A Support Vector Regression Approach for Three–Level Longitudinal Data

Mohammad Moqaddasi Amiri [(1)], *Leili Tapak* [(2)], *Javad Faradmal* [(2)]

(1) Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.
(2) Assistant Professor of Biostatistics, Department of Biostatistics, School of Public Health, Modeling of Noncommunicable diseases Research center, Hamadan University of Medical Sciences, Hamadan, Iran

CORRESPONDING AUTHOR: *Javad Faradmal, Associate Professor of Biostatistics, Department of Biostatistics and Epidemiology, School of Public Health, Modeling of Noncommunicable diseases Research center, Hamadan University of Medical Sciences, Hamadan, Iran; Email: javad.faradmal@umsha.ac.ir, Phone: +98 9124055256*

ABSTRACT

Background: Longitudinal data structure is frequently observed in health science. This introduces correlation to the data that needs to be handled in modelling process. Recently, machine learning approaches have been introduced in the context of longitudinal data for prediction of the response variable purpose. In this paper a mixed-effects least squares support vector regression model is presented for three-level longitudinal data. In the proposed model, multiple random-effect terms are used for considering the existing correlation structures in longitudinal data. The proposed model is flexible in modelling (non-)linear and complex relationships between predictors and response, while it takes into account the hierarchical structure of data and is computationally efficient.

Methods Both random intercept and random trend models with a special correlation structure of errors are illustrated. A real data example on human Brucellosis rate is analysed and two simulation studies are performed to illustrate the proposed model. The fitting and generalisation performance of the proposed model are investigated and compared with the ordinary least squares support vector regression and linear mixed-effects models.

Results: Based on the human Brucellosis rate example and two simulation studies, the proposed models had the best performance in generalisation. Also, the fitting performances of the proposed models were better than that of the classic models.

Conclusion: Our study revealed that in the presence of nonlinear relationship between covariates and outcome, the proposed MLS-SVR model has the best fitting and generalisation performance and can capture correlation of the data.

Key words: Machine Learning, Support Vector Machine, Longitudinal Study, Brucellosis, Linear Models

INTRODUCTION

Longitudinal data analysis are found frequently in clinical research where a subject has multiple correlated observations over time and can be assumed as a two-level data set. This correlated structure needs to be accounted for in the analysis [1,2]. In some cases, the two-level longitudinal data may be nested in the centres or clusters which adds an additional level to the data and forms a three-level longitudinal data structure. The intracluster correlation is then introduced in the three-level longitudinal data, and ignoring this correlation in the analysis procedure can lead to incorrect estimations of the standard errors [2].

Variants of classical models have been developed

and introduced for modelling the two-level and three-level longitudinal data under a variety of names: hierarchical linear models, random-effects models, two-stage models and random regression models [1-3]. In most of these models, the random-effects terms have been considered to account for the existing correlations of the longitudinal data.

Linear mixed-effects models (LMM) [2], generalized linear mixed-models (GLMM) [2] and generalized estimating equations (GEE) [4] models are the most commonly used methods in the longitudinal data analysis. However, fitting these models is not possible in high-dimensional data sets (number of observations < number of variables) by taking into account all covariates in the model at the same time [5]. Also, the relationships between covariates and response variable are considered as linear or limited nonlinear forms and need to specify the functional form of covariates and response for the appropriate transformation of covariates [6]. Therefore, it is not possible to capture unspecific complex nonlinear relationships between variables using these parametric methods.

In recent years, the use of machine learning methods has become very interesting. One promising and useful method of the machine learning is support vector machines (SVMs) introduced by Vapnik [7] for classification and function estimation problems. There are some properties which make the SVM interesting versus other machine learning techniques such as artificial neural network (ANN) and random forest (RF) methods. Unlike ANN, the SVM does not lead to local minima and is less prone to over-fitting [8]. RF provides discrete models for prediction and the range of predictions are limited to the range of response variable over training set [9,10]. Also the RF technique tends to underestimate and overestimate higher and lower values, respectively. Moreover, theoretical analysis of the RF is difficult [9]. ANN and RF require long training time [11]. The SVMs can be fitted in high-dimensional data sets and can consider complex nonlinear relations between covariates and outcome by using kernel functions [12,13]. Suykens [18] presented an extended model of the SVM called least-squares support vector machine (LS-SVM) which uses the least-squares loss function. The optimisation problem in LS-SVM is solved linearly with a unique solution which makes it computationally time-saving [19]. The regression or function estimation form of LS-SVM is called least-squares support vector regression (LS-SVR) [19].

Kernel technique and SVM methods have been used in various studies for both classification and regression problems in longitudinal data. Luts [20] and Chen [21] used LS-SVM and kernel methods for classification of the longitudinal data. Also, a mixed-effects LS-SVR model was proposed by Seok [22] for continuous longitudinal outcomes and developed by Shim[23] for time-varying coefficients. There have been some studies which have used other machine learning techniques in longitudinal or clustered data [24-27]. All these studies have been done in two-level longitudinal data.

In this study, we propose a novel mixed-effects LS-SVR (MLS-SVR) model for predicting continuous outcomes in three-level longitudinal data which can be considered by a random intercept for level-3, both random intercept and trend for level-2, and a correlation structure for errors. Also our proposed method can capture the unspecific non-linear relationships between features and outcome by using a kernel function. To the best of our knowledge, this is the first SVM or kernel method study that considers the centre or cluster (third level) effect in three-level longitudinal data setting. In a recently published study, Moqaddasi et al. developed the LS-SVR method into the three-level setting. However, they considered only the random intercept term in each level [28]. We compare the generalisation performance of the proposed model with that of the ordinary LS-SVR and LMM in a real life data example and two simulation studies with various scenarios.

The rest of this paper is structured as follows. In materials and methods section, we describe the ordinary LS-SVR and present the proposed model with estimation procedure. Also, the numerical studies through a real data analysis and two simulation studies are described in this section. Then we provide the results of the real data example and simulation studies. Finally, discussion and conclusion sections are presented, respectively.

MATERIALS AND METHODS

The special property of the SVM is the nonlinear fitting using the kernel approach. In kernel–based techniques, a nonlinear problem are solved linearly by mapping the feature space into a higher dimension space [14,15]. Nevertheless, the SVM is computationally intensive because of using numerical optimisation algorithms and solving a quadratic optimisation problem. Hinge and ε-insensitive loss functions are two common losses used in the SVM in classification and regression problems, respectively [16,17]. Because of the expensive computation of the SVM we used LS-SVR with least squares loss function. In this section, we introduce the ordinary LS-SVR with the estimation procedure. Then, we propose the three-level random intercept and trend, i.e. MLS-SVR, with an estimation technique for three-level longitudinal data. A generalized cross-validation (GCV) function presented in [23] is used to obtain the optimal regularization parameters.

