

Article

Wavelet Support Vector Censored Regression

Mateus Maia ^{1,*}, Jonatha Sousa Pimentel ², Raydonal Ospina ^{2,3} and Anderson Ara ^{3,4,*}

¹ Hamilton Institute, Mathematics and Statistics, Maynooth University, County Kildare, W23 F2K8 Maynooth, Ireland

² Department of Statistics, Center of Natural and Exact Sciences, Federal University of Pernambuco, Recife 50.740-540, PE, Brazil

³ Statistics Department, Institute of Mathematics and Statistics, Federal University of Bahia, Salvador 40.170-110, BA, Brazil

⁴ Statistics Department, Exact Sciences Sector, Federal University of Paraná, Curitiba 81.531-980, PR, Brazil

* Correspondence: mateus.maiamarques.2021@mumail.ie (M.M.); ara@ufpr.br (A.A.)

Abstract: Learning methods in survival analysis have the ability to handle censored observations. The Cox model is a predictive prevalent statistical technique for survival analysis, but its use rests on the strong assumption of hazard proportionality, which can be challenging to verify, particularly when working with non-linearity and high-dimensional data. Therefore, it may be necessary to consider a more flexible and generalizable approach, such as support vector machines. This paper aims to propose a new method, namely wavelet support vector censored regression, and compare the Cox model with traditional support vector regression and traditional support vector regression for censored data models, survival models based on support vector machines. In addition, to evaluate the effectiveness of different kernel functions in the support vector censored regression approach to survival data, we conducted a series of simulations with varying number of observations and ratios of censored data. Based on the simulation results, we found that the wavelet support vector censored regression outperformed the other methods in terms of the C-index. The evaluation was performed on simulations, survival benchmarking datasets and in a biomedical real application.

Keywords: survival analysis; support vector censored regression; support vector regression; Cox



Citation: Maia, M.; Pimentel, J.S.; Ospina, R.; Ara, A. Wavelet Support Vector Censored Regression. *Analytics* **2023**, *2*, 410–425. <https://doi.org/10.3390/analytics2020023>

Academic Editors: Jose L. Salmeron and Domenico Ursino

Received: 16 February 2023

Revised: 7 April 2023

Accepted: 19 April 2023

Published: 4 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Survival analysis is a branch of statistics widely used by the scientific community of leading areas of knowledge to study the time until the occurrence of an event of interest. The time of failure, also referred to as the event time or survival time, denotes the time at which an event of interest occurs. In medical contexts, for instance, it may indicate the time until a patient's death, cure, or disease relapse. In engineering, this failure time can be the time until a piece of equipment or component becomes defective. In the criminal justice field, the time of failure may denote the time elapsed between a prisoner's release and the occurrence of a subsequent crime, among other possibilities.

The principal distinction between survival analysis techniques and classical statistical methods, such as regression analysis and the design of experiments, lies in the former's ability to incorporate information from censored data into statistical analysis. The incorporation of censored data information, which is partial observation of the response, is the main feature of survival data.

The non-parametric Kaplan–Meier estimator proposed by Kaplan [1] is undoubtedly the most widely used in the literature to estimate survival function. However, non-parametric techniques are limited in their ability to include explanatory variables in the analysis. To address this, appropriate regression models are used for censored data, such as parametric and semi-parametric models. The Cox regression model [2], a semi-parametric approach, offers the flexibility to incorporate time-dependent explanatory variables with

ease and is more flexible compared to parametric models, as it does not assume a specific statistical distribution in survival time. However, the Cox model assumes that hazards are proportional, with the supposition that the risk of failure is proportional over time.

Although the proportional hazard assumption can be easily verified in one dimension, verification of this assumption becomes challenging when working with higher dimensions. Moreover, adjusting a Cox model can be difficult when the number of explanatory variables exceeds the number of individuals since it uses partial probability to estimate the parameters [3]. Given the challenges of verifying the proportional hazard assumption in higher dimensions and the strong assumption inherent in the Cox model, it is necessary to employ a more flexible and high predictive performance approach, such as support vector machines (SVMs). SVMs are based on the theory of statistical learning developed by Vapnik [4]. They are known for their good and theoretical properties and high classification accuracy in classification problems when working with large dimensions [5].

Emerging from the use of the wavelet kernel, the wavelet kernel with the support vector machines [6] has been shown to be quite effective in the approximation process of non-linear functions in order to obtain, in some cases, a predictive capability superior to that of the radial basis function (RBF) kernel [7] in a wide range of applications, such as economics [8], wind speed prediction [9], and carbon price prediction [10]. Examples of its use are commonly seen in conjunction with neural networks for classification problems as a way to improve the predictive capacity of the model [11–14].

