

Performance Evaluation of Support Vector Regression Models for Survival Analysis: A Simulation Study

Hossein Mahjub

Professor of Biostatistics
Department of Biostatistics and Epidemiology
Hamadan, Iran

Javad Faradmal

Associate Professor of Biostatistics
Department of Biostatistics and Epidemiology
Hamadan, Iran

Shahrbanoo Goli

Ph.D. Student of Biostatistics
Department of Biostatistics and Epidemiology
Hamadan, Iran

Ali-Reza Soltanian

Associate Professor of Biostatistics
Department of Biostatistics and Epidemiology
Hamadan, Iran

Abstract—Desirable features of support vector regression (SVR) models have led to researchers extending them to survival problems. In current paper we evaluate and compare performance of different SVR models and the Cox model using simulated and real data sets with different characteristics. Several SVR models are applied: 1) SVR with only regression constraints (standard SVR); 2) SVR with regression and ranking constraints; 3) SVR with positivity constraints; and 4) L1-SVR. Also, a SVR model based on mean residual life is proposed.

Our findings from evaluation of real data sets indicate that for data sets with high censoring rate and high number of features, SVR model significantly outperforms the Cox model. Simulated data sets also show similar results. For some real data sets L1-SVR has a significantly degraded performance in comparison to the standard SVR. Performance of other SVR models is not substantially different from the standard SVR with the real data sets. Nevertheless, the results of simulated data sets show that standard SVR slightly outperforms SVR with regression and ranking constraints.

Keywords—support vector machines; support vector regression; survival analysis; simulation study; Cox model; mean residual life

I. INTRODUCTION

Survival analysis is applied in different fields, such as medicine, public health, biology, epidemiology, engineering, economics, and demography. In survival studies within the medical field, patients are followed over the length of a predefined period. Those patients which experience an event of interest (failure) during the follow-up period are considered as complete (uncensored) observations. An event of interest is defined as some individual occurrence or experience such as death, disease incidence, relapse from remission, etc. The patient, whose exact time of event is not known but is known to occur in a certain period of time, is considered as censored. Right censoring is the most common form of censoring which is focused on in this study. If a patient is right censored at a

given time, he has not experienced the event by that time and the event of interest will occur after wards. For example if the death is considered as the event of interest then the patients who survive the entire follow-up period are considered as right censored [1].

The traditional models such as the Cox proportional hazard model and the accelerated failure-time model are applied in statistical literature for survival prediction [2, 3]. The most common survival model is the Cox proportional hazards regression. This model requires the proportional hazard assumption that is not always realistic. Also Cox is not able to model the nonlinear relations. Some other models such as artificial neural networks (ANN) and support vector machines (SVM) are applied for overcoming these problems [4-6]. SVM models are based on the statistical learning theory and have some beneficial features. They are able to model nonlinear relationships between variables using the kernels. Also they result in globally optimal solutions by solving a convex optimization problem, while contemporary models such as artificial neural networks deal with problems of local minima [7].

SVM first was proposed for solving classification problems [8]. Later, these models were extended to be applicable in regression problems [9]. Support Vector Regression (SVR) has been extensively applied in the literatures for response prediction [10-13]. However, there are few studies which use SVR for survival analysis. This is in part due to the response variable (survival time) in survival analysis including censored observations that the traditional SVR is not able to model. However, the desirable features of SVR have led researchers to extend them to be applicable in survival problems.

Yijun et al. [14] considered the survival time as a categorical variable and used the support vector classification model for survival analysis. The SVR model for survival

analysis is proposed by Shivaswamy et al. [7]. The authors investigated performance of some competing SVR models for real data sets with different censoring percentages. Ding [15] also has discussed possible application of SVM in survival analysis. They applied SVR model with different kernels on some real data sets. Khan and Zubek [16] compared SVR with Cox for five real data sets. Van Belle et al. [17, 18] proposed a SVR approach making use of ranking and regression constraints for right censored data. The authors compared performances of SVR and Cox models for both clinical and micro-array datasets. Also, they discussed a modified SVR model for other types of censorships in [6]. In another study, a new survival modeling technique based on least-squares SVR was proposed. The proposed model was compared with classical techniques on a breast cancer data set [19]. Du and Dua [20] applied SVR with two feature selection methods namely individual feature selection and feature subset forward selection and discussed their effect on performance of Cox and SVR on breast cancer data sets.

A data set has characteristics such as: censoring percentage, number of features and training sample size. To the best of our knowledge, there is no previous survival study which uses simulated data with different characteristics for evaluating and comparing performance of Cox and SVR models. The aim of this study is evaluating and comparing performance of various SVR models and Cox for survival analysis using simulated and real clinical data sets. To this end, different SVR models are applied: 1) SVR with only regression constraints; 2) SVR with regression and ranking constraints; 3) SVR with positivity constraints; 4) L1-SVR. Also, a new SVR model based on mean residual lifetime (MRL) is proposed.

The rest of this paper is organized as follows. SVR models for censored data are explained in section 2. Section 3 gives three performance measures which are used for comparing the models. In section 4 three real data sets are described. Section 5 explains the simulation method which is used to make clinical data sets with different characteristics. In section 6, the results of implemented analysis on artificial data sets as well as real data sets are presented. Finally, discussion is given in Section 7 and Section 8 is conclusion.

II. SVR MODELS

In this section, first, the standard SVR model is described for censored data. Second, a new SVR model is proposed and other SVR models and Cox regression are explained.

