Soft Computing
https://doi.org/10.1007/s00500-023-09047-2

APPLICATION OF SOFT COMPUTING

Evolutionary support vector regression for monitoring Poisson profiles

Ali Yeganeh[1] · Saddam Akber Abbasi[2,3] · Sandile Charles Shongwe[1] · Jean-Claude Malela-Majika[4] · Ali Reza Shadman[5]

Accepted: 21 July 2023
© The Author(s) 2023. This article is published with open access at Springerlink.com

Abstract
Many researchers have shown interest in profile monitoring; however, most of the applications in this field of research are developed under the assumption of normal response variable. Little attention has been given to profile monitoring with non-normal response variables, known as general linear models which consists of two main categories (i.e., logistic and Poisson profiles). This paper aims to monitor Poisson profile monitoring problem in Phase II and develops a new robust control chart using support vector regression by incorporating some novel input features and evolutionary training algorithm. The new method is quicker in detecting out-of-control signals as compared to conventional statistical methods. Moreover, the performance of the proposed scheme is further investigated for Poisson profiles with both fixed and random explanatory variables as well as non-parametric profiles. The proposed monitoring scheme is revealed to be superior to its counterparts, including the likelihood ratio test (LRT), multivariate exponentially weighted moving average (MEWMA), LRT-EWMA and other machine learning-based schemes. The simulation results show superiority of the proposed method in profiles with fixed explanatory variables and non-parametric models in nearly all situations while it is not able to be the best in all the simulations when there are with random explanatory variables. A diagnostic method with machine learning approach is also used to identify the parameters of change in the profile. It is shown that the proposed profile diagnosis approach is able to reach acceptable results in comparison with other competitors. A real-life example in monitoring Poisson profiles is also provided to illustrate the implementation of the proposed charting scheme.

Keywords Control charts · Particle swarm optimization · Poisson profiles · Profile monitoring · Statistical process control · Support vector regression

1 Introduction

A control chart demonstrates a virtual representation of a process over time; hence, they are a valuable tool for monitoring process performance in statistical process control (SPC). Control charts have been widely used in industrial processes since the 1920s in various applications; see, for example, Aslam et al. (2020). Usually, control charting in SPC is implemented in two phases: Phase I (retrospective phase) and Phase II (monitoring phase). The aims of Phase I include estimation of process's stability, the in-control (IC) parameters and control limits through a preliminary retrospective study using historical datasets; while, in Phase II analysis, one uses the IC model, obtained from Phase I, as a base scheme to control the process in real time and detect outliers as quickly as possible (Yeganeh et al. 2022a). So, it is vital to control the process over time in Phase II as changes in the process parameters might be caused by some unnatural patterns which might be due to faults, non-compatible products, low quality raw materials and so forth (Montgomery 2019; Gupta et al. 2006). In on-line monitoring (Phase II analysis), the instability of a process, should be identified as early as possible, is declared with the out-of-control (OOC) signals. This is usually evaluated in terms of the required number of drawn points on a control chart before an OOC signal; it is denoted as average run length (*ARL*). In a fair evaluation, the *ARL* value for the IC process, referred to ARL_0, is adjusted at a constant and desired value and the charts endeavour to provide the minimum *ARL* value in OOC condition, called ARL_1 (Yeganeh et al. 2023). Note that the greater the ARL_1, the weaker the detection ability of a control chart (Montgomery 2019). In addition to the *ARL*,

Extended author information available on the last page of the article

Published online: 20 September 2023

control charts' performance metrics include the, performance comparison index (*PCI*), standard deviation of run length (*SDRL*), relative average run length (*RARL*) and extra quadratic loss (*EQL*) (Riaz et al. 2014).

To construct a control chart in SPC either in Phase I or Phase II, two different approaches can be used: (i) investigation of the distribution function of a single or multiple quality characteristics; (ii) checking the stability of a functional relationship between a dependent (response) and one or several independent (explanatory) variables over time. The term profile monitoring hints using of SPC techniques to investigate the stability of the functional relationship over time instead of monitoring a single or multivariate quality characteristic.

It is noteworthy to mention that this topic has firstly been introduced with '*signature*' term in Gardner et al. (1997) and the use of '*profile*' became more commonplace after the extension of exponentially weighted moving average (EWMA) control charts in monitoring simple linear profiles in which there is a linear relationship between a response and an explanatory variable (Kang and Albin 2000). Since then, many researchers have been focusing on monitoring linear profiles; see, for example, Gupta et al. (2006), Zou et al. (2007), Huwang et al. (2014), Motasemi et al. (2017), Haq (2020), Yeganeh et al. (2021) and the references therein. Also, some researchers focused on monitoring other types of profiles including nonlinear (Williams et al. 2007), roundness (Pacella and Semeraro 2011), exponential (Steiner et al. 2016), circular (Zhao et al. 2020), multi-channel autoregressive (Zhou et al. 2022) and non-parametric (Jones et al. 2020; Zou et al. 2008; Nassar and Abdel-Salam 2021).

As a secondary approach, some researchers considered two restricting assumptions in all the monitoring schemes. First, the response variables might not be continuous and have a discrete form. For example, the response variable may demonstrate the number of defects (percent of defective) per item. Secondly, it is usually assumed that the random error and therefore, the response variables follow the normal distribution. For many real-life applications, the normality assumption can be violated. Considering these limitations, generalized linear models (GLM) constituting a large class of statistical models relating responses to linear combinations of predictor variables, have been extended to the profile monitoring regime. There are two major categories of GLMs that have had the most applications in the literature, these are logistic and Poisson profiles. For more details on the former, readers are referred to the works of Yeh et al. (2009), Shang et al. (2011), Huwang et al. (2016), Alevizakos et al. (2019a) and Mohammadzadeh et al. (2021). Recently, many studies are focused on monitoring Poisson profiles, for instance, Zhou et al. (2012) proposed an EWMA chart with random sample sizes where it was observed that using a novel updating formulation, not only yields a more robust IC and OOC performance, but also a generally more sensitive chart to small and moderate shifts. The Phase I monitoring of Poisson profiles has been carried out by Amiri et al. (2015), where they developed three different schemes based on the likelihood ratio test (LRT), Hotelling's T^2 and F statistics. To extend this work in Phase II, Qi et al. (2016) proposed the weighted LRT (WLRT) scheme by combining the EWMA and LRT statistics. They also evaluated the LRT, LRT-EWMA and multivariate EWMA (MEWMA) control charts in Phase II applications of Poisson profiles under the assumption of fixed and random explanatory variables. The results showed the superiority of WLRT over other competitors. Later, Qi et al. (2017) extended the WLRT approach for autocorrelated processes. A change point statistic was developed by Shadman et al. (2017) which was also applied for efficient monitoring of linear profile parameters (Xu et al. 2012) for the GLM profiles. By considering all samples as the candidate change point, they computed LRT statistic for the two groups of samples including IC (or before change point) as well as OOC (or after change point) for both logistic and Poisson profiles. More recently, the change point approach was also implemented for autocorrelated Poisson profiles in He et al. (2020). In addition to the existence of autocorrelation, they assumed random explanatory variables. Another LRT control chart for profiles with random variables was proposed by Song et al. (2021). In addition to random predictors, they considered profiles with auto-correlation. Similar to Shadman et al. (2017), Shang et al. (2018) extended MEWMA charts for Poisson and logistic profiles by considering no prior information about the process in the OOC situations (i.e. non-parametric models). In other words, only the profile parameters have been changed and the type of relation is fixed in the parametric models; whereas, the relation can be transferred to other types, for example, changing a linear IC profile to a non-linear OOC profile, without any limitations. Some remedial methods in parameter estimations of Poisson profiles and computation of process capability index can also be found in Maleki et al. (2019) and Alevizakos et al. (2019b). A non-parametric approach to generalized likelihood ratio and the EWMA schemes for a real case study could be found in Wang et al. (2022).

The investigation of the related mentioned literature and the existing review papers in this field reveal that there is little attention given to machine learning approaches in comparison to statistical approaches; not only in the GLM profiles, but also in all types of profiles monitoring (Maleki et al. 2018; Woodall 2007). To the best of the authors' knowledge, artificial neural network (ANN) has only been used for profile monitoring in Hosseinifard et al. (2011), Pacella and Semeraro (2011), Yeganeh et al. (2022a) and Yeganeh and Shadman (2020). Li et al. (2019) used support vector regression (SVR) technique for function fitting in the non-linear profile monitoring process. Autoencoders and transfer learning are part of the deep learning techniques that have also been developed in autocorrelated (Chen et al. 2020) and multiple profiles (Fallahdizcheh and Wang 2022). One of the main reasons of this reluctance, may be the weaker performance of machine learning techniques than statistical approaches in terms of *ARL*. For example, the ANN-based control chart proposed by Hosseinifard et al. (2011) were not able to improve on the performance of conventional EWMA control charts for the detection of most of the shifts when there is a simple linear profile model. To remedy this weakness, Yeganeh and Shadman (2020) improved the performance of the Hosseinifard et al. (2011)'s control chart using supplementary run-rules but there was no modification implemented in the structure of the ANN by Hosseinifard et al. (2011). The same can be said about other mentioned machine learning-based control charts; in other words, they used a simple conventional structure of ANN or SVR and it may be one of the possible reasons for the weakness in their performance. Another weakness of machine learning techniques may be related to their complexity. Although machine learning techniques are more complex and considered as a "black box", they can produce more accurate results (Cuentas et al. 2022). On the other hand, with the rapid development of digital technologies, the complexity issue of machine learning techniques, in particular deep learning models, are becoming less important over time in real-world applications, in which several complicated on-line models, such as image processing, computer vision and speech recognition, have been developed that can easily be applied in real applications (Chen et al. 2020). In addition, interpreting the prediction of machine learning methods have been studied recently (Pourpanah et al. 2016). However, it is not the focus of this study, but it is an interesting area that requires further investigations.

Considering the limited use of machine learning techniques in profile monitoring, this paper introduces a novel SVR structure as a control chart for monitoring Poisson profiles in Phase II of SPC. The aim of this study is to develop a control chart with quicker detection ability over conventional schemes in Poisson profile monitoring problem which is equivalent to have a more optimized process with lower non-compatible products, cost, waste and other non-desirable outputs. To achieve this, we first define/extract more informative input features and then fed them into a well-known machine learning technique i.e., SVR for training in an offline manner. Finally, the trained models can be implemented to monitor the process online and detect any OOC situations. It makes considerable contributions not only in the input features of the SVR but also in the training procedure to enhance the sensitivity in detecting OOC situations. In addition to improvement in detection ability of a Poisson process with machine learning, other contributions of this paper can be summarized as follows:

- Develop a novel structure of SVR as a base control chart.
- Introduce a new input layer structure for the proposed and other related schemes.
- Taking advantage of a novel training of the proposed SVR.
- Enhance the detection ability of the proposed charting technique in comparison with ANN.
- Evaluation of the performance of the proposed scheme under parametric and non-parametric scenarios.
- Use of the diagnosis procedure in Poisson profiles with SVR.

The rest of this paper is briefly structured as follows: the fundamental framework of Poisson profile model used in Phase II monitoring and the formulations of the two fundamental control chart concepts are briefly presented in Sect. 2. in addition, a brief introduction to some important concepts about SVR is given in Sect. 2, which include the principles of evolved SVRs and a description of particle swarm optimization (PSO) algorithm. Section 3 provides a full description of the proposed SVR-based control chart. Section 4 investigates the performance of the proposed scheme in terms of the *ARL* and *SDRL*. The comparisons with the existing counterparts are presented in Sect. 5. Section 6 presents the diagnosis procedure of the proposed approach, while Sect. 7 provides an illustrative example. Finally, the conclusion, recommendations and future research works are presented in Sect. 8.

2 Preliminaries

2.1 Phase II Poisson profile monitoring

Assume that n observations are collected for the jth ($j = 1, 2, \ldots$) random profile. Let (x_{ijk}, y_{ij}) represent the pairs of observations on two variables from the jth random profile available in the GLM profiles in the form of $\boldsymbol{X_{ij}} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})$, with $i = 1,2,\ldots,n$ and $k = 1,2,\ldots,p$. Since in fixed design monitoring, the explanatory variables are assumed to be constant in each profile, the j indices are omitted in the explanatory variables then $\boldsymbol{X_{ij}} = \boldsymbol{X}_i \; \forall j$. Thus, they can be written as an $n \times p$ matrix denoted as $\tilde{\boldsymbol{X}}$ and can be defined by

$$\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X_1} \\ \vdots \\ \boldsymbol{X_n} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \tag{1}$$

A GLM model in the jth sample consists of the following three main components:

(i) The $1 \times n$ response vector $\boldsymbol{Y_j} = (y_{1j}, y_{2j}, \ldots, y_{nj})$ under the discrete distribution, with the mean $\mu_{ij} = E\left(y_{ij} | (x_{i1}, x_{i2}, \ldots, x_{ip}\right)$, which belongs to the same distribution from the exponential family (e.g., Poisson or binomial distribution). Considering the independency of observations within and between profiles, we have $\boldsymbol{\mu_j} = (\boldsymbol{\mu}_{1j}, \boldsymbol{\mu}_{2j}, \ldots, \boldsymbol{\mu}_{nj})$ for the jth profile.

(ii) The matrix of the independent variables are the same as those in (1).

(iii) The monotone link function g that connects the mean of the response variable to the combination of the linear predictors in the way that $g(\boldsymbol{\mu_j}) = \boldsymbol{\eta_j} = \tilde{\boldsymbol{X}}\widehat{\beta}_j$ where $\boldsymbol{\eta_j}$ is the linear combination of the jth profile parameters $\begin{pmatrix} \widehat{\beta} \\ j \end{pmatrix}$. From the above framework, a Poisson GLM model is given by:

$$\begin{aligned} \boldsymbol{Y_j} &= Poisson\left(\boldsymbol{\mu_j}\right), j = 1,2,\ldots., \\ log\left(\boldsymbol{\mu_j}\right) &= \tilde{\boldsymbol{X}}\beta_j. \end{aligned} \tag{2}$$

The parameters of model (2) denoted by $\widehat{\boldsymbol{\beta}}_j = \left(\widehat{\beta}_{1j}, \widehat{\beta}_{2j}, \ldots, \widehat{\beta}_{pj}\right)$, are estimated with the iterative weighted least squares (IWLS) algorithm in this paper. However, to save space, this is not included here. For more details readers are referred to Yeh et al. (2009) and Amiri et al. (2015). Hence, the aim of monitoring Poisson profiles is defined as the detection of changes in $\widehat{\beta}_j$ from its IC value, denoted with $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \ldots, \boldsymbol{\beta}_{p0})$. Note that monitoring explanatory variables in GLM profiles is as important as monitoring the response variable; see for example, Shang et al. (2011); note though, this is not the focus of this paper.

