the quality of products by analyzing data obtained from the images. While many studies have focused on applying various statistical methods to

image-based process monitoring, a helpful overview is presented by Megahed et al. (2011) who reviewed the applications of the control charting methods using image data. Duchesne et al. (2012) also presented a thorough review of studies regarding the MIA methods used for process quality control. Prats-Montalbán et al. (2015) proposed a framework in which process monitoring based on image data is possible. Additionally, the location of defects can be identified in this method, in case an inconsistency with the statistical standards occurs. Moreover, Reis & Bauer (2009) and Ottavian et al. (2013) addressed the monitoring problems regarding changes in lighting conditions. A recent comprehensive review of the latest statistical process monitoring methods is provided by He et al. (2019) that suggests novel approaches for feature-based monitoring utilized for fault detection and soft sensor development. Analyzing extracted features from images is mainly achieved through empirical modeling approaches (Duchesne et al., 2012). The desired feature vector of each image is collected in a $X_F(N * K)$ matrix, in which N is the number of process images in each set and K is the number of features in each image. Then, the vector Y is considered as a set of response variables that are used to analyze future sets or their classifications. Moreover, latent variables (LV) are employed to move to a lower dimension for data sets with fewer variables through the new transformed space "T". This transformation is usually achieved by methods such as PCA. The output of the LV model, either T or Y, will then be used to decide on the performance of the process. To incorporate the LV model into process control methods, the two phases involved in image-based process control are approached as follows:
Phase 1: Online process monitoring. This phase starts with building a PCA model. PCA projects the correlated process variables to reduced-dimension space with independent variables called latent. These variables are ordered concerning their variances. In other words, for an $(I \times J)$ X matrix consisting of I multivariate observations and J process, when the PCA is used we have:

$$X = TP^T + E, \tag{1}$$

where $T(I \times R)$ and $P(J \times R)$ are respectively the score and loading matrix of R principle components $(R \leq rank(X))$ and $E(I \times J)$ is the residual matrix of the PCA model. The PCA model is built using one or few images which are taken from the process while it is in control. Then using the in-control loading matrix P, the score of a new image X_{new} is computed using Eq. (2).

$$T_{new} = X_{new} - T_{new}P \tag{2}$$

Hence, the new residuals can be achieved by:

$$Z_{new} = X_{new} - T_{new}P, \tag{3}$$

where T_{new} and Z_{new} are utilized in calculating two statistics T_R^2 and RSS (residual sum of squares) which are different from each other in concept (Kourti & MacGregor 1996). In this paper only the T_R^2-statistic is used. The T^2 control chart was first presented by Hotteling (1947) and investigated and generalized by many researchers including Alt (1977), Alt (1985), Alt & Smith (1988), Ryan (1989), and Jackson (1991) thereafter. The above T_R^2 is the Hotteling T^2-statistic in a reduced dimension with R components instead of the original process variables. Under the normality assumption of the scores, the following upper control limit (UCL) and lower control limit (LCL) are obtained for M samples (Montgomery, 2007):

$$UCL = \frac{j(m-1)(i-1)}{m(i-1)-j+1} F_{\alpha,j,m(i-1)-j+1} \begin{cases} j = 1, \dots J \\ i = 1, \dots I \\ m = 1, \dots M \end{cases} \tag{4}$$

$$LCL = 0. \tag{5}$$

The normality assumption is generally logical as the reference models are built using the in-control images (Nomikos & MacGregor, 1995). Researchers such as Ge et al. (2009) and Phaladiganon et al. (2013) presented a nonparametric method to use PCA-based control charts for nonnormal data that could provide a roadmap in case such processes needed to be monitored.

Phase 2: Online monitoring. After the model is built based on the in-control images, the process can be monitored online by new images taken from the process. The control limits that are used in Phase 2 using the T_R^2 statistic are calculated as follows:

$$UCL = \frac{j(m+1)(i-1)}{m(i-1)-j+1} F_{\alpha,j,m(i-1)-j+1} \begin{cases} i = 1, \dots I \\ j = 1, \dots J \\ m = 1, \dots M \end{cases} \tag{6}$$

$$LCL = 0. \tag{7}$$

As demonstrated in Eq. (6), the upper control limit is exclusive for each level of α. This feature is used in this paper to define several control limits by varying the value of α.

Deciding on whether or not the process is performing under control can be made employing several criteria. In this paper, the conformity criterion is defined and measured using statistical learning methods to provide the means to identify images that indicate the process is performing out of control. Statistical learning methods are a set of tools utilized to understand the data better. These methods fall into two main categories of supervised and unsupervised procedures. In general, supervised statistical learning consists of making statistical models useful to predict or estimate output based on one or more inputs, while in unsupervised statistical learning methods, inputs with no specified supervised output exist. In this study, since there is a distinct output for every input (each image is classified as in-control, out-of-control, or in the warning zone), the problem belongs to the field of supervised methods. Many of statistical learning methods such as linear regression, logistic regression, classification methods, and support vector machines are among supervised methods. This study employs different classification methods to classify the input images in in-control, out-of-control, and in-warning zone classes. A classifier is a mapping function of input attribute vectors $x \in \chi$ to output class labels $y \in \{1, \dots, C\}$. The data classification process consists of the following two steps:

1. *Learning*: In this step, the classifiers are trained using training data and the classification rules are established.
2. *Classification*: If the accuracy of classifiers, when examined by test data, is acceptable, the rules are eligible for classification of a new data set. As one statistical learning method cannot be represented as the best one, normally a couple of methods are tested and the best performing one is selected.

In this paper, the K-fold cross-validation (CV) method is used to calculate false classification rate of each classifier, based on which the superior methods area picked to construct the process control model to be used in Phase 2. In the K-fold CV, one set of observations is used for both the test and training sets as it is divided into K groups of equal sizes. The first group is assumed to be the validation set while the remaining K-1 groups are used for fitting the method. This procedure is repeated K times when each time one of the groups becomes the validation set. Normally K=5 or K=10 is used in this method (James et al., 2013). As s a result, each time an estimate of the misclassified observations (Err_i ; $i = 1, \dots k$) is calculated and the K-fold CV error rate is obtained as:

$$CV_k = \frac{1}{k}\sum_{i=1}^{k} Err_i. \tag{8}$$

Then the classifiers with the lowest rate of misclassified observations can be selected to be used in the next steps. In this study, LDA, QDA, KNN, and SVM are selected among many more to be employed in the process monitoring method. Before going to the next section where the new monitoring approach is proposed, these classifiers are reviewed in more depth:

1. LDA (linear discriminant analysis): LDA is defined as an orthogonal linear transformation that can discriminate 2 or more classes. This method was first introduced by Fisher (1936) for two-class problems and then was extended for more than two classes by Rao (1948). This method is based on Bayes' theorem and seeks to estimate process parameters using some assumptions. Let's assume that $x = (X_1, X_2, \dots, X_p)$ is the attribute vector driven form a multivariate normal distribution (N (μ_k, Σ)) with specific mean vector for each class (k) and a common covariance matrix among all classes. This method assigns observation $X = x$ to a class with maximum $\delta_k(x)$ which is a linear function of x as:
$$\delta_k(x) = x^T \Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T \Sigma^{-1}\mu_k + \log \pi_k, \tag{9}$$
in which π_k is the prior probability that a randomly chosen observation comes from the k^{th} class (James et al., 2013).
2. QDA (quadratic discriminant analysis: In this method, the assumptions are the same as the ones in LDA, unlike LDA however, QDA assumes that each class has its covariance matrix. In other words, observations that belong to the k^{th} class follow $N(\mu_k, \Sigma_k)$. In this case, the observation $X = x$ is assigned to the class in which $\delta_k(x)$ defined in Equation (10) has its largest value:
$$\delta_k(x) = -\frac{1}{2}(x - \mu_k)^T \Sigma_k^T (x - \mu_k) - \frac{1}{2}\log|\Sigma_k| + \log \pi_k. \tag{10}$$
As shown, here $\delta_k(x)$ is a quadratic function of x, hence this method is called quadratic discriminant analysis (James et al., 2013).
3. KNN (K^{th} nearest neighbor): This method has been hugely used for pattern detection since the 1960s when computational complexities were resolved. KNN classifies based on comparing test set with similar training sets. The KNN classifier first finds the closest K points of the training data to test observation x_0 and then calculates the conditional probability for each class. Finally x_0 is assigned to the class with the largest probability. In this method, the "closeness" is defined by a distance matrix. Selecting the best value for k can be done either by trial or by computing error of the model per each k and finally selecting the value with the lowest error. In general, it is better to select larger k for a bigger number of training sets so that the classification can be performed by a larger ratio of sets.
4. SVM (support vector machines): Although this method has a low training pace, it is known to be one of the most accurate ones as it is capable of modeling nonlinear decision boundaries and is less prone to overfitting. In the SVM method, training data are first projected to a higher dimension using nonlinear mapping. Then in the new dimension, this method looks for linear optimal hyperplanes that are the decision boundaries that separate different

classes. This method finds the mentioned hyperplanes based on support vectors and margins. Observation x_0 is classified based on the side of the hyperplane it locates (James et al., 2013).

Table 1 summarizes some of the studies that have used classification and dimension reduction methods in the process monitoring field.

Table 1
Reviewed studies

Row	Scientists	Industry	Transformation technique	Classification method
1	Wang et al. (2016)	Cutting	PCA	-
2	Singh et al. (2016)	Medicine	PCA	KNN, SVM, ANN, Naïve Bayes, Decision tree
3	Dutta et al. (2016b)	Food	Wavelet	SVM
4	Dutta et al. (2016a)	Food	Wavelet	-
5	Dutta et al. (2016c)	Food	Wavelet	SVM
6	Adem et al. (2015)	Packaging	-	Linear Regression
7	Reis (2015)	Various Industries	PCA	-
8	Eldessouki et al. (2014)	Fabric	PCA	ANN
9	Castiñeira et al. (2012)	Combustion	MPCA	SVD
10	Prats-Montalbán & Ferrer (2007)	Various Industries	PCA	SIMCA, FPA, PLS-D
11	Szatvanyi et al. (2006)	Combustion	MPCA	SVD
12	Yu & MacGregor (2004)	Combustion	MPCA	SVD
13	Liang et al. (2009)		PCA	Q-statistics
14	Kistner et al. (2013)	Mining	GLCM, Wavelet, Steerable pyramids	LDA, QDA, KNN
15	Facco et al. (2011)	Nanofiber Membranes	WTD+PLS	-
16	Reis & Bauer (2009)	Paper	Wavelet	WTA-based analysis and classification
17	Pereira et al. (2008)	Food	PCA	A masking approach in the PCA scores space
18	Chatterjee et al. (2010)	Mining	PCA	Multi-layer perceptron neural network (MLP)
19	Chatterjee & Bhattacherjee (2011)	Mining	GA-based feature selection algorithm	NN
20	Reis & Bauer (2010)	Paper	VS, PCA, FDA	LC, QC, SVC, KNN, NN
21	Yan et al. (2015)	Various Industries	UPCA, MPCA, UMPCA, MIA	-
22	Mirschel et al. (2019)	Textile	PLS Regression	-

In the next section, the image-based monitoring methodology proposed in this paper is explained in detail.

3. Methodology

This research takes advantage of the concepts of both the PCA-based T^2 control chart and the classification methods to develop a tool capable of controlling an image-based process. To this aim, the principal components are first determined. Then, the T^2-statistic is calculated using the covariance matrix of the principal components. Next, this statistic is compared with the control limits defined using classification methods. The classifiers assign the obtained samples to in-control, warning zone, or out-of-control classes. The performance accuracy and the speed of four different classifiers, namely LDA, QDA, SVM, and KNN are measured in a simulated process evaluated in both in-control and out-of-control states. The main steps taken are explained as follows.

Step 1. *Image acquisition and preparation*: In the first step, images taken from the process form the matrix of observations. It is assumed that preprocessing and feature extraction of images (such as removing the noise and extracting the region of interest) has been done before. Based on these presumptions and assuming the data population follows a 5-variant normal distribution, 1,000 samples each with 1,000 observations and 5 attributes are randomly generated with known mean, covariance matrix, and predetermined correlations. Then, using PCA the data matrix is projected to principle components' subspace, where loading, score matrix, and the explained variability of each component is calculated for all observations. As it has been observed that the first two principal components explain over 94 percent of the variability of the observations, two types of observation matrix are considered; one that uses all 5 variables (called "the complete" version thereafter) and the other that employs the first 2 principal components (called "the reduced" version thereafter).

