

Research Article

Survival Prediction and Feature Selection in Patients with Breast Cancer Using Support Vector Regression

Shahrbanoo Goli,¹ Hossein Mahjub,² Javad Faradmal,³
Hoda Mashayekhi,⁴ and Ali-Reza Soltanian³

¹Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

²Research Center for Health Sciences, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

³Modeling of Non-Communicable Diseases Research Center, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

⁴Computer and IT Engineering Department, Shahrood University, Shahrood, Iran

Correspondence should be addressed to Hossein Mahjub; h_mahjub@yahoo.com

Received 19 March 2016; Revised 1 August 2016; Accepted 5 September 2016

Academic Editor: Francesco Pappalardo

Copyright © 2016 Shahrbanoo Goli et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Support Vector Regression (SVR) model has been broadly used for response prediction. However, few researchers have used SVR for survival analysis. In this study, a new SVR model is proposed and SVR with different kernels and the traditional Cox model are trained. The models are compared based on different performance measures. We also select the best subset of features using three feature selection methods: combination of SVR and statistical tests, univariate feature selection based on concordance index, and recursive feature elimination. The evaluations are performed using available medical datasets and also a Breast Cancer (BC) dataset consisting of 573 patients who visited the Oncology Clinic of Hamadan province in Iran. Results show that, for the BC dataset, survival time can be predicted more accurately by linear SVR than nonlinear SVR. Based on the three feature selection methods, metastasis status, progesterone receptor status, and human epidermal growth factor receptor 2 status are the best features associated to survival. Also, according to the obtained results, performance of linear and nonlinear kernels is comparable. The proposed SVR model performs similar to or slightly better than other models. Also, SVR performs similar to or better than Cox when all features are included in model.

1. Introduction

In the literature, different models have been proposed for survival prediction, such as Cox proportional hazard model and the accelerated failure-time model [1]. Although Cox proportional hazards model is the most popular survival model, its proportional hazards assumption may not always be realistic. Also, this model is applicable when the factors affect the outcome linearly. To overcome these issues, other models such as artificial neural networks and Support Vector Machines (SVM) are applied in the literature [2, 3].

The Support Vector Regression (SVR) model has been applied extensively for response prediction [4–6]. However,

there are few studies that use SVM for survival analysis. The traditional SVM models require complete response values, whereas survival data usually include the censored response values. Different attempts have been made to extend the standard SVM model making it applicable in survival analysis. Shivaswamy et al. [7] proposed a modified SVR algorithm for survival analysis. Ding [8] discussed possible application of SVR for censored data. Van Belle et al. [3, 9] proposed a new SVR model using ranking and regression constraints. Another proposal regarding application of SVR in survival analysis was made by Khan and Bayer-Zubek [10]. The authors compared SVR with Cox for some clinical data

sets. Van Belle et al. [11] proposed a new survival modeling technique based on least-squares SVR. Their proposed model was compared with different survival models on a clinical BC data set. Mahjub et al. evaluated and compared performance of different SVR models using data sets with different characteristics [12].

SVM does not include automatic feature selection. Feature selection methods for SVM are categorized into two types: filter and wrapper [13]. Wrapper methods are based on SVM, while filter methods are general and not specifically based on SVM.

When dealing with uncensored data, various studies use SVM for feature selection. Many of these studies use feature selection methods for SVM in classification problems [14–18]. For regression problems majority of available approaches propose filter methods [19, 20]. Few studies can be found which propose a wrapper feature selection method for regression [19, 21, 22]. Among these studies is the Recursive Feature Elimination (RFE) method proposed by Guyon et al. [23] for gene selection in cancer diagnosis. Camps-Valls et al. [24] used SVR for prediction and analysis of antisense oligonucleotide efficacy. They applied feature selection using three methods: correlation analysis, mutual information, and SVM-RFE. Becker et al. [18] developed some feature selection methods and used SVM-RFE as a reference model for comparison.

In survival analysis, different feature selection methods are applied in the literature [1, 25–30]. However, using SVR for identifying factors associated with response is rare and to the best of our knowledge, there is no previous study in which SVR is used for evaluating subsets of features. Liu et al. [31] applied SVR with L1 penalty for survival analysis on large datasets. They identified survival-associated prognosis factors for patients with prostate cancer. In contrast to current study, they directly used SVR weights to select the reduced features. Du and Dua [32] discussed the strength of SVR in BC prognosis. They used two feature selection methods neither of which are based on SVR.

The purpose of this study is two-fold. First, we predict the survival time for patients with BC using different SVR models including all available features. We propose a new SVR model using a two-sided loss function for censored data. Performance of Cox model is compared to SVR models by means of three performance measures.

Second, for feature selection, we assess the relationship between every feature and the predicted survival time using relevant statistical tests. We further employ univariate feature selection and the SVM-RFE method to select the best subset of features. Univariate feature selection based on concordance index is proposed in current study. To the best of our knowledge, in the field of survival analysis, there are no previous studies that employ SVM-based feature selection methods. In this paper, all employed feature selection methods are based on SVR.

2. Materials and Methods

In this section, first the BC dataset and two available datasets are introduced. Then, different SVR models, Cox regression

for censored data, and three feature selection methods are described.

2.1. Dataset Description. From April 2004 to April 2011, a historical cohort study was performed on 573 patients with BC, who visited the Oncology Clinic in Institute of Diagnostic and Therapeutic Center orphanage of Hamadan province, in west of Iran. Inclusion criteria were as follows: (i) the patient must be woman, (ii) treated with chemotherapy and radiotherapy, (iii) has undergone one of the related surgery types: breast conservative surgery (lumpectomy), simple mastectomy, radical mastectomy, and partial removal of breast tissue (a quarter or less of the breast) (segmental mastectomy). 31 patients are not eligible and are eliminated from the study. 11 features are selected as independent variables. Table 1 displays the patients' characteristics. Survival time is defined as the time period from diagnosis to patient death. The patients who survived during the entire study period are considered as right censored. During our study, 201 (37.1%) cancer patients died and 341 (62.9%) patients survived.

Two publicly available datasets (data are available at <https://vincentarelbundock.github.io/Rdatasets/datasets.html>) are also used for further analysis. The first data set was derived from one of the trials of adjuvant chemotherapy for colon cancer [33]. The dataset includes data of 929 patients. In this experiment (CD), the event under study is recurrence. The variables are as follows: type of treatment, sex, age, time from surgery to registration, extent of local spread, differentiation of tumor, number of lymph nodes, adherence to nearby organs, perforation of colon, and obstruction of colon by tumor. The second data set was obtained from Mayo Clinic trial in primary biliary cirrhosis of the liver [34]. The data set includes 312 randomized participants. There are two events for each person. In a first experiment (PT), the event is transplantation and in a second one (PD) the event is death. There are 17 variables which are as follows: stage, treatment, alkaline phosphatase, edema, age, sex, albumin, bilirubin, cholesterol, aspartate aminotransferase, triglycerides, urine copper, presence of hepatomegaly or enlarged liver, platelet count, standardized blood clotting time, presence of ascites, and blood vessel malformations in the skin.