2.2 Existing control charts

In this subsection, the details of the two existing fundamental approaches based on the MEWMA and LRT schemes are provided (Qi et al. 2016). To simultaneously control the p-dimensional IC coefficient vector $\boldsymbol{\beta}_0$, its estimate $\left(\widehat{\boldsymbol{\beta}_j}\right)$ is scaled as follows:

$$\boldsymbol{Z}_j = \boldsymbol{S}\left(\widehat{\beta}_j - \boldsymbol{\beta}_0\right), \tag{3}$$

with $\boldsymbol{S} = \left(\tilde{\boldsymbol{X}}' W \tilde{\boldsymbol{X}}\right)^{\frac{1}{2}}$ and $\widehat{\boldsymbol{\beta}}_j = \widehat{\beta}_{1j}, \widehat{\beta}_{2j}, \ldots, \widehat{\beta}_{pj}$ where S is a $p \times p$ symmetrical matrix and $\boldsymbol{\beta}_0$ is the IC p-dimensional parameters vector. Considering $\boldsymbol{\mu}_0 = (\boldsymbol{\mu}_{10}, \boldsymbol{\mu}_{20}, \ldots, \boldsymbol{\mu}_{n0})$ and $log(\mu_{i0}) = \boldsymbol{X}'_i \boldsymbol{\beta_0}$, $\boldsymbol{W}$ is an $n \times n$ diagonal symmetrical matrix with the main diagonal elements given by $\boldsymbol{\mu}_{10}$, $\boldsymbol{\mu}_{20}, \ldots$, $\boldsymbol{\mu}_{(n-1)0}$ and $\boldsymbol{\mu}_{n0}$. It is worth mentioning that $\boldsymbol{S}$ depends on $\tilde{\boldsymbol{X}}$ in Eq. (1) hence there will be the varied $\boldsymbol{S}$ matrix, indeed using of $\boldsymbol{S_j}$ instead of $\boldsymbol{S}$, when the explanatory variable is not constant in each profile.

From (3), the EWMA statistic for the scaled p-dimensional parameters vector is defined as:

$$\boldsymbol{E}_j = \lambda \boldsymbol{Z}_j + (1 - \lambda)\boldsymbol{E}_{j-1}, j = 1, 2, \ldots, \tag{4}$$

where $\boldsymbol{E_0}$ is a vector of zeros with size $(p + 1)$ and λ $(0 < \lambda < 1)$ is the EWMA constant (or smoothing parameter) which we considered to be equal to 0.2. Thus, MEWMA statistic is given by:

$$M_j = \boldsymbol{E}'_j \boldsymbol{E}_j. \tag{5}$$

For more details on the LRT formulation for the Poisson profiles, readers are referred to Amiri et al. (Amiri et al. 2015). By taking the logarithm on both sides of the joint likelihood function of the independent observations, the LRT statistic is constructed as:

$$LRT_j = 2\left(l_j\left(\widehat{\boldsymbol{\beta}}_j\right) - l_j(\boldsymbol{\beta}_0)\right), \tag{6}$$

$$l_j\left(\widehat{\boldsymbol{\beta}}_j\right) = \sum_{i=1}^{n} y_{ij}\log(\mu_{ij}) - \sum_{i=1}^{n}\log(\mu_{ij}) - \sum_{i=1}^{n}\log(y_{ij}!),$$

$$l_j(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} y_{ij}\log(\mu_{i0}) - \sum_{i=1}^{n}\log(\mu_{i0}) - \sum_{i=1}^{n}\log(y_{ij}!),$$

$$\mu_{ij} = e^{\mathbf{X}_i'\widehat{\boldsymbol{\beta}}_j},$$

$$\mu_{i0} = e^{\mathbf{X}_i'\boldsymbol{\beta}_0}.$$

The likelihood function which evaluates the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters is demonstrated by $l_j(\cdot)$ in the LRT scheme. It is calculated by consideration of the IC process and estimated parameters of profiles in (6).

2.3 SVR formulation

In 1995, an innovative machine learning method called support vector machine (SVM) was introduced by Vapnik (1995) in order to rectify the drawbacks of the ANN methods especially in classification problems. The idea of SVM is based on the minimization of the training error with empirical or structural risk decreasing. For this aim, the features of a nonlinear problem is mapped to another hyperplane with the aim of maximization of the geometric margins and minimization of the classification error. Although SVM has been able to present itself as a powerful method in supervised learning classification problems, its general form only solves the binary classification problems and some treatments should be applied in multi-class classifications and regression problems. In this paper, SVM for regression (hereafter, SVR) is used and it is briefly described in this section. For more details, an interested reader is referred to Vapnik (1995), Cortes and Vapnik (1995), Vapnik (1998) and Stoean and Stoean (2014).

Similar to other supervised learning techniques, a train dataset is firstly prepared. For simplicity but without loss of generality, we show inputs and targets with B_g and T_g respectively in a way that g is the indicator of samples' indexes. Also, it is assumed that inputs and targets are continuous values with dimension U and 1, and there are G samples in the training dataset. Hence, we have a $G \times (U + 1)$ dataset and it could be shown with (B_g, T_g); $g = 1, 2, \ldots, G$. Conventional SVM and SVR usually utilise the following formulation for establishing a relationship between inputs and outputs (estimated targets)

$$f(B_g) = w\phi(B_g) + b. \tag{7}$$

In (7), a predefined kernel function $\phi(\cdot)$ in combination with some weights (w) and bias (b) are used to carry out the mapping tasks and generally, the aim of training is to reach the best values for weights and bias. To obtain the weights and bias, a soft margin (i.e., a possible acceptable interval) is defined as

$$\varepsilon - \xi_g^- \le f(B_g) - T_g \le \varepsilon + \xi_g^+, \tag{8}$$

where ε is the absolute acceptable difference between the target values and estimated ones; while ξ_g is the generated loss in the *gth* sample from the training dataset. In other words, the loss function is defined as:

$$Loss(f(B_g), T_g) = \begin{cases} 0 & |f(B_g) - T_g| \le \varepsilon \\ |f(B_g) - T_g| - \varepsilon & \text{otherwise} \end{cases} \tag{9}$$

From (9), considering the principle of structural risk minimization, the following minimization problem leads to an optimum hyperplane or weights:

$$\min_{(w,\xi_g^+,\xi_g^-)} \frac{1}{2}\|w\|^2 + C\sum_{g=1}^{G}\left(\xi_g^+ + \xi_g^-\right)$$
$$\text{subject to} \begin{cases} -f(B_g) + T_g + \varepsilon + \xi_g^+ \ge 0 & \forall g, \\ f(B_g) - T_g + \varepsilon + \xi_g^- \ge 0 & \forall g, \\ \xi_g^- \ge 0, \xi_g^+ \ge 0 & \forall g \end{cases} \tag{10}$$

Because of the complexity of the above model and the need to minimize the bias term (b) in the objective function, the corresponding dual form of (10) is often used in the SVR training instead of the primal model in (10). The new dual optimization problem of which the Karush–Kuhn–Tucker (KKT) conditions are utilized in the constraints is defined as follows:

$$\min_{(\alpha_g^+,\alpha_g^-)} \frac{1}{2}(\alpha' H\alpha) + \tilde{q}\alpha$$
$$\text{subject to} \begin{cases} \sum_{g=1}^{G}\left(\alpha_g^+ - \alpha_g^-\right) = 0, & \\ 0 \le \alpha_g^+ \le C & \forall g, \\ 0 \le \alpha_g^- \le C & \forall g, \end{cases} \tag{11}$$

where $\alpha = [\alpha_1^+, \alpha_2^+, \ldots, \alpha_G^+, \alpha_1^-, \alpha_2^-, \ldots, \alpha_G^-]$ is a $2G \times 1$ vector and $\tilde{q} = [-T_1 + \varepsilon, -T_2 + \varepsilon, \ldots, -T_G + \varepsilon, T_1 + \varepsilon, T_2 + \varepsilon, \ldots, T_G + \varepsilon]$ is a $1 \times 2G$ vector. Also, $H = \begin{pmatrix} h & -h \\ -h & h \end{pmatrix}$ is a $2G \times 2G$ matrix of kernel transformation with $h(a,b) = \phi(B(a),B(b))$. Hence, the above optimization problem in (11) has $2G$ variables including $\alpha = [\alpha_1^+, \alpha_2^+, \ldots, \alpha_G^+, \alpha_1^-, \alpha_2^-, \ldots, \alpha_G^-]^T$. The quadratic programming algorithms such as kernel adatron (KA), sequential minimal optimization (SMO), iterative single data algorithm (ISDA) and so forth are used to solve the

above problem. Then, the weights and bias, or equivalently, the estimation of each observation are obtained using the following relations:

$$\begin{aligned} S &= \left\{ g | 0 < \alpha_g^+ - \alpha_g^- < C \right\}, \\ b &= \sum_{s \in S} T_s - \left(\sum_{s \in S} \left(\alpha_s^+ - \alpha_s^- \right) \phi \left(B_g, B_s \right) - \left(\varepsilon \times sign \left(\alpha_s^+ - \alpha_s^- \right) \right) \right), \\ f \left(B_g \right) &= \sum_{s \in S} \left(\alpha_s^+ - \alpha_s^- \right) \phi \left(B_g, B_s \right) + b, \\ g &= 1, 2, \ldots, G. \end{aligned} \tag{12}$$

In (12), S is the support vector set which is usually a small subset of vectors from the training dataset. The magnitude of S or the number of support vectors still depends on hyperparameters including C, ε, kernel function and the structure of the problem. It adjusts the model accuracy by a trade-off between a high-complexity model (increasing the number of support vectors) which may over-fit the data and a large-margin (decreasing the number of support vectors) which will incorrectly classify some of the training data in the interest of better generalization.

2.4 Evolutionary SVR

The combination of meta-heuristic and evolutionary algorithms (EA) with machine learning techniques has received a great deal of attention in the past decade. The main aim of this hybridisation is to use EA in the training or parameter tuning of a machine learning technique (Ojha et al. 2017). As a pioneer work, Kim and Cho (2008) proposed an evolutionary neural network based on the genetic (GA) algorithm in a way that a speciation-based model was established through fitness sharing and then ANN was incorporated by a behaviour knowledge space method. Due to promising results, several other researches have investigated the performance of different EAs such as extended marine predators algorithm (EMPA), gradient-based optimization (GBO), moth-flame optimization (MFO) and water cycle optimization algorithm (WCA) (Adnan et al. 2021a; Ikram et al. 2022a; Kadkhodazadeh and Farzin 2022). The integration of some EA has been recently extended in the literature; for example, Adnan et al. (2021b) implemented the combination of PSO and grey wolf optimization (GWO) algorithms in the training of extreme learning machine (ELM) technique.

Similarly, the paradigm evolutionary SVR (ESVR) refers to the condition of hybridization between EA and SVR. This has been well-received in the literature and can be categorized in three different groups. The first group used EA in the hyperparameter optimization of SVR, see for instance, Adnan et al. (2022), Wang and Du (2014), Ikram et al. (2022b) and Al-Zoubi et al. (2021). In their studies, the optimum values for C, ε and kernel function are acquired with different EAs. In the second group, the EAs were performed with feature selection in combination with parameter optimization in the SVR training (Al-Zoubi et al. 2018; Ziani et al. 2017). In accordance with this paper's objective, in the third group, some researchers have used either the primal or dual problem (see, Eqs. (10) and (11)), with EAs instead of common quadratic problem solvers. However, Arana-Daniel et al. (2016), Zhang et al. (2016) and Dantas Dias and Rocha Neto (2017) used EAs entailing GA, differential evolution (DE), PSO and simulated annealing (SA) in the support vectors identification. One may ask about the preference or application of EAs in comparison with common quadratic solvers. Arana-Daniel et al. (2016) and Dantas Dias and Rocha Neto (2017) reported that the EAs have less computational complexity and are easier to implement than other techniques. With that being said, it is important to note that there is no definite solution to this question and it depends on the nature of the problem.

2.5 PSO algorithm

Note that EA is used in the third group of the ESVR mentioned in the previous subsection. In other words, solving of the dual problem defined in (11) is performed with PSO as one of the versatile algorithms. By questing the related literature, it was revealed that GA and PSO are the most widespread approaches in this application, with PSO more compatible than GA with continuous variables; for example, the reasonable accuracy of PSO in comparison with some other EAs has been reported in Dantas Dias and Rocha Neto (2017). Also, our simulations reveal that PSO made better accuracy than some other common EA techniques (a brief part of the results are illustrated in the sensitivity analysis section). Therefore, in this paper, PSO is used for the solving of SVR optimization problem.

The PSO idea was extracted from migration of a crowd of birds, where the individual knowledge and performance

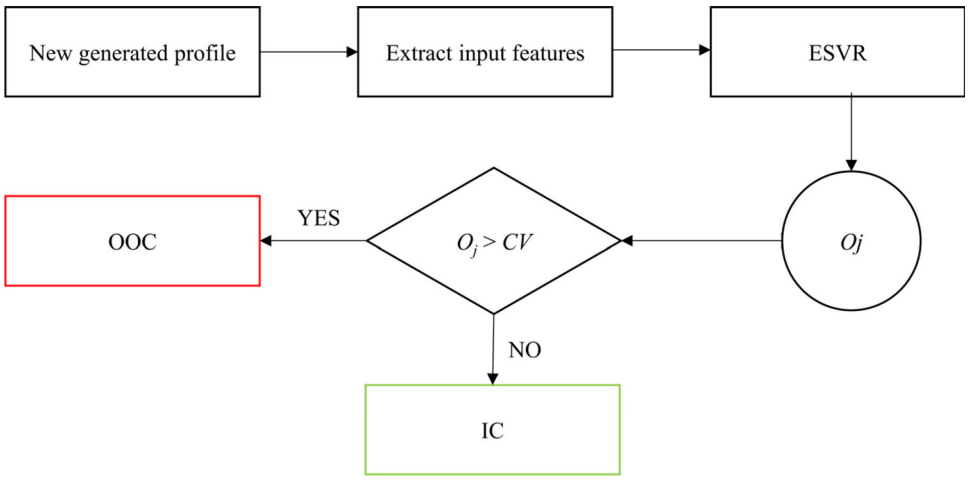

Fig. 1 Basic model of ESVR for determination of process condition at the *jth* generated profile in Phase II

are determined based on the whole population. As a general procedure in metaheuristic algorithms, the best solution is acquired by generation of supreme solutions from a specific population. A specific location and velocity which is determined based on its own best solution, the global current best solution, and some random parameters are considered for each candidate solution (sometimes called particle) in the PSO algorithm. By reaching the best location and velocity, PSO has been made an excellent evolutionary algorithm in continuous variables or nonlinear optimization problem. For more details about the PSO updating relations, the readers are referred to Kennedy (2010). It is inevitable to assign some hyper parameters in the PSO algorithm including (i) the population size, denoted by *npopPSO* here, (ii) iteration numbers (*maxItPSO*), (iii) and (iv) two coefficients for computation of the difference between the current and best position of this solution and best position of all solutions, respectively.