Least-squares support vector regression

In this section, we explain the ordinary LS-SVR. Suppose the training data set is denoted by $G = \left\{ (x_i, y_i) \right\}_{i=1}^{n}$ with feature vector $x_i \in \mathbb{R}^d$ and the response $y_i \in \mathbb{R}$. The relation of the response and feature variables is shown in regression setting as follows

$$y(\boldsymbol{x}) = \omega^T \varphi(\boldsymbol{x}) + b \qquad (1)$$

where b is the bias term. $\varphi(\cdot)$ is the feature mapping function which maps the input space into a higher dimensional feature space and ω is a weight vector of the same dimension as $\varphi(.)$. The optimisation equation (J) of the nonlinear LS-SVR is as follows

$$\min \quad J = \frac{1}{2}\|\omega\|^2 + \frac{\lambda}{2}\sum_{i=1}^{N} e_i^2 \qquad (2)$$

subject to the equality constraints of

$$y_i = \omega^T \varphi(\boldsymbol{x}_i) + b + e_i, \quad i = 1,2,...,\text{n} \qquad (3)$$

here $\lambda > 0$ is a regularization parameter, $\omega \in \mathbb{R}^d$ is the w e i g h t vector $e_i \in \mathbb{R}$ and is the error vector. The primal Lagrange function can be constructed as

$$L = J - \sum_{i=1}^{N} \alpha_i (\omega^T \varphi(\text{x}_i) + b + e_i - y_i) \qquad (4)$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. The optimality conditions are given by

$$\begin{cases} \dfrac{\delta L}{\delta \omega} = 0 \quad \rightarrow \quad \omega = \sum_{i=1}^{N} \alpha_i \varphi(\boldsymbol{x}_i) \\ \dfrac{\delta L}{\delta b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i = 0 \\ \dfrac{\delta L}{\delta e_i} = 0 \quad \rightarrow \quad \alpha_i = \lambda e_i \\ \dfrac{\delta L}{\delta \alpha_i} = 0 \quad \rightarrow \quad \omega^T \varphi(\boldsymbol{x}_i) + b + e_i - y_i = 0 \end{cases} \qquad (5)$$

After eliminating e_i and ω, the optimal values of α_i and b are obtained by solving a linear equation as follows:

$$\begin{pmatrix} 0 & 1_n^T \\ 1_n & K + I/\lambda \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \qquad (6)$$

where $y = [y_1; y_2;...; y_n]$, $\alpha = [\alpha_1;...;\alpha_n]$, I_n is the identity matrix of dimension n and K is the n×n matrix with elements, $\varphi(x_i)^T \varphi(x_i)$, $i, j = 1,...,\text{n}$, 1_n is the vector of ones of dimension n.

After solving the linear equation, we can find the optimal values of the bias. $\hat{b}$. and Lagrange multipliers $\hat{\alpha}_i$. Then the optimal regression function for the given x_0 is obtained as:

$$y(x_0) = \sum_{i=1}^{N} \hat{\alpha}_i K(x_0, x_i) + \hat{b} \qquad (7)$$

Three-level random intercept and trend LS-SVR

Now, we propose the MLS-SVR with random intercept and trend for a three-level longitudinal data. Let the training data set be $G = \{(x_{ijk}, y_{ijk})\}_{i,j,k=1}^{N,n_i,n_{ij}}$, where y_{ijk} is the k-th observation of the response variable of the j-thcentre of the level-2 in i-thcentre of the level-3 corresponding to p×1 fixed-effects covariates x_{ijk}. We assume that y_{ijk} is related to in x_{ijk} a nonlinear regression form as

$$y_{ijk} = \omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + v_{ij}' z_{ijk} + e_{ijk} \qquad i=1,...,N, j=1,...,n_i, k=1,...,n_{ij} \qquad (8)$$

where $\varphi(X_{ijk})$ is a nonlinear feature mapping function, b is the bias term, z_{ijk} is q×1 random effect covariate vector with the random effect parameter v_{ij} for level-2 from , $N(0,\Sigma)$, γ_{0i} is a random effect parameter for level-3 from $N(0,\sigma_r^2)$ and nij×1 error vector .

$$e_{ij} = (e_{ij1}, e_{ij2},...,e_{ijn_{ij}})^t \sim N_{n_{ij}}(0, R_{ij})$$

For known Σ and Rij the optimisation problem of the nonlinear three-level mixed effects LS-SVR can be define as

$$\min \quad J = \frac{1}{2}\|\omega\|^2 + \frac{\lambda_1}{2}\sum_{i=1}^{N}\gamma_{0i}^2 + \frac{\lambda_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{n_i} v_{ij}' \Sigma^{-1} v_{ij} + \frac{\lambda_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k,l=1}^{n_{ij}} e_{ijk} R_{ij\,kl}^{-1} e_{ijl} \qquad (9)$$

subject to equality constraints

$$y_{ijk} = \omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + v_{ij}' z_{ijk} + e_{ijk} \qquad i=1,...,N, j=1,...,n_i, k=1,...,n_{ij} \qquad (10)$$

Here $\lambda_1, \lambda_2, \lambda_3$ are tuning or regularization parameters and $R_{ij\,kl}^{-1}$ is the (k,l)th element of the inverse matrix of Rij, $i=1,...,N, j=1,...,n_i, k=1,...,n_{ij}$ .The Lagrange function can be written as

$$L = J - \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk}(\omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + v_{ij}' z_{ijk} + e_{ijk} - y_{ijk}) \qquad (11)$$

where α_{ijk} are Lagrange multipliers. The conditions for optimality are given by

$$\begin{cases} \dfrac{\delta L}{\delta \omega} = 0 \quad \rightarrow \quad \omega = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk}\varphi(x_{ijk}) \\ \dfrac{\delta L}{\delta b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk} = 0 \\ \dfrac{\delta L}{\delta \gamma_{0i}} = 0 \quad \rightarrow \quad \lambda_1 \gamma_{0i} - \sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk} = 0 \\ \dfrac{\delta L}{\delta v_{ij}} = 0 \quad \rightarrow \quad \lambda_2 \Sigma^{-1} v_{ij} - \sum_{k=1}^{n_{ij}} \alpha_{ijk} z_{ijk} = 0 \\ \dfrac{\delta L}{\delta e_{ijk}} = 0 \quad \rightarrow \quad \lambda_3 \sum_{k=1}^{n_{ij}} R_{ij}^{-1} e_{ijk} - \alpha_{ijk} = 0 \\ \dfrac{\delta L}{\delta \alpha_{ijk}} = 0 \quad \rightarrow \quad \omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + v_{0ij} + e_{ijk} - y_{ijk} = 0 \end{cases} \qquad (12)$$

After eliminating v_{0ij} , γ_{0i} , ω, e_{ijk}, we can obtain the

optimal values of α_{ijk} and b by solving the following linear system equation as

$$\begin{pmatrix} 0 & 1_{N_t}^T \\ 1_{N_t} & K+\frac{1}{\lambda_1}\tilde{\Gamma}+\frac{1}{\lambda_2}Z\tilde{\Sigma}Z^T+\frac{1}{\lambda_3}\tilde{R} \end{pmatrix}\begin{pmatrix} b \\ \alpha \end{pmatrix}=\begin{pmatrix} 0 \\ y \end{pmatrix} \qquad (13)$$

where K is the kernel function in the form of, $K(x_{ijk},x_0)=\varphi'(x_{ijk})\varphi(x_0)$, $N_t=\sum_{i=1}^{n_i}\sum_{j=1}^{n_i}n_{ij}$, 0_{N_t} and 1_{N_t} are the (Nt×1) vector of zeros and ones respectively, $\tilde{\Gamma}=diag(\Gamma_1,....,\Gamma_N)$ where G_i is a square matrix of one by ni dimension, $\tilde{Z}=diag(Z_{11},...,Z_{1n_i},...,Z_{N1},...,Z_{Nn_i})$ is the Nt×Nq block diagonal matrix with $Z_{ij}=(z_{ij1},z_{ij2},...,z_{ijn_{ij}})'$, $\alpha=(\alpha'_1,...,\alpha'_N)$ where $\alpha_i=(\alpha_{i1},...,\alpha_{in_i})'$ and $\alpha_{ij}=(\alpha_{ij1},...,\alpha_{ijn_{ij}})'$, $y=(y'_1,...,y'_N)$ where $y_i=(y_{i1},...,y_{in_i})'$ and $y_{ij}=(y_{ij1},....,y_{ijn_{ij}})'$.

