Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2024;10(1): 98-110

Original Article

A Multi-Method Comparison of Machine Learning in Predicting Pharmacokinetic Parameters: A Simulation Study

Marziyeh Doostfatemeh, Kamal Amini, Elham Haem*

Department of Biostatistics, Shiraz University of Medical Sciences, Shiraz, Iran.

ARTICLE INFO

Received 08.12.2023
Revised 26.01.2024
Accepted 20.02.2024
Published 15.03.2024

Key words:
Population pharmacokinetics;
Machine learning; lasso;
Random forest;
Support vector regression

ABSTRACT

Introduction: One important aim of population pharmacokinetics (PK) and pharmacodynamics (PD) is the identification and quantification of the relationships between the parameter and covariates to improve the predictive performance of the population PK/PD modeling. Several new mathematical methods have been developed in pharmacokinetics in recent years which indicated that the machine learning-based methods are an appealing tool for analyzing PK/PD data.

Methods: This simulation-base study aims to determine whether machine learning methods, including support vector regression (SVR) and Random forest (RF) which are specifically designed for the prediction of blood serum concentration or clearance, could be an effective replacement for the Lasso covariate selection method in nonlinear mixed effect models. Accordingly, the predictive performance of penalized regression Lasso, SVR, and RF regression was compared to detect the associations between clearance and model covariates. PK data was simulated from a one-compartment model with oral administration. Covariates were created by sampling from a multivariate standard normal distribution with different levels of correlation. The true covariates influenced only clearance at different magnitudes. Lasso, RF, and SVR were compared in terms of mean absolute prediction error (MAE).

Results: The results show that SVR performed the best in small data sets, even in those in which a high correlation existed between covariates. This makes SVR a promising method for covariate selection in nonlinear mixed-effect models.

Conclusion: The Lasso method offered a higher MAE, making it less promising than RF and SVR, especially when dealing with a high correlation between covariates and a low number of individuals.

Introduction

Population pharmacokinetics and pharmacodynamics (PK/PD) models have been extensively used to identify how individual factors such as demographics, genotype, phenotypic disease, and environmental factors such as medications or alcohol consumption affect patient exposure to the drug and their further response. The structure of data in PK/

*.Corresponding Author: Elhamhaem@gmail.com

PD studies includes subject demographics, drug administration details, and measurements of drug concentration and effects and they are essential in understanding how drugs behave in the body and their effects on biological systems. Nonlinear mixed effects modeling, as implemented by the popular software, NONMEM, has been widely used in the analysis of PK/ PD data. The most important part of population PK/PD modeling is the evaluation of the relationships between model parameters and covariates. In this regard, the selection of a subset of the covariate relations is often performed via the stepwise method; however, this procedure has limited power to select important covariates in small or moderate data sets which could affect the predictive performance of the model for optimal dosage schedule and amount for patients and in addition this method could be tedious and time-consuming.[1-3] In PK/PD data the small sample sizes are often created due to feasibility, targeted populations, ethical considerations, and variability in drug response. Overcoming this problem, several new mathematical methods have been developed in pharmacokinetics in recent years.[4, 5] Several studies showed that machine learning-based (ML) methods are an appealing tool for analyzing PK/PD data which could improve the predictive performance of population PK/PD modeling.[2, 5-9] These studies demonstrated equivalent or greater accuracy than their statistical forbears, non-linear mixed effect models. Neural networks are the earliest alternative strategy in population PK/ PD data analysis. In this regard, Chow et al.[6] Examined the applicability of using a neural network approach to capture the relationships between the plasma levels and some PD factors and concluded that this method provided comparable predictions to the nonlinear mixed effects model. In another study, Kang et al.[9] Evaluated the ability of neural networks to assess the predictive performance of the PK/PD models. They concluded that this ML method produced more accurate predictions than the nonlinear mixed effects model concerning the data used.