Step 2. *Training classifiers and estimating their accuracy*: As mentioned previously, a unique control limit can be defined per each α using Equation (6). This point is used to propose a new criterion to classify the observations into more than two groups so that several upper control limits can be defined based on each value of α, for which the spaces between the limits can introduce different groups. For instance, $\alpha = 0.05$ and $\alpha = 0.1$ provide two upper control limits. In this case, the observations which fall into space between 0 and the upper control limit initiated by $\alpha = 0.1$ are classified in the first category (in-control state), those which fall between this upper limit and the one obtained using $\alpha = 0.05$ are classified in the second category (warning zone), and finally, those observations that fall out of the upper control limit obtained $\alpha = 0.05$ are classified in the third category (out-of-control state). Thus, in general, n values of α defines $n + 1$ classes. This is what shapes the logic behind the proposed use of a classification method. Next, by comparing T^2 statistic of each observation with the predefined control limits, its class is determined. Here, the principal components are the predictors and the classes are considered as the response variables to training the classifiers. Finally, using the cross-validation method, the data sets are separated into training and test sets to evaluate the accuracy of different classifiers which will assist in choosing the most accurate ones.

Step 3. *Monitoring the process using classifiers and comparing their performance in an in-control state*: in this step, new data are taken from the process and are monitored and controlled by the supervised classification methods. To compare the performance of these methods in the "in-control" state, the average run length (ARL) criterion is used. This measure is calculated by generating 1,000 "in-control" samples each having 1,000 observations. The classes of the observations are then specified using the classifiers, based on which the in-control RL (RL_0) is calculated for each sample. This process is repeated 1,000 times, for which the in-control average run length (ARL_0) is calculated for each classifier.

Step 4. *Monitoring the process using classifiers and comparing their performance in an out-of-control state*: In the last step, to investigate the performance of the classifiers in an "out-of-control" state, %5 of the variance of each variable is added to its mean to generate new data sets with new mean values. The classes of these simulated data are then identified by the classifiers in 30 iterations, each having 1,000 samples. Similar to the previous step, the out-of-control run lengths (RL_1) are obtained to determine the out-of-control average run length (ARL_1) of the utilized classifier.

The above steps are followed for different values of α. For a better comparison, however, $\alpha_1 = 0.1$ remains constant to calculate the first *UCL* and α_2 takes the values of $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09\}$ to determine the second *UCL*. This way, by changing α_2, the space between two control limits gets narrower and the performance of different classifiers can be investigated in 9 conditions. It should be noted that as the classification criterion changes when α_2 changes, the classifiers need to be trained per each α_2. This results in a huge amount of calculations and derived data. Thus, only some of the charts are shown in the next section to simplify their analysis.

4. Results

In this section, suitable classifiers are first found based on their accuracy. Then, the performance of the chosen classifiers is assessed in both in-control and out-of-control conditions.

4.1 Selecting suitable classifiers

The initial step of this research is to find proper classification methods based on their accuracy. To accomplish this, the accuracy of 22 classification methods is evaluated using the MATLAB software based on 1,000 samples each with four combinations of (1) having five classes, (2) having two classes, (3) in the presence of all five components, and (4) in the presence of two principal components. The above 22 classifiers fall into five major categories of (1) tree classifiers, (2) discriminant classifiers, (3) SVM classifiers, (4) nearest-neighbors classifiers, and (5) ensemble classifiers. The cross-validation method is deployed to measure the accuracy of these classifiers. As mentioned in Section 2, the K-fold cross-validation is one of the re-sampling methods which can train classifiers and calculate their accuracy using a limited number of samples. As such, data are first divided into "K" groups. Then, the classifier is trained on "$K-1$" groups, and its capability to classify the K^{th} group is measured. This cycle is repeated for $k = \{1, 2, \dots, K\}$ times and the final accuracy is calculated using the accuracy obtained in each iteration. To find a suitable K value, $K = \{3, 5, 10\}$ are used based on which $K = 10$ has been chosen due to its better performance. The comparison results of these classifiers are demonstrated in Tables 2-5. Based on the results in these tables LDA, QDA, KNN (with $K = 10$), and SVM work generally better in different combinations of the number of classes and the number of components and hence are picked as the best performing classifiers.

Table 2

Accuracy of classifiers in the complete version when 5 classes are used (α={0.05, 0.1, 0.15, 0.2})

Row	Method	Accuracy	Row	Method	Accuracy
1	Complex Tree	70.8	**12**	Fine KNN	68.7
2	Medium Tree	77.2	**13**	Medium KNN	80.1
3	Simple Tree	79.4	**14**	Coarse KNN	80.2
4	Linear Discriminant	79	**15**	Cosine KNN	80.1
5	Quadratic Discriminant	74.5	**16**	Cubic KNN	80
6	Linear SVM	80.2	**17**	Weighted KNN	77.6
7	Quadratic SVM	79.7	**18**	Ensemble- Boosted Trees	78.8
8	Cubic SVM	7.5	**19**	Ensemble- Bagged Trees	78.4
9	Fine Gaussian SVM	80.2	**20**	Ensemble- Subspace Discriminant	79.8
10	Medium Gaussian SVM	80.2	**21**	Ensemble-Subspace KNN	74
11	Coarse Gaussian SVM	80.2	**22**	Ensemble-RUS Boosted Trees	13.7

Table 3
Accuracy of classifiers in the reduced version when using 5 classes (α= {0.05, 0.1, 0.15, 0.2})

Row	Method	Accuracy	Row	Method	Accuracy
1	Complex Tree	74	12	Fine KNN	64.2
2	Medium Tree	78.8	13	Medium KNN	80.1
3	Simple Tree	80	14	Coarse KNN	80.2
4	Linear Discriminant	80	15	Cosine KNN	80.2
5	Quadratic Discriminant	77.4	16	Cubic KNN	80.1
6	Linear SVM	41.4	17	Weighted KNN	67.9
7	Quadratic SVM	5.7	18	Ensemble- Boosted Trees	79.6
8	Cubic SVM	47.1	19	Ensemble- Bagged Trees	65
9	Fine Gaussian SVM	80.2	20	Ensemble- Subspace Discriminant	80
10	Medium Gaussian SVM	80.2	21	Ensemble-Subspace KNN	64.2
11	Coarse Gaussian SVM	80.2	22	Ensemble-RUS Boosted Trees	10

Table 4
Accuracy of classifiers in the complete version when using 2 classes (α= {0.05, 0.1})

Row	Method	Accuracy	Row	Method	Accuracy
1	Complex Tree	84.7	12	Fine KNN	81.1
2	Medium Tree	87.8	13	Medium KNN	89.2
3	Simple Tree	88.6	14	Coarse KNN	89.2
4	Linear Discriminant	89.2	15	Cosine KNN	89.2
5	Quadratic Discriminant	72.8	16	Cubic KNN	89.2
6	Linear SVM	89.2	17	Weighted KNN	87.5
7	Quadratic SVM	88.8	18	Ensemble- Boosted Trees	88.8
8	Cubic SVM	8.9	19	Ensemble- Bagged Trees	88.7
9	Fine Gaussian SVM	89.2	20	Ensemble- Subspace Discriminant	89.2
10	Medium Gaussian SVM	89.2	21	Ensemble-Subspace KNN	86.3
11	Coarse Gaussian SVM	89.2	22	Ensemble-RUS Boosted Trees	31.3