This paper aims to propose a new method, namely wavelet support vector censored regression (WSVCR), and compare the Cox model with traditional support vector regression (SVR) and traditional support vector regression for censored data (SVCR) models. Section 2 provides an overview of the survival analysis, the Cox proportional hazards model, and the C-index metric used for evaluation. Section 3 presents a brief overview of support vector models and the proposition of WSVCR. The simulation study is presented in Section 4. Benchmarking analysis and real biomedical data application are given in Section 5. Finally, Section 6 concludes the paper with the final considerations.

2. Survival Analysis

A characteristic of survival data is the presence of censorship. In the context of censored data, the time of occurrence of the event of interest may be to the right of the recorded time (“right-censoring”), or the recorded time may be greater than the failure time (“left-censoring”). The following notations are used throughout the text. The censored observation is defined as $(\mathbf{x}_i, u_i, l_i) \quad \forall i = 1, \dots, n$ where \mathbf{x}_i is a p -dimensional explanatory variable vector, l_i is the lower bound and u_i is the upper bound. The tuple representing observation with left- and right-censoring is given by $(\mathbf{x}_i, -\infty, u_i)$ and $(\mathbf{x}_i, l_i, +\infty)$, respectively, $\forall i = 1, \dots, n$. If the data are uncensored, the observation is given by $(\mathbf{x}_i, y_i) \quad \forall i = 1, \dots, n$ where y_i is the corresponding survival time. In some situations, the indicator δ_i is used to denote whether an event is observed ($\delta_i = 1$) or if the observation is censored ($\delta_i = 0$).

2.1. Cox Proportional Hazards Model

The Cox model is the most popular statistical approach to survival analysis. The basic assumption with using this model is that hazards are proportional. This assumption is reflected in the formula [15],

$$h_i(t) = \exp\left(\sum_{j=1}^p X_{ij}\beta_j\right)h_0(t),$$

where $h_i(t)$ is the hazard function for the i th individual, X_{ij} is the explanatory variable j for individual i , β_j is the slope term for the j th explanatory variable, and $h_0(t)$ refers to a hazard function for an individual with zeros for all features. The coefficients of the Cox regression model (β 's) are obtained via maximum likelihood estimation. Violation of the assumption of proportional hazards may lead to bias in estimating the model coefficients.

2.2. C-Index

The concordance index (C-index) is a performance evaluation metric widely used in survival analysis, which can be viewed as a generalization of area under the curve for continuous response variables. It assesses a model's performance by comparing the predicted survival time and the true survival time for pairs of observations and determining the proportion where they share the same ordering (i.e., they are concordant). Chen et al. [16], Shivaswamy et al. [17], Van Belle et al. [18] and Fouodo et al. [3] employed the C-index to evaluate the performance of the proposed survival models. In general, this measure estimates the model's performance in the classification of the survival times and evaluates the degree of agreement between the model's predictions or risk estimates and the observed survival outcomes.

The C-index can be defined as the probability of agreement between two randomly chosen observations, that is, the observation that has the shortest lifetime must also have the highest risk score and vice versa [19,20]. The C-index may be expressed as

$$\text{C-index}(v) = \frac{\sum_{i=1}^n \sum_{j \neq i} I[(v(\mathbf{x}_i) - v(\mathbf{x}_j)) \times ((y_i - y_j)) \geq 0]}{\sum_{i=1}^n \sum_{j \neq i} \text{comp}(i, j)},$$

where $v(\mathbf{x}_i)$ is an estimate function of \mathbf{x}_i , for example, the risk score estimated by the Cox model, I is the indicator function, and $\text{comp}(i, j)$ is a comparability indicator for a pair of observations $\{(\mathbf{x}_i, y_i, \delta_i), (\mathbf{x}_j, y_j, \delta_j)\}$ defined as

$$\text{comp}(i, j) = \begin{cases} 1 & \text{if } \delta_i = 1 \text{ and } \delta_j = 1, \\ & \delta_i = 1 \text{ and } \delta_j = 0 \text{ and } y_i \leq y_j, \\ 0 & \text{otherwise.} \end{cases}$$

A C-index with value 1 indicates perfect agreement. On the other hand, 0.5 means a random prediction. Survival data usually yield values between 0.6 and 0.7. The performance of some analysts in the stock market has a C-index between 0.3 and 0.4.