Another powerful ML method for nonlinear relationships is support vector Machines(SVMs),[10] also, are useful analytical tools for population PK/PD modeling.[5] Medical literature has shown equivalent accuracy of neural networks and SVM models.[11-13] Some others, however, have shown that SVM is more accurate than neural networks.[12-14] In all of this literature, the predictability of the nonlinear mixed effects model under the stepwise covariate selection methods was compared to ML methods. In other words, none of them have compared nonlinear mixed effects model via other covariate selection methods such as Least Absolute Shrinkage and Selection Operator (Lasso) which was proposed by Tibshirani[14] which could deal with the problem of selection bias and poor predictive performance of the stepwise methods, especially in low sample studies.[3] Nonetheless, studies comparing the use of multiple ML methods, such as SVMs and random forest (RF),[15] advanced and effective prediction methods, are very scarce in pharmacokinetic data. These techniques are potentially useful for PK/PD data analysis if their accuracy exceeds or is equivalent to that of a nonlinear mixed effects model. RF regression is great for accuracy and handling complex data but lacks interpretability. Lasso Regression excels in feature selection and interpretability but assumes linearity. SVM is powerful for high-dimensional data but requires

careful tuning and can be less interpretable. Choosing the right algorithm depends on the specific problem, dataset characteristics, and the importance of interpretability versus accuracy.[16, 17] To the best of our knowledge, there is no simulation-based study that compares these ML models in pharmacokinetic modeling. Therefore, our research question was to determine whether these ML methods, including SVM and RF specifically designed to predict blood serum concentration or clearance, could be an effective replacement even for the Lasso covariate selection method in the nonlinear mixed effects model. Accordingly, the predictive performance of penalized regression Lasso, SVM regression, and RF regression was compared, to detect the associations between clearance, which is estimated by the nonlinear mixed effects model, and model covariates.

Methods

Modeling nonlinear mixed effects model with Lasso covariate selection

The population PK model utilizes the nonlinear mixed effects models which are hierarchical and comprised of three sub-models: structural, statistical, and covariate. The structural model, which uses fixed effects parameters, describes the overall trend in the individual data. The statistical model accounts for inter-individual and intra-individual (residual) random effect; and the covariate model for the relationships between covariates and the PK parameters. In the first step of data analysis, the individual drug concentration parameter is modeled at a certain time point j as follows:

$$C_{obs,ij}=f(x_i,P_i)+\varepsilon_{ij} \quad i=1,\ldots,N \quad j=1,\ldots,n \quad (1)$$

here N is the number of individuals, n is the number of observations per individual, $C_{obs,ij}$ is the jth observed concentration for individual *i*, *f(.)* is a nonlinear function of individual parameter-vector, P_i, and covariate-vector, x_i, which describes the prediction of concentrations and the term ε_{ij} that forms the random error of the prediction model which is normally distributed with zero mean and variance σ^2.

In predicting individual concentration, different PK models can be applied. The current study used a one-compartment model with a single oral administration: (one-compartment model with first-order input for oral administration)

$$\hat{C}_{ij}=\frac{Dose\times(Ka)_i}{V_i\times\left((Ka)_i-\frac{CL_i}{V_i}\right)}\left(e^{-\frac{CL_i}{V_i}\times t_j}-e^{-(Ka)_i\times t_j}\right),$$

where $\hat{C}_{ij}$ is the predicted concentration for the *i*th individual at time point *tj* and CL_i, V_i and Ka_i are the random effects parameters which indicate "clearance", "volume of distribution" and "absorption rate" for the ith individual, respectively.

In the second step of the analysis, the values of CL_i, V_i and Ka_i are obtained as:

$$CL_i=\exp\left(\ln(TVCL_i)+\eta_{CL_i}\right), \quad (3)$$

$$V_i=exp\left(\ln(TVV_i)+\eta_{V_i}\right) \quad (4)$$

$$Ka_i=(TVKa)_p.\exp\left(\eta_{Ka_i}\right), \quad (5)$$

Where $TVCL_i$, $TVVi$ and $TVka_i$ are the typical value of CL_i, V_i and Kai for the ith individual and the η_{CLi}, η_{Vi} and η_{Kai} are the random errors which are normally distributed with zero mean and variance ω_2.