Some artificial experiments were also conducted to evaluate the effect of number of features included in the model. These experiments are generated with different number of features: 10, 20, ..., 120. 50 datasets with 200 training and 500 test observations are generated for each feature value. Features values are simulated from a standard normal distribution. Similar methods of dataset generation have been previously applied in the literature [9]. The event time is exponentially distributed with a mean equal to $w^T x$, where x is the feature vector and w is the weight vector which is zero for half of features and is generated from a standard normal distribution for the rest of the features [31, 35, 36]. The censoring time is also exponentially distributed with a mean equal to $cw^T x$, where c determines the censoring rate. Other studies applied similar methods of dataset simulation [35, 36].

TABLE 1: Comparison of characteristics of dead and alive patients in the BC dataset, and the relation between the features and the estimated prognostic based on standard linear SVR.

Independent variables	Category	Characteristics of patients in dataset				Estimated prognostic index in SVR model			p value
		Alive		Dead		Mean prognostic index	SD prognostic index	Pearson correlation (with prognostic index)	
Age of diagnosis		Count	Mean (SD)	Count	Mean (SD)				-0.22
		341	45.33 (10.26)	201	47.30 (11.64)				
Tumor size (cm)	≤2	Count	Percent	Count	Percent	-0.22	12.59	0.313	
	2-5	230	70.12	98	29.88				
	>5	103	56.28	80	43.72				
Involved lymph nodes (n)	<4	Count	Percent	Count	Percent	-0.22	11.87	<0.001***	
	4	8	25.80	23	74.2				
	≥4	198	69.72	86	30.28				
Marital status	Unknown	Count	Percent	Count	Percent	-0.22	11.32	<0.001***	
	Married	103	62.05	63	37.95				
	Single	40	43.48	52	56.52				
Family history	Yes	Count	Percent	Count	Percent	-0.22	12.21	0.354	
	No	326	62.69	194	37.31				
	Unknown	15	68.18	7	31.82				
Type of surgery	Lumpectomy	Count	Percent	Count	Percent	-0.22	12.93	0.418	
	Radical mastectomy	300	62.11	183	37.89				
	Segmental mastectomy	41	69.49	18	30.51				
Histological type	simple mastectomy	Count	Percent	Count	Percent	-0.22	12.13	0.031*	
	Ductal	18	64.24	10	35.76				
	Lobular	250	61.27	158	38.73				
Metastases status	Present	Count	Percent	Count	Percent	-0.22	15.79	0.022*	
	Absent	26	63.41	15	36.59				
	Unknown	47	72.30	18	27.70				
HER2	Positive	Count	Percent	Count	Percent	-0.22	10.69	<0.001***	
	Negative	301	62.57	180	37.43				
	Unknown	21	61.76	13	38.24				
ER status	Positive	Count	Percent	Count	Percent	-0.22	10.81	<0.001***	
	Negative	19	70.37	8	29.63				
	Unknown	51	36.42	89	63.58				
PR status	Positive	Count	Percent	Count	Percent	-0.22	11.62	<0.001***	
	Negative	290	72.13	112	27.87				
	Unknown	110	65.47	58	34.53				
PR status	Positive	Count	Percent	Count	Percent	-0.22	10.09	<0.001***	
	Negative	84	60.00	56	40.00				
	Unknown	147	62.82	87	37.18				
PR status	Positive	Count	Percent	Count	Percent	-0.22	12.35	0.148	
	Negative	106	63.85	60	36.15				
	Unknown	126	68.10	59	31.90				
PR status	Positive	Count	Percent	Count	Percent	-0.22	12.06	<0.001***	
	Negative	109	57.06	82	42.94				
	Unknown	85	59.44	58	40.56				
PR status	Positive	Count	Percent	Count	Percent	-0.22	11.57	<0.001***	
	Negative	142	71.71	56	28.29				
	Unknown	114	56.71	87	43.29				

* p value < 0.05, ** p value < 0.01, *** p value < 0.001.

2.2. *Standard SVR for Censored Data (Model 1)*. Throughout the current text, x_i denotes a d -dimensional independent variable vector, y_i is survival time, and δ_i is the censorship status. If an event occurs δ_i is 1, and if the observation is right censored δ_i is 0. In standard SVR, the prognostic index, that is the model prediction, is formulated as

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

where $w^T \varphi(x)$ denotes a linear combination of a transformation of the variables $\varphi(x)$ and b is a constant. To estimate the prediction, SVR is formulated as an optimization problem and loss function is minimized subject to some constraints [9]. Shivaswamy et al. [7] extended the SVR model for censored data by modifying the constraints of standard SVR. In this paper, the standard SVR model for censored data is named model 1 and formulated as follows.

Description of the Formulas Used in SVR Models

Model 1

$$\begin{aligned} \min_{w,b,\epsilon,\epsilon^*} \quad & \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*), \\ \text{subject to} \quad & \forall i = 1, \dots, n, \\ & w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ & -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ & \epsilon_i \geq 0, \\ & \epsilon_i^* \geq 0. \end{aligned} \quad (2)$$

Model 2

$$\begin{aligned} \min_{w,b,\epsilon,\epsilon^*,\xi} \quad & \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*) + \mu \sum_{i=1}^n \xi_i, \\ \text{subject to} \quad & \forall i = 1, \dots, n, \\ & w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ & -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ & w^T (\varphi(x_i) - \varphi(x_{j(i)})) \geq y_i - y_{j(i)} - \xi_i \\ & \epsilon_i \geq 0, \\ & \epsilon_i^* \geq 0, \\ & \xi_i \geq 0. \end{aligned} \quad (3)$$

SVR-MRL

$$\begin{aligned} \min_{w,b,\epsilon,\epsilon^*,\xi} \quad & \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\epsilon_i + \epsilon_i^*) + \mu \sum_{i=1}^n \xi_i, \\ \text{subject to} \quad & \forall i = 1, \dots, n, \\ & w^T \varphi(x_i) + b \geq y_i - \epsilon_i, \\ & -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \epsilon_i^*, \\ & (\delta_i - 1) (w^T \varphi(x_i) + b) \\ & \geq (\delta_i - 1) (y_i + \text{MRL}_i) - \xi_i, \\ & \epsilon_i \geq 0, \\ & \epsilon_i^* \geq 0, \\ & \xi_i \geq 0, \end{aligned} \quad (4)$$

where γ and μ are positive regularization constants, ϵ_i, ϵ_i^* , and ξ_i are slack variables, and n is sample size.

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

$$\hat{u}(x^*) = \sum_i (\alpha_i - \delta_i \alpha_i^*) \varphi(x_i)^T \varphi(x^*) + b, \quad (5)$$

where α_i and α_i^* indicate the Lagrange multipliers. $\varphi(x_i)^T \varphi(x_j)$ can be formulated as a positive definite kernel:

$$k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j). \quad (6)$$

Commonly used kernels are linear, polynomial, and RBF [9]. Recently, an additive kernel has been used for clinical data defined as $k(x, z) = \sum_{p=1}^d k_p(x^p, z^p)$ in which $k_p(\cdot, \cdot)$ is separately defined for continuous and categorical variables [9, 37].