Suppose our proposed method denoted by ESVR hereafter is available after the training procedure. Generally, some features or characteristics of a generated profile are extracted and imported to the ESVR and the condition of the process as an IC or OOC is identified by its output. In this paper, the ESVR output denoted by O_j is compared with a predefined cutting value (*CV*) which is considered as upper control limit (*UCL*) in common control charts. By our training procedure, it is not needed to define the lower control limit (*LCL*); this adaptation is consistent with some previous works; for example see Hosseinifard et al. (2011) in which they adjusted *LCL* at 0. For better understanding, Fig. 1 depicts the conceptual model of ESVR for deciding about the process condition.

The determination of input features would be the next key step; however, not much attention was given to it in the previous machine learning based control charts. For example, the estimation of parameters were only imported as the input features in Hosseinifard et al. (2011); note though, adding other input features would be significant for better detection ability of OOC profiles in general. The proper input features not only should be indicator of the process parameters properties but also have to embrace the effect of the former samples in the contemporaneous statistic. As one of the contributions of this study, the input features of the *jth* generated profile consist of four major groups:

- **The normalized estimated parameters**: After estimation of the parameters by IWLS algorithm and considering a *p*-dimensional normal distribution as $\widehat{\underset{j}{\beta}} \sim N_p\left(\boldsymbol{\beta}_0, \left(\tilde{\boldsymbol{X}}' W \tilde{\boldsymbol{X}}\right)^{-1}\right)$ in Yeh et al. (2009), the estimation of the parameters are scaled through (13) (more details are provided in Sect. 4.2 of Johnson and Wichern (2007):

$$\boldsymbol{\beta}'_{\boldsymbol{j}} = \left(\tilde{\boldsymbol{X}}' W \tilde{\boldsymbol{X}}\right)^{-\frac{1}{2}} \times \left(\widehat{\boldsymbol{\beta}_{\boldsymbol{j}}} - \boldsymbol{\beta}_0\right)^T, \quad \boldsymbol{\beta}'_{\boldsymbol{j}} = \left(\beta'_{1j}, \beta'_{2j}, \ldots, \beta'_{pj}\right). \tag{13}$$

- **The normalized average of responses parameters**: Considering (2), we have $y_{ij} \sim Poisson(\boldsymbol{\mu}_{ij})$ for $i = 1,2,\ldots,n$. Although the exact distribution of the average of the responses in the *jth* profile is not known, the central limit theorem enables us to scale it as:

$$\overline{y}'_j = \frac{\overline{y}_j - \frac{\sum_{i=1}^{n} \mu_{0i}}{n}}{\frac{\sqrt{\sum_{i=1}^{n} \mu_{0i}}}{n}}, \tag{14}$$

in such a way that

$$\overline{y}_j = \frac{\sum_{i=1}^{n} y_{ij}}{n}, \tag{15}$$

where $\log(\boldsymbol{\mu_0})$ and $\boldsymbol{\mu_0}$ have been defined after (3). It is better to import the above parameters as an EWMA form; in other words, an EWMA form of $p+1$ parameters $\left(EWMA_{Pj} = \left[\beta'_{1j}, \beta'_{2j}, ..., \beta'_{pj}, \overline{y}'_j\right]\right)$ are computed in each generated profile as in (4) with the initial values $[0,0,\ldots,0]_{p+1}$.

- **The ratio of MEWMA statistics**: The better detection ability of the runs-rules monitoring scheme proposed by Yeganeh and Shadman (2020) and Yeganeh et al. (2021) led us to adopt a similar approach in this paper. They applied the ratio of points as a supplementary tool to increase the chart's performance. Because of the complexity in the design of run-rules schemes, it is recommended to use the ratio of points in the input features. To this end, the UCL_{MEWMA} are obtained by specifying a desired ARL_0 considering only the MEWMA chart (this has been reported in Table 2 of Qi et al. (2016) for ARL_0 equal to 370). Then, MEWMA statistics are computed using (5) in the *jth* profile and the number of samples in the three different regions entailing $\left(0, \frac{UCL_{MEWMA}}{2}\right)$, $\left(\frac{UCL_{MEWMA}}{2}, UCL_{MEWMA}\right)$ and beyond the control limits $(UCL_{MEWMA}, +\infty)$ which are each denoted as $d^{(1)}_{MEWMA}$, $d^{(2)}_{MEWMA}$ and $d^{(3)}_{MEWMA}$, respectively. The 1×4 vector $\left(\left[\frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}, \frac{d^{(3)}_{MEWMA}}{j}, M_j\right]\right)$ is imported to the ESVR to incorporate the effect of previous samples, in similar fashion as run-rules.
- **The ratio of LRT statistics**: Because LRT chart has superior performance than MEWMA chart for detection of large shifts of Poisson profiles (see for example, Table 3 of Qi et al. (2016)), this statistic is also added in the input features by definition of UCL_{LRT}. Hence, using a similar approach as in the previous point and using (6), the 1×4 vector $\left(\left[\frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}, \frac{d^{(3)}_{LRT}}{j}, LRT_j\right]\right)$ is computed and imported to the ESVR.

By these definitions, ESVR has a $(p+1+4+4)$-dimensional input vector $\left(I_j = \left[EWMA_{Pj}, \frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}, \frac{d^{(3)}_{MEWMA}}{j}, M_j, \frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}, \frac{d^{(3)}_{LRT}}{j}, LRT_j\right]\right)$. Many investigations were conducted to reach the above four groups which are the best input combinations for reaching minimum ARL_1. It is discussed more in the sensitivity analysis by comparing with other ESVRs that have different input structures.

Also, there are some interesting results about the proposed $(p+9)$-dimensional input vector (i.e., I_j) of ESVR that were obtained based on simulations (these results are not reported here to conserve space). First, to estimate W defined after (3), it has been suggested to consider the current profile information instead of the IC values as it was proposed by a number of researchers (see, e.g. Huwang et al. (2016)); whereas, the simulations study revealed better results regarding the IC values; hence, we computed W with IC model instead of current profiles. The same can be said about the estimation of $\overline{y}'_j$ in (14). The second point is about the ratio of samples beyond the control limits $\left(\frac{d^{(3)}_{MEWMA}}{j}, \frac{d^{(3)}_{LRT}}{j}\right)$. The main reason of including it in the proposed method is to improve the robustness of ESVR to large simultaneous shifts. In other words, the effect of MEWMA ratios $\left(\frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}\right)$, LRT ratio shifts $\left(\frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}\right)$, and beyond the control limit ratios $\left(\frac{d^{(3)}_{MEWMA}}{j}, \frac{d^{(3)}_{LRT}}{j}\right)$ are present in small, large single and large simultaneous shifts, respectively. As a third point, one may suggest to use WLRT statistics instead of LRT or MEWMA because of its superiority in performance. This is also possible in this paper and may improve the results; but, WLRT needs more computational time because of its full use of all available samples up to the current time point j, especially for the detection of small shifts.

To conduct the simulations, a $(p+9)$-dimensional input vector of ESVR (I_j) is computed in each generated profile and imported to the ESVR. Then, O_j is compared with CV and an OOC signal is triggered when $CV < O_j$ (see Fig. 1). To compute the ARL_1 and $SDRL_1$, this procedure iterates in several Monte Carlo simulations and the signalling times are stored as the run lengths (RL). For a better illustration, the process of obtaining ARL_1 and $SDRL_1$ for a desired shift with $MaxIt$ iterations is illustrated in Pseudocode 1.

Pseudocode 1. The ARL and SDRL Monte Carlo simulation procedure in ESVR

Compute UCL_{MEWMA} and UCL_{LRT};
Consider a specific shift in the profile parameters; RL = [];
for *iter* = 1:*MaxIt*
$j = 0$; OOC = 0; $d^{(1)}_{MEWMA} = d^{(2)}_{MEWMA} = d^{(3)}_{MEWMA} = d^{(1)}_{LRT} = d^{(2)}_{LRT} = d^{(3)}_{LRT} = 0$;
while OOC == 0
$j = j + 1$;
Generate random profile with sample size n;
Estimate profile parameter;
Calculate $EWMA_{pj}$;
Calculate M_j and LRT_j;
if ($M_j > UCL_{MEWMA}$) $d^{(3)}_{MEWMA} = d^{(3)}_{MEWMA} + 1$;
elseif ($M_j > 0.5UCL_{MEWMA}$) $d^{(2)}_{MEWMA} = d^{(2)}_{MEWMA} + 1$;
else $d^{(1)}_{MEWMA} = d^{(1)}_{MEWMA} + 1$;
end if
if ($LRT_j > UCL_{LRT}$) $d^{(3)}_{LRT} = d^{(3)}_{LRT} + 1$;
elseif ($LRT_j > 0.5UCL_{LRT}$) $d^{(2)}_{LRT} = d^{(2)}_{LRT} + 1$;
else $d^{(1)}_{LRT} = d^{(1)}_{LRT} + 1$;
end if
Calculate I_j;
Import I_j to ESVR and obtain O_j;
if ($O_j > CV$) OOC = 1; **end if**
end while
RL = [RL; j];
end for
$ARL_1 = mean$(RL); $SDRL_1 = std$(RL); % Note that with IC profiles it is equal to ARL_0 and $SDRL_0$

3 Training of the proposed method

In the previous section, it has been assumed that the ESVR has been trained. To train it, a $G \times (p + 10)$ training dataset similar to the one in Hosseinifard et al. (2011) is generated. To this end, *0.5G* IC profiles and *0.5G* OOC profiles (with some desired shifts) are generated and a $(p + 9)$-dimensional input vector is computed for each generated profile. The target values for IC and OOC profiles are 0 and 1, respectively, or $T_1 = T_2 = ... = T_{\frac{G}{2}} = 0; T_{1+\frac{G}{2}} = T_{2+\frac{G}{2}} = ... = T_G = 1$.

As mentioned previously, the dual problem in (11) is solved using PSO algorithm and the optimum (minimum) value of objective function is reached by assigning the values of existing $2G$ variables $\left(\alpha = \left[\alpha_1^+, \alpha_2^+, ..., \alpha_G^+, \alpha_1^-, \alpha_2^-, ..., \alpha_G^-\right]\right)$. But, as one of the contributions of this paper, some scanty changes are added to the objective function of (11). This model was introduced by Zhang et al. (2016) in which an additional coefficient is added to objective function of primal problem.

Before the identification of the additional terms to objective function, it is better to be aware of the challenge on the relation between common accuracy criteria entailing mean square error (MSE) and *ARL* when designing control charts based on machine learning techniques. In a usual ANN, SVM, SVR and other machine learning techniques, the training process continues until reaching a desired threshold; whereas in the case of classical Phase II, control charts are evaluated in terms of the *ARL* and there is no

direct relationship between the two approaches. This challenge has been discussed in details by Yeganeh and Shadman (2020) who suggested heuristic solution similar to the design of experiment approach for ANN training which is not the focus in this paper.

Since the 0 and 1 values have been assigned for the IC and OOC profiles as the target values, respectively, and the process condition is identified through the *CV*, it was observed that the higher performance (i.e., lower ARL_1 for a desired ARL_0) occurs when the difference between OOC and IC estimated target values is at the maximum value. In other words, in a common situation, the outputs of ANN or SVR tend towards 0 and 1 in IC and OOC profiles and the *CV* is obtained closer to 1 to reach a desired ARL_0 (see Hosseinifard et al. (2011)) while the higher the difference between the outputs, the lower the ARL_1 value. Thus, some criteria are needed to depict the significance of the difference between IC and OOC ESVR outputs.

To this end, the output of each input in the dataset is obtained using (12). Suppose that it is denoted by $\hat{T}_g; g = 1, 2, ..., G$ which is equivalent to $f(B_g)$ given in (12) where $g = 1, 2, ..., \frac{G}{2}$ are the predicted IC values and others are for OOC profiles. Therefore, the dual problem can be revised as follows:

$$\begin{aligned}
&\min_{\left(\alpha_g^+, \alpha_g^-\right)} \quad MSE + DAVE + DR \\
&\text{subject to} \begin{cases} \sum_{g=1}^{G} \left(\alpha_g^+ - \alpha_g^-\right) = 0, \\ 0 \le \alpha_g^+ \le C \qquad \forall g, \\ 0 \le \alpha_g^- \le C \qquad \forall g, \end{cases} \\
&MSE = \frac{\sum_{g=1}^{G} \left(T_g - \hat{T}_g\right)^2}{G}, \\
&DAVE = \frac{\sum_{g=1}^{\frac{G}{2}} \left(\hat{T}_g\right)}{\frac{G}{2}} - \frac{\sum_{g=1+\frac{G}{2}}^{G} \left(\hat{T}_g\right)}{\frac{G}{2}}, \\
&DR = \underset{g=1,2,...,\frac{G}{2}}{range} \left(\hat{T}_g\right) - \underset{g=1+\frac{G}{2},2+\frac{G}{2},...,G}{range} \left(\hat{T}_g\right).
\end{aligned} \tag{16}$$

The proposed objective function consists of three components. The first one is the common MSE criteria which is frequently utilized in machine learning applications but as mentioned, it cannot lead to minimum ARL_1; hence, the two other components entailing the difference of averages (DAVE) and the difference between the range (i.e., maximum–minimum) of outputs (DR) are appended to the objective function. It is proved in simulation studies that the proposed training approach converges to a solution with minimum ARL_1 values or equivalently, quicker OOC detection ability. Based on our simulations, the effect of these is that the IC and OOC outputs have the maximum difference which will lead to minimum ARL_1 while MSE scales the outputs and hinders increasing the output values. Note that, the obtained *CV* can be increased above 1 to reach a specific ARL_0.