According to Brentnall and Cuzick [21], the C-index was used for the first time to estimate the degree to which an observation chosen at random from one distribution was superior to that chosen independently from another distribution so that it can also be understood as a generalization of the Wilcoxon–Mann–Whitney statistics [22,23]. In this case, the concordance index can be written as [24]

$$c(D, G, f) = \frac{1}{|\mathcal{E}|} \sum_{\varepsilon_{ij}} I_{f(\mathbf{x}_i) < f(\mathbf{x}_j)},$$

where in the context of the order graph, which represents the ordering of survival times for two subjects in a graphical form (refer to Steck et al. [24] for further details), the function $I_{a < b}$ is defined as 1 if $a < b$, and 0 otherwise. Here, $|\mathcal{E}|$ denotes the number of edges in the graph, and $f(\mathbf{x}_i)$ represents the model-predicted survival time for subject i .

Considering the Cox proportional-hazards model, the C-index may be written as

$$c = \frac{1}{\text{comp}(i, j)} \sum_{i: \delta_i=1} \sum_{j: y_i < y_j} I_{\mathbf{x}_i \hat{\beta} > \mathbf{x}_j \hat{\beta}},$$

where $\mathbf{X}_i \hat{\beta}$ measures the risk. Furthermore, considering the joint survival or failure time, the C-index is given by

$$c = \frac{1}{\text{comp}(i, j)} \sum_{i: \delta_i=1} \sum_{j: y_i < y_j} I_{\hat{Y}_i < \hat{Y}_j},$$

where \hat{Y}_i represents the predicted value, that is, a shorter failure time means a lower predicted value.

3. Support Vector Regression Approach

The support vector machine, proposed by [25], is used to create an optimal hyperplane that can separate a set of observations by maximizing the distance between the margins. Originally devised for the separation of two classes, the model can be used for linear and non-linear cases through the different kernel functions, and has become a widely used model, given its power of generalization and theoretical foundation.

3.1. Traditional Support Vector Regression

Support vector regression machines [26] were proposed as generalized versions of support vector machines for regression tasks. Instead of creating a hyperplane with a maximum margin in feature space, support vector regression builds an optimal hypertube which considers the observations inside the hypertube as constraints, as well as minimizing the distance among the observations and the center of the hypertube.

Let a dataset be $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, n$, where n is the number of observations, $y_i \in \mathcal{R}$ is the response variable, $\mathbf{x}_i \in \mathcal{R}^p$ are the explanatory variables, $\mathbf{w} \in \mathcal{R}^p$ is the parameter vector on the hypertube as well as the hyperparameters ϵ and $C \in \mathcal{R}^+$. Vapnik [26] showed that considering slack variables $\mathcal{E}, \mathcal{E}^*$ and the optimization of convex objective function is given by

$$\min_{\mathbf{w}, \mathcal{E}, \mathcal{E}^*} \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^n (\mathcal{E}_i + \mathcal{E}_i^*),$$

$$\text{subject to} \begin{cases} y_i - f(\mathbf{x}_i) \leq \mathcal{E}_i + \epsilon, \\ f(\mathbf{x}_i) - y_i \leq \mathcal{E}_i^* + \epsilon, \\ \mathcal{E}_i \geq 0, \mathcal{E}_i^* \geq 0, \end{cases}$$

where $f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i$. Hyperparameter C represents the regularization constant responsible for imposing weight to minimize errors. Additionally, the hypertube may be expressed in the dual form optimization given by the following Equation (1) and its respective constraints [26]:

$$\max_{\alpha, \alpha^*} \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_i \cdot \mathbf{x}_j + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \epsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) \right),$$

$$\text{subject to} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \sum_{j=1}^n (\alpha_j - \alpha_j^*) = 0 \\ 0 < \alpha_i, \alpha_i^* \leq C. \end{cases} \quad (1)$$

This approach of SVR works successfully in linear regression problems. However, there are cases where non-linearity exists among the explaining and predictor variables. In these cases, the kernel trick may be used, based on Mercer's theorem [27] in order to deal with non-linearity. Using kernel methods, rather than considering the input space, higher feature spaces are considered, where the observations could be linearly separable through the function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$, where ϕ is a map function and replaces the inner product in Equation (1). These functions are defined as semi-definite kernel functions [28].