Table 5
Accuracy of classifiers in the reduced version when using 2 classes (α= {0.05, 0.1})

Row	Method	Accuracy	Row	Method	Accuracy
1	Complex Tree	84.8	12	Fine KNN	80.1
2	Medium Tree	88	13	Medium KNN	89.2
3	Simple Tree	88.8	14	Coarse KNN	89.2
4	Linear Discriminant	89.2	15	Cosine KNN	89.2
5	Quadratic Discriminant	89.2	16	Cubic KNN	89.2
6	Linear SVM	89.2	17	Weighted KNN	86.1
7	Quadratic SVM	53.2	18	Ensemble- Boosted Trees	88.4
8	Cubic SVM	45.8	19	Ensemble- Bagged Trees	88.4
9	Fine Gaussian SVM	89.2	20	Ensemble- Subspace Discriminant	89.2
10	Medium Gaussian SVM	89.2	21	Ensemble-Subspace KNN	86.8
11	Coarse Gaussian SVM	89.2	22	Ensemble-RUS Boosted Trees	35.8

After selecting the classifiers, enhancing their classification accuracy and speed is essential. Much higher classification accuracy and speed is achieved by replacing the score matrix with the T^2 statistics of the principal components. In fact, after this substitution, the misclassification error and the speed of the training process of the classifiers using one million samples reached to less than 2 percent and 1 minute, respectively. The results for both complete and reduced versions are demonstrated in Table 6 and Table 7, when $\alpha_2 = 0.05$. These results reflect that a much higher speed can be obtained in the reduced version with a negligible reduction in the accuracy of all classifiers.

Table 6
Accuracy and speed of classification methods in the second round of training for $\alpha_2 = 0.05$ in the complete version

Row	Method	Accuracy	Speed (seconds)
1	LDA	%98.2	14.771
2	QDA	%99.9	4.912
3	KNN	%100	19.162
4	SVM	%100	27.319

Table 7
Accuracy and speed of classification methods in the second round of training for $\boldsymbol{\alpha_2 = 0.05}$ in the reduced version

Row	Method	Accuracy	Speed (seconds)
1	LDA	%98	5.535
2	QDA	%99.5	9.365
3	KNN	%100	26.141
4	SVM	%100	23.587

The enhanced classifiers are then trained 9 times for $\alpha_2 = \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09\}$ while α_1 constantly remains 0.1.

4.2. Performance evaluation of the classifiers in in-control state

As mentioned in Step 3 of the proposed methodology, trained classifiers are employed to classify 1,000 simulated in-control samples in both complete and reduced conditions. Then, ARL_0 is calculated for each α_2 in $\{0.01,\ 0.02,\ 0.03,\ 0.04,\ 0.05,\ 0.06,\ 0.07,\ 0.08,\ 0.09\}$ while α_1 is 0.1. The results are shown in Table 8 and Table 9for complete and reduced versions, respectively. In these tables, $UCLI_1$ is a border between Class 1 and 2 that has remained the same while $UCLI_2$ which is the border between Class 2 and 3 changes per each α_2 value. As shown in Table 8 and Table 9, the performance of the SVM and the KNN methods are very good as their ARL_0s are very close to the desired ARL_0 in both complete and reduced versions. However, whilst QDA shows a good performance in most cases, LDA does not work very well.

Table 8
Comparison of classifiers' ARL_0 with the desired ARL_0 in the complete versio

α_1	α_2	$UCLI_1$	$UCLI_2$	Desired ARL_0	LDA ARL_0	QDA ARL_0	KNN ARL_0	SVM ARL_0
0.1	0.01	4.6098	9.2196	100	126.2231	96.7348	100.8596	100.8678
0.1	0.02	4.6098	7.8319	50	65.5225	51.4500	50	50.2500
0.1	0.03	4.6098	7.0202	33.3333	42.6933	34.2667	33.3667	33.5000
0.1	0.04	4.6098	6.4442	25	20.0228	25.5250	25.1500	25.1250
0.1	0.05	4.6098	5.9975	20	26	20.3448	20.2730	20.2820
0.1	0.06	4.6098	5.6325	16.6666	21.0490	17.1167	16.7833	16.8000
0.1	0.07	4.6098	5.3239	14.2857	14.9342	14.5714	14.3857	14.3857
0.1	0.08	4.6098	5.0565	12.5	13.0004	12.7875	12.6000	12.5375
0.1	0.09	4.6098	4.8207	11.1111	11.6786	11.4000	11.1222	11.2222

Table 9
Comparison of classifiers' ARL_0 with the desired ARL_0 in the reduced version

α_1	α_2	$UCLI_1$	$UCLI_2$	Desired ARL_0	LDA ARL_0	QDA ARL_0	KNN ARL_0	SVM ARL_0
0.1	0.01	4.6098	9.2196	100	139.2270	102.3025	101.1290	101.5071
0.1	0.02	4.6098	7.8319	50	65.0369	52.5715	50.6571	50.6686
0.1	0.03	4.6098	7.0202	33.3333	44.7925	35.5308	33.4667	33.5023
0.1	0.04	4.6098	6.4442	25	33.7505	25.9522	25.2941	25.3610
0.1	0.05	4.6098	5.9975	20	27.8053	21.4202	20.2197	20.2805
0.1	0.06	4.6098	5.6325	16.6666	23.9260	17.9107	16.8333	16.9853
0.1	0.07	4.6098	5.3239	14.2857	16.2857	14.6143	14.3945	14.3986
0.1	0.08	4.6098	5.0565	12.5	15.3750	12.7500	12.6125	12.6201
0.1	0.09	4.6098	4.8207	11.1111	13.3752	11.4556	11.1581	11.1596