The ideal values of MSE, DAVE and DR values are 0, -1 and -1, respectively, when $\hat{T}_1 = \hat{T}_2 = ... = \hat{T}_{\frac{G}{2}} = 0; \hat{T}_{1+\frac{G}{2}} = \hat{T}_{2+\frac{G}{2}} = ... = \hat{T}_G = 1$; thus, the best value of the proposed objective function is -2. Note that the first term of primary objective function in (11) (i.e. $0.5(\alpha' H \alpha) + \tilde{q}\alpha$) may also be included in (16); however, this is not recommended since it is a redundant term with the same outcome in this condition and it only makes it more complex.

Considering a training dataset with G elements, the optimization problem in (16) with $2G$ variables should be solved by PSO. To this end, initial solutions with size *npopPSO* are randomly generated and updated with PSO algorithm in a way that each member of the population is a vector with size $2G$. By changing the variables' values $\left(\alpha_g^+, \alpha_g^-\right)$ in (16) during the PSO implementation, the output of each input $\left(\hat{T}_g; g = 1, 2, ..., G\right)$ is computed by (12); in other words, function evaluation process (obtaining objective function) is carried out by computation of $\hat{T}_g$ for $g = 1, 2, \ldots, G$. The process terminates when the objective function reaches its ideal value or *BestSol* (i.e., -2) or the iteration number exceeds *maxItPSO*. The framework of ESVR training is illustrated in Pseudocode 2.

Pseudocode 2. The ESVR training procedure

Assume the predefined value of G, *npopPSO* and *maxItPSO*; % Hyper parameters

Set *iterPSO* = 1 and *BestSol* = *bestsolution* = -2;

Create a train dataset with size $G\times(p+10)$;

Generate a population with size *npopPSO* from the train dataset in such a way that each member consists $2G$ elements;

while *bestsolution* != *BestSol* || *iterPSO* < *maxItPSO*

Obtain the G predicted values ($\hat{T}_g; g=1,2,...,G$) for each member of population;

% For each of the *npopPSO* member, there are G predicted values at the end of this step

Compute the loss function in (16) for each member based on the predicted values;

% There are *npopPSO* loss values at the end of this step

Assign the *bestsolution* as the minimum of the computed loss functions;

iterPSO = *iterPSO* + 1;

update each member of population;

% We have *npopPSO* members at this step

end while

bestsolution is the final weights and biases of ESVR;

After training the ESVR, the *CV* value is adjusted such that the desired ARL_0 is reached. It is done using the algorithm provided in Pseudocode 1, on the condition that there are no shifts in the profile generation. Note that UCL_{MEWMA} and UCL_{LRT} are constant during the training.

Table 1 The preassigned model parameters

C	*ε*	*Kernel function*	
SVR parameters			
1	0.15	Linear	
npopPSO	*maxItPSO*	c_1	c_2
PSO parameters			
100	300	1.5	2
MaxIt	ARL_0	*Software*	
Simulation parameters			
10000	370	MATLAB 2018	
G	*λ*		
Training parameters			
2400	0.2		

4 Performance comparisons

Motivated by Qi et al. (2016) and Shang et al. (2018), three different OOC situations entailing parametric model with fixed explanatory variable, parametric model with nonfixed explanatory variable and non-parametric model have been simulated in this section to evaluate the performance of the proposed approach (the competitors' results are extracted from the above mentioned references). The model parameters are provided in Table 1.

Based on the Qi et al. (2016), the IC model for the fixed and random design points was assumed as:

$$\boldsymbol{\beta_0} = [1\ 1], \quad \tilde{\boldsymbol{X}}' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0.1 & 0.2 & \cdots & 1 \end{pmatrix}, \quad n = 10, p = 2. \tag{17}$$

The OOC profiles parameters (denoted as $\boldsymbol{\beta_{OOC}}$) were generated such that

$$\boldsymbol{\beta_{OOC}} = \boldsymbol{\beta_0} + \Delta, \quad \sigma_1 = 0.3518, \sigma_2 = 0.5095, \tag{18}$$

where $\Delta = (\delta_1\sigma_1, \delta_2\sigma_2)$ represents the magnitude of shifts in terms of standard deviation. To generate training dataset,

Table 2 Comparison of ARL_1 ($SDRL_1$) for the Poisson profiles with fixed design points

Shift size	Method				
(δ_1, δ_2)	LRT	MEWMA	LRT-EWMA	WLRT	ESVR
(0.2,0)	201.0 (199.0)	365.0 (366.0)	153.0 (151.0)	44.8 (39.6)	**30.1 (38.0)**
(0,0.2)	202.0 (209.0)	265.0 (255.0)	154.0 (152.0)	45.1 (39.5)	**28.7 (35.2)**
(0,0.25)	151.0 (152.0)	130.0 (123.0)	102.0 (99.7)	27.5 (22.4)	**16.7 (19.5)**
(0.31,0)	106.0 (107.0)	70.8 (61.6)	62.9 (58.5)	17.5 (13.2)	**11.0 (12.3)**
(0.2,0.2)	64.0 (65.7)	28.7 (21.2)	33.2 (29.3)	10.8 (7.1)	**7.1 (6.2)**
(0.5,0)	33.9 (33.8)	14.1 (8.2)	16.3 (12.5)	6.9 (3.7)	**4.6 (3.7)**
(0.32,0.32)	16.1 (15.4)	8.3 (3.6)	8.5 (5.6)	4.8 (2.2)	**2.9 (2.1)**
(0,0.7)	10.6 (10.2)	6.9 (2.5)	6.2 (3.7)	4.0 (1.6)	**2.5 (1.5)**
(0.44,0.44)	5.3 (4.7)	4.8 (1.5)	3.9 (2.1)	3.0 (1.1)	**1.8 (1.1)**
(0.59,0.59)	2.0 (1.5)	3.2 (0.7)	2.1 (1.0)	2.1 (0.6)	**1.2 (0.5)**
(1,1)	1.0 (0.1)	2.0 (0.2)	1.0 (0.2)	1.1 (0.3)	**1.0 (0.0)**

in addition to 1200 IC profiles, three different conditions with 400 OOC profiles with the magnitude of shifts $\delta_1 = 0.2$, $\delta_2 = 0.2$ and $\delta_1 = \delta_2 = 0.2$ were generated when $G = 2400$.

4.1 ARL_1 values for the fixed design points condition

Four competitors (namely: LRT, MEWMA, LRT-EWMA and WLRT schemes) were compared with the proposed ESVR method. After the training procedure, the *CV* was set equal to 2.12 by simulation to reach $ARL_0 = 370$ (i.e., implementation of Pseudocode 1's procedure with no shift). Table 2 reports the values of ARL_1 and $SDRL_1$ in parentheses at different shift magnitudes. It is noteworthy to mention that the boldfaced values denote the best performing scheme.

The proposed ESVR scheme yielded lower values of ARL_1 and $SDRL_1$ regardless of the size of the shifts which gave an advantage to ESVR over other competitors. Tangible reduction in the values of ARL_1 and $SDRL_1$ can be seen for most of the shifts; for example, the ARL_1 in the first row were obtained as 30.1, 44.8, 153, 201 and 365 for ESVR, WLRT, LRT-EWMA, LRT and MEWMA schemes, respectively. Comparable performance in a wide range of shifts indicated that the training procedure of the ESVR would be very good as it made ESVR a robust control chart over different types of shifts. In other words, although only one shift magnitude has been taught to the ESVR during the training, it detected other OOC shifts as quick as possible.

As another finding, the MEWMA and LRT schemes have weaker performances than the WLRT for most of the shifts, which reveals that the combination of two or more control charts (such as combination of EWMA and LRT for construction of WLRT and LRT-EWMA) could increase the performances of the resulting methods (Qi et al. 2017). It is more evident that the WLRT performs

Table 3 Comparison of ARL_1 ($SDRL_1$) for the Poisson profiles with random design points

Shift size	Method				
(δ_1, δ_2)	LRT	MEWMA	LRT-EWMA	WLRT	ESVR
(0,0)	368.0 (371.0)	319.0 (308.0)	368.0 (370.0)	369.0 (371.0)	378.9 (538.5)
(0.2,0)	213.0 (216.0)	460.0 (458.0)	164.0 (159.0)	51.0 (**45.2**)	**33.8** (46.3)
(0,0.2)	208.0 (212.0)	312.0 (305.0)	162.0 (157.0)	49.8 (**44.5**)	**38.2** (53.5)
(0,0.25)	157.0 (158.0)	161.0 (157.0)	111.0 (106.0)	31.2 (**25.8**)	**19.4** (26.3)
(0.31,0)	114.0 (113.0)	91.0 (83.1)	70.8 (66.3)	19.9 (15.3)	**12.2 (13.3)**
(0.2,0.2)	71.0 (70.9)	34.8 (27.0)	38.2 (33.7)	12.1 (8.1)	**6.5 (5.8)**
(0.5,0)	38.6 (38.4)	16.4 (10.3)	18.8 (14.7)	7.8 (**4.2**)	**5.0 (4.2)**
(0.32,0.32)	18.7 (18.2)	9.2 (4.4)	9.8 (6.7)	5.3 (2.5)	**3.1 (2.1)**
(0,0.7)	12.2 (12.0)	7.3 (2.9)	7.1 (4.4)	4.4 (1.9)	**2.6 (1.7)**
(0.44,0.44)	6.1 (5.4)	5.2 (1.7)	4.3 (2.4)	3.3 (1.2)	**1.8 (1.1)**
(0.59,0.59)	2.3 (1.7)	3.5 (0.8)	2.3 (1.1)	2.3 (0.7)	**1.2 (0.5)**
(1,1)	1.0 (0.1)	2.0 (0.2)	1.0 (0.2)	1.2 (0.4)	**1.0 (0.0)**

much better, especially for small and moderate shifts; for instance, when ($\delta_1 = 0.2$, $\delta_2 = 0$) the ARL_1 and $SDRL_1$ of the WLRT are 5 times less than that of the LRT. This idea can be extended to the ESVR schemes to increase their performances. Based on this fact, we can conclude that one of the main reasons for the superior performance of ESVR could be the combination of the LRT and MEWMA statistics.

4.2 ARL_1 values for the random design points condition

Similar to Qi et al. (2016), random explanatory variables with $n = 9$ were generated using the same IC model. To this end, a random number from a discrete uniform distribution over the integers from 1 to 10 were selected and afterwards deleted to construct a random design point with $n = 9$. To make a more robust scheme, a new ESVR was not trained in this case and only the data generation procedure was changed based on the random design points. The results of this case are gathered in Table 3.

Since the same control limits from the previous subsection were used, the ARL_0 is not exactly equal to 370 (the results of IC situation are shown in the first row). The OOC results revealed the superiority of the newly proposed ESVR in case of random design points over other competitors with similar properties as the one for fixed design points. Regardless of the size of the shift, the ESVR performs better followed by the WLRT chart in terms of the ARL values. On the other hand, the performance in term of SDRL was higher as compared to that of its competitors for moderate and large shifts; while, the WLRT performed better among all competing charts for small shifts. We can refer, for instance, to the shift with magnitude ($\delta_1 = 0.2$, $\delta_1 = 0$) in which the minimum $SDRL_1$ i.e., 45.2 was related to WLRT. The same conclusions can be drawn for smaller shifts. Shang et al. (2011) and Song et al. (2021) hinted to complexity of motioning profiles with random design points. Comparing the results of Tables 2 and 3, it can be seen that both ARL_1 and $SDRL_1$ are much larger for a random design, which confirms Shang et al. (2011) and Song et al. (2021)' s findings.

Table 4 Comparison of ARL_1 ($SDRL_1$) for the non-parametric Poisson profiles in scenario I

Shift size	Method		
δ_3	LRT	MEWMA	ESVR
0.10	324.9 (322.2)	**135.3 (122.4)**	141.4 (191.9)
0.20	138.0 (127.7)	43.9 (**25.7**)	**33.4** (43.0)
0.30	72.6 (69.3)	23.0 (11.0)	**10.1 (10.1)**
0.40	34.8 (29.4)	15.8 (6.5)	**6.2 (5.7)**
0.50	20.5 (18.5)	11.3 (3.9)	**4.0 (3.2)**
0.60	9.0 (9.4)	8.4 (2.1)	**2.5 (1.8)**
0.70	7.1 (7.7)	7.3 (1.6)	**2.1 (1.3)**
0.80	3.4 (3.1)	6.1 (1.3)	**1.6 (0.9)**
0.90	2.4 (2.2)	5.2 (0.9)	**1.4 (0.7)**
1.00	1.6 (1.1)	4.5 (0.7)	**1.1 (0.5)**

Table 5 Comparison of ARL_1 ($SDRL_1$) for the non-parametric Poisson profiles in scenario II

Shift size	Method		
δ_3	LRT	MEWMA	ESVR
0.10	313.2 (342.5)	339.8 (307.9)	**156.6 (228.4)**
0.20	228.9 (222.4)	98.5 (78.2)	**51.0 (68.8)**
0.30	131.2 (133.9)	41.5 (24.1)	**19.3 (29.9)**
0.40	73.5 (73.1)	22.9 (9.8)	**9.4 (8.9)**
0.50	45.2 (42.5)	15.7 (5.4)	**6.1 (5.4)**
0.60	23.3 (22.0)	12.0 (4.0)	**4.0 (3.7)**
0.70	12.6 (11.3)	9.5 (2.2)	**2.7 (2.2)**
0.80	7.7 (7.7)	7.7 (1.8)	**2.3 (1.4)**
0.90	4.4 (3.8)	6.7 (1.3)	**1.8 (1.1)**
1.00	3.0 (2.5)	5.8 (1.0)	**1.4 (0.8)**

4.3 ARL_1 values for the non-parametric condition

Non-parametric monitoring refers to the OOC conditions where the type of OOC profile is not known and the IC model can transform to any possible shape. Due to whole changing in the profile relationship, the OOC situation is usually denoted with some scenarios in non-parametric conditions (Zou et al. 2008; Shang et al. 2018; Abbasi et al. 2022). Note that there is no research in non-parametric Poisson profiles so all the values have been achieved by our simulations. To simulate this situation, two different OOC scenarios were investigated with the fixed design points and the IC model of (17). Equations (19) and (20) represent the OOC model in each scenario. Equation (19) presents the OOC model of scenario I.

$$\begin{aligned} \boldsymbol{Y}_j &= Poisson\left(\boldsymbol{\mu}_j\right), \\ & j = 1,2,\ldots, \\ log\left(\boldsymbol{\mu}_j\right) &= \tilde{\boldsymbol{X}}\beta_j + \delta_3 \boldsymbol{cos}\left(2\pi \tilde{\boldsymbol{X}}'_{pure}\right), \end{aligned} \tag{19}$$

where $\tilde{\boldsymbol{X}}'_{pure}$ is the indicator of the explanatory variables without intercept term or

$$\tilde{\boldsymbol{X}}'_{pure} = (0.1 \quad 0.2 \quad \cdots \quad 1)$$

. The results of ARL_1 ($SDRL_1$) for this scenario are displayed in Table 4.