Many types of kernel functions are used in distinct regression examples. The choice of specific kernel functions creates unique non-linear mappings, and the efficiency of the resulting SVR usually depends on the correct choice of the kernel [29]. There are several kernel functions in the general SVR framework, some of the most common of which are considered in this paper. Table 1 displays these kernels for hyperparameters $\gamma \in \mathbb{R}^+$ and $d \in \mathbb{N}^*$.

Table 1. Kernel functions.

Kernel	$K(\mathbf{x}, \mathbf{y})$	Parameters
Linear	$\gamma(\mathbf{x} \cdot \mathbf{y})$	γ
Polynomial	$(\gamma(\mathbf{x} \cdot \mathbf{y}))^d$	γ, d
Gaussian	$\exp\{-\gamma\ \mathbf{x} - \mathbf{y}\ ^2\}$	γ
Exponential	$\exp\{-\gamma\ \mathbf{x} - \mathbf{y}\ \}$	γ
Sigmoidal	$\tanh(\gamma(\mathbf{x} \cdot \mathbf{y}))$	γ
Cauchy	$\frac{1}{1 + \exp\{2\gamma\ \mathbf{x} - \mathbf{y}\ ^2\}}$	γ
Laplacian	$\exp\{-\sqrt{2\gamma}\ \mathbf{x} - \mathbf{y}\ \}$	γ

The trick is precisely to find the appropriate kernel to build the model for a given set of data so that there is no single appropriate kernel function for every type of database. The decision of the most appropriate kernel function may be determined by grid search [30], as well the choice of the best hyperparameters.

3.2. Support Vector Censored Regression

As mentioned in Section 2, the censored observation is defined as $(\mathbf{x}_i, u_i, l_i) \forall i = 1, \dots, n$ where l_i is the lower bound and u_i is the upper bound. In the censored data context, there exist two different censors: the data can be left-censored, with the tuple that represents that observation being given by $(\mathbf{x}_i, -\infty, u_i) \forall i = 1, \dots, n$, and the data can be right-censored, with $(\mathbf{x}_i, l_i, +\infty) \forall i = 1, \dots, n$. If the data are not censored, the observation is given by the tuple $(\mathbf{x}_i, y_i) \forall i = 1, \dots, n$.

In order to develop a model that can use all available information, even when the data are censored, Shivaswamy et al. [17] developed the support vector censored regression (SVCR) approach. The proposal is generalized for both censoring types [17]. However, in their work, only left-censored data were explored, as well as a few kernel functions, linear and polynomial.

In the SVCR, the loss function is defined by

$$L(f(\mathbf{x}_i), u_i) = \max(0, l_i - f(\mathbf{x}_i), f(\mathbf{x}_i) - u_i).$$

Let $I_u = \{i \mid l_i > -\infty, u_i < +\infty\}$, $I_r = \{i \mid u_i = +\infty\}$ and $I_l = \{i \mid l_i = -\infty\}$. Then, I_u is the index of uncensored samples, while I_r and I_l are the index of the samples which are right- and left-censored, respectively. Letting $U = I_u \cup I_l$ and $L = I_u \cup I_r$, the optimization of the convex objective function for SVCR is given by

$$\begin{aligned} & \min_{\mathbf{w}, \mathcal{E}, \mathcal{E}^*} \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \left(\sum_{i \in U} \mathcal{E}_i + \sum_{i \in L} \mathcal{E}_i^* \right), \\ & \text{subject to } \begin{cases} \hat{f}(\mathbf{x}_i) - u_i \leq \mathcal{E}_i^*, \quad \forall i \in U, \\ l_i - \hat{f}(\mathbf{x}_i) \leq \mathcal{E}_i, \quad \forall i \in L, \\ \mathcal{E}_i \geq 0 \quad \forall i \in U, \mathcal{E}_i^* \geq 0 \quad \forall i \in L. \end{cases} \end{aligned}$$

Expanding this formulation through the Lagrangian multipliers, the dual formulation [26] considering the censored data becomes

$$\begin{aligned} & \max_{\alpha, \alpha^*} \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_i \cdot \mathbf{x}_j - \sum_{i \in L} y_i \alpha_i + \sum_{i \in U} y_i \alpha_i^* \right), \\ & \text{subject to } \begin{cases} \sum_{i \in L} \alpha_i - \sum_{i \in U} \alpha_i^* = 0, \\ 0 < \alpha_i, \alpha_i^* \leq C. \end{cases} \end{aligned} \quad (2)$$

With the SVCR approach, the dot product also can be replaced for the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$ to obtain a non-linear function mapping. Some works already demonstrated the use of the support vector regression approach to censored data only use three types of kernel functions: linear, polynomial and Gaussian [17,18]. This paper presents a proposition of a new SVCR, as well as presenting a comparison with other types of kernel functions in order to evaluate how the SVCR algorithm would behave for these other cases.