The optimal regression function for the given (x0,z0) is as follows

$$y(x_0,z_0)=\hat{b}+\sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}}\hat{\alpha}_{ijk}K(x_{ijk},x_0)+\hat{\gamma}_{0i}+\hat{v}'_{ij}z_0 \qquad (14)$$

The solution of the linear system equation depends on the **S** and R_{ij}. For this problem, we use an iterative procedure presented in [23] for the estimation of (α,b) and (Σ,R_{ij}) for given regularization parameters as follows:

1. Start with the initial values $\tilde{s}=I_q$ and $\hat{R}_{ij}=I_{n_u}$.

2. Calculate $(\hat{\alpha},\hat{b})$ based on equation (13) using the $\tilde{\mathbf{S}}$ and $\hat{R}_{ij}$ obtained in the previous step.

3. Calculate the estimates $\hat{\mathbf{S}}$ and $\hat{R}_{ij}$ using the obtained in the previous step as follows:

$\hat{\Sigma}=\frac{1}{N-1}\sum_{i=1}^{N}\hat{v}_i\hat{v}'_i$ and $R_{ijkl}=\hat{\sigma}^2\hat{\rho}^{|k-l|}$

where $\hat{\sigma}^2=\frac{1}{m}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}\hat{e}_{ij}^2$ and $\hat{\rho}=\frac{1}{m}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{\sum_{k=2}^{n_{ij}}\hat{e}_{ij}\hat{e}_{i,j-1}}{\sum_{k=2}^{n_{ij}}\hat{e}_{i,j-1}^2}\right)$

with $\hat{e}_{ij}=y_{ij}-\hat{y}(x_{ij},z_{ij})$.

4. Iterate steps until $\frac{1}{N_t}\|\alpha^{(k)}-\alpha^{(k+1)}\|^2<\varepsilon$

where k is the kth iteration and ε denotes the tolerance level. We consider ε=0.1 in this paper.

Regularization parameters

The functional structure of the MLS-SVR is characterised by tuning parameters, the regularization parameters $\lambda_1,\lambda_2,\lambda_3$ and parameters of the kernel function. To obtain optimal values of these tuning parameters, we use the GCV function. The inverse of the leftmost matrix in (13) can be partitioned into follow submatrices as

$$S=\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \qquad (15)$$

where S_{11} is a scalar, S_{12} is a (1×Nt) vector, S_{21} is a (Nt×1) vector and S_{22} is a (Nt×Nt) matrix.

Then, $\hat{y}$ can be expressed as $\hat{y}=Sy$, where $S=1S_{12}+(K+\frac{1}{\lambda_1}\tilde{\Gamma}+\frac{1}{\lambda_2}Z\tilde{\Sigma}Z')S_{22}$ The GCV function can be obtained by applying the leave-one-out method and the first order Taylor expansion as follows

$$GCV(\theta)=\frac{N_t y'(I-S)'\tilde{R}^{-1}(I-S)y}{(N_t-tr(S))^2} \qquad (16)$$

where θ is a set of tuning parameters. The optimal values of the tuning parameters are those that minimise the GCV function.

Human Brucellosis Rate Example

The Brucellosis data contains monthly frequency of subjects diagnosed with human Brucellosis. The data was related to cities of five provinces in the west of Iran (Hamadan, Kordestan, Ilam, Lorestan and Kermanshah) in 2016-2017. We transformed the monthly frequency of Brucellosis (a count variable) to the logarithm of the rate (a continuous response variable) as follows

$$y_{continuous}=\log\left(\frac{y_{count}+0.04}{\text{city population}}\right). \qquad (15)$$

The constant 0.04 was added to brucellosis count in equation (15) for obtaining the non-zero rates. The box-plot of the logarithm of brucellosis rate in each month was plotted in figure 1.

The ratio of rural to urban population, livestock (sheep, cattle, and goat) population (frequency), area of croplands, forest, and grassland (hectare) and some climatic variables such as monthly average of temperature, monthly sum of sunshine hours, rainfall, humidity and wind speed were considered as covariates. These covariates had extracted from Statistical Yearbook of Iran in 2016-2017 and Iran Meteorological Organisation website [29,30].

There were five provinces (level-3), 54 cities nested in these provinces (level-2) and 12 observations in each city (level-1). So, we had 648 observations in total. We used the first 10 observations of each city as the training and the last two observations as the testing data sets. So, the sizes of the training and testing data sets were 540 and 108 for this example respectively. We fitted MLS-SVR with random intercept and trend (MLS-SVR.it), MLS-SVR with random intercept (MLS-SVR.i), ordinary LS-SVR, LMM with random intercept and trend (LMM.it), LMM with random

intercept (LMM.i) on training data set. The testing data set was used to assess the generalisation performance of each model. Mean square error (MSE) and mean absolute error (MAE) were considered for comparison of fitted models generalisation capabilities.

Simulation study1

We performed a Monte Carlo simulation study to investigate the performance of different methods in prediction of observations.

Let y_{ijk} be the response of k th subject within j thcentre of the level-2 within ith centre of the level-3 related to covariates $x_{ijk}^{(p)}$. We generated data from the regression equation as

$$\mu_{ijk} = (Beta * \tau_{ijk}) + b + \gamma_{0i} + v'_{ij}z_{ijk} + e_{ijk} \qquad (16)$$

where $\tau_{ijk} = (x_{ijk}^{(1)}, \frac{\sin(\pi x_{ijk}^{(p)})}{x_{ijk}^{(p)}}), \; p = 2,...,13, \; i = 1,...,N, j = 1,...,n_i, k = 1,...,n_{ij}$, $x_{ijk}^{(1)}$ is the time variable (month=1,…,12), b= -11, Beta=(0.08,0.18,0.26,0.02,0.14,0.61,0.1,-0.16,0.65,-0.18,0.12,-0.1,-0.08), the covariates $x_{ijk}^{(p)}$ $p=2,...,13$ generated from normal distribution with zero mean and unit variance. The coefficients obtained from the real data example. The correlation between the first five covariates was

$$\begin{bmatrix} 1 & -0.6 & 0.0 & 0.4 & -0.6 \\ & 1 & -0.6 & -0.9 & 0.9 \\ & & 1 & 0.7 & -0.6 \\ & & & 1 & -0.8 \\ & & & & 1 \end{bmatrix} \qquad (17)$$

.

We considered two different values (5 and 10) for the number of centres at level-2 and level-3. The sample size of level-1 was 12 (number of the month) for each centreof level-2 in all scenarios. The distribution of the random effect was considered as normal with zero mean and standard deviation equal1.2 for level-3. The random effect of level-2 simulated from $N(0,\Sigma)$ where $\Sigma = \begin{bmatrix} 0.5 & \rho(\sqrt{0.5 \times 0.05}) \\ \rho(\sqrt{0.5 \times 0.05}) & 0.05 \end{bmatrix}$. We considered two levels for ρ as (-0.15, 0.7). Also the standard deviation of the error variable with standard normal distribution was considered in two levels (1.3, 0.3).There were 16 settings in this simulation study. In all conditions, we used the first ten observations of each centre of the second level as training data set and the last two observations were placed in the testing data set. We fitted the MLS-SVR.it, M LS-SVR.i, ordinary LS-SVR, LMM.it, and LMM.i on the training data and estimated the parameters of each model. In the MLS-SVR and LS-SVR methods, the radial basis function (RBF) was used as the kernel function. The fitting and generalisation performance of each model were evaluated using two measures of MSE and MAE in training and testing data sets, respectively. This procedure was repeated 1000 times in each scenario.