A. Standard SVR for survival analysis

SVM models are able to incorporate nonlinearity relations by different kernels. SVM do not use standard statistical approaches for estimation of model parameters. In these models the empirical risk of misranking two instances with regard to their event time, is minimized. We used some notations throughout the current text. x_i denotes a d-dimensional vector of independent variables, y_i is the corresponding survival and δ_i is censorship status. δ_i is 1, if an event has occurred, and δ_i is 0, if the observation is right censored. The prognostic index, i.e. the prediction of the model in SVR, is formulated as:

$$u = w^T \varphi(x) + b \quad (1)$$

In (1), w denotes the weight vector, $\varphi(x)$ denotes a transformation of the variables and b is a constant. To estimate the parameters, SVR is formulated as an optimization problem and a loss function is minimized subject to some constraints [6]. Shivaswamy et al. [7] proposed a modified algorithm for employing SVR to survival problems. This algorithm modifies the constraints of standard SVR. In this paper, standard survival SVR, is called SSVR and is formulated as:

SSVR:

$$\min_{w,b,\epsilon,\epsilon^*} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*) \quad (2)$$

subject to ($\forall i = 1, \dots, n$):

$$\begin{cases} w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ \epsilon_i \geq 0, \\ \epsilon_i^* \geq 0. \end{cases}$$

In (2) n is sample size and the parameter γ is a positive regularization constant. ϵ_i and ϵ_i^* are slack variables and allow the errors in the prediction. The large values of slack variables are penalized by the loss function. The prognostic index, the prediction of the model, for a new point x^* is computed as:

$$\hat{u}(x^*) = \sum_i (\alpha_i - \delta_i \alpha_i^*) \varphi(x_i)^T \varphi(x^*) + b, \quad (3)$$

Where α_i and α_i^* are the Lagrange multipliers. $\varphi(x_i)^T \varphi(x_j)$ is formulated as a positive definite kernel:

$$k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j), \quad (4)$$

Kernels often used for survival data are: linear, polynomial, RBF and clinical [21]. The linear kernel is formulated as:

$$k(x, z) = x^T z \quad (5)$$

Linear kernel was employed for all experiments in this paper.

B. A new SVR model using MRL function

The survival SVR model discussed in previous section uses a one-sided loss function for errors arising from prediction of censored observations. This loss function penalizes the model only when the censored observations are predicted smaller than their censoring time. A new SVR model is proposed which uses a two-sided loss function.

This model assumes that the event time for a censored observation, is equal to sum of its censoring time and the MRL. For individuals of age x , MRL measures their expected remaining lifetime [1] and is calculated using the following formula:

$$MRL(x) = \frac{\int_x^\infty S(t) dt}{S(x)} \quad (6)$$

where $S(x)$ is survival function. A standard estimator of the survival function is Kaplan–Meier estimator which is used in current study. Therefore, the model is also penalized when censored observations are predicted greater than sum of censoring time and the MRL. This model is called SSVR-MRL in this paper and is formulated as follows.

SSVR-MRL:

$$\min_{w,b,\epsilon,\epsilon^*,\xi} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*) + \mu \sum_{i=1}^n \xi_i, \quad (7)$$

subject to $(\forall i = 1, \dots, n)$:

$$\begin{cases} w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ -\delta_i(w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ (\delta_i - 1)(w^T \varphi(x_i) + b) \geq (\delta_i - 1)(y_i + MRL_i) \\ \epsilon_i \geq 0, \\ \epsilon_i^* \geq 0, \\ \xi_i \geq 0, \end{cases}$$

The prognostic index, for a new point x^* is found as:

$$\hat{u}(x^*) = \sum_{i=1}^n (\alpha_i - \delta_i \alpha_i^* + (\delta_i - 1) \beta_i) \varphi(x_i)^T \varphi(x^*) + b, \quad (8)$$

Where α_i , α_i^* and β_i are the Lagrange multipliers. For more information, please refer to [18].

C. Linear survival-SVR model with positivity constraints

This model is used for feature selection. In this model, a constraint is added the SSVR model to ensure positivity of weights [17].

Feature selection is included in this model by restricting the weights w to accept positive values. In this method, a preprocessing step on the dataset is required before training the model. Suppose x^p presents the p^{th} feature of input data. The concordance between each x^p and the event time is calculated, and each x^p with a concordance less than 0.5 is changed to $-x^p$. This model is called SSVRP in this paper. After modifying the data set, model is trained on data set. The estimation is obtained by solving the following optimization problem:

SSVRP:

$$\min_{w,b,\epsilon,\epsilon^*} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*), \quad (9)$$

subject to $(\forall i = 1, \dots, n, \forall p = 1, \dots, d)$:

$$\begin{cases} w^T x_i + b \geq y_i - \epsilon_i, \\ -\delta_i(w^T x_i + b) \geq -\delta_i y_i - \epsilon_i^*, \\ w_p \geq 0, \\ \epsilon_i \geq 0, \\ \epsilon_i^* \geq 0. \end{cases}$$

This constraint leads to the estimated weights be close to zero for irrelevant variables and be higher for relevant variables.

D. L1-SVR method

L1-SVR is another SVR model that is used for feature selection. In this model, the L1 penalty, $\sum_{i=1}^d |w_i|$ is used instead of the term $w^T w$. This model results in sparse solutions. Therefore, it selects fewer features than standard SVR [6]. In this paper, L1-SVR for survival analysis is called L1-SSVR.

E. A survival-SVR using ranking and regression constraints

Van Belle et al. [17, 18] proposed an SVR approach which makes use of ranking and regression constraints. This model is called SSVR2 in this paper. The standard SVR method (SSVR) includes only the regression constraints but SSVR2 includes the both regression and ranking constraints. In this method, the observations are arranged according to their event or censoring times. Then comparable pairs of observations are identified. A data pair is defined to be comparable whenever the order of their event times is known. For example, if patient A is censored in time a and patient B is uncensored with the related event occurring at time b ($a < b$), they are not comparable as it is not known that which has occurred earlier.