Equation (20) provides the OOC model of scenario II. The results of ARL_1 ($SDRL_1$) for this scenario are displayed in Table 5.

$$\begin{aligned} \boldsymbol{Y}_j &= Poisson\left(\boldsymbol{\mu}_j\right), \\ j &= 1,2,\ldots., \\ log\left(\boldsymbol{\mu}_j\right) &= \tilde{\boldsymbol{X}}\beta_j + \frac{\delta_3}{\tilde{\boldsymbol{X}}'_{pure}}, \\ \tilde{\boldsymbol{X}}'_{pure} &= (0.1 \quad 0.2 \quad \cdots \quad 1). \end{aligned} \tag{20}$$

Comparing two scenarios, sooner OOC detection was raised in the first scenario for nearly all control charts; for example, the ARL_1 ($SDRL_1$) for LRT scheme were 138.0 (127.7) and 228.9 (222.4), respectively, in each scenario. It indicates that the cyclic patterns based on the Eq. (19) can be detected easier in comparison with OOC term in Eq. (20). The ESVR scheme turned out to be the best among all other competing schemes. In general, compared with the LRT and MEWMA schemes, the ESVR scheme has both robustness and sensitivity to the complete changes in the profile type for different shifts; whereas, the MEWMA has lower $SDRL_1$ for small shifts of scenario I. That said, its performance was not as comparable as in the second scenario, which is an indicator of lack of robustness in terms the size of the shift. One of the main reasons for this phenomenon is that most of the existing statistical control charts have been extended based on some fundamental assumptions about the properties of the process (Montgomery 2019; Gupta et al. 2006). While the process condition completely follows these assumptions, the statistical control charts can perform well but it deteriorates its detection ability in other situations such as non-parametric models, complex relation forms and so forth. In these conditions, several researches have stated that machine learning techniques could be superior than statistical methods and had more robustness (Yeganah et al. 2022a; b; Pacella and Semeraro 2011; Mohammadzadeh et al. 2021; Chen et al. 2020). As it was expected that ESVR scheme, as a machine learning technique, would outperform the statistical approaches due to its nature, the lowest values of ARL_1 and $SDRL_1$ have belonged to it.

5 Sensitivity analysis

This section provides six different sensitivity analyses. First, the effect of the proposed input structure and training algorithm against other machine learning techniques are evaluated. Secondly, the detection ability with other desired ARL_0 is reported. Thirdly, a sensitivity analysis for different n values is performed. In the fourth part, the detection ability is increased with some run-rules. The effect of PSO in training of ESVR is investigated in the fifth part and finally, the best performance of the proposed input structure is depicted on the last subsection. All the simulations have the same setups as in Sect. 5.1.

5.1 ARL_1 comparisons under different machine learning techniques

To show the capability of the proposed input layer and training method, four different scenarios were designed with ANN and usual SVRs. In the first state, a common ANN with back-propagation algorithm called ANN-BP1 was trained similar with Hosseinifard et al. (2011) i.e. the inputs were estimation of coefficients. Then, the proposed input layer structure was also trained in a similar way with ESVR (i.e., assigning 11 neurons in input layer) called ANN-BP2. Both techniques have been trained with '*feedforwardnet*' function in MATLAB 2018 and have two

Table 6 Comparison of ARL_1 ($SDRL_1$) for different machine learning schemes in fixed design points

Shift size	Method				
(δ_1, δ_2)	ANN-BP1	ANN-BP2	SVR1	SVR2	ESVR
(0.2,0.0)	44.2 (27.7)	28.6 (34.0)	**24.1 (13.9)**	33.7 (29.8)	30.1 (38.0)
(0.0,0.2)	36.2 (21.9)	24.0 (26.6)	**22.6 (15.4)**	32.4 (29.8)	28.7 (35.2)
(0.0,0.25)	25.2 (12.7)	18.0 (18.1)	**16.5 (8.1)**	26.6 (27.7)	16.7 (19.5)
(0.31,0.0)	21.2 (9.5)	14.2 (16.2)	14.0 (6.7)	18.1 (20.2)	**11.0 (12.3)**
(0.2,0.2)	14.4 (5.3)	8.4 (8.2)	9.3 (3.7)	14.1 (14.4)	**7.1 (6.2)**
(0.5,0.0)	10.7 (3.0)	6.0 (6.2)	8.0 (3.0)	9.1 (11.8)	**4.6 (3.7)**
(0.32,0.32)	7.7 (2.0)	5.2 (8.2)	5.6 (1.7)	5.0 (4.8)	**2.9 (2.1)**
(0.0,0.70)	6.6 (1.5)	5.0 (10.3)	4.8 (1.4)	4.5 (4.5)	**2.5 (1.5)**
(0.44,0.44)	5.4 (1.2)	7.7 (31.7)	3.9 (1.0)	2.4 (2.1)	**1.8 (1.1)**
(0.59,0.59)	4.0 (0.7)	–	2.8 (0.6)	1.5 (1.1)	**1.2 (0.5)**
(1.0,1.0)	2.4 (0.5)	–	1.7 (0.5)	**1.0 (0.0)**	**1.0 (0.0)**

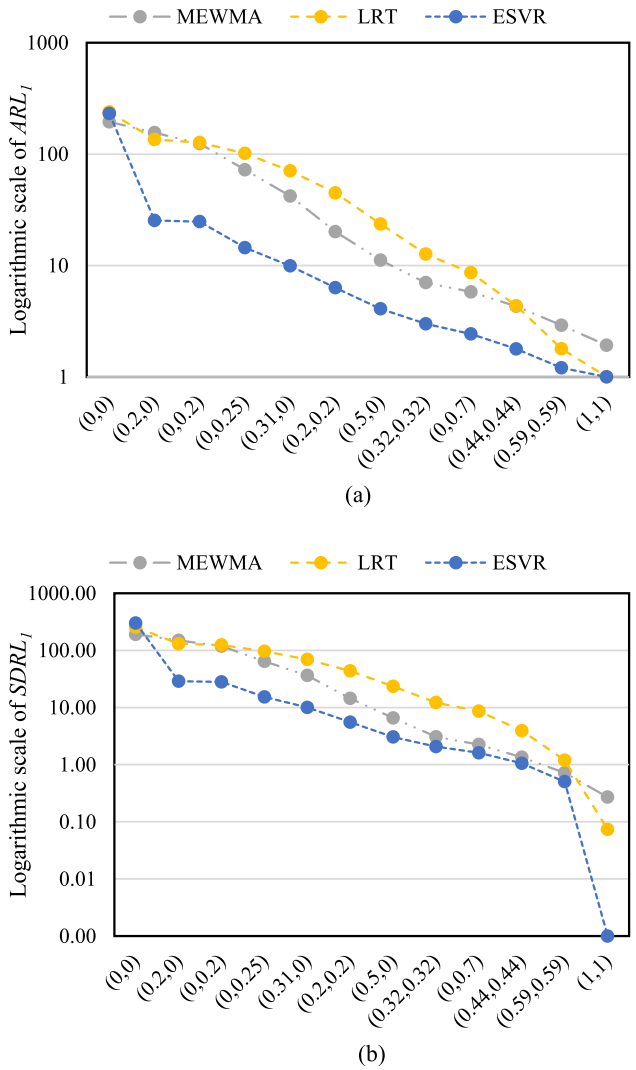

Fig. 2 The results of **a** ARL_1 and **b** $SDRL_1$ values for MEWMA, LRT and ESVR methods when $ARL_0 = 200$

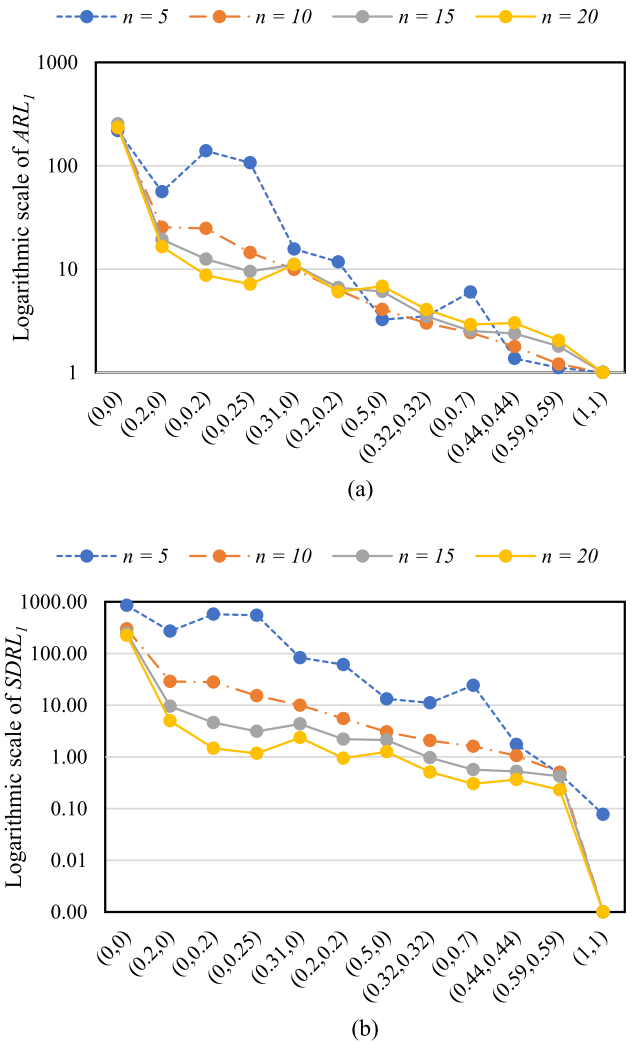

Fig. 3 The results of **a** ARL_1 and **b** $SDRL_1$ values for different n when $ARL_0 = 200$

hidden layers. Moreover, two SVRs called SVR1 and SVR2 have been trained with similar inputs with ANN-BP1 and ANN-BP2 with '*fitrsvm*' function. By these adjustments, ANN-BP1 and SVR1 assessed the proposed input structure while the performance of training method was evaluated under ANN-BP2 and SVR2. The results of these setups are displayed in Table 6.

As it can be seen, the ESVR outperformed ANN-BP1, ANNBP2 and SVR2 for most of the shifts; whereas, for small single positive shifts, it had weaker performance than SVR1. Although the ESVR had lower ARL_1 values than SVR1 for most of the shifts, the simple training of SVR1 might violate the efficiency of ESVR. By more concentration in ARL_1 of different shifts including negative shift and simultaneous positive and negative shifts (the results were not shown), it was revealed that SVR1 came across with the bias effect which means the inability of detecting some shifts (this is also shown for ANN-BP2 in the last two rows of Table 6). As pointed out by Huwang et al. (2014), machine learning techniques suffers from bias effect which means that for such control charts, the OOC signals are not triggered for some shifts; thus, some remedial actions should be considered. However, this bias and poor detection ability did not exist in the case of ESVR scheme.

5.2 ARL_1 comparisons under an ARL_0 value of 200

In Fig. 2, the simulation adjustments were carried out to have $ARL_0 = 200$ under the fixed design point and the ARL_1 and $SDRL_1$ are reported in Panels (a) and (b), respectively. Figures 2a, b illustrate that the ESVR scheme remains superior in the new condition, which reveals the apparent complete robustness of ESVR scheme in terms of the type I error and/or ARL_0 (Abbas

et al. 2016). The comparative analysis in Fig. 2 remains valid for other different values of ARL_0 but they are not reported for the sake of brevity. Note that the results were obtained using the previous ESVR training where the CV was decreased to reach the desired ARL_0 value of 200.

5.3 ARL_1 Comparisons under different n values

To study the effect of different sample sizes with ARL_0 = 200, we also set n = 5, 15 and 20 with the step 0.1; for example, $\tilde{X}'_{pure}$ = (0.1 0.2 … 1.5) when n = 15. Figures 3a, b illustrate some comparisons for different values of n in terms of ARL_1 and $SDRL_1$, respectively. The results conducted without a new training were predictable in $SDRL_1$ while the ARL_1 values had outstanding pattern for some of the moderate shifts because it was expected that the greater the value of n, the lower the ARL_1 for a specific shift. Another strange pattern occurred in the third and fourth shift when n = 5 (blue line). In this state, the larger shifts had larger ARL_1 and $SDRL_1$. It may be because of small sample sizes and biased estimated parameters. It was mentioned by Montgomery (2019), Haq (2020) and Abbasi et al. (2022) that detection of OOC condition in process with variable sample size provides some challenges such as variance inflation and it is necessary to employ some adaptive schemes for statistical control charts to reduce this effect. The results indicated that ESVR can perform better as an adaptive scheme in case of variable sample size. Conceptually speaking, the value of n is known in Phase II so control charts are extended with permanent n values and it is more practical to use ESVR in this situation with specific training for each n. However, this is not the focus of this paper, due to brevity and will be reported in the future studies.