Simulation Study2

Another simulation study was performed to assess the performance of the proposed MLS-SVR in the presence of a correlation structure in the error variables. We compared our proposed model with the LS-SVR and the LMM. The data generated from

FIGURE 1. Box-plot of the Brucellosis rate for each month in logarithmic scale (red points are the mean of the logarithm of brucellosis rate)
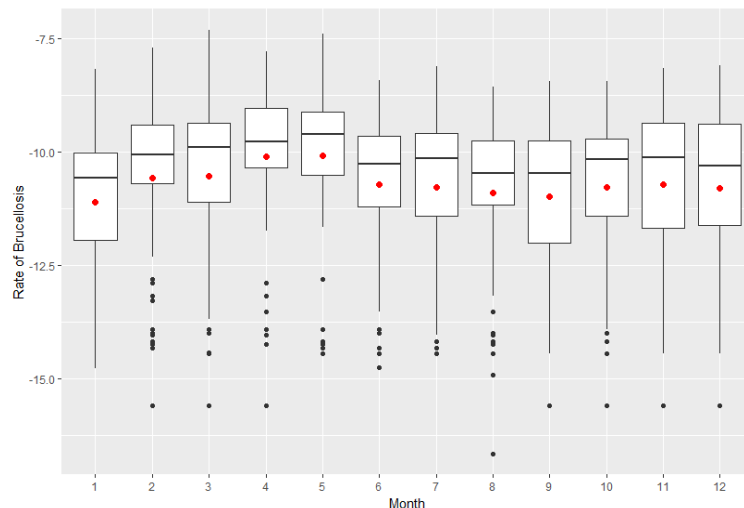
$$\mu_{ijk} = (Beta * \tau_{ijk}) + b + \gamma_{0i} + \nu_{0ij} + e_{ijk}. \qquad (18)$$

Here the Beta, τijk, b,γ0i, and the number of centres in each level are the same as simulation example 1. Also ν_{0ij}, simulated from, $N(0,0.25)$ $\gamma_{0i} \sim N(0,1.4)$ and $e_{ij} = (e_{ij1}, e_{ij2}, ..., e_{ijn_q}) \sim N(0, R_{ij})$ with $r_{ij_{kl}} = 0.5^{|k-l|}$ for k,l=1,2,…12. Also, $x_{ijk}^{(p)}$ $p=1,...,13$ were the same as those of simulation study 1. In this simulation study, the number of observations in each centre of the level-2 was 12 which the first ten observations were used as the training data and the last two observations were considered as the testing data. We fitted MLS-SVR with random intercept using an autoregressive structure for the errors (MLS-SVR.i.cor), MLS-SVR with random intercept (MLS-SVR.i), ordinary LS-SVR, LMM with random intercept and autoregressive structure for the errors (LMM.i.cor), LMM with random intercept (LMM.i) on the training data set. Then we investigated the generalisation performance of each model using the testing data set. Like the simulation study 1, we used the RBF as the kernel function in the LS-SVR methods. Each scenario of this simulation study was repeated 1000 times. After dividing the generated data sets into training and testing sets, we had 1000 training data set and 1000 testing data set for each scenario.

Computational Support

We used the R software version 3.5.1 [31]. The package "nlme" [32] was used for fitting LMMs. Also some packages such as "kernlab" [33], "Matrix" [34], and "MASS" [35] were used for coding ordinary LS-SVR and proposed MLS-SVR methods in simulation studies and real data example. The R codes of MLS-SVR are available upon the request from the first author.

RESULTS

Human Brucellosis rate example results

There were five provinces which Kermanshah had more cities than the others (14 cities). Lorestan and Ilam had the most and the least rate of brucellosis (per 100,000) with 80.7 and 18.8, respectively. The most livestock and lands belonged to Lorestan province with 1,899,697 people and 2,829,591 hectare, respectively. The descriptive statistics of the used variables in the human brucellosis data were shown in Table 1.

The results of fitting various models (MLS-SVR.it, MLS-SVR.i, LS-SVR, LMM.it, and LMM.i) in human Brucellosis rate example were presented in Table 2. According to the Table 2, the ordinary LS-SVR had smallest MSE (0.97) and MAE (0.72) in training data, but the generalisation performance of the proposed MLS-SVR.it model was better than the others based on both MSE (1.59) and

MAE(0,98) in testing set. The MSE and MAE criteria for MLS-SVR.it and MLS-SVR.i were smaller than their classical counterparts (LMM.it and LMM.i) for both training and testing sets. Also, the generalisation performances of the three LS-SVR methods in Table 2 were better than the LMMs. Furthermore, by fitting the ordinary LS-SVR and ignoring the random effect terms, the generalisation performance criteria got worse in testing set. Also, the fitting and generalisation performances of the models with random intercept and trend (MLS-SVR.it and LMM.it) were better than those of the methods which only considered the random intercept term (MLS-SVR.i and LMM.i).

We plotted the observed values of training and testing sets versus the prediction values obtained by the MLS-SVR.it and LMM.it in Figure 2. Based on Figure 2, the MLS-SVR.it had a better performance than the LMM.it in prediction of response variable over both training and testing sets (the points were closer to the bisector line).

Simulation study 1 results

In this simulation study, both random intercept and trend effects were considered in data generation process. The results of the simulation study 1 were given in Table 3, 4 (MSE and MAE along with their standard errors (SE) over 1000 replications) for the training and testing data, respectively. These tables describe the performance of the proposed MLS-SVR.it along with that of the MLS-SVR.i, LS-SVR, LMM.it, and LMM.i. According to the results, the proposed method outperformed other models in terms of the used criteria. In this regard, the MLS-SVR.it yielded the smallest mean of MSEs and MAEs in both fitting and generalisation performance. The results of Table 3 and Table 4 showed that the fitting and generalisation performance of the MLS-SVR.it and MLS-SVR.i were better than their LMM counterparts (LMM.it and LMM.i) based on MSE and MAE criteria for all scenarios. Also, ignoring the random effect terms had a great influence on the prediction performance over both training and testing sets (see the columns for LS-SVR method in Table 3 and Table 4). Moreover, the prediction performance of the models which considered both random intercept and random trend (MLS-SVR.it, LMM.it) were better than those which only considered the random intercept (MLS-SVR.i, LMM.i) (Table 3 and Table 4).