Since the event time for a censored observation is not known, a data pair is comparable if both observations are uncensored, or only one of them is uncensored with the censoring time of the other observation being later than event time of the uncensored observation.

The SVR method with ranking constraints involves a penalization for each comparable pair of observations for which the order in the prediction of model (prognostic index) differs from the observed order. The number of comparisons is reduced by comparing each observation i with the comparable neighbor with the largest survival time smaller than y_i , which will be indicated with $y_{j(i)}$. This model for censored data is formulated as:

$$\min_{w,b,\epsilon,\epsilon^*,\xi} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*) + \mu \sum_{i=1}^n \xi_i, \quad (10)$$

subject to $(\forall i = 1, \dots, n)$:

$$\begin{cases} w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ -\delta_i(w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ w^T (\varphi(x_i) - \varphi(x_{j(i)})) \geq y_i - y_{j(i)} - \xi_i \\ \epsilon_i \geq 0, \\ \epsilon_i^* \geq 0, \\ \xi_i \geq 0, \end{cases}$$

The parameters γ , μ in SVR models were tuned using the three-fold cross-validation criterion. Different SVR models were used for survival analysis using artificial and real data sets.

The prognostic index, for a new point x^* is computed as:

$$\hat{u}(x^*) = \sum_{i=1}^n \left(\alpha_i (\varphi(x_i) - \varphi(x_{j(i)}))^T + (\beta_i - \delta_i \beta_i^*) \varphi(x_i)^T \right) \varphi(x^*) + b, \quad (11)$$

Where α_i , β_i and β_i^* are the Lagrange multipliers [18].

F. Cox proportional hazard (PH) model

The Cox PH model is formulated as:

$$h(x, t) = h_0(t) \exp(w^T x) \quad (12)$$

$h(x, t)$ denotes the hazard rate, $h_0(t)$ is a baseline hazard rate, x is a specific feature vector and t is the time at which the

hazard is to be calculated. The hazard rate is the instantaneous risk to occur the event now, knowing that the event did not happen before. The Cox PH model is based on the proportional hazards assumption. This assumption presumes that the ratio of hazard rates for each two individuals in study is constant. The hazard of an observation with covariates x_i is associated to $w^T x_i$. The parameters w are estimated by maximizing the partial likelihood function [1, 3]. Cox model was used for comparison with SVR models.

III. PERFORMANCE MEASURES

Three performance measures were used for performance evaluation of survival models [17, 18, 21].

The researchers are usually interested in groups of patients with higher or lower risk profiles. Therefore, the concordance index (c-index) is used as a first performance measure to assess the concordance between the model results and the observed survival. For the second measure, patients are divided into two risk groups according to median prognostic index. The median prognostic index is used as threshold to identify the two groups. The second measure is the logrank test χ^2 statistic that measures the difference in survival between the two groups. To obtain a third measure, first the estimated prognostic indexes are normalized. Then a univariate Cox model is fitted to the normalized prognostic indexes. The estimated hazard ratio for this model is reported as the third measure. For all used performance measures, higher measures indicate better performance.

IV. REAL DATA SETS

The first data set concerns a historical cohort study which was performed on 197 heart attack patients, who visited the hospitals of Bushehr port, in South of Iran, from April 1997 to April 2001. Inclusion criteria were as follows: (i) the patient must be living in Bushehr, and (ii) the patient has not had a heart attack previously. In this experiment (HA), the event is death. The patient status and event time were attained by visiting the patients. The data set contains information on sex, age, cholesterol, LDL, HDL, systolic blood pressure and diastolic blood pressure.

We also used two publicly available data sets¹ for four other experiments. The first data set is from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984 [22]. This data set contains 312 PBC patients that have participated in the randomized trial. A total of 276 patients remain after removing missing values. In a first experiment (PD), the event is death while in a second experiment (PT) the event is transplantation. The variables in this data set are as follows: age, sex, stage, treatment, alkaline phosphatase, aspartate aminotransferase, albumin, bilirubin, cholesterol, triglycerides, urine copper, edema, platelet count, presence of ascites, presence of hepatomegaly or enlarged liver, standardized blood clotting time and blood vessel malformations in the skin.

The other data set concerns one of the first successful trials of adjuvant chemotherapy for 929 patients with colon cancer

[23]. 888 patients remain after removing missing values. There are two events in this data set. In a first experiment (CD), the event is death while in a second experiment (CR) the event is recurrence. This data set contains 10 variables which are as follow: treatment, sex, age, obstruction of colon by tumor, perforation of colon, adherence to nearby organs, number of lymph nodes with detectable cancer, differentiation of tumor, extent of local spread and time from surgery to registration.

V. SIMULATION METHOD

In simulated experiments of this study, continuous features were generated from a normal distribution with zero mean and unit variance. The correlations between the first ten features are zero except for the first and second variable, the third and fourth, the fifth and sixth, and the seventh and eighth, having correlation coefficients 0.7, 0.3, -0.7 and -0.3, respectively. The correlations between the second ten features, the third ten features up to end are similar to the first ten features. Similar methods of data set generation has been previously applied in the literature [18]. To generate weight vector of features, w , half of weights are set to zero and the rest are simulated from a normal distribution with zero mean and unit variance. There are some previous studies which have used similar methods of data set generation [6, 24, 25]. The event (failure) time follows the exponential distribution with parameter equal to $w^T x$, where x is the feature vector. So the survival time is associated to half of features through the prognostic index, $w^T x$. The censoring time also follows an exponential distribution with parameter equal to $cw^T x$ where the coefficient c is used to control the censoring percentage in the training and test sets. Similar methods of data set generation has been previously applied in the literature [6, 26].