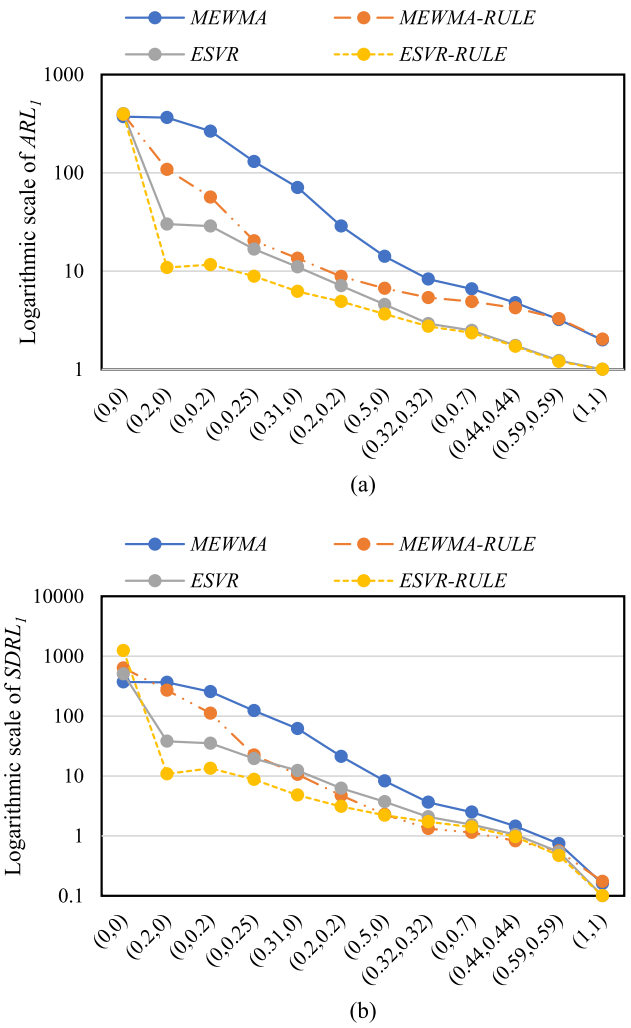

Fig. 4 The results of **a** ARL_1 and **b** $SDRL_1$ values for combination of ESVR and MEWMA with run-rules

5.4 Increasing the detection ability with run-rules

To improve the sensitivity of control charts, many techniques (or methods) including runs-rules, adaptive methods variable sampling designs and mixed procedures have been recommended in the area of profile monitoring (Haq 2020; Mohammadzadeh et al. 2021). For instance, adding runs-rules to the basic control chart can increase its ability to quickly detect shifts of different magnitudes. Employing of variable sampling interval (VSI) technique in which it is allowed to take samples with shorter intervals in case existing potential for a shift in IC model, while samples are taken with longer interval in the routine situations, has been utilised in the area of profile monitoring. Also, some other adaptive methods including modified successive sampling, ranked based and set approaches have been considered in profile monitoring (Maleki et al. 2018; Woodall 2007) but as a more profitable technique, run-rules was extended by Yeganeh et al. (2021), Yeganeh and Shadman (2020) and Yeganeh and Shadman (2021) in which they were used in form of a rule matrix. To construct a rule matrix, the IC region is divided into several regions by a heuristic approach. After defining these, then the number of regions and the ratio of points in them are reconciled with prespecified values (i.e., thresholds) and an OOC signal is obtained when they locate beyond these thresholds.

To investigate its effect in combination of ESVR, in this paper, the proposed ESVR method supplemented with rule matrix (denoted as ESVR-RULE). For this aim, the ratio of ESVR statistic (O_j) in the run-rule regions was computed and compared with the limits of rule matrix. The details of rule matrix were not given for the sake of brevity. For comparison purposes, the combination of rule matrix and MEWMA scheme (denoted with MEWMA-RULE) was

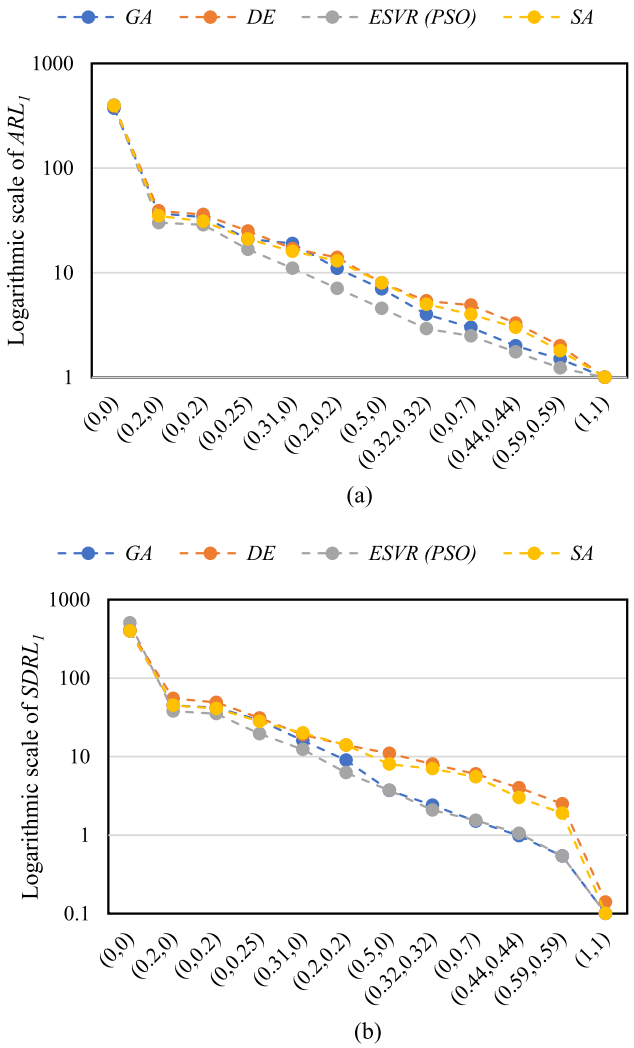

Fig. 5 The results of **a** ARL_1 and **b** $SDRL_1$ values for different EA in training of ESVR

also provided. Figures 4a, b illustrate the results of ESVR, ESVR-RULE, MEWMA and MEWMA-RULE.

From Fig. 4, it is observed that run-rules could improve the performance of both MEWMA and ESVR in terms of ARL_1 and $SDRL_1$ under small and moderate shifts except for the IC condition. The $SDRL_0$ of the MEWMA-RULE and ESVR-RULE were tangibly greater than MEWMA and ESVR, respectively; however, this may cause some problems in increasing the number false alarms in specific conditions. The second finding of this simulation study was that the run-rules could not be effective for large shifts. This finding is rational as run-rules usually improve detection ability of small shifts (Montgomery 2019). As the last finding, the ESVR-RULE performed better than the MEWMA-RULE which revealed a superior detection ability of the proposed method.

5.5 Effect of EA in training of ESVR

As mentioned in the previous section, the optimization problem in (16) is solved with PSO in ESVR. To show the superiority of PSO over other common EAs, three other well-known EA entailing GA, DE and SA was also employed. For this aim, all the designing steps between this method were similar and only they were utilised in solving of Eq. (16). Figure 5 depicts the ARL_1 and $SDRL_1$ values for our proposed method or ESVR (PSO), DE, GA and SA.

It is obvious that ESVR (PSO) had the best detection ability in all the shifts in terms of ARL_1 while it was also the best approach in small and moderate shifts in terms of $SDRL_1$. But GA had a very small superiority over ESVR (PSO) in large shifts in term of $SDRL_1$. Hence, these and some other similar simulations justified choosing of PSO in our proposed method. However, as stated in the literature (Adnan et al. 2021a; Ikram et al. 2022a; Kadkhodazadeh

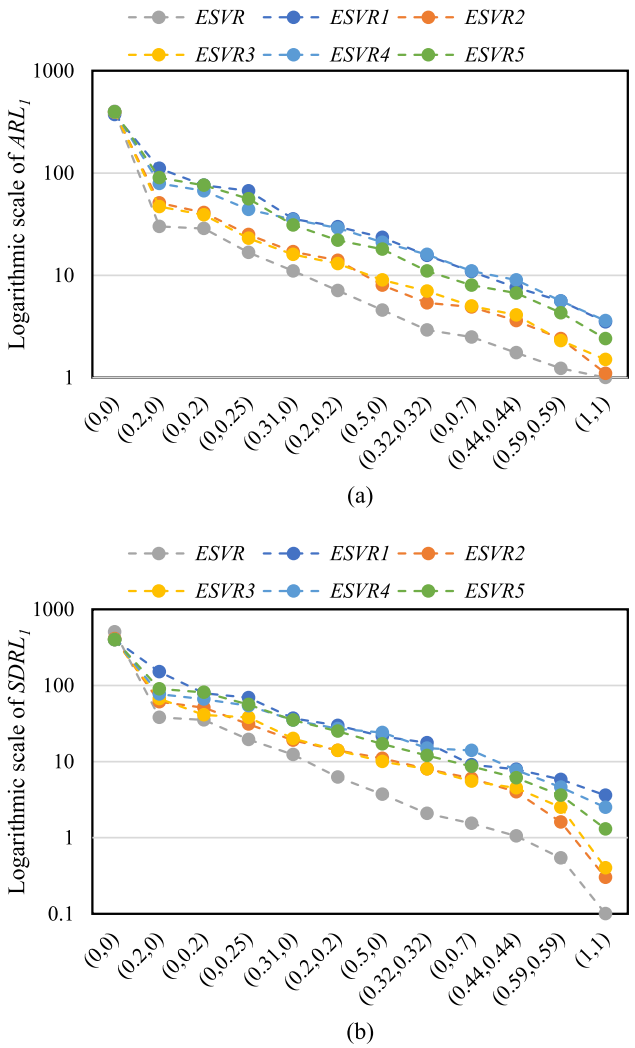

Fig. 6 The results of **a** ARL_1 and **b** $SDRL_1$ values for different input combinations

and Farzin 2022), some EAs such as EMPA, GBO, MFO, WCA and GWO may be highly sensitive to the initial parameters and adjustments. Hence, reaching superior performance over PSO may occur with these approaches under some sensitivity analysis. This idea can be investigated in the future by interested researchers.

5.6 Effect of input features

To show the best performance of our proposed input structure $\left(I_j = \left[EWMA_{Pj}, \frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}, \frac{d^{(3)}_{MEWMA}}{j}, M_j, \frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}, \frac{d^{(3)}_{LRT}}{j}, LRT_j\right]\right)$, some other input combinations were defined as follows:

- ESVR1: $I_j = [EWMA_{Pj}]$.
- ESVR2: $I_j = \left[EWMA_{Pj}, \frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}, \frac{d^{(3)}_{MEWMA}}{j}, M_j\right]$.
- ESVR3: $I_j = \left[EWMA_{Pj}, \frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}, \frac{d^{(3)}_{LRT}}{j}, LRT_j\right]$.
- ESVR4: $I_j = \left[\frac{d^{(1)}_{MEWMA}}{j}, \frac{d^{(2)}_{MEWMA}}{j}, \frac{d^{(3)}_{MEWMA}}{j}, M_j\right]$.
- ESVR5: $I_j = \left[\frac{d^{(1)}_{LRT}}{j}, \frac{d^{(2)}_{LRT}}{j}, \frac{d^{(3)}_{LRT}}{j}, LRT_j\right]$.

The training procedure for each of the above input combinations was the same as ESVR and the difference was only related to the input size. Consequently, the dimension of inputs in the train data was $p + 1$, $p + 5$, $p + 5$, 4 and 4 for ESVR1, ESVR2, ESVR3, ESVR4 and ESVR5 respectively. Figure 6 depicts the performance of ESVR approach regarding to each predefined input combination.

The superiority of ESVR over other input combinations are obvious from Fig. 6. By reduction of each part of input features, we can notice some decreases in the performance of ESVR scheme in term of ARL_1 and $SDRL_1$. The ability to identify OOC situations are more apparent for large shifts such as $\delta_1 = \delta_2 = 0.59$ of which ESVR1 was nearly five (ten) times less quick than ESVR in term of ARL_1 ($SDRL_1$). As another finding, it could be inferenced that combination control chart statistics and EWMA form of estimated parameters had a strong effect in detection ability due to superiority of ESVR2 and ESVR3 over ESVR1, ESVR4 and ESVR5. It means that the LRT and MEWMA statistics were not solely able to increase detection ability and they required some characteristics of the process to reduce ARL_1 and $SDRL_1$. The weak performance of ESVR1 utilized only EWMA form of estimated parameters has also confirmed this argument.

6 Diagnosis aid

In some real cases, the practitioner is interested to identify the parameters that have shifted after an OOC signal has been detected; however, this approach called profile diagnosis has gotten scant attention in the literature of profile monitoring. For example, some statistics have been proposed by Zou et al. (2007), Zou et al. (2008) and Huwang et al. (2016) for the diagnosis of the shift causes in linear, non-parametric and logistic profiles, respectively. Yeganeh and Shadman (2020) introduced a different approach using ANN with signalling rules as a tool for profile diagnosis. However, to the best of the authors' knowledge, there is no research work on diagnosis for Poisson profiles. In this paper, a novel structure based on a set of SVRs is proposed for diagnosis actions in Poisson profiles.

6.1 Requirements for profile diagnosis actions

There are two key points in the profile diagnosis simulations. Firstly, since profile diagnosis is usually implemented after the change point estimation. This is not part of this paper as we have assumed that all the shifts are considered from the onset of the process. Then, the IC estimated profiles need to be removed or ignored (see, for example, Fig. 9 in Yeganeh and Shadman (2020)) so that the machine learning procedure is based only on OOC samples to identify the parameters that have changed. Secondly, for a fair judgement, it is assumed that the control chart has the same signalling method. For example, similar diagnosis techniques can yield different results because of different signalling methods; see for example, Zou et al. (2007) and Huwang et al. (2014). In this paper, the diagnosis actions are implemented after triggering a signal by ESVR control chart.