Simulation study 2 results

We considered the random intercept effect and an autoregressive correlation structure for the error term in the data generating process of this simulation study. Table 5 shows the results of the simulation study 2 (MSE and MAE along with their standard errors (SE) over 1000 replications) for the training and testing data sets, respectively. The

TABLE 1. Description of the Brucellosis data variables in each provience (Mean, Min and Max are based on the monthly observations in all cities of each province)

	Province	Ilam	Kordestan	Kermanshah	Lorestan	Hamedan
Population	Number of Cities	10	10	14	11	9
	Brucellosis count	109	972	841	1420	882
	Urban (people)	395,263	1,134,229	1,468,615	1,134,908	1,097,217
	Rural (people)	184,444	468,768	478,444	623,896	639,005
	Brucellosis rate (per 100,000)	18.8	60.6	43.2	80.7	50.8
Livestock	Cattle (heads)	27,679	93,609	73,600	102,585	85514
	Sheep (heads)	559,033	811,423	979,895	1,316,334	1,023,724
	Goat (heads)	252,886	129,729	171,577	480,778	57,548
Lands	Croplands (hectare)	336,365	1,173,280	949,000	727,922	708,409
	Forest (hectare)	641,667	374,084	527,000	1,217,314	40,617
	Grassland (hectare)	784,694	1,414,000	1,193,000	884,355	822,000
Temperature (°C)#	Mean (SD*)	22.4 (10.4)	12.2 (9.8)	16.5 (9.8)	15.6 (9.4)	12.6 (9.4)
	Min**	3.4	-5.7	-1.3	-1.8	-7
	Max***	41	28.6	35.9	37.2	28.8
Rainfall (mm)##	Mean (SD)	45.5 (69.8)	38.9 (51.4)	43.9 (58.0)	45.2 (62.1)	34.8 (45.6)
	Min	0	0	0	0	0
	Max	339.1	227.3	288.8	244	212.5
Humidity (%)	Mean (SD)	36.7 (16.8)	48.2 (18.8)	43.1 (18.1)	43.0 (16.8)	45.4 (16.7)
	Min	1	17	13.9	15	19
	Max	68	83	76.7	77	76.1
Wind Speed (m/s)###	Mean (SD)	2.7 (0.9)	2.5 (0.5)	2.6 (0.9)	2.6 (0.8)	2.6 (0.9)
	Min	1	1.5	0.1	1	1
	Max	5.2	4.4	6.2	5.4	6
Sunshine (hours)	Mean (SD)	260.8 (66.4)	256.2 (79.7)	255.2 (68.2)	263.6 (71.5)	263.5 (70.7)
	Min	141.7	81.3	129.6	98	143.7
	Max	372.8	382.5	367	384.8	374.1

*standard deviation; **minimum; ***maximum; # celcius scale; ## millimetre; ### metre per second

FIGURE 2. Comparison of MLS-SVR.it and LMM.it methods in prediction of the training and testing sets of brucellosis rate example. Top panel (a) presents observed versus prediction values of the training set. Bottom panel (b) presents observed versus prediction values of the testing set.
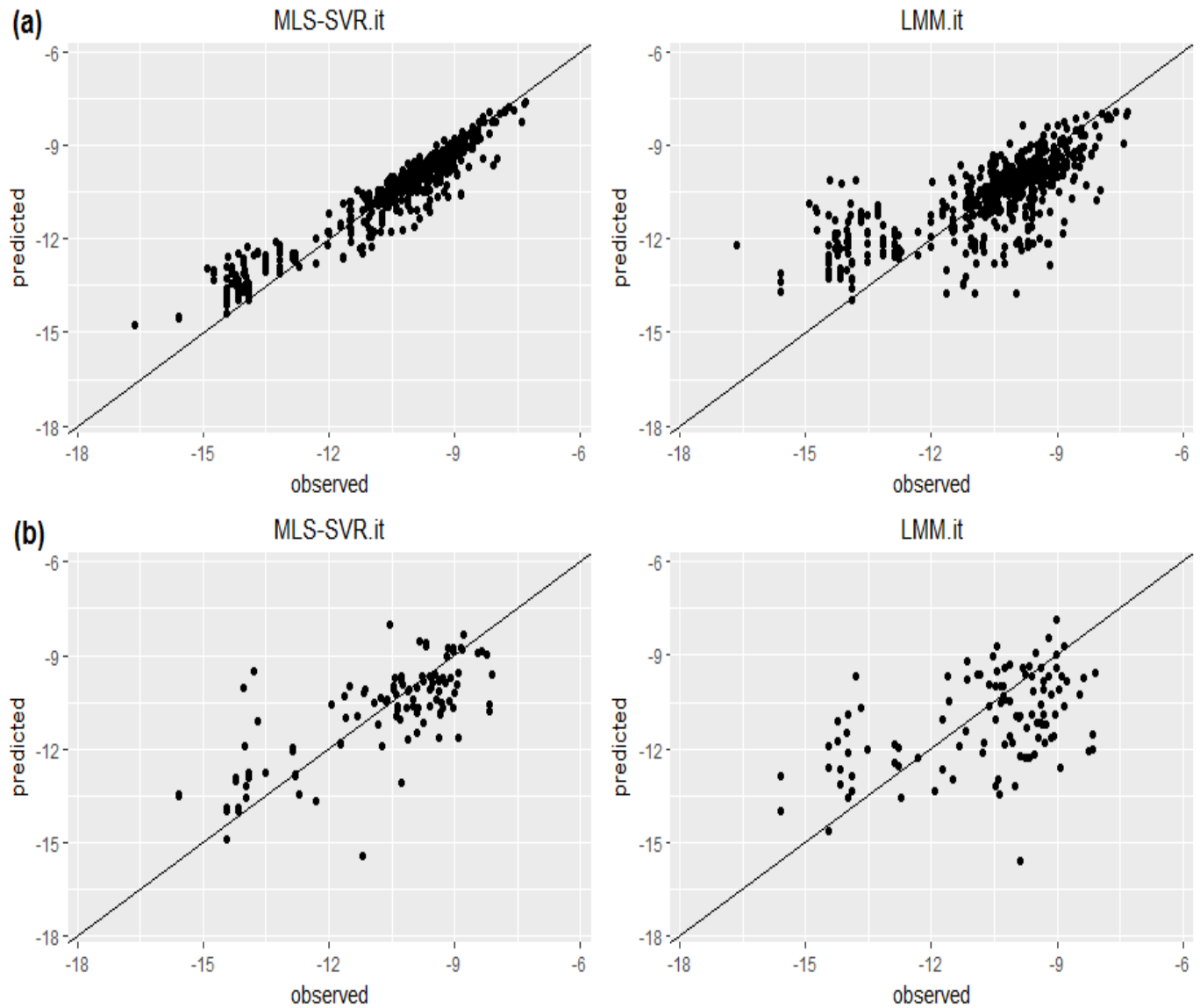
TABLE 2. Performance comparison of the proposed MLS-SVRs, ordinary LS-SVR, and LMMs for Brucellosis rate example

Data Type	Measure	MLS-SVR.it	MLS-SVR.i	LS-SVR	LMM.it	LMM.i
Training Data	MSE*	1.24	1.44	**0.97**	1.42	1.45
	MAE**	0.81	0.87	**0.72**	0.87	0.87
Testing Data	MSE	**1.59**	1.80	2.51	3.22	3.42
	MAE	**0.98**	1.05	1.24	1.44	1.48

*mean squares error; **mean absolute error*

prediction performance of various models (MLS-SVR.i.cor, MLS-SVR.i, LS-SVR, LMM.i.cor, and LMM.i) had been provided in this table. According to the MES and MAE measures reported in Table 5, the proposed MLS-SVR models had the best fitting performance among the fitted models. Also, the generalisation performance of the proposed MLS-SVR.i.cor was better than that of the other methods based on criteria shown in Table 5. The prediction performances of the MLS-SVR.i.cor and MLS.SVR.i were better than their LMM counterparts based on the MSE and MAE in both training and testing data sets. Furthermore, the MSE and MAE measures increased after eliminating the random effect term and the autoregressive structure of the error terms (see the columns for LS-SVR method in table 5). Although, the fitting performance of the LMM.i.cor was better than the LMM.i in training set, but its generalisation performance was worse over testing set. Nevertheless, both fitting and generalisation performances of the MLS-SVR.i.cor were better compared to the MLS-SVR.i.