2.3. *Survival-SVR Using Ranking and Regression Constraints (Model 2)*. The SVR model proposed by Van Belle et al. [3, 9] which includes both ranking and regression constraints is called model 2 in this paper and is detailed in “Description of the Formulas Used in SVR Models”. Recall that the previous model (model 1) includes only regression constraints. In model 2, comparable pairs of observations are identified. A data pair is defined to be comparable whenever the order of their event times is known. Data in a pair are comparable if both of them are uncensored, or only one is censored with the censoring time being later than the event time. The number of comparisons is reduced by comparing each observation i with its comparable neighbor that has the largest survival time smaller than y_i , which is indicated with $y_{j(i)}$ [9].

2.4. *A New SVR Model Using Mean Residual Lifetime*. The survival SVR models, discussed in previous subsections, uses a one-sided loss function for errors arising from prediction of censored observations. We propose a new SVR model 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 mean residual lifetime (MRL). For individuals of age y , MRL measures their expected remaining lifetime using the following formula [1]:

$$\text{MRL}(y) = \frac{\int_y^{\infty} S(t) dt}{S(y)}, \quad (7)$$

where $S(y)$ is the survival function. Kaplan–Meier estimator is a standard estimator of the survival function which is used in the current study. Suppose the i th individual is censored and y_i shows her/his censoring time. Therefore, it is only known that she/he is alive until y_i . Since MRL_i measures her/his expected remaining lifetime from y_i onwards, event time for this individual can be estimated using sum of censoring time and MRL. Thus, the model is also penalized for censored observations when they are predicted to be greater than this sum. In this paper, this model is called SSVR-MRL and is formulated in “Description of the Formulas Used in SVR Models”.

The parameters of SVR models described in this subsection and previous subsections are tuned using the three-fold cross-validation criterion. We compare different SVR models with the Cox model [1].

The different survival models are compared using three performance measures: concordance index (c-index) [3, 7, 9], log rank test χ^2 statistic [3, 9, 37], and hazard ratio [3, 9, 10].

Standard SVR (model 1) and SVR with ranking and regression constraints (model 2) based on different kernels are trained on all datasets. In each experiment 2/3 of the dataset is used as the training set and the rest is used as the test set. Each experiment is repeated 100 times with random division of training and testing subsets. The training set is used for estimating different models. The performance of models is evaluated based on the test data.

2.5. Feature Selection. The first feature selection method proposed in current study is not a feature subset selection method. Considering that the prognostic index in SVR regression is directly correlated to survival time the prognostic index is calculated for all patients and the relationship between this index and each feature is assessed using a relevant statistical test.

The second feature selection method used in this study is univariate feature selection, which employs a univariate standard linear SVR for evaluating the relationship between each feature and survival time. After implementing experiments for all features, they are arranged in increasing order of c-index. Then, we pick out the p top ranked variables and include them in a linear standard SVR model. In the BC dataset, to select the best value of p , both SVR and Cox models are fitted for every value of p (1 – 11), and a subset with highest performance is selected.

The SVM-RFE algorithm, proposed by Guyon et al. [23], is also used for feature selection. This method identifies a subset with size p among m features ($p < m$) which maximizes the performance measures of model. In this algorithm, the square of feature weight values in linear SVR was used as the ranking criterion [13, 23].

3. Results

All reported performance measures are obtained using the median performance on 100 randomizations between training and test sets for prediction and 50 randomizations for feature selection. All SVR models are implemented in Matlab using the Mosek optimization toolbox for Matlab and Yalmip toolbox. Also we use “R”, Version 3.1.2, for running the Cox model and calculating some performance measures.

3.1. Prediction of Survival. All measures for the BC dataset indicated that SVR with linear kernel outperformed SVR with nonlinear kernels. The linear SVR models outperformed other models. Using the Wilcoxon rank sum test, statistically significant differences between linear SVR (model 1) and other models are indicated in Table 2. SVR-MRL slightly outperformed linear SVR (model 1), but differences of performances between the two models were not significant.

As the linear standard SVR model significantly outperformed other models, the prognostic indices were obtained based on this model. By comparing the mean value of prognostic index of subgroups of variables, the following results were obtained (Table 1). The survival time decreased as the age increased. The patients who developed metastasis survived less than other patients. The mean of prognostic index in subgroup with smaller tumor size was higher than the mean of subgroup with larger tumor size. The relationship between other variables and survival time can be evaluated similarly.

The results of CD experiment showed that SVR-MRL and SVR with clinical kernel outperformed other models but the difference between their performances and linear SVR (model 1) was not significant. In the PD experiment, SVR-MRL and SVR with polynomial kernel slightly outperformed other models but their performance differences with Linear SVR were not significant. Linear SVR significantly outperformed Cox. In the PT experiment, performances of all models were almost similar except that Cox and SVR with clinical kernel performed significantly worse than linear SVR.

3.2. Feature Selection. In the BC dataset, based on the first feature selection method, the relationship between the prognostic index and each feature was evaluated and p values of the tests are presented in Table 1. Seven variables were significant (p value < 0.05). For univariate feature selection, Figure 1 indicates that the subset with five features has the best performance. In SVR-RFE method, eleven replications were implemented for finding all subset of features. According to the performance measures of subsets, the subset with three features was selected as the best. In this figure, it is clearly shown that when the model involves large number of features, SVR outperforms Cox.

In the CD experiment, RFE and univariate methods selected subsets with six and five features respectively, which had five common features (Figure 1). In PT experiment, univariate method selected the subset of five features and RFE method selected the subset of three features among which two features were similar (Figure 2). In the PD experiment, univariate method selected the subset with eight features.

TABLE 2: Performance measures of Cox and SVR models using different kernels and datasets when all features are included in the model. Statistical significant differences between SVR based model 1 (indicated in *>italic*) and the other models are indicated based on the Wilcoxon rank sum test.