We were interested to evaluate the effect of censoring percentage, number of model features and the sample size on the model performance. In the first setting, to evaluate the effect of censoring percentage, we generated datasets with different censoring rates: 0.1, 0.2, ..., 0.9. For each censoring percentage, we generated 50 datasets with 20 continuous features. These data sets included 200 training and 1000 test observations [18].

A clinical data set often includes both categorical and continuous features. So, to investigate the impact of categorical features on model performance, in the second setting, we generated data sets similar to previous setting except that 16 features were generated from a Bernoulli distribution with different nonzero means and 4 features were generated from a normal distribution with zero mean and unit variance.

To evaluate effectiveness of number of features, in the third setting, we generated data sets with different number of features: 10, 20, ..., 120. For each given number of features, 50 data sets with censoring rate equal to 0.5 including 200 training and 1000 test observations were generated.

In the fourth setting, we generated data sets with different training sample sizes: 50, 100, 200, 350, 500, 750, 1000 and 1500. The testing sample size was set equal to 4 times the training sample size. In this setting, 50 data sets with 20 continuous features and censoring rate equal to 0.5 were generated for each given sample size.

1. Data are available at
<https://vincetarelbundock.github.io/Rdatasets/datasets.html>

VI. RESULTS

Three real data sets were used for evaluating performance of SVR models. Each experiment was repeated 100 times with random partitioning of training and testing such that in each experiment 2/3 of the data set was used for training set and the rest was used for test. In all experiments, real and artificial data sets, the training set was used for learning the models and tuning the parameters. Models performance was evaluated based on the test data.

All SVR models were implemented in Matlab using the Mosek optimization toolbox for Matlab and Yalmip toolbox. Also we used 'R', Version 3.1.2 for implementing Cox model and calculating some performance measures.

TABLE I presents the censoring percentage, number of features and total sample size for five experiments with real data sets. For each experiment, performance measures of Cox and SVR models are displayed in TABLE II.

Statistically significant differences between SSVR and other models are indicated based on the Wilcoxon rank sum test. In experiments PT and PD, SVCR significantly outperformed Cox. These experiments had the highest censoring percentage and number of features among the five real experiments. In the HA data set, censoring times for all censored observations were similar and were equal to the follow-up period and we were not able to compute MRL for this data set. In the rest of the experiments, SSVR-MRL performed slightly better than other models but these differences were not significant. Real data sets did not indicate significant differences between SSVR and SSVR2 or SSVRP. In some experiments, performance of SSVR model was significantly better than L1-SSVR.

Figures indicate performance measures of Cox and SVR models using artificial data sets. In these figures, reported measures are median performance of 50 simulated experiments.

Fig. 1 shows performance measures of Cox and SSVR models for data sets with different censoring percentages. The left plots are related to data sets with only continuous features and the right plots concern the data sets which also have categorical features. All plots show that Cox outperforms SSVR for lower censoring percentages and when censoring percentage is high, SSVR outperforms Cox. Performance of SSVR and SSVR-MRL models are compared in fig. 2. This figure indicates that SSVR-MRL slightly outperforms SSVR for almost different censoring percentages but this difference is very small. In this figure similar to Fig.1, the results for two data sets, the data sets containing categorical features and data sets with only continuous features are similar except that plots for data sets with only continuous features exhibit smoother curves. Performance measures of two models, SSVR and SSVR2, for data sets with different censoring percentages are displayed in Fig. 3. All plots show SSVR outperforms SSVR2 for all censoring percentages.

TABLE I. SOME CHARACTERISTICS OF REAL DATA SETS

Data sets	Sample size	Number of features	Censoring percentage
-----------	-------------	--------------------	----------------------

HA	197	7	86.3
PD	276	17	59.78
PT	276	17	93.48
CD	888	10	51.58
CR	888	10	49.78

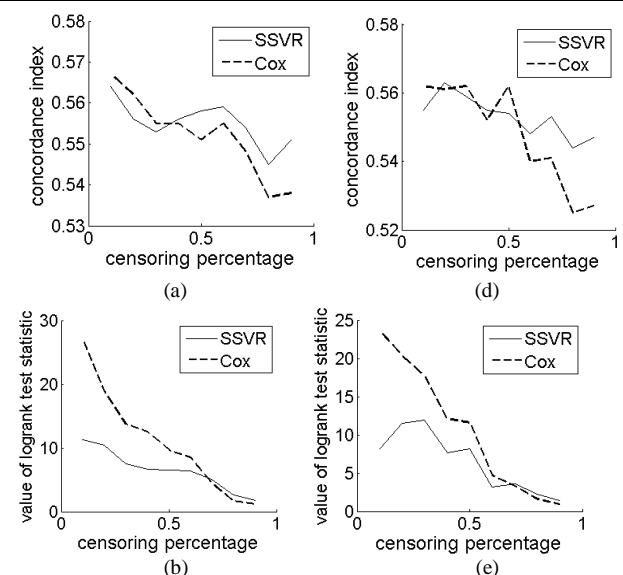
TABLE II. COMPARISON OF DIFFERENT SURVIVAL MODELS ON SIX REAL EXPERIMENTS. MEDIAN AND MEDIAN ABSOLUTE DEVIATION OF 100 RANDOM SPLITS INTO TRAIN-TEST SETS ARE GIVEN. STATISTICAL SIGNIFICANT DIFFERENCES BETWEEN MODEL SSVR AND THE OTHER MODELS ARE INDICATED BASED ON THE WILCOXON RANK SUM TEST