3.3. Wavelet Kernel in Support Vector Models

Wavelets are a tool used to decompose functions, operators, or data into various frequency components, which enables the study of each component individually. In this context, wavelet support vector machine was introduced by Zhang et al. [6] based on the aggregation of the wavelet kernel to the support vector machine theory. The wavelet kernel is a kind of multidimensional wavelet function which approximates arbitrary non-linear functions [31]. In this sense, the main idea behind the wavelet analysis is to approximate a function or a signal by a family of functions generated by dilations or contractions (scaling) and translations of a function called the mother wavelet.

The wavelet transformation of a function $g(x)$ for $x \in \mathcal{R}$ may be written as

$$W_{A,B}(f) = g(x) \cdot h_{A,B}(x),$$

where

$$h_{A,B}(x) = |A|^{-1/2} \left(\frac{x-B}{A} \right),$$

where $h(x)$ is the mother wavelet function, A and $B \in \mathcal{R}$, A is a dilation factor, and B is a translation factor.

This approximation in finite t terms is given by

$$\hat{g}(x) = \sum_{i=1}^t W_i h_{A_i, B_i}(x),$$

with $\hat{g}(x)$ is the approximation of $g(x)$.

The Mercer theorem gives the conditions that a dot product kernel must satisfy, and these properties are extended to the support vector regression and support vector censored regression as cited in Section 3. The choice of the wavelet plays a crucial role in the analysis, detection, and localization of power quality disturbances. In this context, using properties of translation-invariant kernels and multidimensional wavelet functions, Zhang et al. [6] construct a translation-invariant wavelet kernel by the following wavelet mother function:

$$h(x) = \cos(1.75x) \exp(-0.5x),$$

which is the cosine–Gaussian wavelet, which is the real part of the Morlet wavelet [32].

Wavelet support vector censored regression (SVCR) is the importation of the wavelet kernel in Equation (2), given the optimization of the convex objective function in the dual formulation for the censored data expressed in Equation (3) on the same constraints:

$$\max_{\alpha, \alpha^*} \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \prod_{k=1}^p h \left(\frac{\mathbf{x}_i^k - \mathbf{x}_j^k}{A} \right) - \sum_{i \in L} l_i \alpha_i + \sum_{i \in U} u_i \alpha_i^* \right). \quad (3)$$

where \mathbf{x}_i^k the denotes the k th component of the explanatory variable of the i th training observation.

WSVCR and SVCR were implemented in the R software [33], and the dual Lagrangian problem was solved using the quadratic programming approach [34]. Additionally, in order to speed up the SVCR approach, the same algorithm can use the sequential minimal optimization (SMO) [35] to solve faster the dual optimization problem presented in Equations (1)–(3).

4. Simulation Study

In order to evaluate the performance of different kernel functions used in the support vector censored regression approach to survival data, we firstly explored simulated data, for which it is possible to create a controlled environment for the experiment. The data generation model is defined over three explanatory variables X_1 , X_2 and X_3 , which follow the distributions below:

- $X_1 \sim \mathcal{N}(1, 1)$;
- $X_2 \sim \mathcal{N}(2, 2)$;
- $X_3 \sim \text{Exp}(1)$.

To capture the time behavior from observations in survival data analysis, we use the Weibull distribution to generate the time of each observation since several works already demonstrated that it is appropriate to generate this kind of data [36–38]. The time generation is given by $Y \sim \text{Weibull}(\lambda, k)$, where the scale parameter and shape parameter are given by

$$\lambda = 5 \times \exp\left(-\left(X_1 \ln |X_2| + \sin(X_3^2)\right)\right), \quad (4)$$

$$k = 3,$$

respectively. Equation (4) was proposed in order to represent non-linear behavior among explanatory variables. In this case, the simulation of this artificial dataset is based on varying sample sizes and proportions of censored observations, along with a non-linear relationship between the explanatory variables and the true generation function. This approach is necessary, as most traditional survival approaches fail to meet these assumptions as the case Cox model, which assumes hazard proportionality. For this reason, SVR or SVCR models can handle with this non-linear relationship, and our focus is on studying these models using the wavelet kernel in comparison with the other common kernels.