The logarithm of typical value of the clearance for the *i*th individual, $Ln(\boldsymbol{TVCL}_i)$, have a linear

relationship with the individual covariates as:

$$n\left(TVCL_i\right) = ln(\theta_{CL}) + \sum_{K=1}^{N_{cov}} \beta_k x_{ik}, \quad (6)$$

Where θ_{CL} is the typical value of clearance for the population, β_k is the covariate coefficient, x_{ik} is the value of kth covariate for ith individual and NCOV is the number of covariates. $Ln(TVV_i)$ and $ln(TVka_i)$ can be a linear function of covariates; however, for simplicity, they are considered here to be only a linear function of $Ln(\theta_V)$ and $Ln(\theta_{kai})$ which θ_V and θ_{kai} are the typical population values of volume of distribution and absorption rate, respectively.

After modeling concentration with a nonlinear mixed effects model in a set of data, to explain the variability between the subject, the relationships between individual parameters and covariates are assumed to be modeled as the logarithm of the typical value of clearance according to Eq. (6). Therefore, the next important step of the analysis is using the Lasso as a covariate selection method. The Lasso uses a penalized estimation technique for linear models.[14] The Lasso estimates are defined as:

$$\hat{\beta} = argmin\left\{\sum_{i=1}^{N}\left(ln\left(TVCL_i\right) - \sum_{K=1}^{N_{cov}} \beta_k x_{ik}\right)^2\right\}, \quad (7)$$

Subject to the restriction of $\sum_{k=1}^{N_{cov}} |\beta_k| \le t$, where t is a tuning parameter that shrinks the coefficient β_k toward zero so that any estimate within the prespecified cut-point is forced to be zero. The value of t is estimated via cross-validation. For more details, refer to Ref. 3 and 15.[3, 15]

SVM methods

Vapnik[10] developed SVM foundations and are gaining popularity due to many interesting characteristics and desirable empirical performance. Although SVMs were originally proposed to solve classification problems, progressively they have been extended to the domain of regression problems.[18] In the present study, the support vector regression (SVR) method was implemented for the prediction of clearance. The idea of SVR is based on its ability to model the non-linearity of the data and map them into a higher dimensional feature space (here features are the covariates that affect pharmacokinetic parameters). SVR could obtain a function that approximates the hidden relationships of the pharmacokinetic data. This method attempts to find those features which minimize the generalization error.[19]

The theory of SVR has been widely described.[13, 18] Therefore, only a brief description is presented here. Suppose we have the training data $= \left(\boldsymbol{x}_1, CL_1\right), \ldots, \left(\boldsymbol{x}_N, CL_N\right) = \subset R^d \times R$ where $x_i \epsilon R^d$, $CL_i \epsilon R$. These might be, for instance, clearance, CL_i, which is measured for ith individuals with corresponding covariate vectors, x_i. In SVR, the goal is to obtain a function that has at most ε deviation from the obtained clearance CL_i for all the training data. In other words, we do not care about errors as long as they are less than ε, however, any deviation larger than ε would not be accepted. For instance, this may be important if somebody wants to be sure that the serum blood concentration level or clearance does not exceed a certain value, ε, after drug consumption. In the SVR algorithm, a kernel function, $K(x_i, x_k)$, is used to map the support vectors into a higher dimensional feature space, and linear regression is then performed in this space. The optimal regression function can be represented by:

$$\widehat{CL}_i(\boldsymbol{x}) = \sum_{i=1}^{N}\left(\alpha - \alpha^*\right)K\left(x, x_i\right) + b, \qquad (8)$$

Where the coefficient α, α^* and b are obtained by maximizing the following expression:

$$-\varepsilon\sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right)CL_i$$
$$-\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\alpha_i + \alpha_i^*\right)\left(\alpha_k + \alpha_k^*\right)\times\left(x_i, x_k\right)$$

subject to

$$0 \le \alpha_i, \alpha_i^* \le C, \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0, \qquad (9)$$

Where C is a penalty for training excessive error ε. For more details refer to Smola et.al.[18]

Random Forest regression

The Random Forests (RF) algorithm, which is introduced by Breiman as an extension of classification or regression tree,[14] is an increasingly popular ML algorithm within pharmacokinetic studies,[20-24] In this regard, RF regression is an ensemble of unpruned regression trees that are grown using a bootstrap sample of training data. This technique is capable of modeling a large number of covariates in the presence of a small number of observations and achieving good prediction performance. This ML approach is even able to account for high correlation and complex interaction structures among covariates, as the two common events in PK data. A major strength of random forests is its capability to provide variable importance metrics which could be used to substantially reduce the number of covariates used in the forest in high-dimensional problems. To construct a predictive model via RF regression in a pharmacokinetic study, consider a set of N subjects, $(\boldsymbol{x}_1, CL_1), \ldots, (\boldsymbol{x}_N, CL_N)$, for model training, where x_i and CL_i are the covariate vector and clearance level for an ith individual, respectively. the following steps should be executed:

First, from the training data, draw a bootstrap sample of size N (i.e., random sample, with replacement). Second, for each bootstrap sample, grow a tree. In the last step, the outputs of all trees are aggregated to produce one final prediction which is the average of the individual tree predictions (For classification problems, the final prediction is the class predicted by the majority of trees). For more theoretical details, refer to Breiman and Ziegler et al.[14, 25]

Simulation study

The PK data was simulated from a one-compartment model with a single oral administration identified in eq. 1. To reflect the effect of influential factors on the predictive performance of multiple methods of ML, different scenarios were considered. They are composed of the combination of several study individuals, the number of covariates and the value of correlation among covariates. Tables 1 and 2 include further details about the model simulation, particularly regarding the value of other parameters that remain constant throughout the scenarios. The following steps were taken to simulate a data set in every scenario;

1. Vector of covariates are created from a multivariate standard normal distribution with no (0.0), low (0.2), and moderate (0.5) correlations.
2. The vector of true coefficients β with various magnitudes is utilized to obtain $Ln(TVCL_i)$ as

Table 1. The simulation model used for different scenarios

True coefficients	Number of covariates	Correlation between covariates	Number of subjects
β = (0.25, 0.15, 0, 0, 0)	5	0, 0.2, 0.5	30, 60, 100
β = (0.25, 0.15, 0.1, 0.05,0,…,0)	10	0, 0.2, 0.5	30, 60, 100
β = (0.25, 0.25, 0.15, 0.15,0.1, 0.05, 0,…,0)	15	0, 0.2, 0.5	30, 60, 100

Table 2. Population values (θ) and inter-individual variability (ω) for the model parameters used in the simulation study

Parameters	θ	ω
θ_{CL}	0.5	0.2
LnTVV	0.2	0.2
LnTVka	10	0.2
σ	0.1	-
Dose	1	-

θCL, LnTVV and LnTVka are typical population values of clearance, volume of distribution and absorption rate, respectively.

in Eq. (6).

3. Individual parameter values are obtained as in Eqs. (3), (4) and (5) in which ω, θCL, $LnTVV_i$ and $LnTVka_i$ are considered to be 0.2, 0.5, 0.2 and 10, respectively. (Table 2)

4. A one-compartment model with a single oral administration with a Dose = 1 is used as in Eq. (2) in order to estimate individual concentrations.

5. The observation times are fixed to three-time points, 0.1, 2, 3.5.

6. The observed concentrations $C_{obs,ij}$ are generated using Eq. (1) where σ is consider to be 0.1.

7. The datasets that were simulated in the previous step, are then fitted with the base model (Eq. 2) without considering covariates.

8. The individual clearances *CLi* are estimated from the base model and used as the PK outcomes in covariate analysis and model fitting.

All scenarios were replicated 100 times. Training data was generated for each simulation scenario and for every replication and then used for model fitting. Finally, a large data set of 5000 subjects was generated for every simulation scenario, as the validation data. The number of observations per subject in the validation data was the same as that of the training. The validation data was used to evaluate the methods.

Evaluation of methods

The Lasso, SVR, and RF regression were implemented in R and their performance was investigated for all scenarios. To evaluate the performance of the Lasso, SVR, and RF regression mean absolute prediction error (MAE) was computed for the validation data sets. A low value of this index indicates optimal performance, reflecting a closer alignment between the predicted and actual values. MAE for a method in a given scenario was calculated as:

$$MAE = \frac{1}{100}\sum_{v=1}^{100}\left(\frac{1}{N_v}\sum_{n=1}^{N_v}\left|\widehat{CL}_{nv} - CL_{nv}\right|\right)$$

Where CL_{nv} and are $\widehat{CL}_{nv}$ the observed and predicted nth clearance out of N_v observations in the validation data set.