Data sets	method	c-index	logrank test ststistic	Hazard ratio
HA	Cox	0.62 ± 0.05	1.24 ± 1.11	1.27 ± 0.17**
	SSVR	0.62 ± 0.05	1.15 ± 1.04	1.51 ± 0.30
	SSVRP	0.61 ± 0.05	1.21 ± 1.07	1.46 ± 0.24
	L1-SSVR	0.61 ± 0.05	0.96 ± 0.80	1.46 ± 0.24
	SSVR2	0.60 ± 0.05*	1.08 ± 0.92	1.39 ± 0.23
PD	Cox	0.82 ± 0.02***	23.08 ± 5.22**	2.66 ± 0.30***
	SSVR	0.84 ± 0.01	26.25 ± 5.11	3.10 ± 0.56
	SSVR-MRL	0.84 ± 0.01	26.12 ± 5.78	3.13 ± 0.53
	SSVRP	0.84 ± 0.01	26.12 ± 4.22	3.10 ± 0.55
	L1-SSVR	0.82 ± 0.02**	25.40 ± 5.87	2.95 ± 0.51*
PT	Cox	0.70 ± 0.05**	2.13 ± 1.50**	1.41 ± 0.24***
	SSVR	0.74 ± 0.06	3.18 ± 2.24	2.11 ± 0.49
	SSVR-MRL	0.76 ± 0.06	3.91 ± 2.40	2.12 ± 0.49
	SSVRP	0.75 ± 0.06	3.38 ± 2.07	2.17 ± 0.53
	L1-SSVR	0.69 ± 0.06***	1.85 ± 1.65**	1.95 ± 0.51
CD	Cox	0.64 ± 0.01	20.81 ± 5.50	1.56 ± 0.06
	SSVR	0.64 ± 0.01	21.59 ± 5.53	1.49 ± 0.07
	SSVR-MRL	0.66 ± 0.01	22.63 ± 5.51	1.50 ± 0.09
	SSVRP	0.64 ± 0.01	21.93 ± 5.84	1.49 ± 0.07
	L1-SSVR	0.64 ± 0.01	21.34 ± 4.97	1.45 ± 0.06***
CR	Cox	0.64 ± 0.01	18.87 ± 4.96	1.57 ± 0.06
	SSVR	0.65 ± 0.01	19.71 ± 3.63	1.47 ± 0.08
	SSVR-MRL	0.65 ± 0.01	22.63 ± 5.51	1.50 ± 0.09
	SSVRP	0.65 ± 0.01	20.51 ± 4.38	1.47 ± 0.08
	L1-SSVR	0.64 ± 0.01*	19.24 ± 5.28	1.49 ± 0.08
SSVR2	0.65 ± 0.01	21.42 ± 4.49	1.45 ± 0.09	

* p-value < 0.05 (Wilcoxon rank sum test).

** p-value < 0.01 (Wilcoxon rank sum test).

*** p-value < 0.001 (Wilcoxon rank sum test).



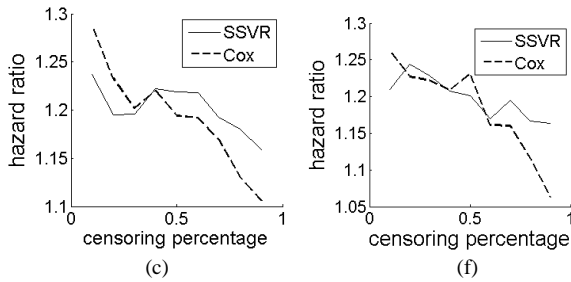


Fig. 1. (a), (b), (c): performance measures of Cox and SSVR models for artificial data sets with 20 continuous features and different censoring percentages, (d), (e), (f): performance measures of Cox and SSVR model for artificial data sets with categorical and continuous features and different censoring percentages. These measures are obtained using median measures of 50 artificial train-test sets

Fig. 4 shows performance measures of three models, SSVR, SSVRP and L1-SSVR for data sets with different censoring percentages. In this figure, SSVR is compared to SSVRP and L1-SSVR. The plots indicate that performance of SSVR, SSVRP and L1-SSVR are comparable. The left plots of Fig. 5 display performance measures of SSVR and Cox for data sets with different number of features. The plots indicate that performance of two models, Cox and SSVR, decrease as number of model features increases but the amount of reduction for SSVR is lower than Cox. Therefore, for data sets with higher number of features, SSVR outperforms Cox. Performance measures of SSVR and Cox for data sets with different sample sizes are shown in right plots of fig. 5. These plots indicate that performance of both SSVR and Cox improves as training sample size increases. Also two plots in this figure indicate that for lower training sample size, performance of SVR is a little better than Cox. When training sample size slightly increases, Cox outperforms SVR and for large training sample size, two models perform similarly.

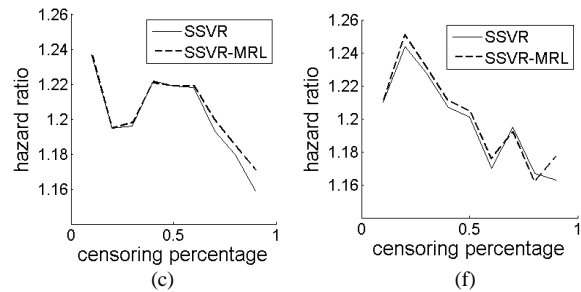
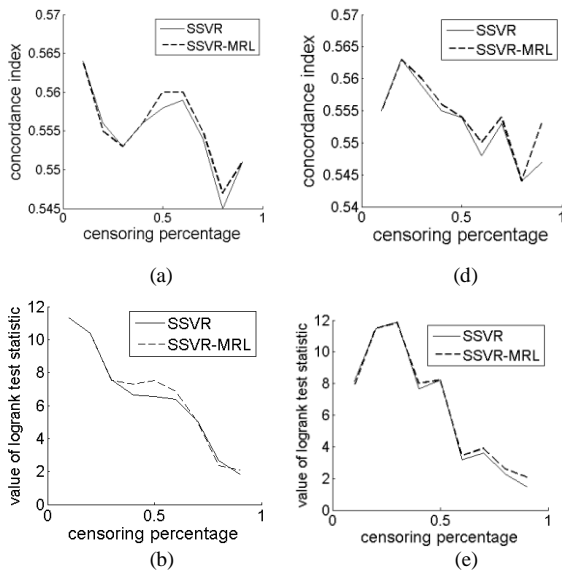