The setting of a simulated dataset was generated, varying $n \in \{30; 100; 300\}$ and the ratio of censored observations $r \in \{0.05; 0.25; 0.5; 0.95\}$. To evaluate the results among the different approaches, the C-index was used. The measures were calculated over a repeated holdout validation with 100 repetitions and a split ratio of 70–30% of training–test, respectively [39,40]. Additionally, the tuning process on SVR and SVCR was realized through a grid search, varying the parameters $C \in \{0.25; 0.5; 1; 2; 4\}$, γ and $A \in \{0.125; 0.25; 0.5; 1; 2; 4; 8\}$. The results are summarized by mean in Table 2, where the bold values identify the best concordance index for each scenario. The results summarize about 673,200 fitted models over 1200 Monte Carlo samples. In the Appendix A, a more detailed summary is provided by median, minimum, maximum and standard deviation. The results corroborate with the outperforming of wavelet SVCR in the centrality and the dispersion of the best concordance.

Table 2 displays that the wavelet SVCR generally outperforms the other methods, considering the C-index. Additionally, the SVCR with other kernels functions still greater than the traditional Cox model, which is similar to SVR with the linear kernel.

Table 2. Summary with the mean of C-index for all simulated databases and methods.

<i>n</i>		30			100			300			Mean
<i>r</i>	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.95	
COX	0.635	0.609	0.633	0.664	0.681	0.697	0.648	0.659	0.682	0.606	0.652
SVCR											
wavelet	0.795	0.819	0.885	0.717	0.743	0.776	0.718	0.734	0.750	0.820	0.776
Laplacian	0.772	0.793	0.813	0.720	0.734	0.748	0.712	0.724	0.740	0.773	0.753
Cauchy	0.780	0.794	0.810	0.723	0.740	0.760	0.719	0.731	0.751	0.786	0.759
exponential	0.783	0.806	0.833	0.727	0.745	0.768	0.720	0.735	0.755	0.792	0.766
Gaussian	0.794	0.820	0.840	0.726	0.743	0.773	0.723	0.736	0.754	0.809	0.772
sigmoidal	0.705	0.709	0.793	0.658	0.688	0.720	0.628	0.640	0.593	0.518	0.665
polynomial	0.648	0.655	0.691	0.597	0.602	0.624	0.644	0.647	0.683	0.631	0.642
linear	0.646	0.618	0.655	0.656	0.681	0.711	0.644	0.657	0.680	0.658	0.661
SVR											
wavelet	0.795	0.826	0.870	0.716	0.738	0.779	0.718	0.732	0.743	0.810	0.773
Laplacian	0.777	0.795	0.819	0.713	0.730	0.739	0.707	0.720	0.722	0.732	0.745
Cauchy	0.781	0.795	0.806	0.720	0.732	0.747	0.715	0.727	0.732	0.724	0.748
exponential	0.782	0.803	0.830	0.720	0.737	0.755	0.716	0.729	0.737	0.763	0.757
Gaussian	0.795	0.821	0.838	0.723	0.736	0.770	0.719	0.732	0.740	0.782	0.766
sigmoidal	0.703	0.702	0.805	0.663	0.685	0.717	0.626	0.641	0.647	0.743	0.693
polynomial	0.643	0.687	0.707	0.584	0.589	0.590	0.646	0.654	0.685	0.680	0.646
linear	0.640	0.586	0.651	0.653	0.673	0.683	0.643	0.651	0.670	0.647	0.650

5. Real Data Applications

5.1. Benchmarking Data

In addition to the evaluation of the SVCR using simulated data, its performance with various kernel functions was also assessed on real data benchmarks that are featured in several survival analysis publications [17,41]. There are a total of seven datasets, each with a distinct number of observations, explanatory variables, and ratio of censored instances. Table 3 displays the general statistics of the datasets, where n is the sample size, p the number of input variables and r the censoring rate. This was done to increase diversity among the datasets on which the models were built. To determine the best model, the C-index was selected as the primary parameter, where a larger C-index indicates a better model. The parameters used in the grid search for the tuning process were the same in the simulation— $C = \{0.125; 0.25; 0.5; 1; 2; 4; 8\}$ and $\gamma = \{0.125; 0.25; 0.5; 1; 2; 4; 8\}$ —based on the C-index maximization. The validation technique was also the same: repeated holdout with 30 repetitions and a split ratio of 70–30%.

Table 3. Real benchmarking datasets.