Computational time

Using a computer with CPU: corei3 and 2.13 GHz, RAM: 4.00 GB, for known regularization parameters of machine learning techniques, the computation times of fitting different methods in Brucellosis rate data were 1.87, 1.20, 0.39, 2.53, and 0.26 seconds for the MLS-SVR.it, MLS-SVR.i, LS-SVR, LMM.it, and LMM.i, respectively. The computational time of the machine learning methods were different when the regularization parameters were unknown. The GCV function was used in this case which changed the computation time of the MLS-SVR.it, MLS-SVR.i, and LS-SVR to 88.00, 41.22, and 37.14 seconds, respectively.

DISCUSSION

A novel multi-level LS-SVR model was proposed in this paper for three-level longitudinal data. The proposed model used multiple random-effect terms to account for the correlated structures of the levels in the longitudinal data. Our proposed model considered both random intercept and random trends. Also, a correlation structure can be considered in the proposed MLS-SVR model for the error terms. Like the LS-SVR, the proposed MLS-SVR is computationally efficient due to a need to solve a linear system and it is useful for modelling the highly unbalanced data sets (non-fixed time points and unequal number of observations) with (non-)linear complex relationships between covariates and response.

In the real data example related to the human Brucellosis rate, the performance of the proposed MLS-SVR.it and ordinary LS-SVR were the best in terms of generalisation and fitting, respectively. Moreover, the MSE and MAE of the MLS-SVR.it were smaller than those

of the LMM.it. The performances of the proposed models (MLS-SVR.it and MLS-SVR.i) were better compared with the LMMs in both training and testing data sets of this example. There might be nonlinear relationships between covariates and response which the classical approaches (LMM.it and LMM.i) could not capture them well as machine learning techniques (MLS-SVR.it and MLS-SVR.i). Also, the generalisation performance decreased after eliminating the random effect terms and fitting the ordinary LS-SVR showing the influence of random effect terms in prediction performance. Both MLS-SVR.it and LMM.it, which considered the random intercept and trend, had a better performance in prediction than the MLS-SVR.i and LMM.i, that considered only the random intercept, in training and testing data sets. It shows the importance of taking into account the random trend term which should be considered in the models for a better predictions.

Based on the simulation study 1, the generalisation and fitting performance of the proposed MLS-SVR models were the best among the fitted methods in all scenarios. According to the results of the first simulation study, the MSE and MAE measures of the MLS-SVR.it and MLS-SVR.i models were smaller than their LMM counterparts in both training and testing sets. The reason of the better performance of the MLS-SVR method might be due to the created nonlinear relationships (using the sine function in data generation process) between the features and outcome. Also, it seems that the random intercept and random trend effects have important roles in prediction performance in the first simulation study and must be considered in the modelling.

The proposed MLS-SVR models had the best performance in fitting and generalisation among the fitted models in the simulation study 2. The best generalisation performance belonged to the MLS-SVR.i.cor in all scenarios. Also, the prediction performances of the MLS-SVR.i.cor and MLS-SVR.i were better than their LMM counterparts in training and testing sets. Like the first simulation study, the created nonlinear relationships between covariates and outcome might be the reason of this better performance. Therefore, in presence of nonlinear or complex relationships between covariates and response, the proposed MLS-SVR method could be a better choice for prediction problems. The random intercept term and autoregressive structure of the error terms were influential in prediction performance and ignoring them (fitting ordinary LS-SVR) increased the MSE and MAE of the models. Therefore, they should be considered in the modelling process. Moreover, although the performance of the LMM.it was better than that of the LMM.i in the training set, but the LMM.i had a better prediction performance than LMM.it in the testing set. Also, by increasing the sample size, the MSE and MAE measures of the MLS-SVR.i.cor, MLS-SVR.i, LMM.i.cor, and LMM.i tended to be decreased.

Several studies have been done which used the LS-SVR technique in longitudinal or clustered data. Seok

TABLE 3. Fitting performance comparison of the proposed MLS-SVRs, ordinary LS-SVR and LMMs on training data set for simulation study 1 (standard error in parenthesis)

Correlation of Month and level-2	number of Level-3 clusters	number of Level-2 clusters	SD* of error	MLS-SVR.it		MLS-SVR.i		LS-SVR		LMM.it		LMM.i	
				MSE#	MAE##	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
-0.15	5	5	1.3	**1.22** (0.017)	**0.86** (0.007)	1.98 (0.038)	1.08 (0.009)	4.28 (0.065)	1.60 (0.013)	3.08 (0.010)	1.40 (0.002)	3.58 (0.010)	1.51 (0.002)
			0.3	**0.25** (0.006)	**0.37** (0.005)	1.02 (0.029)	0.74 (0.009)	2.87 (0.050)	1.30 (0.012)	1.73 (0.006)	1.06 (0.002)	2.18 (0.007)	1.19 (0.002)
	10	5	1.3	**1.36** (0.012)	**0.92** (0.005)	3.12 (0.191)	1.25 (0.021)	4.76 (0.048)	1.71 (0.009)	3.17 (0.008)	1.43 (0.002)	3.69 (0.008)	1.54 (0.002)
			0.3	**0.19** (0.004)	**0.33** (0.003)	1.18 (0.031)	0.80 (0.010)	3.21 (0.037)	1.40 (0.008)	1.77 (0.004)	1.08 (0.001)	2.25 (0.005)	1.21 (0.002)
	5	10	1.3	**1.38** (0.012)	**0.93** (0.004)	3.01 (0.147)	1.25 (0.019)	4.59 (0.051)	1.68 (0.010)	3.19 (0.008)	1.43 (0.002)	3.70 (0.008)	1.54 (0.002)
			0.3	**0.19** (0.004)	**0.33** (0.003)	1.20 (0.031)	0.80 (0.010)	3.20 (0.041)	1.39 (0.009)	1.77 (0.004)	1.08 (0.001)	2.26 (0.005)	1.21 (0.001)
	10	10	1.3	**1.26** (0.007)	**0.89** (0.002)	2.27 (0.159)	1.14 (0.011)	5.15 (0.035)	1.79 (0.007)	3.22 (0.006)	1.44 (0.001)	3.74 (0.006)	1.55 (0.001)
			0.3	**0.13** (0.001)	**0.29** (0.001)	0.70 (0.013)	0.62 (0.005)	3.44 (0.033)	1.45 (0.007)	1.79 (0.003)	1.08 (0.001)	2.28 (0.004)	1.21 (0.001)
0.7	5	5	1.3	**1.20** (0.017)	**0.85** (0.007)	2.12 (0.047)	1.12 (0.010)	5.79 (0.080)	1.86 (0.014)	3.07 (0.010)	1.40 (0.002)	3.55 (0.011)	1.51 (0.003)
			0.3	**0.26** (0.006)	**0.38** (0.005)	1.22 (0.037)	0.80 (0.010)	4.44 (0.071)	1.62 (0.013)	1.72 (0.005)	1.06 (0.002)	2.17 (0.007)	1.18 (0.002)
	10	5	1.3	**1.34** (0.012)	**0.91** (0.004)	2.92 (0.130)	1.26 (0.016)	6.42 (0.063)	1.98 (0.010)	3.16 (0.007)	1.42 (0.002)	3.65 (0.008)	1.53 (0.002)
			0.3	**0.20** (0.004)	**0.34** (0.003)	1.22 (0.029)	0.81 (0.009)	4.87 (0.055)	1.71 (0.010)	1.79 (0.004)	1.08 (0.001)	2.24 (0.005)	1.20 (0.001)
	5	10	1.3	**1.35** (0.012)	**0.91** (0.004)	3.39 (0.206)	1.31 (0.021)	6.29 (0.065)	1.96 (0.011)	3.17 (0.007)	1.43 (0.002)	3.65 (0.008)	1.53 (0.002)
			0.3	**0.20** (0.004)	**0.34** (0.003)	1.28 (0.037)	0.83 (0.010)	4.65 (0.057)	1.67 (0.011)	1.80 (0.004)	1.08 (0.001)	2.24 (0.005)	1.20 (0.001)
	10	10	1.3	**1.25** (0.006)	**0.89** (0.002)	2.24 (0.029)	1.17 (0.006)	6.85 (0.042)	2.06 (0.007)	3.23 (0.005)	1.44 (0.001)	3.72 (0.006)	1.54 (0.001)
			0.3	**0.13** (0.001)	**0.29** (0.001)	0.82 (0.017)	0.67 (0.006)	5.22 (0.041)	1.78 (0.007)	1.83 (0.003)	1.09 (0.001)	2.28 (0.004)	1.21 (0.001)