Fig. 2. (a), (b), (c): performance measures of SSVR and SSVR-MRL models for artificial data sets with 20 continuous variables and different censoring percentages, (d),(e), (f): performance measures of SSVR and SSVR-MRL models for artificial data sets with categorical and continuous features and different censoring percentages. These measures are obtained using median measures of 50 artificial train-test sets

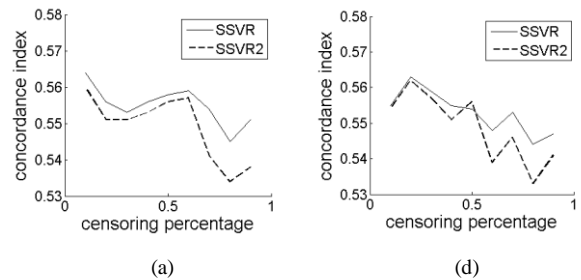
VII. DISCUSSION

The results of current study indicated that performance measures of SSVR and Cox decreased as censoring percentage increased, but the amount of reduction for SSVR was lower than Cox. For data sets with a high censoring percentage, SVR models outperformed the traditional Cox model.

Shivaswamy et al. [7] also evaluated the effect of censoring rate on performance of SVR model. They did not use simulated data and changed survival times in some real data sets to obtain data sets with different censoring percentages. They did not use Cox model and compared SVR with a survival model based on Gaussian process. They similar to current study found that performance of two mentioned models decreased as censoring rate increased and the amount of reduction for SVR is lower than the Gaussian process model.

Van Belle et al. [17, 18] compared performance of some survival models on six clinical data sets and three high dimensional data sets. They compared performance of some survival models with the SVCR2 model (the model with both regression and ranking constraints) and found that the differences of performance measures between SVCR2 and Cox were not significant.

In our study, SVCR model was compared with other survival models and the findings based on two real data sets indicated that SVCR significantly outperformed Cox model.



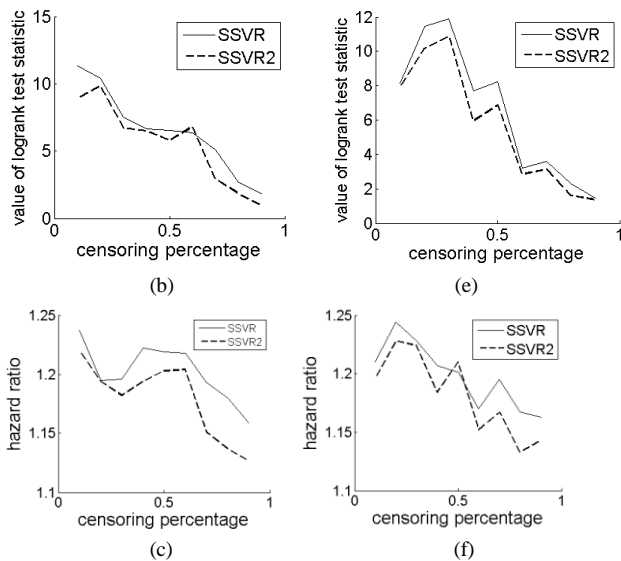


Fig. 3. figures (a), (b), (c): performance measures of SSVR and SSVR2 models for artificial data sets with 20 continuous variables and different censoring percentages, (d), (e), (f): performance measures of SSVR and SSVR2 models for artificial data sets with categorical and continuous features and different censoring percentages

Van Belle et al. [18] using clinical data sets found that differences between SVCR2 and SVCR were not significant. This result is Similar to our results for real data sets. They also yielded that for some high dimensional data sets that SVCR significantly outperformed SVCR2. Our simulated data sets also showed that SVCR slightly outperformed SVCR2. SSVR is a simpler model with one parameter while SSVR2 has two parameters and requires much time for tuning the parameters of model. Therefore, SSVR has better performance and less time complexity than SSVR2.

Some studies used a two-sided loss function for uncensored observations and a one-sided one for censored observations [6, 7, 17, 18]. In current study a tow-sided loss function was used for all observations. The results indicated that this loss function improves model performance slightly but the amount of improvement is not significant. Khan et al. [16] also applied a two-sided loss function for censored observations but their method is different from used method in this study. They entered some parameters in model to consider a two-sided loss functions for all observations and different losses for errors of underestimation and overestimation. The results of their study using five real clinical data sets indicated that SVR outperformed Cox. In contrast to current study, their model contains many parameters and requires much time for tuning the parameters of model.