Dataset	CGD	Lung	Ovarian	PBC	Rats2	Retinopathy	Veteran
<i>n</i>	203	167	26	276	253	394	137
<i>p</i>	8	7	4	17	2	6	6
<i>r</i>	0.63	0.28	0.54	0.53	0.16	0.61	0.07

The result is summarized in Table 4, which presents the mean C-index over the 30 holdout samples. The maximum C-index values for each dataset are shown in bold values.

Table 4. Summary of C-index mean values over the benchmarking datasets.

Method	CGD	Lung	Ovarian	PBC	Rats2	Retinopathy	Veteran	Mean
COX	0.669	0.616	0.715	0.788	0.627	0.589	0.676	0.669
SVCR								
wavelet	0.754	0.646	0.807	0.786	0.648	0.610	0.693	0.706
Laplacian	0.724	0.627	0.736	0.755	0.631	0.598	0.691	0.680
Cauchy	0.718	0.630	0.752	0.786	0.641	0.602	0.695	0.689
exponential	0.733	0.640	0.768	0.817	0.638	0.606	0.701	0.700
Gaussian	0.727	0.628	0.774	0.694	0.645	0.607	0.690	0.681
sigmoidal	0.466	0.612	0.849	0.656	0.632	0.494	0.656	0.623
polynomial	0.659	0.596	0.701	0.540	0.621	0.578	0.617	0.616
linear	0.651	0.614	0.724	0.747	0.626	0.485	0.679	0.646
SVR								
wavelet	0.684	0.644	-	0.798	0.637	0.562	0.693	0.670
Laplacian	0.651	0.631	-	0.800	0.614	0.527	0.695	0.653
Cauchy	0.648	0.629	-	0.800	0.628	0.535	0.697	0.656
exponential	0.662	0.640	-	0.806	0.625	0.534	0.707	0.662
Gaussian	0.649	0.631	-	0.786	0.635	0.547	0.696	0.657
sigmoidal	0.604	0.643	-	0.774	0.630	0.534	0.660	0.641
polynomial	0.612	0.577	-	0.757	0.608	0.534	0.612	0.617
linear	0.622	0.622	-	0.752	0.622	0.510	0.679	0.634

The results presented in Table 4 emphasize the great efficiency of the SVCR method. It is evident that this approach outperforms both the SVR and Cox models. It is important to notice that the wavelet kernel was the best-performing kernel function, with the largest value in 4 out of 7 in the benchmark dataset. Additionally, 70% of kernel functions that achieved the best result were among the most commonly used kernel functions in studies employing the SVCR method, i.e., linear, polynomial, and Gaussian. Therefore, the results indicate the importance of carefully selecting the correct kernel function since the exponential, Cauchy and Laplacian kernels produced the highest C-index in most cases.

The result is also summarized in Figure 1, which schematically plots the C-index for all datasets to confirm the superiority of the SVCR approach compared with other models. Figure 2, in turn, shows the rate of times each model was considered the winner over the 30 repetitions performed during the experiments. In this way, it is possible to observe, once again, the superiority of the wavelet kernel over the other kernels, either for the pure SVR approach, or mainly for the SVCR approach.

5.2. Biomedical Real Data

The dataset is based on research conducted by the Heart Institute at the “das Clínicas” Hospital, Faculty of Medicine, Universidade de São Paulo, Brazil (<https://www.hc.fm.usp.br>, accessed on 18 January 2023), with the purpose of comparing the length of stay for heart disease patients undergoing cardiac surgery [42]. This study considered 145 patients and 7 variables, of which 5 are the explanatory variables, as well as the response variable and the censoring indicator. The data were recently analyzed by [43], using the Weibull regression model and the random survival forest (RSF) model.

The response variable, T , represents the amount of time (in hours) from the patient’s admission to their discharge from the surgical ward and is accompanied by an indicator variable, δ . If $\delta = 0$, there is censoring, meaning the exact length of stay is unknown. If δ is non-zero, the length of stay is known exactly. Additionally, the following explanatory variables were taken into account: the age of the patient in years (X_1), the type of protocol (X_2), which can be either conventional (0) or fast track (1), race (X_3), which is divided in the Brazilian classification into white (1), black (2), and yellow (3), sex (X_4), which is divided

between female (0) and male (1), and the type of patient (X_5), separated into congenital (0) and coronary (1).