6.2 The proposed SVR structure in profile diagnosis

Following Yeganeh and Shadman (2020) model, SVR is used in this paper for profile diagnosis actions. Yeganeh and Shadman (2020) used the EWMA statistics with estimated parameters as the inputs of ANN. Due to consideration of previous samples information, the EWMA statistic considers the change point effect automatically. Thus, after an OOC signal, the $EWMA_{Pj}$ $\left(EWMA_{Pj} = \left[\beta'_{1j}, \beta'_{2j}, \ldots, \beta'_{pj}, \bar{y}'_j\right]\right)$ of the last sample (i.e., the j^{th} sample in this case is equal to signalling sample; but hereafter its index is omitted) is considered as the input vector of profile diagnosis. But there is a fundamental difference between ANN and SVR. It is possible to assign different neurons for each parameter in the output layer of

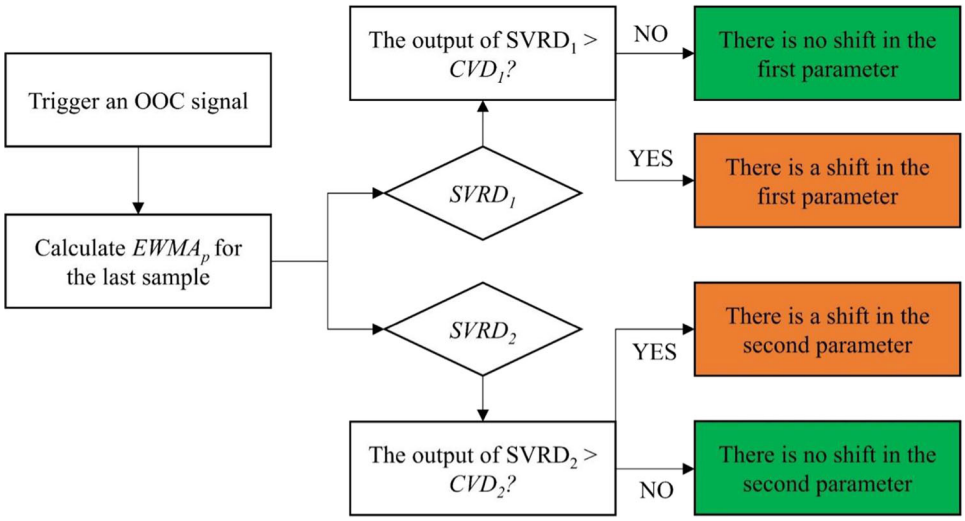

Fig. 7 The profile diagnosis procedure of SVRS approach after an OOC signal by ESVR when $p = 2$

ANN while SVR could only generate one output. This is the major challenge that arises when conducting profile diagnosis with SVR as a classification problem. To overcome this problem, we used an approach denoted by SVRS (SVR Set) in a way that one SVR called SVRD (SVR Diagnosis) is trained for each possible change and the existence of each shift is identified by SVRD. In this approach, there are in total p SVRDs in diagnosis process. In other words, SVRS consist of p SVRDs, i.e., $SVRD_1$, $SVRD_2$, …, $SVRD_p$. For example, if we have two parameters in the IC profile, two SVRDs identify the shifted parameters such that the first and second SVRDs are indicators of the shifts in the first and second parameters, respectively. Naturally, the identifications of a shift by both parameters indicate a simultaneous shift.

Also, a limit, called *CVD* (Cutting Value Diagnosis) is assigned for each SVRD to identify the change in the parameters. To conduct diagnosis actions after an OOC signal by ESVR control chart, the $1 + p$ vector $EWMA_p$ is computed using the last (i.e., current) sample and then it is considered as the input of each SVRD. In other words, $SVRD_1$, $SVRD_2$, … and $SVRD_p$ have the same input. The outputs of SVRDs are compared with their *CVD* to identify the shifted parameters.

Considering the IC model given in (17) with $p = 2$, SVRS includes $SVRD_1$ and $SVRD_2$ with CVD_1 and CVD_2 such that the shift in the first and second parameters are identified, respectively. To better illustrate the above diagnosis, Fig. 7 depicts the diagnosis procedure of SVRS after an OOC signal in the jth profile is detected, using the IC model defined in (17) when $p = 2$.

To train each SVRD in SVRS, some OOC profiles are generated until an OOC signal is obtained by ESVR. Then, the $EWMA_p$ is considered as the inputs of the training dataset. The targets are defined such that the target value of the shift in the pth parameter is 1 and the other values are 0 in $SVRD_p$. Other training aspects and assigned limits are the same as the ones in Yeganeh and Shadman (2020). Hence, the details are not reported here for brevity.

6.3 The accuracy of the proposed SVRS structure in profile diagnosis

Due to the lack of research for profile diagnosis in Poisson profiles, four other competitors entailing three machine learning techniques and one statistic method are provided in this paper. First, we use multiclass SVM (the details and concept are ignored here to save space) with '*fitcecoc*' function in MATLAB called MSVM. For a better comparison, ANN training with back-propagation and entropy-based training algorithm ('*feedforward*' and '*patternnet*' functions in MATLAB) denoted by ANN-BP and *Patternnet* are also carried out for profile diagnoses. In addition to these methods, the Wald statistic proposed in Huwang et al. (2016) is also computed as the last competitor (denoted as Wald Test). The OOC profiles are generated from (17) with $p = 2$. The simulation procedure for obtaining accuracies in profile diagnoses with SVRS when there is shift in the intercept is described by Pseudocode 3.

Pseudocode 3. The Monte Carlo simulation procedure in SVRS

```
Compute UCL_MEWMA and UCL_LRT;
Consider a specific shift in the intercept; Corrected = 0;
 for iter = 1:MaxIt
 OOC = 0;
  while OOC == 0
  Generate random profile with sample size n;
  Calculate EWMA_p;
  if (ESVR trigger an OOC signal) OOC = 1; end if
  end while
 Import EWMA_p to SVRD1 and SVRD2 and Compute the outputs;
  if (output of SVRD1 > CVD1 & output of SVRD2 < CVD2)   % Shift in intercept
  Corrected = Corrected + 1; end if
 end for
Accuracy = Corrected / MaxIt;
```

Note that Pseudocode 3 changes to the situations that OOC shifts have been occurred in slope and simultaneous by replacing the following codes with the last ***if***.

- **If** (output of $SVRD_1 < CVD_1$ & output of $SVRD_2 > CVD_2$) % Shift in slope
- **If** (output of $SVRD_1 > CVD_1$ & output of $SVRD_2 > CVD_2$) % Shift in intercept and slope

The results of diagnosis accuracy are reported in Table 7 for Poisson profiles based on 10,000 iterations (*MaxIt*), where the boldfaced values denote the best performing scheme at that particular shift size. For example, the accuracy of SVRS in the first shift is 0.52 which means that "*Corrected*" value in Pseudocode 3 is 5200. From Table 7, it can be seen that SVRS and Patternnet are preferred over others in terms of the average of accuracies while ANN-BP is the best method in terms of the standard deviation of accuracies. MSVM and Wald Test have biased performances; that is, for some shifts, they are not able to detect any shifts while they have good accuracies for others shifts.

Table 7 Profile diagnosis accuracy

Shift size	Method				
(δ_1, δ_2)	ANN-BP	Patternnet	MSVM	Wald Test	SVRS
(0.2,0)	0.465	0.482	0.130	**0.521**	0.520
(0,0.2)	0.435	0.318	0.000	**0.643**	0.640
(0,0.25)	0.431	0.353	0.000	**0.683**	0.500
(0.31,0)	0.447	0.490	0.190	0.546	**0.550**
(0.2,0.2)	0.396	0.476	**0.710**	0.070	0.330
(0.5,0)	0.406	0.525	0.280	**0.584**	0.480
(0.32,0.32)	0.490	0.553	**0.780**	0.090	0.360
(0,0.7)	0.507	0.309	0.000	**0.738**	0.560
(0.44,0.44)	0.516	0.631	**0.740**	0.109	0.500
(0.59,0.59)	0.556	0.719	**0.760**	0.119	0.530
(1,1)	0.484	0.666	**0.990**	0.238	0.570
Average	0.467	0.502	0.416	0.394	**0.504**
Std	**0.049**	0.137	0.380	0.268	0.090

7 Illustrative example

A real-life application of Poisson profiles in the airline industry is provided here from Chatterjee and Hadi (2013) and Alevizakos et al. (2019b). The aim of this example is to check the relationship between the number of injury incidents and the proportion of total flights over time. Naturally, it is expected that the probability of accidents will increase with an increase in the proportion of total flights.

To this end, the accidents and injuries of nine major USA airlines were studied in these references. If all the airlines have equally safe performance in a specific period, the injury incidents can be explained by the IC model considering the number of flights of each airline as a percentage of the total number of the airlines as an explanatory variable and the injury incidents as a response variable. Following Eq. (2), the IC Poisson model is established by the relationship between the explanatory and response variables:

Table 8 The OOC profiles characteristics of the illustrative example

j	$[\widehat{\beta}_j \ \bar{y}_j]$			$[\beta'_j \bar{y}'_j]$			$EWMA_{pj}$		M_j	LRT_j	The ratio of samples							
1	1.67	3.53	8.00	1.53	− 10.86	1.00	0.31	− 2.17	0.20	0.23	6.98	1	0	0	0.23	0	1	0	6.98
2	1.03	8.31	7.89	0.08	− 0.45	0.88	0.26	− 1.83	0.34	0.22	0.75	1	0	0	0.22	0.50	0.50	0	0.75
3	0.54	12.13	8.44	− 1.08	7.86	1.50	− 0.01	0.11	0.57	0.20	5.33	1	0	0	0.20	0.33	0.67	0	5.33
4	0.92	8.33	7.11	0.05	− 0.38	0.00	0.00	0.01	0.45	0.13	0.01	1	0	0	0.13	0.50	0.50	0	0.01
5	0.26	12.28	6.56	− 1.17	8.26	− 0.62	− 0.23	1.66	0.24	0.14	3.10	1	0	0	0.14	0.60	0.40	0	3.10
6	0.53	11.04	7.11	− 0.78	5.53	0.00	− 0.34	2.44	0.19	0.28	1.35	1	0	0	0.28	0.67	0.33	0	1.35
7	1.20	8.28	9.33	0.11	− 0.56	2.50	− 0.25	1.84	0.65	0.40	5.70	1	0	0	0.40	0.57	0.43	0	5.70
8	1.21	7.56	8.56	0.32	− 2.11	1.63	− 0.14	1.05	0.85	0.53	2.71	1	0	0	0.53	0.63	0.38	0	2.71
9	1.31	6.55	8.22	0.62	− 4.30	1.25	0.02	− 0.02	0.93	0.60	2.45	1	0	0	0.60	0.67	0.33	0	2.45
10	1.07	7.09	7.00	0.43	− 3.07	− 0.12	0.10	− 0.63	0.72	0.37	0.44	1	0	0	0.37	0.70	0.30	0	0.44
11	1.22	8.35	9.67	0.10	− 0.41	2.88	0.10	− 0.59	1.15	0.94	7.43	0.91	0.09	0	0.94	0.64	0.36	0	7.43

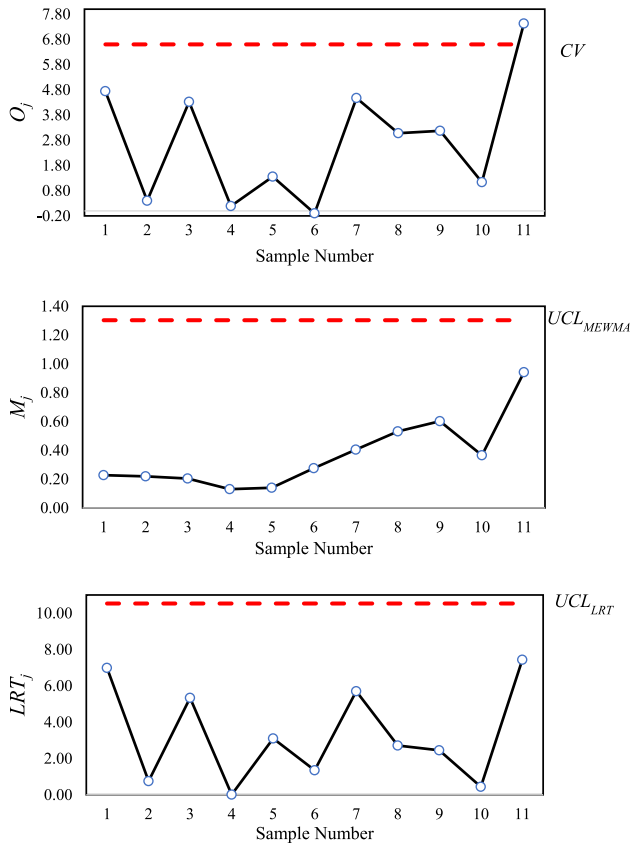

Fig. 8 The statistics of the ESVR, MEWMA and LRT control charts for 11 OOC generated profiles in the illustrative example

To reach the ARL_0 equal to 200, the values of UCL_{MEWMA}, UCL_{LRT} and CV are obtained as 1.303, 10.53 and 6.61, respectively. As a common approach in Phase II applications, the intercept changes to 0.965, as an artificial shift, to reach an OOC signal. Table 8 gathers the details of *11* OOC generated profiles. It provides the details of estimated parameters (first part), normalized parameters (second part), EWMA statistic of the normalized parameters (third part), MEWMA and LRT statistics (fourth part) and the ratio of samples (fifth part), respectively. Note that the input vector of ESVR in this example is defined with length 11 ($p + 9$); for example, the input in the first sample ($j = 1$) is [0.31 − 2.17 0.2 1 0 0 0.23 0 1 0 6.98].

The outputs of ESVR for the above inputs are depicted in Fig. 8. The MEWMA and LRT statistics are also added to this figure to visualised their trends as compared to the ESVR. It can be observed from Fig. 8 that ESVR triggered an OOC signal at the *11th* sample while LRT and MEWMA procedures were not able to detect this shift.

To identify the shifted parameter, $EWMA_P$ = [0.1 − 0.59 1.15] is incorporated into the $SVRD_1$ and $SVRD_2$ and the outputs are 1.16 and − 0.27, respectively. This is an indication of shifts in the first parameter because CVD_1 and CVD_2 are 0.39 and 0.42 (i.e., $1.16 > 0.39$ and $-0.27 < 0.42$).

$$\boldsymbol{\beta_0} = [0.8945\ 8.5018],$$
$$\tilde{\boldsymbol{X}}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.0503 & 0.054 & 0.0629 & 0.075 & 0.095 & 0.1292 & 0.1382 & 0.1920 & 0.2078 \end{pmatrix}, \quad (21)$$
$$n = 9, p = 2.$$

From this example, it is observed that the O_j exceeds its control limit sooner than MEWMA and LRT schemes and the results under the existing case study are in accordance with the simulation results. This excursion suggests that the proposed ESVR has excellent abilities in practical applications of Phase II SPC problems in comparison to other competitors. Also, these findings act as evidence that the proposed diagnosis approach encompass a significant impact on the detection of shifted parameters. Therefore, the ESVR control chart is found to be more efficient in Poisson profile monitoring.

8 Conclusions

A novel use of SVR as a control chart is extended to monitor Poisson profiles in Phase II. This method, equipped with new input features and evolutionary training procedure based on the PSO algorithm, is able to quickly detect the OOC situations due to the advantage of an evolutionary training framework either in parametric or non-parametric monitoring where the OOC model can be unknown. To design a more efficient method for identification of small and moderate shifts, the proposed scheme is incorporated with additional run-rules. Finally, a diagnostic procedure is used with SVR structures and ANN. Both the SVR and ANN approaches showed a better detection ability. The contributions of this study include: firstly, implementation of SVR as a control chart in monitoring Poison profiles. Secondly, utilizing a novel input feature corresponding to the ratio of MEWMA and LRT statistics. Lastly, the training of the SVR using evolutionary PSO algorithm.