Software

This study was conducted in RStudio 1.3.1073 using R 3.5.0. All the data were simulated in R environment.[26] The nlmixr package was used for the implementation of nonlinear mixed effects model parameter estimation; glmnet, e1017, and randomForest packages were used to implement the lasso, SVR, and RF regression, respectively.[27-30]

Results

Figure 1 shows the MAE of the Lasso, RF, and SVR versus the number of subjects in combination with different numbers of covariates and different levels of correlation between covariates. It is evident that SVR had better predictive performance (lower MAE) than RF and Lasso when the number of subjects was low, whatever the number of covariates and the level of correlation between covariates, but as the number of subjects increased, the difference between the methods decreased so that when the size of the sample gets into one hundred individuals, no obvious distinction could be made between the three methods. The result also indicated that when 15 covariates were applied, poorer performance of the Lasso was observed compared to the other methods,

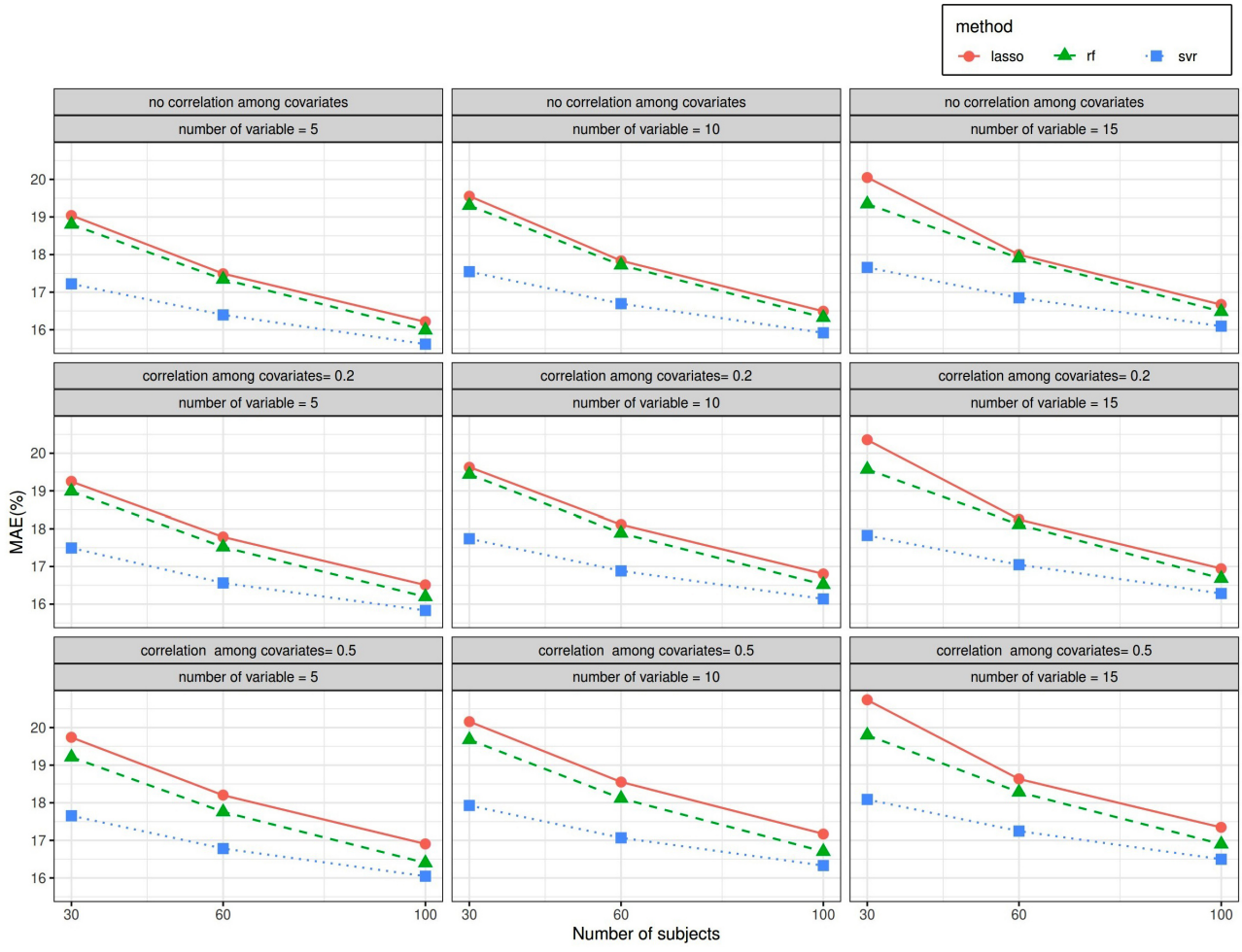

Figure 1. Comparison of the three methods (Lasso, RF, and SVR) with MAE criteria

particularly in the case of a small study sample (N=30). The difference in MAE between the Lasso and RF increased as the level of correlation increased and Lasso performed the worst, whatever the number of covariates.

Discussion

The present study provides a simulation-based framework to investigate the predictive performance of several machine learning methods in pharmacokinetic data. The performance of Lasso as a popular covariate selection method in PK was assessed and compared to RF and SVR in terms of prediction error (MAE). Ribbing et al.[5] showed that Lasso could outperform the stepwise procedure for a small data set; however, the advantage was insignificant in large data sets. The advantage of ML methods over other statistical modeling is achieving influential covariates when the number of study individuals is not sufficiently large relative to the number of covariates. It is important to specify which model performs best for small data sets since, as the size of the data set increases, the performance of common procedure such as stepwise becomes similar to more advanced modeling, like RF and SVR. Moreover, previous studies have shown that although Lasso is widely used to deal with large covariate, low sample size data situations, it tends to aggressively skip covariates correlated with already selected ones,[31-33] which is completely contrary to RF and SVR.[33-35] Furthermore, some research also indicated that RF and SVR performed better than Lasso in capturing the nonlinearity of the data.[33, 36] Admittedly, Lasso is not designed for the data with collinear covariate and nonlinear response variables.