Dataset	Model	Type of kernel	c-index	Log rank test statistic	Hazard ratio
BC	Cox model	—	0.61 ± 0.03	1.38 ± 1.10	1.22 ± 0.11**
	<i>SVR based model 1</i>	<i>Linear</i>	<i>0.61 ± 0.03</i>	<i>1.85 ± 1.48</i>	<i>1.33 ± 0.14</i>
	SVR-MRL based model 1	Linear	0.62 ± 0.03	1.95 ± 1.44	1.33 ± 0.15
	SVR based model 2	Linear	0.60 ± 0.03	1.76 ± 1.33	1.31 ± 0.14
	SVR based model 1	RBF	0.56 ± 0.03***	0.52 ± 0.46***	1.13 ± 0.07***
	SVR based model 2	RBF	0.54 ± 0.05***	0.64 ± 0.64***	1.14 ± 0.11***
	SVR based model 1	Polynomial	0.59 ± 0.04*	1.56 ± 1.56	1.23 ± 0.17**
	SVR based model 2	Polynomial	0.57 ± 0.08***	0.88 ± 0.88***	1.19 ± 0.23***
	SVR based model 1	Clinical	0.60 ± 0.02*	1.47 ± 1.11	1.26 ± 0.14
	SVR based model 2	Clinical	0.60 ± 0.04*	1.14 ± 0.97**	1.27 ± 0.14
CD	Cox model	—	0.64 ± 0.01	20.81 ± 5.33	1.53 ± 0.07
	<i>SVR based model 1</i>	<i>Linear</i>	<i>0.64 ± 0.01</i>	<i>21.59 ± 5.53</i>	<i>1.49 ± 0.07</i>
	SVR based model 2	Linear	0.64 ± 0.01	22.17 ± 6.41	1.51 ± 0.07
	SVR-MRL based model 1	Linear	0.66 ± 0.01	22.60 ± 5.51	1.50 ± 0.09
	SVR based model 1	RBF	0.61 ± 0.01***	12.13 ± 3.49***	1.46 ± 0.08**
	SVR based model 2	RBF	0.61 ± 0.01***	11.98 ± 3.98***	1.43 ± 0.07***
	SVR based model 1	Polynomial	0.63 ± 0.01*	18.94 ± 7.05*	1.56 ± 0.11
	SVR based model 2	Polynomial	0.63 ± 0.02**	14.72 ± 9.65**	1.54 ± 0.17
	SVR based model 1	Clinical	0.65 ± 0.01	22.63 ± 4.80	1.63 ± 0.07
	SVR based model 2	Clinical	0.65 ± 0.01	22.28 ± 5.12	1.65 ± 0.07
PT	Cox model	—	0.70 ± 0.05**	2.13 ± 1.50**	1.41 ± 0.24***
	<i>SVR based model 1</i>	<i>Linear</i>	<i>0.74 ± 0.06</i>	<i>3.18 ± 2.24</i>	<i>2.08 ± 0.49</i>
	SVR based model 2	Linear	0.74 ± 0.06	3.14 ± 1.92	2.14 ± 0.48
	SVR-MRL based model 1	Linear	0.76 ± 0.06	3.91 ± 2.40	2.12 ± 0.49
	SVR based model 1	RBF	0.68 ± 0.06***	0.81 ± 0.75***	1.45 ± 0.30***
	SVR based model 2	RBF	0.67 ± 0.05***	0.84 ± 0.74***	1.45 ± 0.28***
	SVR based model 1	Polynomial	0.74 ± 0.05	3.69 ± 2.03	2.15 ± 0.52
	SVR based model 2	Polynomial	0.76 ± 0.06	3.76 ± 2.32	2.35 ± 0.49
	SVR based model 1	Clinical	0.71 ± 0.07**	1.85 ± 1.67**	1.83 ± 0.46**
	SVR based model 2	Clinical	0.70 ± 0.07**	1.60 ± 1.42***	1.83 ± 0.48**
PD	Cox model	—	0.82 ± 0.02***	23.11 ± 5.24**	2.67 ± 0.31***
	<i>SVR based model 1</i>	<i>Linear</i>	<i>0.84 ± 0.01</i>	<i>26.31 ± 5.19</i>	<i>3.12 ± 0.55</i>
	SVR based model 2	Linear	0.83 ± 0.01	27.40 ± 5.59	3.07 ± 0.54
	SVR-MRL based model 1	Linear	0.84 ± 0.01	26.09 ± 5.82	3.14 ± 0.52
	SVR based model 1	RBF	0.84 ± 0.02	27.93 ± 4.46	3.02 ± 0.55
	SVR based model 2	RBF	0.84 ± 0.02	28.51 ± 4.61	3.01 ± 0.59
	SVR based model 1	Polynomial	0.84 ± 0.02	26.58 ± 4.52	3.02 ± 0.52
	SVR based model 2	Polynomial	0.84 ± 0.02	26.61 ± 4.85	3.12 ± 0.49
	SVR based model 1	Clinical	0.83 ± 0.01**	23.92 ± 4.80*	3.21 ± 0.56
	SVR based model 2	Clinical	0.83 ± 0.01**	25.11 ± 5.23*	3.14 ± 0.54

* p value < 0.05, ** p value < 0.01, *** p value < 0.001 (Wilcoxon rank sum test).

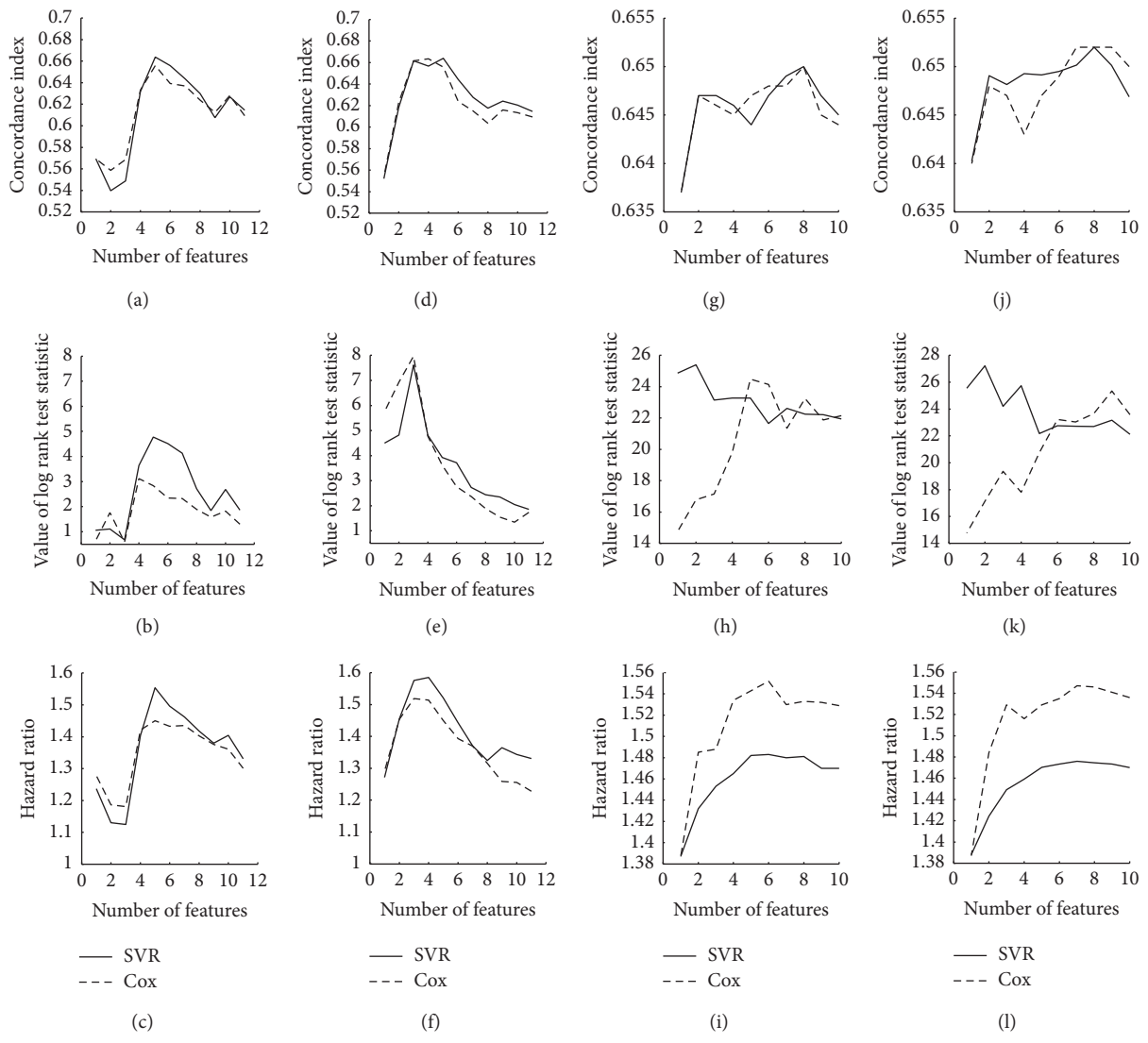
RFE selected a subset with five features all of which were observed in the subset selected by RFE. However, in these experiments, the best subset may be not unified. The first feature selection method found most features as significantly associated with the survival time. The selected features had noticeable overlap with two other methods.