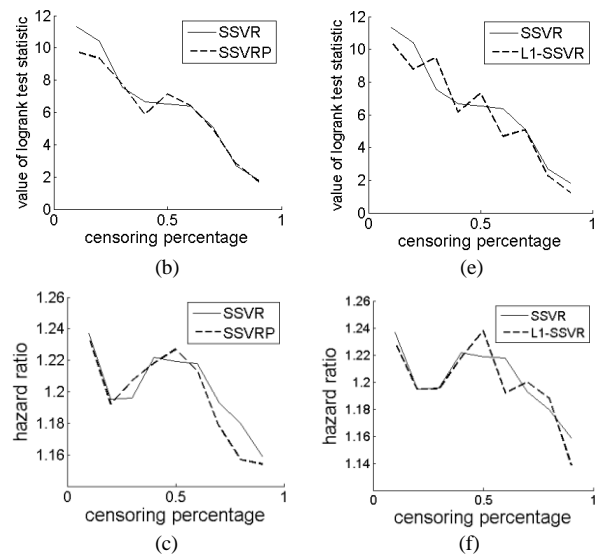
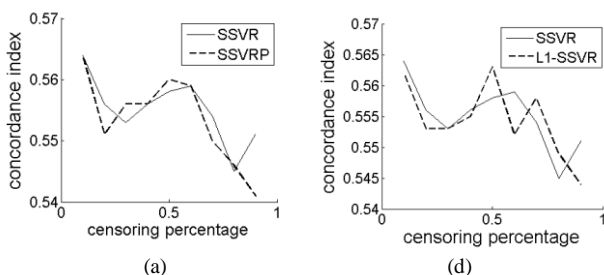


Fig. 4. (a), (b), (c): performance measures of SSVR and SSVRP models for artificial data sets with 20 continuous variables and censoring percentage equal to 0.5, (d), (e), (f): performance measures of SSVR and L1-SSVR2 models for artificial data sets with categorical and continuous features and censoring percentage equal to 50

Current study indicated that when the number of features was large, SVR outperformed Cox and when number of model features compared with the sample size was large enough, Cox was not able to be trained. Due and Dua [20], also using a breast cancer data set, yielded that feature selection improved performance of Cox and SVR and the amount of improvement for Cox was more than SVR. They reported that SVR outperformed Cox on the initial data set. After using feature selection, fewer features were included in the model and performances of Cox and SVR was similar. Some studies also reported good performance of SVR in dealing with high-dimensional data [7, 17].

There are limited papers which used artificial survival data sets for training survival SVR models. Shiao and Cherkassky[26] proposed two SVM methods to apply in survival analysis. The authors evaluated their methods using real data sets and artificial data sets. In contrast to current study, they used SVM models for classification of survival data. Van Belle et al. [6, 18] used limited artificial data sets for evaluating the performance of a SVR model with only ranking constraints. Liu et al. [25] also used limited artificial data sets to evaluate a novel survival L1-SVR method for large scale data sets. Goldberg and Kosorok [27] proposed a novel SVR method for censored data and used a simulation study for evaluating their method. The proposed survival SVR model in their study is completely different from methods studied in current paper and other studies in the literature. Apart from this study, to the best of our knowledge, there is no previous survival study which uses simulated data sets for comparing performance of Cox and SVR.

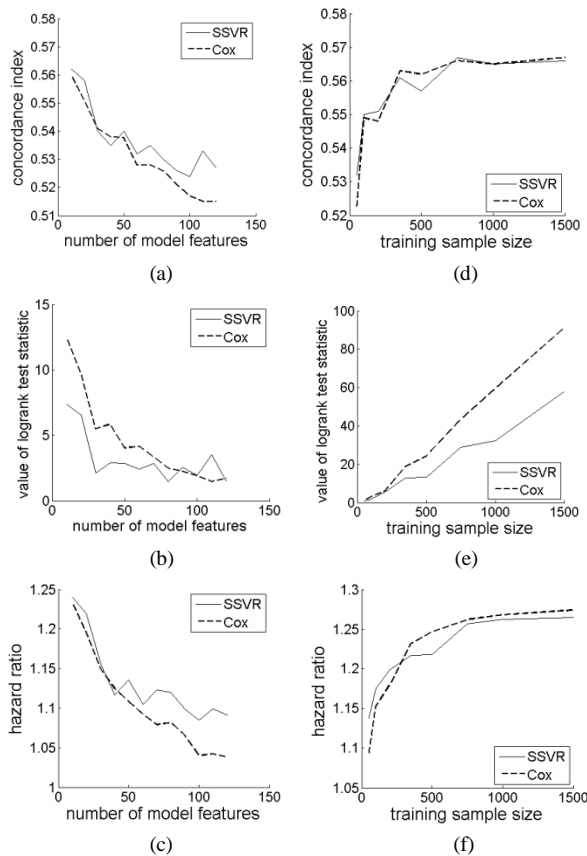


Fig. 5. (a), (b), (c): performance measures of SSVR and Cox models for artificial data sets with different number of model features, (d),(e) and (f): performance measures of SSVR and Cox models for artificial data sets with different training sample size. Censoring percentage of datasets are equal to 50

L1-SSVR and SSVRP (model with positivity constraints) are employed for feature selection. This study indicated that there were not significant differences between performances of SSVR and SSVRP. However for some real data sets it was found that SSVR significantly outperformed L1-SSVR. Cox model is a semiparametric model that needs to check the proportional hazard assumption. This model is only able to expose linear effects of features on hazard while some features may have a non-linear effect. For using of SVR models one does not have to check such assumptions. Also, the findings showed that for data with high censoring percentage or many features, SVR models have desirable performance. Due to good performance of SVR models for survival analysis, It is suggested that in future studies, these models are extended to other survival subjects such as competitive risks and analysis of recurrent events data.

VIII. CONCLUSION

The results of current study for two real data sets showed that if the censoring percentage of the clinical data sets is high and the model includes many features, SVR significantly outperforms traditional Cox model. Experiments with the artificial data sets in current study indicated that when censoring percentage of a clinical data set is high, SVR outperforms Cox. If the censoring percentage is low, Cox has a

better performance. However, SVR has the advantage of not requiring the proportional hazard assumption. Also, when the data set includes many features, SVR outperforms Cox. In addition, if the training set size is large enough, two models perform similarly.