*standard deviation; #meaan squares error; ## mean absolute error

et al. proposed a mixed-effects LS-SVR and used it for pharmacokinetic (PK) and pharmacodynamic (PD) longitudinal data sets. The performance of their proposed LS-SVR method was better than that of the standard method for the analysis of population PK and PD data in both training and testing data [22]. In another study, the prediction performance of the proposed mixed-effects LS-SVR was better compared to the LMM in the two real data examples and two simulation studies for both training and testing sets [23]. The LS-SVR methods in the above two studies were proposed for two-level longitudinal data. Moqaddsi et al. extended the LS-SVR technique and proposed three-level mixed-effects LS-SVR for count data. The prediction performance of their proposed model was better compared to the ordinary LS-SVR, LMM, (zero-inflated) poisson mixed-effects model, and (zero-inflated) negative binomial mixed-effects model in a simulation study. Also, they used their proposed model in a real data example which outperformed other techniques [28].

We used the RBF in MLS-SSVR and LS-SVR models as the kernel function. Other kernels like polynomial can be used in the proposed model. Although the numerical studies were performed on the balanced data sets, the proposed MLS-SVR can be used for the very unbalanced data sets with high-dimensionality. An autoregressive structure for the error term was used in simulation study1. Other correlation structures for error terms can be considered in the proposed MLS-SVR method.

CONCLUSIONS

It seems that, in presence of unspecific nonlinear relationships between covariates and outcome, the proposed MLS-SVR model can be more useful. Also, the

TABLE 4. Generalization performance comparison of the proposed MLS-SVRs, ordinary LS-SVR and LMMs on testing data set for simulation study 1 (standard error in parenthesis)

Correlation of Month and level-2	Number of Level-3 clusters	number of Level 2 clusters	SD* of error	MLS-SVR.it		MLS-SVR.i		LS-SVR		LMM.it		LMM.i	
				MSE#	MAE##	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
-0.15	5	5	1.3	**5.10** (0.037)	**1.80** (0.007)	6.45 (0.047)	2.03 (0.008)	11.76 (0.099)	2.73 (0.012)	5.30 (0.035)	1.84 (0.006)	6.64 (0.044)	2.06 (0.007)
			0.3	**2.79** (0.020)	**1.34** (0.005)	4.43 (0.034)	1.68 (0.007)	9.92 (0.087)	2.51 (0.011)	3.14 (0.020)	1.42 (0.005)	4.55 (0.031)	1.71 (0.006)
	10	5	1.3	**4.91** (0.024)	**1.77** (0.005)	6.23 (0.032)	1.99 (0.005)	11.82 (0.068)	2.75 (0.008)	5.07 (0.022)	1.80 (0.004)	6.41 (0.030)	2.03 (0.005)
			0.3	**2.51** (0.013)	**1.27** (0.003)	4.19 (0.023)	1.63 (0.005)	10.19 (0.062)	2.55 (0.008)	2.96 (0.013)	1.39 (0.003)	4.46 (0.022)	1.69 (0.004)
	5	10	1.3	**4.86** (0.023)	**1.76** (0.004)	6.20 (0.031)	1.99 (0.005)	11.73 (0.068)	2.73 (0.008)	5.02 (0.022)	1.79 (0.004)	6.41 (0.030)	2.02 (0.005)
			0.3	**2.49** (0.013)	**1.27** (0.003)	4.21 (0.024)	1.64 (0.005)	10.05 (0.063)	2.53 (0.008)	2.98 (0.013)	1.39 (0.003)	4.48 (0.022)	1.70 (0.004)
	10	10	1.3	**4.64** (0.016)	**1.72** (0.003)	5.99 (0.021)	1.96 (0.004)	11.76 (0.048)	2.73 (0.006)	4.96 (0.016)	1.78 (0.003)	6.39 (0.021)	2.02 (0.004)
			0.3	**2.24** (0.008)	**1.20** (0.002)	4.01 (0.017)	1.60 (0.003)	9.94 (0.047)	2.51 (0.006)	2.90 (0.009)	1.37 (0.002)	4.44 (0.016)	1.69 (0.003)
0.7	5	5	1.3	**4.96** (0.035)	**1.78** (0.007)	6.58 (0.046)	2.05 (0.007)	15.35 (0.122)	3.12 (0.013)	5.11 (0.034)	1.81 (0.006)	6.61 (0.044)	2.06 (0.007)
			0.3	**2.68** (0.019)	**1.31** (0.005)	4.52 (0.035)	1.70 (0.007)	13.73 (0.125)	2.95 (0.014)	2.95 (0.018)	1.38 (0.005)	4.55 (0.032)	1.71 (0.006)
	10	5	1.3	**4.76** (0.023)	**1.75** (0.004)	6.37 (0.032)	2.02 (0.005)	15.48 (0.090)	3.14 (0.010)	4.87 (0.021)	1.77 (0.004)	6.39 (0.029)	2.02 (0.005)
			0.3	**2.39** (0.013)	**1.24** (0.003)	4.19 (0.023)	1.63 (0.005)	14.04 (0.085)	2.98 (0.010)	2.81 (0.013)	1.35 (0.003)	4.44 (0.021)	1.69 (0.004)
	5	10	1.3	**4.74** (0.024)	**1.74** (0.005)	6.42 (0.033)	2.02 (0.006)	15.25 (0.089)	3.12 (0.010)	4.84 (0.022)	1.76 (0.004)	6.43 (0.029)	2.03 (0.005)
			0.3	**2.40** (0.013)	**1.24** (0.004)	4.18 (0.023)	1.63 (0.005)	13.56 (0.085)	2.94 (0.010)	2.82 (0.012)	1.35 (0.003)	4.44 (0.021)	1.69 (0.004)
	10	10	1.3	**4.53** (0.016)	**1.70** (0.003)	6.20 (0.022)	1.99 (0.004)	15.43 (0.064)	3.13 (0.007)	4.76 (0.015)	1.75 (0.003)	6.38 (0.021)	2.02 (0.003)
			0.3	**2.19** (0.008)	**1.19** (0.002)	4.08 (0.017)	1.61 (0.004)	13.76 (0.062)	2.96 (0.007)	2.76 (0.009)	1.34 (0.002)	4.44 (0.015)	1.69 (0.003)

standard deviation of the error term; #meaan squares error; ## mean absolute error