It seemed that in subsets with large number of features, SVR outperformed Cox. The results of simulated datasets

also indicated similar results so that when the number of features was more than 60, SVR significantly outperformed Cox (Table 3).

4. Discussion

The results indicated that, in the BC dataset, linear SVR models outperformed SVR with nonlinear kernel. For other



BC dataset

CD dataset

Subsets using Univariate method:

- 1 = {A}
- 2 = {A, P}
- 3 = {A, P, L}
- 4 = {A, P, L, M}
- 5 = {A, P, L, M, H}
- 6 = {A, P, L, M, H, T}
- 7 = {A, P, L, M, H, T, F}
- 8 = {A, P, L, M, H, T, F, MS}
- 9 = {A, P, L, M, H, T, F, MS, TB}
- 10 = {A, P, L, M, H, T, F, MS, TB, E}
- 11 = {A, P, L, M, H, T, F, MS, TB, E, TS}

Subsets using RFE-SVR method:

- 1 = {M}
- 2 = {M, P}
- 3 = {M, P, H}
- 4 = {M, P, H, L}
- 5 = {M, P, H, L, A}
- 6 = {M, P, H, L, A, TS}
- 7 = {M, P, H, L, A, TS, TB}
- 8 = {M, P, H, L, A, TS, TB, MA}
- 9 = {M, P, H, L, A, TS, TB, MA, E}
- 10 = {M, P, H, L, A, TS, TB, MA, E, T}
- 11 = {M, P, H, L, A, TS, TB, MA, E, T, F}

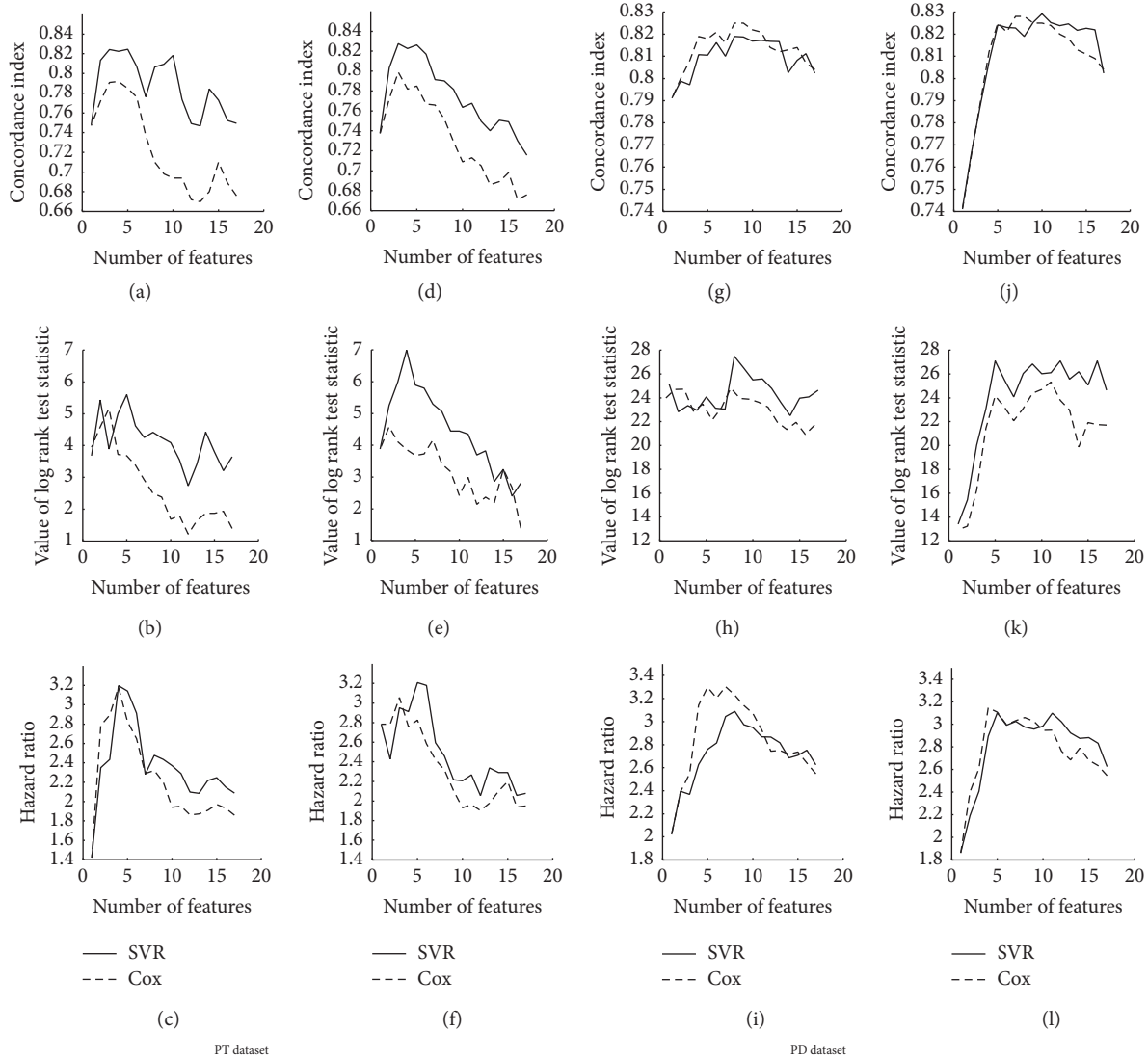
Subsets using Univariate method:

- 1 = {N}
- 2 = {N, E}
- 3 = {N, E, D}
- 4 = {N, E, D, T}
- 5 = {N, E, D, T, TS}
- 6 = {N, E, D, T, TS, O}
- 7 = {N, E, D, T, TS, O, AN}
- 8 = {N, E, D, T, TS, O, AN, A}
- 9 = {N, E, D, T, TS, O, AN, A, P}
- 10 = {N, E, D, T, TS, O, AN, A, P, S}

Subsets using RFE-SVR method:

- 1 = {N}
- 2 = {N, E}
- 3 = {N, E, T}
- 4 = {N, E, T, AN}
- 5 = {N, E, T, AN, D}
- 6 = {N, E, T, AN, D, TS}
- 7 = {N, E, T, AN, D, TS, O}
- 8 = {N, E, T, AN, D, TS, O, A}
- 9 = {N, E, T, AN, D, TS, O, A, S}
- 10 = {N, E, T, AN, D, TS, O, A, S, P}

FIGURE 1: Performance measures of standard linear SVR and Cox for different subsets of features. Horizontal axis shows the number of features included in the model. (a), (b), and (c): performance measures of univariate method for BC dataset. (d), (e), and (f): performance measures of RFE for BC dataset. (g), (h), and (i): performance measures of univariate method for CD dataset. (j), (k), and (l): performance measures of RFE for CD dataset. BC dataset’s abbreviations: M: metastasis, P: PR status, H: HER2, L: number of involved lymph nodes, A: age, T: tumor size, TB: histological type of BC, MS: marital status, E: ER status, TS: type of surgery, and F: family history. CD dataset’s abbreviations: S: sex, P: perforation of colon, A: age, AN: adherence to nearby organs, O: obstruction of colon by tumor, TS: time from surgery to registration, T: treatment, D: differentiation of tumor, E: extent of local spread, and N: number of lymph nodes.