That said, owing to the requirements in evolutionary computations for training of ESVR, the proposed approach requires more computations than the statistical approaches such as MEWMA and LRT. Note though, this challenge commonly occurs in the machine learning applications but due to rapid extension of artificial intelligence related technology from the software and hardware aspects, the importance of this challenge has decreased in recent years.

For future research purpose, the investigation of other evolutionary algorithms with different IC profile types and sample sizes would be preferable. Moreover, the proposed method can conveniently and effectively tackle the problem of other non-parametric profile monitoring; for example, applying the proposed approach with linear IC model can be a good idea for future research.

Authors contributions AY: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing—Original Draft, Visualization. SAA: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—Review & Editing, Supervision, Project administration. SCS: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing—Review & Editing, Supervision, Project administration. J-CM-M: Formal analysis, Investigation, Resources, Writing—Review & Editing, Supervision, Project administration. ARS: Formal analysis, Investigation, Resources, Writing—Review & Editing, Supervision, Project administration.

Funding Open Access funding provided by the Qatar National Library.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest I certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on me or on any organization with which I am associated and, I certify that all financial and material supports for this research and work are clearly identified in the title page of the manuscript.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbas T, Qian Z, Ahmad S, Riaz M (2016) On monitoring of linear profiles using Bayesian methods. Comput Ind Eng 94:245–268

Abbasi SA, Yeganeh A, Shongwe SC (2022) Monitoring non-parametric profiles using adaptive EWMA control chart. Sci Rep 12(1):14336

Adnan RM, Mostafa RR, Islam ARMT, Kisi O, Kuriqi A, Heddam S (2021a) Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. Comput Electron Agric 191:106541

Adnan RM, Mostafa RR, Kisi O, Yaseen ZM, Shahid S, Zounemat-Kermani M (2021b) Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. Knowl-Based Syst 230:107379

Adnan RM, Kisi O, Mostafa RR, Ahmed AN, El-Shafie A (2022) The potential of a novel support vector machine trained with modified mayfly optimization algorithm for streamflow prediction. Hydrol Sci J 67(2):161–174

Alevizakos V, Koukouvinos C, Lappa A (2019a) Comparative study of the Cp and Spmk indices for logistic regression profile using different link functions. Qual Eng 31(3):453–462

Alevizakos V, Koukouvinos C, Castagliola P (2019b) Process capability index for Poisson regression profile based on the Spmk index. Qual Eng 31(3):430–438

Al-Zoubi AM, Faris H, Alqatawna J, Hassonah MA (2018) Evolving support vector machines using whale optimization Algorithm for spam profiles detection on online social networks in different lingual contexts. Knowl-Based Syst 153:91–104. https://doi.org/10.1016/j.knosys.2018.04.025

Al-Zoubi AM, Hassonah MA, Heidari AA, Faris H, Mafarja M, Aljarah I (2021) Evolutionary competitive swarm exploring optimal support vector machines and feature weighting. Soft Comput 25(4):3335–3352

Amiri A, Koosha M, Azhdari A, Wang G (2015) Phase I monitoring of generalized linear model-based regression profiles. J Statist Comput Simul 85(14):2839–2859

Arana-Daniel N, Gallegos AA, López-Franco C, Alanís AY, Morales J, López-Franco A (2016) Support vector machines trained with evolutionary algorithms employing kernel adatron for large scale classification of protein structures. Evol Bioinform Online 12:285–302

Aslam M, Bantan RAR, Khan N (2020) Design of NEWMA np control chart for monitoring neutrosophic nonconforming items. Soft Comput 24(21):16617–16626

Chatterjee S, Hadi AS (2013) Regression analysis by example, 5th edn. Wiley, Hoboken

Chen S, Yu J, Wang S (2020) Monitoring of complex profiles based on deep stacked denoising autoencoders. Comput Ind Eng 143:106402

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Cuentas S, García E, Peñabaena-Niebles R (2022) An SVM-GA based monitoring system for pattern recognition of autocorrelated processes. Soft Comput 26(11):5159–5178

Dantas Dias ML, Rocha Neto AR (2017) Training soft margin support vector machines by simulated annealing: a dual approach. Expert Syst Appl 87:157–169

Fallahdizcheh A, Wang C (2022) Profile monitoring based on transfer learning of multiple profiles with incomplete samples. IISE Trans 54(7):643–658

Gardner MM et al (1997) Equipment fault detection using spatial signatures. IEEE Trans Compon Packag Manuf Technol Part C 20(4):295–304

Gupta S, Montgomery DC, Woodall WH (2006) Performance evaluation of two methods for online monitoring of linear calibration profiles. Int J Production Res 44(10):1927–1942

Haq A (2020) Adaptive MEWMA charts for univariate and multivariate simple linear profiles. Commun Statist Theory Methods 51(16):5383–5411. https://doi.org/10.1080/03610926.2020.1839100

He S, Song L, Shang Y, Wang Z (2020) Change-point detection in phase I for autocorrelated Poisson profiles with random or unbalanced designs. Int J Prod Res 59(14):4306–4323. https://doi.org/10.1080/00207543.2020.1762017

Hosseinifard SZ, Abdollahian M, Zeephongsekul P (2011) Application of artificial neural networks in linear profile monitoring. Expert Syst Appl 38(5):4920–4928

Huwang L, Wang Y-HT, Xue S, Zou C (2014) Monitoring general linear profiles using simultaneous confidence sets schemes. Comput Ind Eng 68:1–12

Huwang L, Wang Y-HT, Yeh AB, Huang Y-H (2016) Phase II profile monitoring based on proportional odds models. Comput Ind Eng 98:543–553

Ikram RMA, Ewees AA, Parmar KS, Yaseen ZM, Shahid S, Kisi O (2022a) The viability of extended marine predators algorithm-based artificial neural networks for streamflow prediction. Appl Soft Comput 131:109739

Ikram RMA, Dai H-L, Ewees AA, Shiri J, Kisi O, Zounemat-Kermani M (2022b) Application of improved version of multi verse optimizer algorithm for modeling solar radiation. Energy Rep 8:12063–12080

Johnson RA, Wichern DW (2007) Applied Multivariate Statistical Analysis. Pearson Prentice Hall, Hoboken

Jones CL, Abdel-Salam A-SG, Mays DA (2020) Practitioners guide on parametric, nonparametric, and semiparametric profile monitoring. Qual Reliab Eng Int 37(3):857–881. https://doi.org/10.1002/qre.2770

Kadkhodazadeh M, Farzin S (2022) A novel hybrid framework based on the ANFIS, discrete wavelet transform, and optimization algorithm for the estimation of water quality parameters. J Water Clim Change 13(8):2940–2961

Kang L, Albin SL (2000) On-line monitoring when the process yields a linear profile. J Qual Technol 32(4):418–426

Kennedy J (2010) Particle swarm optimization. In: Sammut C, Webb GI (eds) Encyclopedia of machine learning. Springer, Boston, pp 760–766

Kim K-J, Cho S-B (2008) Evolutionary ensemble of diverse artificial neural networks using speciation. Neurocomputing 71(7):1604–1618

Li C-I, Pan J-N, Liao C-H (2019) Monitoring nonlinear profile data using support vector regression method. Qual Reliab Eng Int 35(1):127–135

Maleki MR, Amiri A, Castagliola P (2018) An overview on recent profile monitoring papers (2008–2018) based on conceptual classification scheme. Comput Ind Eng 126:705–728

Maleki MR, Castagliola P, Amiri A, Khoo MBC (2019) The effect of parameter estimation on phase II monitoring of poisson regression profiles. Commun Statist Simul Comput 48(7):1964–1978

Mohammadzadeh M, Yeganeh A, Shadman A (2021) Monitoring logistic profiles using variable sample interval approach. Comput Ind Eng 158:107438

Montgomery DC (2019) Introduction to statistical quality control, 8th edn. Wiley, New York

Motasemi A, Alaeddini A, Zou C (2017) An area-based methodology for the monitoring of general linear profiles. Qual Reliab Eng Int 33(1):159–181

Nassar SH, Abdel-Salam A-SG (2021) Semiparametric MEWMA for Phase II profile monitoring. Qual Reliab Eng Int 37(5):1832–1846

Ojha VK, Abraham A, Snášel V (2017) Metaheuristic design of feedforward neural networks: a review of two decades of research. Eng Appl Artif Intell 60:97–116

Pacella M, Semeraro Q (2011) Monitoring roundness profiles based on an unsupervised neural network algorithm. Comput Ind Eng 60(4):677–689

Pourpanah F, Lim CP, Saleh JM (2016) A hybrid model of fuzzy ARTMAP and genetic algorithm for data classification and rule extraction. Expert Syst Appl 49:74–85

Qi D, Wang Z, Zi X, Li Z (2016) Phase II monitoring of generalized linear profiles using weighted likelihood ratio charts. Comput Ind Eng 94:178–187

Qi D, Li Z, Zi X, Wang Z (2017) Weighted likelihood ratio chart for statistical monitoring of queueing systems. Qual Technol Quant Manag 14(1):19–30

Riaz M, Abbasi SA, Ahmad S, Zaman B (2014) On efficient phase II process monitoring charts. Int J Adv Manuf Technol 70(9):2263–2274

Shadman A, Zou C, Mahlooji H, Yeh AB (2017) A change point method for phase II monitoring of generalized linear profiles. Commun Statist Simul Comput 46(1):559–578

Shang Y, Tsung F, Zou C (2011) Profile monitoring with binary data and random predictors. J Qual Technol 43(3):196–208

Shang Y, Wang Z, Zhang Y (2018) Nonparametric control schemes for profiles with attribute data. Comput Ind Eng 125:87–97

Song L, He S, Zhou P, Shang Y (2021) Empirical likelihood ratio charts for profiles with attribute data and random predictors in the presence of within-profile correlation. Qual Reliab Eng Int 38(1):153–173. https://doi.org/10.1002/qre.2965

Steiner S, Jensen WA, Grimshaw SD, Espen B (2016) Nonlinear profile monitoring for oven-temperature data. J Qual Technol 48(1):84–97

Stoean C, Stoean R (2014) Support vector machines and evolutionary algorithms for classification: Single or together? Springer International Publishing, Cham

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

Vapnik VN (1998) Statistical learning theory. Wiley-Interscience, New York

Wang F-K, Du T (2014) Implementing support vector regression with differential evolution to forecast motherboard shipments. Expert Syst Appl 41(8):3850–3855

Wang Y, Li J, Ma Y, Song L, Wang Z (2022) Nonparametric monitoring schemes in Phase II for ordinal profiles with application to customer satisfaction monitoring. Comput Ind Eng 165:107931

Williams JD, Woodall WH, Birch JB (2007) Statistical monitoring of nonlinear product and process quality profiles. Qual Reliab Eng Int 23(8):925–941

Woodall WH (2007) Current research on profile monitoring. Production 17(3):420–425

Xu L, Wang S, Peng Y, Morgan JP, Reynolds MR, Woodall WH (2012) The monitoring of linear profiles with a GLR control chart. J Qual Technol 44(4):348–362

Yeganeh A, Shadman A (2020) Monitoring linear profiles using Artificial Neural Networks with run rules. Expert Syst Appl 168:114237

Yeganeh A, Shadman A (2021) Using evolutionary artificial neural networks in monitoring binary and polytomous logistic profiles. J Manuf Syst 61:546–561. https://doi.org/10.1016/j.jmsy.2021.10.007

Yeganeh A, Shadman A, Amiri A (2021) A novel run rules based MEWMA scheme for monitoring general linear profiles. Comput Ind Eng 152:107031

Yeganeh A, Shadman A, Abbasi SA (2022a) Enhancing the detection ability of control charts in profile monitoring by adding RBF ensemble model. Neural Comput Appl 34(12):9733–9757

Yeganeh A, Abbasi SA, Pourpanah F, Shadman A, Johannssen A, Chukhrova N (2022b) An ensemble neural network framework for improving the detection ability of a base control chart in non-parametric profile monitoring. Expert Syst Appl 204:117572

Yeganeh A, Johannssen A, Chukhrova N, Abbasi SA, Pourpanah F (2023) Employing machine learning techniques in monitoring autocorrelated profiles. Neural Comput Appl 35(22):16321–16340. https://doi.org/10.1007/s00521-023-08483-3

Yeh AB, Huwang L, Li Y-M (2009) Profile monitoring for a binary response. IIE Trans 41(11):931–941

Zhang F, Deb C, Lee SE, Yang J, Shah KW (2016) Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique. Energy Build 126:94–103

Zhao C, Du S, Deng Y, Li G, Huang D (2020) Circular and cylindrical profile monitoring considering spatial correlations. J Manuf Syst 54:35–49

Zhou Q, Zou C, Wang Z, Jiang W (2012) Likelihood-based EWMA charts for monitoring Poisson count data with time-varying sample sizes. J Am Statist Assoc 107(499):1049–1062

Zhou P, Liu P, Wang S, Zhang C, Zhang J, Li S (2022) Functional state-space model for multi-channel autoregressive profiles with application in advanced manufacturing. J Manuf Syst 64:356–371

Ziani R, Felkaoui A, Zegadi R (2017) Bearing fault diagnosis using multiclass support vector machines with binary particle swarm

optimization and regularized Fisher's criterion. J Intell Manuf 28(2):405–417

Zou C, Tsung F, Wang Z (2007) Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. Technometrics 49(4):395–408

Zou C, Tsung F, Wang Z (2008) Monitoring profiles based on nonparametric regression methods. Technometrics 50(4):512–526

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ali Yeganeh[1] · Saddam Akber Abbasi[2,3] · Sandile Charles Shongwe[1] · Jean-Claude Malela-Majika[4] · Ali Reza Shadman[5]

✉ Saddam Akber Abbasi
sabbasi@qu.edu.qa; saddamabbasi@yahoo.com

✉ Sandile Charles Shongwe
ShongweSC@ufs.ac.za

Ali Yeganeh
yeganeg.ali1369@gmail.com

Jean-Claude Malela-Majika
malela.mjc@up.ac.za

Ali Reza Shadman
a.shadman@ferdowsi.um.ac.ir

[1] Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein 9301, South Africa

[2] Statistics Program, Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, 2713, Doha, Qatar

[3] Statistical Consulting Unit, College of Arts and Sciences, Qatar University, 2713, Doha, Qatar

[4] Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Hatfield 0002, South Africa

[5] Department of Industrial Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, P.O. Box 91775-1111, Mashhad, Iran