TABLE 5. Prediction performance comparison of the proposed MLS-SVRs, ordinary LS-SVR, and LMMs for simulation study 2 on training and testing data sets (standard error in parenthesis)

	number of Level-3 clusters	number of Level-2 clusters	MLS-SVR.i.cor		MLS-SVR.i		LS-SVR		LMM.i.cor		LMM.i	
			MSE #	MAE ##	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Training Set	5	5	**0.73** (0.014)	**0.64** (0.008)	1.09 (0.053)	0.75 (0.012)	1.97 (0.034)	1.09 (0.009)	2.43 (0.009)	1.25 (0.003)	2.49 (0.008)	1.27 (0.002)
	10	5	**0.82** (0.012)	**0.70** (0.006)	2.24 (0.128)	1.20 (0.007)	2.31 (0.027)	1.20 (0.007)	2.51 (0.007)	1.27 (0.002)	2.56 (0.006)	1.29 (0.002)
	5	10	**0.82** (0.012)	**0.70** (0.006)	2.33 (0.130)	0.98 (0.022)	2.15 (0.029)	1.15 (0.008)	2.54 (0.007)	1.28 (0.002)	2.58 (0.006)	1.29 (0.002)
	10	10	**0.94** (0.006)	**0.77** (0.003)	1.17 (0.044)	0.81 (0.010)	2.35 (0.025)	1.21 (0.006)	2.58 (0.005)	1.29 (0.001)	2.61 (0.004)	1.30 (0.001)
Testing Set	5	5	**3.30** (0.023)	**1.46** (0.006)	3.33 (0.024)	1.46 (0.006)	4.62 (0.040)	1.72 (0.008)	3.47 (0.024)	1.50 (0.005)	3.41 (0.022)	1.48 (0.005)
	10	5	**3.11** (0.016)	**1.41** (0.004)	3.16 (0.017)	1.42 (0.004)	4.62 (0.030)	1.72 (0.006)	3.31 (0.015)	1.46 (0.004)	3.29 (0.014)	1.46 (0.003)
	5	10	**3.06** (0.015)	**1.40** (0.004)	3.10 (0.016)	1.41 (0.004)	4.38 (0.033)	1.67 (0.006)	3.27 (0.014)	1.45 (0.003)	3.24 (0.014)	1.45 (0.003)
	10	10	**2.88** (0.011)	**1.36** (0.003)	2.92 (0.011)	1.37 (0.003)	4.36 (0.025)	1.67 (0.005)	3.19 (0.011)	1.44 (0.002)	3.20 (0.010)	1.44 (0.002)

#meaan squares error; ## mean absolute error

prediction performance decreases by ignoring the random effect terms in longitudinal or clustered data sets. So the multilevel structure must be considered in the analysis of these types of data.

Acknowledgments

References

1. Diggle PJ, Heagerty P, Liang K-Y, et al. Analysis of longitudinal data. New York: Oxford University Press; 2002.
2. Hedeker D, Gibbons RD. Longitudinal data analysis. Vol. 451. New Jersey: John Wiley & Sons; 2006.
3. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer Science & Business Media; 2009.
4. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. Organizational research methods. 2004;7(2):127-150.
5. Bühlmann P, Van De Geer S. Statistics for high-dimensional data: methods, theory and applications. New York: Springer Science & Business Media; 2011.
6. Kutner MH, Nachtsheim CJ, Neter J, et al. Applied linear statistical models. Vol. 5. New York: McGraw-Hill Irwin Boston; 2005.
7. Vapnik VN. An overview of statistical learning theory. IEEE Trans Neural Netw. 1999;10(5):988-99.
8. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge university press; 2000.
9. Kashyap P. Machine Learning for Decision Makers. New York: Apress; 2017.
10. Rebala G. An Introduction to Machine Learning. New York: Springer; 2019.
11. Nisbet R, Elder J, Miner G. Handbook of statistical analysis and data mining applications. Academic Press; 2009.
12. Joachims T, editor Text categorization with support vector machines: Learning with many relevant features. European conference on machine learning; 1998: Springer.
13. Pappu V, Pardalos PM. High-dimensional data classification. Clusters, Orders, and Trees: Methods and Applications: Springer; 2014. p. 119-150.
14. Rohwer JA, Abdallah CT. One-vs-one multiclass least squares support vector machines for direction of arrival estimation. SANDIA NATIONAL LABS ALBUQUERQUE NM; 2003.
15. Wang L. Support vector machines: theory and applications. Vol. 177. New York: Springer Science & Business Media; 2005.
16. Rennie JD, Srebro N, editors. Loss functions for preference levels: Regression with discrete ordered labels. Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling; 2005: Kluwer Norwell, MA.
17. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and computing. 2004;14(3):199-222.
18. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural processing letters. 1999;9(3):293-300.
19. Suykens JA. Least squares support vector machines. New Jersey: World Scientific; 2002.
20. Luts J, Molenberghs G, Verbeke G, et al. A mixed effects least squares support vector machine model for classification of longitudinal data. Computational Statistics & Data Analysis. 2012 2012/03;56(3):611-628.
21. Chen T, Zeng D, Wang Y. Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration. Biometrics. 2015;71(4):918-928.
22. Seok KH, Shim J, Cho D, et al. Semiparametric mixed-effect least squares support vector machine for analyzing pharmacokinetic and pharmacodynamic data. Neurocomputing. 2011 2011/10;74(17):3412-3419.
23. Shim J, Sohn I, Hwang C. Kernel-based random effect time-varying coefficient model for longitudinal data. Neurocomputing. 2017;12;267:500-507.
24. Capitaine L, Genuer R, Thiébaut R. Random forests for high-dimensional longitudinal data. arXiv preprint arXiv:190111279. 2019.
25. Hajjem A, Bellavance F, Larocque D. Mixed-effects random forest for clustered data. Journal of Statistical Computation and Simulation. 2014;84(6):1313-1328.
26. Speiser JL, Wolf BJ, Chung D, et al. BiMM tree: a decision tree method for modeling clustered and longitudinal binary outcomes. Communications in Statistics-Simulation and Computation. 2018:1-20.
27. Speiser JL, Wolf BJ, Chung D, et al. BiMM forest: A random forest method for modeling clustered and longitudinal binary outcomes. Chemometrics and Intelligent Laboratory Systems. 2019;185:122-134.
28. Moqaddasi Amiri M, Tapak L, Faradmal J. A mixed-effects least square support vector regression model for three-level count data. Journal of Statistical Computation and Simulation. 2019:1-12. Statistical Center of Iran, Statistical Yearbook of Iran in 2016-2017. 2016-2017 [cited 2018]. Available from: https://www.amar.org.ir
29. IRAN Meteorological Organization. 2016-2017 [cited 2018]. Available from: http://irimo.ir
30. R Core Team. R: A Language and Environment for Statistical Computing. 2018.
31. Pinheiro J, Bates D, DebRoy S, et al. nlme: Linear and Nonlinear Mixed Effects Models. 2018.
32. Karatzoglou A, Smola A, Hornik K, et al. kernlab – An S4 Package for Kernel Methods in R. Journal of Statistical Software. 2004;11(9):1-20.
33. Bates D, Maechler M. Matrix: Sparse and Dense Matrix Classes and Methods. 2018.
34. Venables WN, Ripley BD. Modern Applied Statistics with S. New York: Springer; 2002.