Subsets in Univariate method:

- 1 = {B}
- 2 = {B, A}
- 3 = {B, A, U}
- 4 = {B, A, U, ST}
- 5 = {B, A, U, ST, AA}
- 6 = {B, A, U, ST, AA, H}
- 7 = {B, A, U, ST, AA, H, C}
- 8 = {B, A, U, ST, AA, H, C, CT}
- 9 = {B, A, U, ST, AA, H, C, CT, P}
- 10 = {B, A, U, ST, AA, H, C, CT, P, AL}
- 11 = {B, A, U, ST, AA, H, C, CT, P, AL, BV}
- 12 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG}
- 13 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG, AP}
- 14 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG, AP, PA}
- 15 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG, AP, PA, E}
- 16 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG, AP, PA, E, T}
- 17 = {B, A, U, ST, AA, H, C, CT, P, AL, BV, TG, AP, PA, E, T, S}

PT dataset

Subsets in RFE-SVR method:

- 1 = {A}
- 2 = {A, B}
- 3 = {A, B, ST}
- 4 = {A, B, ST, AA}
- 5 = {A, B, ST, AA, U}
- 6 = {A, B, ST, AA, U, BV}
- 7 = {A, B, ST, AA, U, BV, H}
- 8 = {A, B, ST, AA, U, BV, H, AL}
- 9 = {A, B, ST, AA, U, BV, H, AL, T}
- 10 = {A, B, ST, AA, U, BV, H, AL, T, S}
- 11 = {A, B, ST, AA, U, BV, H, AL, T, S, P}
- 12 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG}
- 13 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG, AP}
- 14 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG, AP, E}
- 15 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG, AP, E, PA}
- 16 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG, AP, E, PA, C}
- 17 = {A, B, ST, AA, U, BV, H, AL, T, S, P, TG, AP, E, PA, C, CT}

PD dataset

Subsets in Univariate method:

- 1 = {B}
- 2 = {B, U}
- 3 = {B, U, CT}
- 4 = {B, U, CT, ST}
- 5 = {B, U, CT, ST, AL}
- 6 = {B, U, CT, ST, AL, E}
- 7 = {B, U, CT, ST, AL, E, AA}
- 8 = {B, U, CT, ST, AL, E, AA, A}
- 9 = {B, U, CT, ST, AL, E, AA, A, BV}
- 10 = {B, U, CT, ST, AL, E, AA, A, BV, H}
- 11 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG}
- 12 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA}
- 13 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA, P}
- 14 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA, P, AP}
- 15 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA, P, AP, C}
- 16 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA, P, AP, C, S}
- 17 = {B, U, CT, ST, AL, E, AA, A, BV, H, TG, PA, P, AP, C, S, T}

Subsets in RFE-SVR method:

- 1 = {U}
- 2 = {U, AL}
- 3 = {U, AL, ST}
- 4 = {U, AL, ST, B}
- 5 = {U, AL, ST, B, A}
- 6 = {U, AL, ST, B, A, H}
- 7 = {U, AL, ST, B, A, H, E}
- 8 = {U, AL, ST, B, A, H, E, C}
- 9 = {U, AL, ST, B, A, H, E, C, BV}
- 10 = {U, AL, ST, B, A, H, E, C, BV, P}
- 11 = {U, AL, ST, B, A, H, E, C, BV, P, AA}
- 12 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA}
- 13 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA, TG}
- 14 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA, TG, CT}
- 15 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA, TG, CT, S}
- 16 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA, TG, CT, S, T}
- 17 = {U, AL, ST, B, A, H, E, C, BV, P, AA, PA, TG, CT, S, T, AP}

FIGURE 2: Performance measures of linear SVR and Cox for different subsets of features. (a), (b), and (c): performance measures of univariate method for PT dataset. (d), (e), and (f): performance measures of RFE for PT dataset. (g), (h), and (i): performance measures of univariate method for PD dataset. (j), (k), and (l): performance measures of RFE for PD dataset. B: bilirubin, A: age, U: urine copper, ST: stage, AA: aspartate aminotransferase, H: presence of hepatomegaly, C: cholesterol, CT: standardized blood clotting time, P: platelet count, AL: albumin, BV: blood vessel malformations in the skin, TG: triglycerides, AP: alkaline phosphatase, PA: presence of ascites, E: edema, T: treatment, and S: sex.

TABLE 3: Performance measures of SVR and Cox models for simulated experiments with different number of features.

Number of features	Model	c-index	Log rank test statistic	Hazard ratio
10	SVR	0.562 ± 0.015	12.281 ± 4.955	1.240 ± 0.063
	Cox	0.560 ± 0.013	10.662 ± 5.234	1.245 ± 0.060
20	SVR	0.552 ± 0.013	8.589 ± 5.557	1.203 ± 0.048
	Cox	0.552 ± 0.013	9.289 ± 5.990	1.194 ± 0.054
30	SVR	0.540 ± 0.010	5.070 ± 2.275	1.156 ± 0.042
	Cox	0.541 ± 0.011	5.510 ± 3.907	1.154 ± 0.041
40	SVR	0.535 ± 0.011	4.900 ± 2.907	1.117 ± 0.029
	Cox	0.538 ± 0.011	4.367 ± 2.573	1.125 ± 0.035
50	SVR	0.540 ± 0.014	5.234 ± 3.557	1.136 ± 0.054
	Cox	0.537 ± 0.013	3.874 ± 2.681	1.120 ± 0.047
60	SVR	0.532 ± 0.012	3.474 ± 2.748	1.105 ± 0.039
	Cox	0.528 ± 0.015	3.723 ± 2.662	1.112 ± 0.052
70	SVR	0.535 ± 0.012	4.134 ± 2.304	1.123 ± 0.036
	Cox	0.528 ± 0.016	2.774 ± 1.925*	1.096 ± 0.044*
80	SVR	0.530 ± 0.010	3.171 ± 2.597	1.120 ± 0.036
	Cox	0.526 ± 0.010*	1.687 ± 1.323	1.079 ± 0.041**
90	SVR	0.526 ± 0.013	2.846 ± 2.219	1.098 ± 0.030
	Cox	0.521 ± 0.012*	1.720 ± 1.492*	1.070 ± 0.041**
100	SVR	0.524 ± 0.012	1.594 ± 1.386	1.085 ± 0.045
	Cox	0.517 ± 0.014**	1.226 ± 1.033	1.054 ± 0.030*
110	SVR	0.535 ± 0.012	3.738 ± 3.335	1.101 ± 0.048
	Cox	0.515 ± 0.008***	0.870 ± 0.800***	1.057 ± 0.024***
120	SVR	0.527 ± 0.012	1.863 ± 1.581	1.091 ± 0.035
	Cox	0.515 ± 0.015**	1.092 ± 0.964*	1.058 ± 0.033**