Another important point is that when the significance level increases, the performance of all methods, even LDA, is improved. Nonetheless, it is the other way around for LDA when α_2 has a much smaller amount such as 0.01. Of course, the other methods also have less compliance with the desired ARL_0 in these cases. The discrepancies among the performance of LDA, QDA, KNN, and SVM can be sought through their accuracy in classifying an out-of-control sample in Class 3, correctly. To better understand the differences between these methods and why they perform as they do, investigating their confusion matrix and ROC curve can be helpful. The confusion matrix demonstrates how many of the samples are classified in their correct class and how many are misclassified. The ROC curve simultaneously shows two error rates. This curve is used for the classification methods with two classes; the positive and the negative ones. In the investigated samples with at least 3 classes, one class is assumed as the positive class and the others as negative ones. In the ROC curve, the true positive rate and the false positive rate are depicted. The true positive rate shows the rate to which samples have been classified in their correct class and the false positive rate shows the rate of the samples that have not been classified in their right class. The more this curve is aligned to the north-west corner, and hence the more the area under this curve (AUC) is larger, the classifier performs better. Fig. 1 and Fig. 4 depict the confusion matrices of the classifiers for $\alpha_1 = 0.1$ and $\alpha_2 = 0.05$. As shown in these figures, LDA classifies the samples of the third class to another class in 25% of the times and its false-negative rate for new data is 22% while this measure is 1 % and less for the other classifiers. Besides, the ROC curves of these classifiers for $\alpha_1 = 0.1$ and $\alpha_2 = 0.05$ can be observed in Figs. 5-8. Note that as there are 3 classes, in this case, three ROC curves are obtained for each class. It is evident from these figures that the false-positive rates of KNN and SVM are so trivial that one can assume that their accuracy is 100% as their AUC is 1 for all 3 classes.

4.3. Evaluating the performance of the classifiers in an out-of-control state

To examine the performance of the classifiers in an out-of-control state, 1,000 samples, each with 1,000 observations are generated by adding a coefficient of the observations' variance to their mean values. The samples are gradually moved out of control while their classes are specified by each classifier trained for the specified significance level. Each time the mean value is shifted, the ARL_1 is calculated and plotted in a chart.