The use of a two-sided loss function using MRL did not improved performance of SVR model. Real data sets did not indicate significant differences between SSVR (model with only regression constraint) and SSVR2 (model with regression and ranking constraints). However, SSVR had a better performance compared to SSVR2 using simulated data sets. For two real data sets, performance of L1-SSVR significantly was worse than the SSVR model.

ACKNOWLEDGMENT

The current study is a part of PhD thesis in Biostatistics. Therefore, the authors thank the Vic-chancellor of Research and Technology of Hamadan University of Medical Sciences, Iran, for approving the project and providing financial resources (grant number: 9308204015).

REFERENCES

- [1] J.P. Klein and M.L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated data*. 2003: Springer Science & Business Media.
- [2] J. Faradmal, A. Kazemnejad, R. Khodabakhshi, M.-R. Gohari, and E. Hajizadeh, Comparison of three adjuvant chemotherapy regimes using an extended log-logistic model in women with operable breast cancer. *Asian Pac J Cancer Prev*, 2010. 11(2): pp. 353-8.
- [3] J. Faradmal, A. Talebi, A. Rezaianzadeh and H. Mahjub, Survival analysis of breast cancer patients using cox and frailty models. *Journal of research in health sciences*, 2012. 12(2): pp. 127-130.
- [4] J. Faradmal, A.R. Soltanian, G. Roshanaei, R. Khodabakhshi, and A. Kasaiean, Comparison of the performance of log-logistic regression and artificial neural networks for predicting breast cancer relapse. *Asian Pacific journal of cancer prevention: APJCP*, 2014. 15(14): pp. 5883-5888.
- [5] A. Giordano, M. Giuliano, M. De Laurentiis, A. Eleuteri, F. Iorio, et al., *Artificial neural network analysis of circulating tumor cells in metastatic breast cancer patients*. *Breast cancer research and treatment*, 2011. 129(2): pp. 451-458.
- [6] V. Van Belle, K. Pelckmans, J.A. Suykens and S.V. Huffel, Learning transformation models for ranking and survival analysis. *The Journal of Machine Learning Research*, 2011. 12: pp. 819-862.
- [7] P.K. Shivaswamy, W. Chu and M. Jansche. A support vector approach to censored targets. in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*. 2007. IEEE.
- [8] B.E. Boser, I.M. Guyon and V.N. Vapnik. A training algorithm for optimal margin classifiers. in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992. ACM.
- [9] A.J. Smola and B. Schölkopf, A tutorial on support vector regression. *Statistics and computing*, 2004. 14(3): pp. 199-222.
- [10] J. Che, J. Wang and G. Wang, An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. *Energy*, 2012. 37(1): pp. 657-664.
- [11] Kazem, E. Sharifi, F.K. Hussain, M. Saberi, and O.K. Hussain, Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing*, 2013. 13(2): pp. 947-958.
- [12] J.G. Silva Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, et al., Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan. *Progress in photovoltaics: Research and applications*, 2012. 20(7): pp. 874-882.

- [12] J. Wang, L. Li, D. Niu and Z. Tan, An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy*, 2012. 94: pp. 65-70.
- [13] H. Yijun, Y. Shu, H. Minghuang, Y. Xiaowei, Q. Fang, et al., Application of support vector machine to predict 5-year survival status of patients with Nasopharyngeal Carcinoma after treatment. *The Chinese-German Journal of Clinical Oncology*, 2006. 5(1): pp. 8-12.
- [14] Z. Ding. The application of support vector machine in survival analysis. in *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, 2011 2nd International Conference on. 2011. IEEE.
- [15] F.M. Khan and V.B. Zubek. Support vector regression for censored data (SVRC): a novel tool for survival analysis. in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. 2008. IEEE.
- [16] V. Van Belle, K. Pelckmans, S. Van Huffel and J.A. Suykens, Improved performance on high-dimensional survival data by application of Survival-SVM. *Bioinformatics*, 2011. 27(1): pp. 87-94.
- [17] V. Van Belle, K. Pelckmans, S. Van Huffel and J.A. Suykens, Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artificial Intelligence in Medicine*, 2011. 53(2): pp. 107-118.
- [18] V. Van Belle, K. Pelckmans, J.A. Suykens and S. Van Huffel, Additive survival least-squares support vector machines. *Statistics in Medicine*, 2010. 29(2): pp. 296-308.
- [19] DU AND S. DUA, CANCER PROGNOSIS USING SUPPORT VECTOR REGRESSION IN IMAGING MODALITY. *WORLD JOURNAL OF CLINICAL ONCOLOGY*, 2011. 2(1): pp. 44.
- [20] V. Van Belle, K. Pelckmans, J.A. Suykens and S. Van Huffel. On the use of a clinical kernel in survival analysis. in *ESANN*. 2010.
- [21] T.M. Therneau and P.M. Grambsch, *Modeling survival data: extending the Cox model*. 2000: Springer Science & Business Media.
- [22] C.G. Moertel, T.R. Fleming, J.S. Macdonald, D.G. Haller, J.A. Laurie, et al., Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 1990. 322(6): pp. 352-358.
- [23] H.-T. Shiao and V. Cherkassky. SVM-based approaches for predictive modeling of survival data. in *The 2013 International Conference on Data Mining*. 2013.
- [24] Z. Liu, D. Chen, G. Tian, M.-L. Tang, M. Tan, et al., Efficient support vector machine method for survival prediction with SEER data, in *Advances in Computational Biology*Springer. 2010. pp. 11-18.
- [25] H.-T. Shiao and V. Cherkassky. SVM-based approaches for predictive modeling of survival data. in *Proceedings of the International Conference on Data Mining (DMIN)*. 2013. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- [26] Y. Goldberg and M.R. Kosorok. Support vector regression for right censored data. *arXiv preprint arXiv:1202.5130*, 2012.