* p value < 0.05, ** p value < 0.01, *** p value < 0.001 (Wilcoxon rank sum test).

datasets, linear SVR and SVR with nonlinear kernels were comparable. The SVR-MRL model performed similar to or slightly better than other models. This model is based on additivity assumption. The results of this study indicated that SVR model performed similar to or better than Cox when all features were included in model. When feature selection was used, the performance of SVR and Cox was comparable. Although Cox model is based on proportional hazards assumption, an extension of the model was proposed for time-dependent variables when this assumption is not satisfied [1, 38, 39]. However, the additivity of the proposed MRL-method is also a restriction.

Using BC, PT, and PD datasets, the results showed that when the number of included features in model was large, SVR outperformed the Cox model. The results of simulated datasets also indicated that, in models with large number of features, SVR significantly outperformed the Cox model. Du and Dua [32], also using a breast cancer dataset, indicated that SVR performs better than Cox on initial dataset and performs similar to Cox when feature selection is conducted. Van Belle et al. [3, 9] also indicated that the SVR model outperforms the Cox model for high-dimensional data, while for clinical data the models have similar performance.

The experiments indicated that the difference between the simple standard SVR model (model 1) and the complicated SVR model (model 2) was not significant. Using five clinical

data sets, Van Belle et al. [9] also found that the differences between two models are not significant.

The results of BC dataset indicated that feature selection methods selected subsets of features with different sizes. However, they were subsets of each other. The three features selected by all methods were as follows: metastasis status, progesterone receptor (PR) status, and human epidermal growth factor receptor 2 (HER2) status. The results indicated that the patients that developed metastasis had shorter life time compared to patients who did not develop metastasis. Some other studies also indicated similar results [25, 27, 40, 41]. Also the results indicated patients with negative HER2 status survived longer than patients with positive HER2 status, which is consistent with results of some previous studies [40, 41]. In addition, the present study yielded that patients with positive PR status had less survival time than the patients with negative PR status. Some other studies indicated contrary result [40, 41] which may be due to the missing values for this variable. However, Faradmal et al. [28] presented similar result for the PR status.

There are other techniques for feature selection. Some studies used statistical characteristics such as correlation coefficients or Fisher score for feature ranking [15, 24, 32]. These criteria are simple, neither being based on training algorithm. In the current study, square of feature weight is used as the ranking criterion which is based on SVR. This

criterion was previously applied in other studies [13, 23]. Chang and Lin [15] also ranked each feature by considering how the performance (accuracy) is affected without that. Their method is more time-consuming compared to the feature ranking method used in our study. Also, they found that feature ranking using weights of linear SVR outperforms the feature expelling method in terms of the final model performance. There are some techniques which ranked features based on their influence on the decision hyperplane and were able to also use nonlinear kernels for feature selection [14, 16].

In this study, subset selection methods were used to assess subsets of variables according to their performance in the SVR model. These methods are often criticized due to requiring massive amounts of computation. However, they are robust against overfitting [42]. Also, the better performance of SVM-RFE for feature selection compared with correlation coefficient and mutual information was shown in [23, 24].

Some feature selection methods incorporate variable selection as part of the training using a penalty function [18, 31]. Becker et al. [18] proposed a combination of SCAD and ridge penalties and found it being a robust tool for feature selection. Most of the mentioned techniques were applied for classification problems. It is suggested that, for future studies, these techniques are extended to survival regression models and different feature selection methods are used, evaluated, and compared.

To the best of our knowledge, there is no previous study employing feature selection based on SVR for survival analysis. The univariate method used in this study is a performance-based method which has the advantage of being applicable to all training algorithms. As opposed to this method, SVR-RFE takes into account possible correlation between the variables. For the BC dataset, SVR-RFE method yielded 11 subsets of features one of which was exactly the same as the subset obtained by the first method used in this study. RFE-SVR is an iterative procedure which requires much time for feature selection but the proposed feature selection method is trained once and the results can simultaneously be used for two purposes, prediction, and feature selection. SVR models have some limitations, as well. They require more time for tuning the parameters of the model and the required time for analysis increases with increasing the number of parameters of the model.

5. Conclusion

The results for BC data indicated that SVR with linear kernel outperformed SVR with nonlinear kernels. For other datasets, linear SVR and SVR with nonlinear kernels were comparable. Performance of the model improved slightly but not significantly by applying a two-sided loss function. However, the additivity of this model is a restriction.

SVR performed similar or better than Cox when all features were included in model. When feature selection was used, performance of two models was comparable. The result of simulated datasets indicated that when the number of features included in model was large, SVR model significantly outperformed Cox.

Univariate feature selection based on concordance index was proposed in current study which has the advantage of being applicable to all training algorithms. Another method (a combination of SVR and statistical tests) in contrast to univariate and RFE methods is not iterative and can be used for pilot feature selection. For all datasets, the employed methods selected subsets of features with different size, but with high overlaps. Based on the three feature selection methods, metastasis status, progesterone receptor status, and human epidermal growth factor receptor 2 status were the best features associated to survival.

Disclosure

The current study is a part of Ph.D. thesis in biostatistics field.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors wish to thank the Vic-Chancellor of Research and Technology of Hamadan University of Medical Sciences, Iran, for approving the project and providing financial supports for conducting this research (Grant no. 9308204015).

References

- [1] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, Springer Science & Business Media, New York, NY, USA, 2003.
- [2] A. Giordano, M. Giuliano, M. De Laurentiis et al., "Artificial neural network analysis of circulating tumor cells in metastatic breast cancer patients," *Breast Cancer Research and Treatment*, vol. 129, no. 2, pp. 451–458, 2011.
- [3] V. Van Belle, K. Pelckmans, S. van Huffel, and J. A. K. Suykens, "Improved performance on high-dimensional survival data by application of survival-SVM," *Bioinformatics*, vol. 27, no. 1, pp. 87–94, 2011.
- [4] C.-F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing Journal*, vol. 12, no. 2, pp. 807–818, 2012.
- [5] A. 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*, vol. 13, no. 2, pp. 947–958, 2013.
- [6] D. Tien Bui, B. Pradhan, O. Lofman, and I. Revhaug, "Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models," *Mathematical Problems in Engineering*, vol. 2012, Article ID 974638, 26 pages, 2012.
- [7] P. K. Shivaswamy, W. Chu, and M. Jansche, "A support vector approach to censored targets," in *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM '07)*, pp. 655–660, Omaha, Neb, USA, October 2007.
- [8] Z. X. Ding, "The application of support vector machine in survival analysis," in *Proceedings of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC '11)*, pp. 6816–6819, IEEE, Dengfeng, China, August 2011.