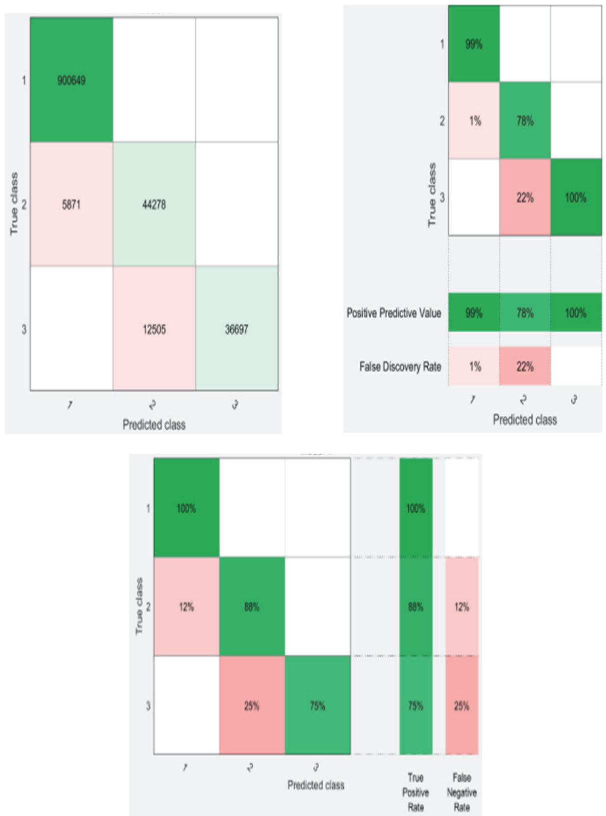

Fig. 1. The confusion matrix of LDA method for $\alpha_2 = 0.05$

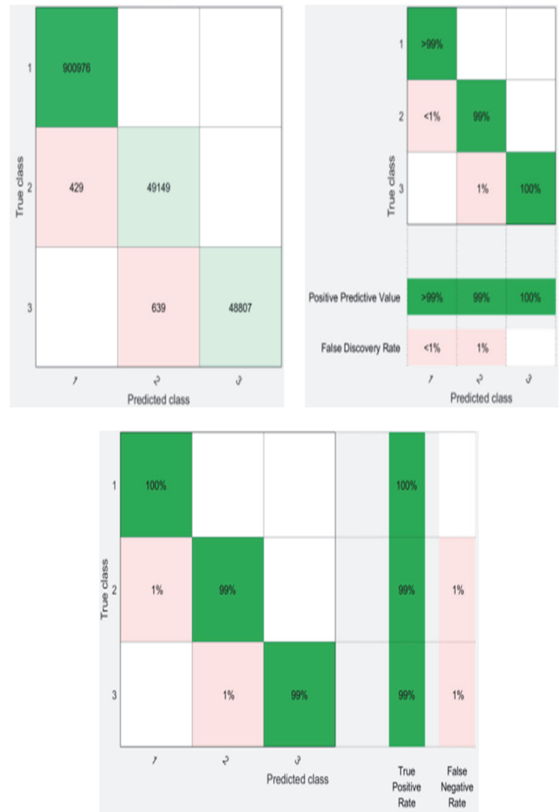

Fig. 2. The confusion matrix of QDA method for $\alpha_2 = 0.05$

Fig. 3. The confusion matrix of KNN method for $\alpha_2 = 0.05$

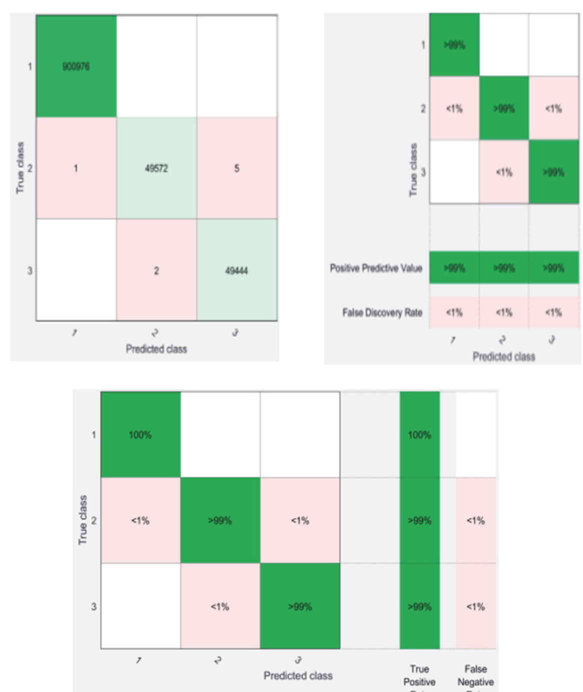

Fig. 4.The confusion matrix of SVM method for $\alpha_2 = 0.05$

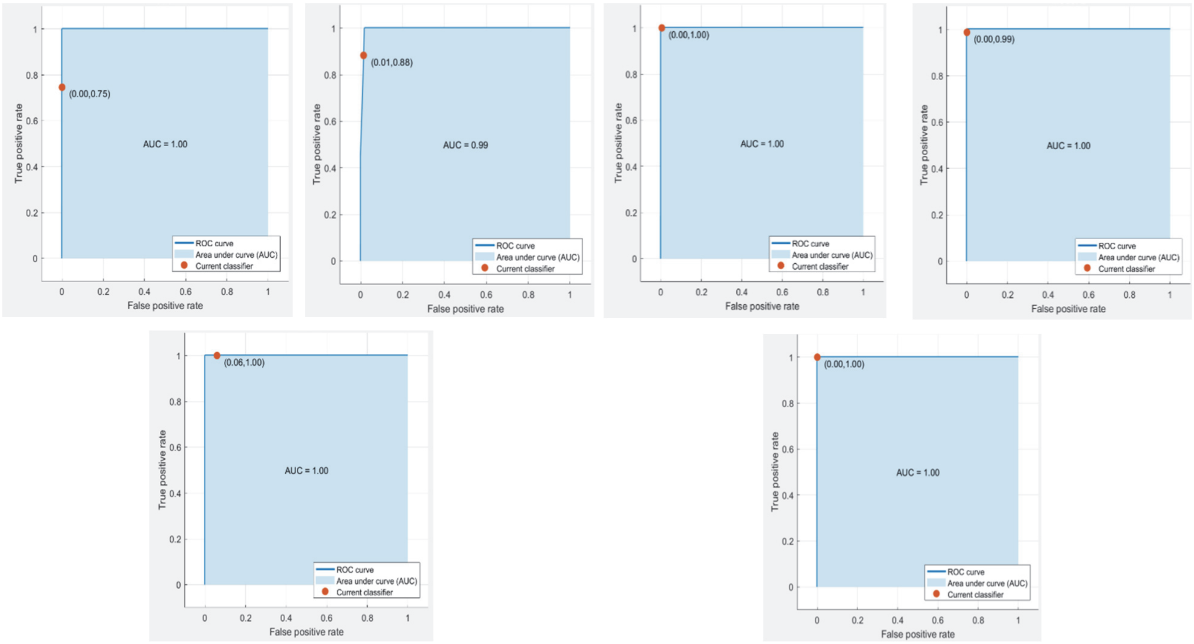

Fig. 5. ROC curve of LDA for $\alpha_2 = 0.05$. Up-left: positive class (PC):1, Negative class (NC):2, 3. Up-right: PC: 2, NC: 1, 3. Bottom: PC: 3, NC: 1, 2.

Fig. 6. ROC curve of QDA for $\alpha_2 = 0.05$. Up-left: positive class (PC):1, Negative class (NC):2, 3. Up-right: PC: 2, NC: 1, 3. Bottom: PC: 3, NC: 1, 2.

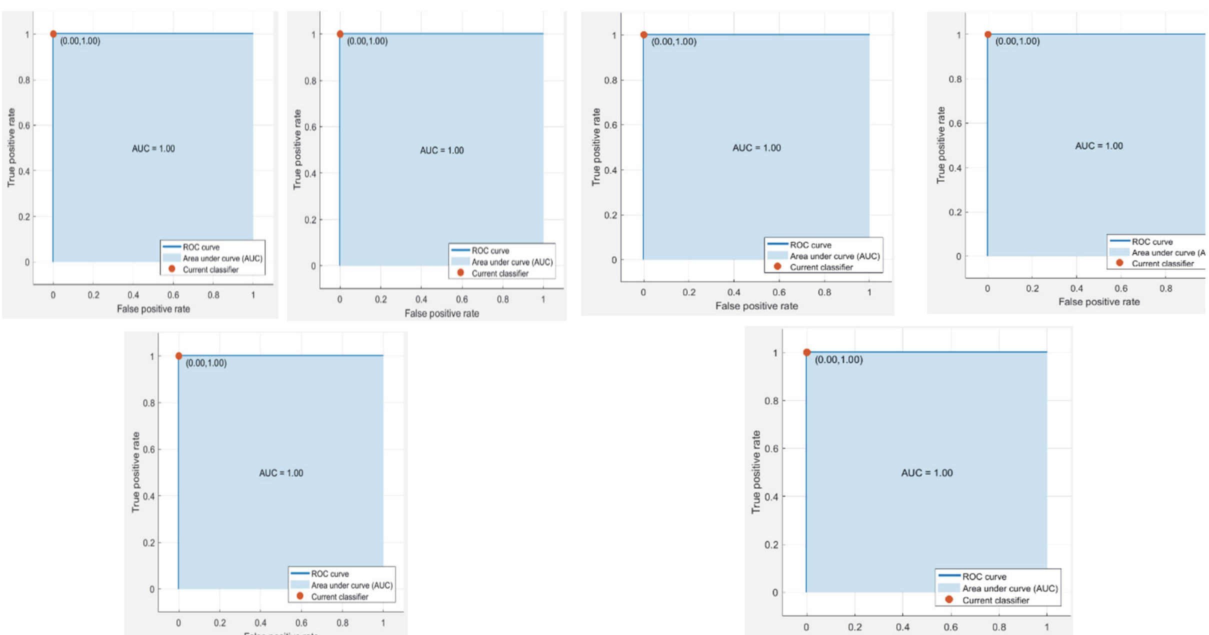

Fig. 7. ROC curve of KNN for $\alpha_2 = 0.05$. Up-left: positive class (PC):1, Negative class (NC):2, 3. Up-right: PC: 2, NC: 1, 3. Bottom: PC: 3, NC: 1, 2.

Fig. 8. ROC curve of SVM for $\alpha_2 = 0.05$. Up-left: positive class (PC):1, Negative class (NC):2, 3. Up-right: PC: 2, NC: 1, 3. Bottom: PC: 3, NC: 1, 2.

To fairly compare the performance of the classifiers, the starting point of these charts must be the same. One of the contributions of the current work is that it can specify warning zones to define rules similar to Western Electrics. These rules can differ based on the type of the process and their sensitivity. In this study, two types of rules called "exit strategies" are defined as follows:

Rule 1: If one observation is classified in the third class, the sample is assumed to be out-of-control.
Rule 2: If one observation is classified in the third class or 3 consecutive observations are classified in the second class (in warning regions), the sample is assumed to be out-of-control.