These findings are consistent with the current study results; a comparison of the predictive performance of the three methods indicated that the performance of SVR is satisfactorily better than Lasso and RF specially for small data sets. Moreover, RF had the second-best performance whatever the number of covariates, sample size, and the level of correlations. In the case of a large sample size, any notable difference cannot be observed between methods. Bonate[37] indicated that when the collinearity between covariates in nonlinear mixed effects models increased, the estimation of parameters became increasingly biased and their standard error increased markedly. This finding is by the results of the present study. It was shown that increasing the level of correlation between covariates increases the MAE of all methods. However, in this study, Lasso had the poorest performance over the other methods. Nevertheless, SVR was not principally influenced by related covariates over the other two methods.

It is worth emphasizing that Lasso and other ML methods such as RF and SVR were proposed in a high-dimensional study with a very large number of unrelated covariates compared to the number of individual samples and when the number of covariates was larger than the study sample, as in gene selection in DNA microarray data.[38] However, in the present study, the number of covariates was not very large but was rational for population PK models.

One notable finding is that as the number of covariates increased from 5 to 15 the value of MAE of the three methods also increased; SVR showed the least and Lasso indicated the most increase in MAE particularly for 15 covariates. The rationale for this increase in MAE is that increasing the number of covariates decreases

the value of the tuning parameter which in turn decreases the estimated coefficients. These results are in line with the slightly increasing negative log-likelihood for Lasso in the generalized linear mixed-effect model reported by Schelldorfer et al.[39]

Major differences between the current study and other ML studies in the area of PK data[2, 9, 40] should be noted. First, the obtained results of the present study were based on simulation whereas that study's findings relied on the real PK data, therefore, relative to our study results, the generalization of their findings is more limited. Second, in the present study, the dependent variable for modeling Lasso, RF, and SVR was the empirical Bayes estimate of the individual PK parameters, clearance, which was estimated by the nonlinear mixed effect model in the simulation, whereas, in the mentioned ML studies the dependent variable was the observed concentration. This issue could be account for one of the limitations of this study, however, in the number of simulation-based pharmacogenetic studies the association between pharmacokinetics and genetic covariates was evaluated based on some penalized methods, and the dependent variable in these studies was also the clearance.[41, 42] In the present study, a one-compartment model with oral administration was used; however, this simple structural model cannot be considered as an important limitation to the results presented. The strategy for assessing model predictions and analyzing covariate relations in the previous studies[41, 42] has not used the complexity of the underlying structural model as a determinant factor for the applicability of a special method. There is, thus, no reason for the findings of the present study to be considered extraneous for more complex structural models.