- [9] V. Van Belle, K. Pelckmans, S. Van Huffel, and J. A. K. Suykens, "Support vector methods for survival analysis: a comparison between ranking and regression approaches," *Artificial Intelligence in Medicine*, vol. 53, no. 2, pp. 107–118, 2011.
- [10] F. M. Khan and V. Bayer-Zubek, "Support vector regression for censored data (SVRC): a novel tool for survival analysis," in *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM '08)*, pp. 863–868, December 2008.
- [11] V. Van Belle, K. Pelckmans, J. A. Suykens, and S. Van Huffel, "Additive survival least-squares support vector machines," *Statistics in Medicine*, vol. 29, no. 2, pp. 296–308, 2010.
- [12] H. Mahjub, S. Goli, H. Mahjub, J. Faradmal, and A.-R. Soltanian, "Performance evaluation of support vector regression models for survival analysis: a simulation study," *International Journal of Advanced Computer Science & Applications*, vol. 1, no. 7, pp. 381–389, 2016.
- [13] I. Iguyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [14] O. L. Mangasarian and G. Kou, "Feature selection for nonlinear kernel support vector machines," in *Proceedings of the 7th IEEE International Conference on Data Mining Workshops (ICDMW '07)*, IEEE, Omaha, Neb, USA, October 2007.
- [15] Y.-W. Chang and C.-J. Lin, "Feature ranking using linear SVM," in *WCCI Causation and Prediction Challenge*, 2008.
- [16] C.-T. Su and C.-H. Yang, "Feature selection for the SVM: an application to hypertension diagnosis," *Expert Systems with Applications*, vol. 34, no. 1, pp. 754–763, 2008.
- [17] M. H. Nguyen and F. De la Torre, "Optimal feature selection for support vector machines," *Pattern Recognition*, vol. 43, no. 3, pp. 584–591, 2010.
- [18] N. Becker, G. Toedt, P. Lichter, and A. Benner, "Elastic SCAD as a novel penalization method for SVM classification tasks in high-dimensional data," *BMC Bioinformatics*, vol. 12, no. 1, article 138, 2011.
- [19] S. Hochreiter and K. Obermayer, "Nonlinear feature selection with the potential support vector machine," in *Feature Extraction*, pp. 419–438, Springer, 2006.
- [20] H.-X. Zhao and F. Magoulés, "Feature selection for support vector regression in the application of building energy prediction," in *Proceedings of the 9th IEEE International Symposium on Applied Machine Intelligence and Informatics (SAMII '11)*, pp. 219–223, Smolenice, Slovakia, January 2011.
- [21] G.-Z. Li, H.-H. Meng, M. Q. Yang, and J. Y. Yang, "Combining support vector regression with feature selection for multivariate calibration," *Neural Computing and Applications*, vol. 18, no. 7, pp. 813–820, 2009.
- [22] J.-B. Yang and C.-J. Ong, "Feature selection for support vector regression using probabilistic prediction," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 343–351, ACM, Washington, DC, USA, July 2010.
- [23] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [24] G. Camps-Valls, A. M. Chalk, A. J. Serrano-López, J. D. Martín-Guerrero, and E. L. L. Sonnhammer, "Profiled support vector machines for antisense oligonucleotide efficacy prediction," *BMC Bioinformatics*, vol. 5, no. 1, article 135, 2004.
- [25] 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*, vol. 12, no. 2, pp. 127–130, 2012.
- [26] H. Kim and M. Bredel, "Feature selection and survival modeling in The Cancer Genome Atlas," *International Journal of Nanomedicine*, vol. 8, supplement 1, pp. 57–62, 2013.
- [27] J. Faradmal, A. R. Soltanian, G. Roshanaei, R. Khodabakhshi, and A. Kasaeian, "Comparison of the performance of log-logistic regression and artificial neural networks for predicting breast cancer relapse," *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 14, pp. 5883–5888, 2014.
- [28] 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 Pacific Journal of Cancer Prevention*, vol. 11, no. 2, pp. 353–358, 2010.
- [29] V. Lagani and I. Tsamardinos, "Structure-based variable selection for survival data," *Bioinformatics*, vol. 26, no. 15, Article ID btq261, pp. 1887–1894, 2010.
- [30] I. Choi, B. J. Wells, C. Yu, and M. W. Kattan, "An empirical approach to model selection through validation for censored survival data," *Journal of Biomedical Informatics*, vol. 44, no. 4, pp. 595–606, 2011.
- [31] Z. Liu, D. Chen, G. Tian, M.-L. Tang, M. Tan, and L. Sheng, "Efficient support vector machine method for survival prediction with SEER data," in *Advances in Computational Biology*, vol. 680 of *Advances in Experimental Medicine and Biology*, pp. 11–18, Springer, Berlin, Germany, 2010.
- [32] X. Du and S. Dua, "Cancer prognosis using support vector regression in imaging modality," *World Journal of Clinical Oncology*, vol. 2, no. 1, pp. 44–49, 2011.
- [33] C. G. Moertel, T. R. Fleming, J. S. Macdonald et al., "Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma," *The New England Journal of Medicine*, vol. 322, no. 6, pp. 352–358, 1990.
- [34] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer Science & Business Media, 2000.
- [35] V. Van Belle, K. Pelckmans, J. A. K. Suykens, and S. Van Huffel, "Learning transformation models for ranking and survival analysis," *The Journal of Machine Learning Research*, vol. 12, pp. 819–862, 2011.
- [36] 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 '13)*, The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), July 2013.
- [37] V. Van Belle, K. Pelckmans, J. A. K. Suykens, and S. Van Huffel, "On the use of a clinical kernel in survival analysis," in *Proceedings of the 18th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN '10)*, pp. 451–456, Bruges, Belgium, April 2010.
- [38] C. A. Bellera, G. MacGrogan, M. Debled, C. T. De Lara, V. Brouste, and S. Mathoulin-Pélissier, "Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer," *BMC Medical Research Methodology*, vol. 10, article 20, 2010.
- [39] O. Hartmann, P. Schuetz, W. C. Albrich, S. D. Anker, B. Mueller, and T. Schmidt, "Time-dependent Cox regression: serial measurement of the cardiovascular biomarker proadrenomedullin improves survival prediction in patients with lower respiratory

- tract infection,” *International Journal of Cardiology*, vol. 161, no. 3, pp. 166–173, 2012.
- [40] E. Bilal, J. Dutkowski, J. Guinney et al., “Improving breast cancer survival analysis through competition-based multidimensional modeling,” *PLoS Computational Biology*, vol. 9, no. 5, Article ID e1003047, 2013.
- [41] A. Karimi, A. Delpisheh, K. Sayehmiri, H. Saboori, and E. Rahimi, “Predictive factors of survival time of breast cancer in kurdistan province of Iran between 2006–2014: a cox regression approach,” *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 19, pp. 8483–8488, 2014.
- [42] J. Reunanen, “Overfitting in making comparisons between variable selection methods,” *Journal of Machine Learning Research*, vol. 3, pp. 1371–1382, 2003.