Adjusted by the above two rules, all of the steps are repeated for the complete and reduced conditions (with the presence of all principal components versus two components that explained more than 90% of the variability). Fig. 9 and Fig. 10demonstrate the performance of the classifiers in reduced and complete cases using the first rule as the exit strategy. One of the most significant points discovered when comparing these two charts is the better performance of the classifiers in the complete condition as opposed to the reduced one. It is also noticeable that in both complete and reduced conditions, KNN and SVM have the best performance. QDA also performs close to these two; however, LDA does not have as good functionality as the others especially when the shift from the controlled state is not large. In Fig. 11 and Fig. 12, ARL curves per one unit change in the mean value is depicted for complete and reduced conditions respectively when the second rule is used as the exit strategy. It is visible that the points mentioned for the previous assumption are also true in these charts. To understand the impacts of defining warning zones and utilizing them to design rules, it is needed to compare Fig. 9 and Fig. 10 with Fig. 11 and Fig. 12, respectively. What is doubtlessly recognizable is that all of the classifiers perform more accurate in classifying out-of-control samples in their correct class when the warning zones are used in the exit strategies.

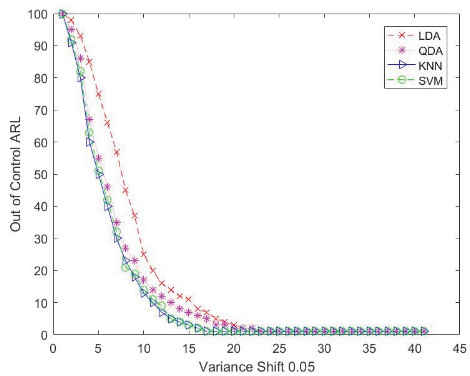

Fig. 9. ARL_1 curves in out-of-control state for 1000 simulations in 40 stages while $\boldsymbol{\alpha = 0.01}$ in complete version. Exit strategy: Rule 1

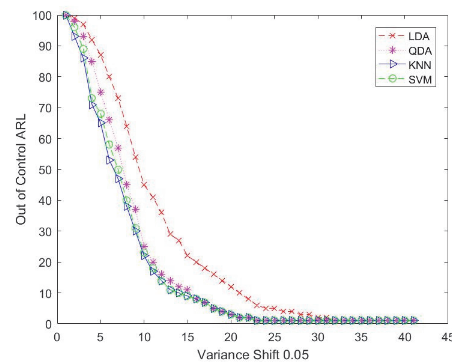

Fig. 10. ARL_1 curves in out-of-control state for 1000 simulations in 40 stages while $\boldsymbol{\alpha = 0.01}$ in reduced version. Exit strategy: Rule 1

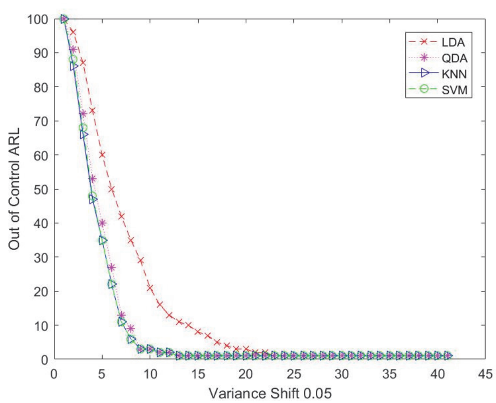

Fig. 11. ARL_1 curves in out-of-control state for 1000 simulations in 40 stages while $\boldsymbol{\alpha = 0.01}$ in complete version. Exit strategy: Rule 2

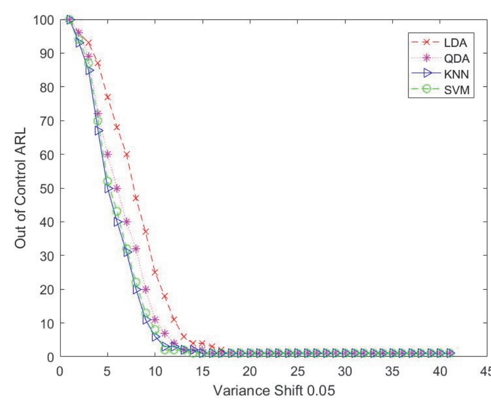

Fig. 12. ARL_1 curves in out-of-control state for 1000 simulations in 40 stages while $\boldsymbol{\alpha = 0.01}$ in reduced version. Exit strategy: Rule 2

5. Conclusion and future research directions

In this paper, a methodology was proposed for controlling image-based processes using supervised statistical learning methods which contributed to higher sensitivity and accuracy level of monitoring processes by defining warning zones. The performance of this methodology was examined on simulated data. The result of such evaluations summarizes into the following points:

1. While the proposed model is mainly dependent on the classification methods, choosing proper methods relies on the type of the process and the structure of data derived from it. In this paper, the classification logic was based on comparing the T^2 statistics of the principal components with the calculated control limits of the Hotelling T^2 method. One of the main shortcomings of the T^2 control chart is the use of the normality assumption. Hence, the simulated data

in this study assumed to follow a multivariate normal distribution with known mean vector and covariance matrix. Even though SVM and KNN methods are not reliant on the distribution of data, LDA and QDA methods assume Gaussian distribution of data in their classification process. As a result, the performance of the chosen classifiers was not influenced by this limitation of the T^2 control chart.

2. Another drawback of the T^2 control chart is its inefficiency to detect small shifts in the mean value. Using the model proposed in this paper, we were able to define sensitivity rules using the warning zones to enhance the model's sensitivity to small shifts.
3. To tackle the high dimension of the data, the reduced model was developed and compared with the complete version in which all of the principal components were present. While the complete version has higher accuracy due to employing 100% of data, more complexity occurs and thus the training process of the classifiers and online monitoring will be highly time-consuming. On the other hand, with the expense of losing little variability of data and gaining less accuracy in calculations, online process monitoring took place with fewer principle components which yielded more than 90% of the information. The results obtained after comparing these two conditions showed that even though the performance of the methods in the reduced version is not as good as the complete one, recovered time and released complexity while maintaining an acceptable performance offers a decent opportunity for online monitoring of the process in less time.
4. The performances of the selected four classifiers were investigated in 3 steps. In the first step, their speed and accuracy in training were measured through cross-validation, confusion matrix, and ROC chart which led us to realize that while the accuracy of all four classifiers was pretty good, LDA's performance was weaker than the others. The reason can be sought in the limiting assumptions of the LDA classifier. One of these assumptions is considering equal variance for each class, which is not necessarily met in simulated data. Besides, the KNN and SVM classifiers and with a distance QDA showed desirable ARL in the in-control state, whilst LDA did not perform as good. Nonetheless, its performance was improved when the significance level increased. The same is true for the out-of-control state, meaning LDA was incapable of recognizing small shifts from the mean value whereas the other methods performed better. It is noticeable that when a warning zone was considered, the outcome was enhanced in all four methods.