Conclusion

In summary, the performance of Lasso as a penalized method for improving the predictive performance of the PK model in small data sets was compared with two other ML methods, RF and SVR, in a simulated one-compartment model with oral administration. The method performed poorly than SVR and RF, especially in the presence of high multicollinearity. The Lasso method offered a higher MAE, making it less promising than RF and SVR, especially when dealing with a high correlation between covariates and a low number of individuals.

List of abbreviations

PK: Population pharmacokinetics
PD: Pharmacodynamics
SVR: support vector regression
RF: Random forest
MAE: Mean absolute prediction error
ML: Machine learning
SVM: support vector Machine

Ethics approval and consent to participate

The study was approved by the local ethics committee of Shiraz University of Medical Sciences.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during the current study available from the corresponding author on

reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Grant Number 97-01-01-18512 from Shiraz University of Medical Sciences Research Council.

Authors' contributions

Marziyeh Doostfatemeh analyzed and wrote the manuscript and researched the data, Kamal amini analyzed and research the data, Elham Haem analyzed the data and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by Shiraz University of Medical Sciences Research Council. This article was extracted from Kamal Amini's Master of Sciences thesis.

References

1. Steyerberg EW, Eijkemans MJC, Habbema JDF. Stepwise Selection in Small Data Sets: A Simulation Study of Bias in Logistic Regression Analysis. Journal of Clinical Epidemiology. 1999;52(10):935-42.

2. Ribbing J, Nyberg J, Caster O, Jonsson EN. The lasso—a novel method for predictive covariate model building in nonlinear mixed effects models. Journal of Pharmacokinetics and Pharmacodynamics. 2007;34(4):485-517.

3. Seok KH, Shim J, Cho D, Noh G-J, Hwang C. Semiparametric mixed-effect least squares support vector machine for analyzing pharmacokinetic and pharmacodynamic data. Neurocomputing. 2011;74(17):3412-9.

4. Durisová M, Dedík L. New mathematical methods in pharmacokinetic modeling. Basic & clinical pharmacology & toxicology. 2005;96(5):335-42.

5. Haem E, Harling K, Ayatollahi SMT, Zare N, Karlsson MO. Adjusted adaptive Lasso for covariate model-building in nonlinear mixed-effect pharmacokinetic models. Journal of Pharmacokinetics and Pharmacodynamics. 2017;44(1):55-66.

6. Chow H-H, Tolle KM, Roe DJ, Elsberry V, Chen H. Application of Neural Networks to Population Pharmacokinetic Data Analysis. Journal of Pharmaceutical Sciences. 1997;86(7):840-5.

7. Poynton M, Choi B, Kim Y, Park I, Noh G, Hong S, et al. Machine learning methods applied to pharmacokinetic modelling of remifentanil in healthy volunteers: a multi-method comparison. Journal of International Medical Research. 2009;37(6):1680-91.

8. Longjun D, Xibing L, Ming X, Qiyue L. Comparisons of Random Forest and Support Vector Machine for Predicting Blasting Vibration Characteristic Parameters. Procedia Engineering. 2011;26:1772-81.

9. Tolle KM, Chen H, Chow H-H. Estimating drug/plasma concentration levels by applying neural networks to pharmacokinetic data sets. Decision Support Systems. 2000;30(2):139-51.

10. Vapnik V, Vapnik V. Statistical learning theory Wiley. New York. 1998;1:624.

11. Dreiseitl S, Ohno-Machado L, Kittler H, Vinterbo S, Billhardt H, Binder M. A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions. Journal of Biomedical Informatics. 2001;34(1):28-36.

12. GÜler NF, Koçer S. Use of Support Vector Machines and Neural Network in Diagnosis of Neuromuscular Disorders. Journal of Medical Systems. 2005;29(3):271-84.

13. Wei L, Yang Y, Nishikawa RM, Jiang Y. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE transactions on medical imaging. 2005;24(3):371-80.

14. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.

15. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.

16. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and computing. 2004;14(3):199-222.

17. Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V, editors. Feature selection for SVMs. Advances in neural information processing systems; 2001.

18. Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V, editors. Support vector regression machines. Advances in neural information processing systems; 1997.

19. Yap CW, Li ZR, Chen YZ. Quantitative structure–pharmacokinetic relationships for drug clearance by using statistical learning methods. Journal of Molecular Graphics and Modelling. 2006;24(5):383-95.

20. DUREJA H, GUPTA S, MADAN AK. Topological Models for Prediction of Pharmacokinetic Parameters of Cephalosporins using Random Forest, Decision Tree and Moving Average Analysis. Scientia Pharmaceutica. 2008;76(3):377-94.

21. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR Modeling: Where Have You Been? Where Are You Going To? Journal of Medicinal Chemistry. 2014;57(12):4977-5010.

22. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. Journal of Chemical Information and Computer Sciences. 2003;43(6):1947-58.

23. Lombardo F, Obach RS, DiCapua FM, Bakken GA, Lu J, Potter DM, et al. A Hybrid Mixture Discriminant Analysis− Random Forest Computational Model for the Prediction of Volume of Distribution of Drugs

in Human. Journal of Medicinal Chemistry. 2006;49(7):2262-7.

24. Ziegler A, König IR. Mining data with random forests: current options for real-world applications. WIREs Data Mining and Knowledge Discovery. 2014;4(1):55-63.

25. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

26. Fidler M. nlmixr: an R package for population PKPD modeling. 2019.

27. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, et al. Package 'e1071'. The R Journal. 2019.

28. Liaw A, Wiener M. The randomforest package. R news. 2002;2(3):18-22.

29. Krämer N, Schäfer J, Boulesteix A-L. Regularized estimation of large-scale gene association networks using graphical Gaussian models. BMC Bioinformatics. 2009;10(1):384.

30. Savin I. A comparative study of the lasso-type and heuristic model selection methods. Jahrbücher für Nationalökonomie und Statistik. 2013;233(4):526-49.

31. Lu F, Petkova E. A comparative study of variable selection methods in the context of developing psychiatric screening instruments. Statistics in medicine. 2014;33(3):401-21.

32. Zhu X-W, Xin Y-J, Ge H-L. Recursive random forests enable better predictive performance and model interpretation than variable selection by LASSO. Journal of chemical information and modeling. 2015;55(4):736-46.

33. Xie Z-X, Hu Q-H, Yu D-R, editors. Improved feature selection algorithm based on SVM and correlation. International symposium on neural networks; 2006: Springer.

34. Song X, Halgamuge SK, Chen D, Hu S, Jiang B. The optimized support vector machine with correlative features for classification of natural spearmint essence. International Journal of Innovative Computing, Information and Control. 2010;6(3):1089-99.

35. Hassan SS, Farhan M, Mangayil R, Huttunen H, Aho T. Bioprocess data mining using regularized regression and random forests. BMC systems biology. 2013;7(1):1-7.

36. Bonate PL. The effect of collinearity on parameter estimates in nonlinear mixed effect models. Pharmaceutical research. 1999;16(5):709-17.

37. Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. PLoS Genet. 2008;4(7):e1000130.

38. Schelldorfer J, Meier L, Bühlmann P. Glmmlasso: an algorithm for high-dimensional generalized linear mixed models using ℓ1-penalization. Journal of Computational and Graphical Statistics. 2014;23(2):460-77.

39. Kang SH, Poynton MR, Kim KM, Lee H, Kim DH, Lee SH, et al. Population pharmacokinetic and pharmacodynamic models of remifentanil in healthy volunteers

using artificial neural network analysis. British Journal of Clinical Pharmacology. 2007;64(1):3-13.

40. Combes F, Retout S, Frey N, Mentré F. Powers of the likelihood ratio test and the correlation test using empirical Bayes estimates for various shrinkages in population pharmacokinetics. CPT: pharmacometrics & systems pharmacology. 2014;3(4):1-9.

41. Tessier A, Bertrand J, Chenel M, Comets E. Combined Analysis of Phase I and Phase II Data to Enhance the Power of Pharmacogenetic Tests. CPT: pharmacometrics & systems pharmacology. 2016;5(3):123-31.