After what was achieved and acknowledged through this study, the following points can be suggested to improve the model proposed in this paper:

1. One of the problems that are often present in industrial processes, is the unknown statistical distribution of the data obtained. This issue highlights the importance of employing nonparametric methods for a more efficient investigation of industrial sectors.
2. While the shift in the process and removing defective products are important, diagnosing the location of the deficit can have the same importance. To make this attainable, reversible transformation can be adopted.
3. The parameters of the models are often pre-assumed in many studies, likewise, the parameters used to train KNN and SVM were chosen by trial and error in this paper. One of the points that can be considered in future research is the use of the optimization methods to define such parameters more accurately.
4. Many classifiers do not have a desirable accuracy; therefore, a solution to improve it can be using unsupervised methods such as clustering for the initial separation of data and then using classifiers to train them. With this approach, different concepts can be combined to increase the practical functionality of the classifiers.

References

He, Q. P., Wang, J., & Shah, D. (2019). Feature space monitoring for smart manufacturing via statistics pattern analysis. *Computers & Chemical Engineering 126*, 321-331.

Hotteling, H. (1947). *Multivariate quality control illustrated by the air testing of sample bombsites*. New York: McGraw-Hill: 111.

Jackson, J. E. (1991). *PCA with more than two variables. A user's guide to principal components*. Wiley series in probability and mathematical p. statistics. Applied probability and statistics, 26-62.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning.* Springer.

Kistner, M., Jemwa, G. T., & Aldrich, C. (2013). Monitoring of mineral processing systems by using textural image analysis. *Minerals Engineering, 52*, 169-177.

Kourti, T., & MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology, 28*(4), 409-428.

Ledley, R. S. (1964). High-speed automatic analysis of biomedical pictures. *Science 146* 3641), 216-223.

Liang, D.-T., Deng, W.-Y. Wang, X.-Y., & Zhang, Y. (2009). Multivariate image analysis in Gaussian multi-scale space for defect detection. *Journal of Bionic Engineering 6*(3), 298-305.

Malamas, E. N., Petrakis, E. G., Zervakis, M., Petit, L., & Legat, J.-D. (2003). A survey on industrial vision systems, applications and tools. *Image and Vision Computing 21*(2), 171-188.

Megahed, F. M., Woodall, W. H., & Camelio, J. A. (2011). A review and perspective on control charting with image data. *Journal of Quality Technology 43*(2), 83-98.

Mirschel, G., Daikos, O., & Scherzer, T. (2019). In-line monitoring of the thickness distribution of adhesive layers in black textile laminates by hyperspectral imaging. *Computers & Chemical Engineering, 124*, 317-325.

Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.

Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics, 37*(1), 41-59.

Ottavian, M., Barolo, M., & García-Muñoz, S. (2013). Maintenance of machine vision systems for product quality assessment. Part I. Addressing changes in lighting conditions. *Industrial & Engineering Chemistry Research, 52*(35), 12309-12318.

Pereira, A. C., Reis, M. S., & Saraiva, P. M. (2008). Quality control of food products using image analysis and multivariate statistical tools. *Industrial & Engineering Chemistry Research 48*(2), 988-998.

Phaladiganon, P., Kim, S. B., Chen, V. C., & Jiang, W. (2013). Principal component analysis-based control charts for multivariate nonnormal distributions. *Expert Systems with Applications, 40*(8), 3044-3054.

Prats-Montalbán, J. M., & Ferrer, A. (2007). Integration of colour and textural information in multivariate image analysis: defect detection and classification issues. *Journal of Chemometrics, 21*(1-2), 10-23.

Prats-Montalbán, J. M., & Ferrer, A. (2014). Statistical process control based on multivariate image analysis: A new proposal for monitoring and defect detection. *Computers & Chemical Engineering, 71*, 501-511.

Prats-Montalbán, J., De Juan, A., & Ferrer, A. (2011). Multivariate image analysis: a review with applications. *Chemometrics and Intelligent Laboratory Systems, 107*(1), 1-23.

Prats-Montalbán, J. M., Cocchi, M., & Ferrer, A. (2015). N-way modeling for wavelet filter determination in multivariate image analysis. *Journal of Chemometrics, 29*(6), 379-388.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society*. *Series B (Methodological) 10*(2), 159-203.

Reis, M. S. (2015). An integrated multiscale and multivariate image analysis framework for process monitoring of colour random textures: MSMIA. *Chemometrics and Intelligent Laboratory Systems, 142*, 36-48.

Reis, M. S., & Bauer, A. (2009). Wavelet texture analysis of on-line acquired images for paper formation assessment and monitoring. *Chemometrics and Intelligent Laboratory Systems, 95*(2), 129-137.

Reis, M. S., & Bauer, A. (2010). Image-based classification of paper surface quality using wavelet texture analysis. *Computers & Chemical Engineering, 34*(12), 2014-2021.

Ryan, T. P. (1989). *Statistical methods for quality control*. John Wiley and Sons, New York.

Shewhart, W. A. (1924). Some applications of statistical methods to the analysis of physical and engineering data. *Bell Labs Technical Journal, 3*(1), 43-87.

Singh, A., Dutta, M. K., ParthaSarathi, M., Uher, V., & Burget, R. (2016). Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. *Computer Methods and Programs in Biomedicine, 124*, 108-120.

Strauss, H. W., Zaret, B. L., Hurley, P. J., Natarajan, T., & Pitt, B. (1971). A scintiphotographic method for measuring left ventricular ejection fraction in man without cardiac catheterization. *The American Journal of Cardiology, 28*(5), 575-580.

Szatvanyi, G., Duchesne, C., & Bartolacci, G. (2006). Multivariate image analysis of flames for product quality and combustion control in rotary kilns. *Industrial & Engineering Chemistry Research, 45*(13), 4706-4715.

Wang, T., Xu, R., Han, X., Chen, Y.-W., Ishizaki, Y., Miyamoto, M., & Hattori, T. (2016). A principal component analysis based method to automatically inspect wear of throw-away tips. *Journal of Intelligent & Fuzzy Systems, 31*(2), 903-913.

Wójcik, W., & Kotyra, A. (2009). Combustion diagnosis by image processing. *Photonics Letters of Poland 1*(1), 40-42.

Yan, H., Paynabar, K., & Shi, J. (2015). Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering, 12*(1), 216-227.

Yu, H., & MacGregor, J. F. (2003). Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemometrics and Intelligent Laboratory Systems, 67*(2), 125-144.

Yu, H.,& MacGregor, J. F. (2004). Monitoring flames in an industrial boiler using multivariate image analysis. *AIChE Journal 50*(7), 1474-1483.