Table 5 presents the descriptions of some of these variables along with their corresponding descriptive statistics. It was observed that for patients with congenital heart disease who were monitored under the traditional protocol, their age ranged from 0.3 to 49 years, while for those under the fast-track protocol, their age ranged from 0.8 to 38 years. For patients with coronary heart disease who were monitored under the traditional protocol, their age ranged from 18 to 81 years, and for those under the fast-track protocol, their age ranged from 38 to 79 years. The censoring rate (r) for these data is 0.041.

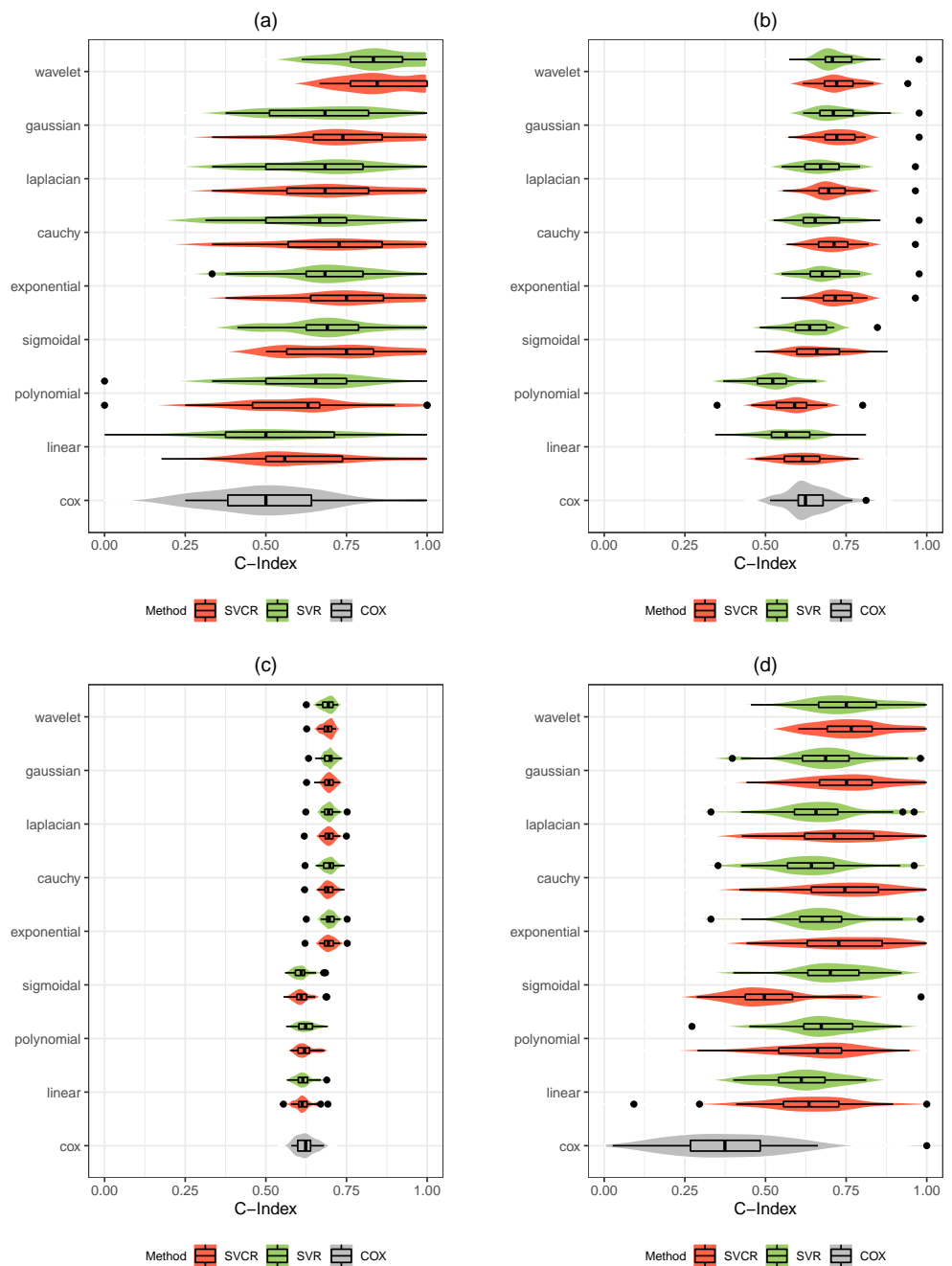


Figure 1. Boxplots of the C-index. (a) $n = 30$ and $r = 0.50$, (b) $n = 100$ and $r = 0.50$, (c) $n = 300$ and $r = 0.05$, (d) $n = 300$ and $r = 0.95$.

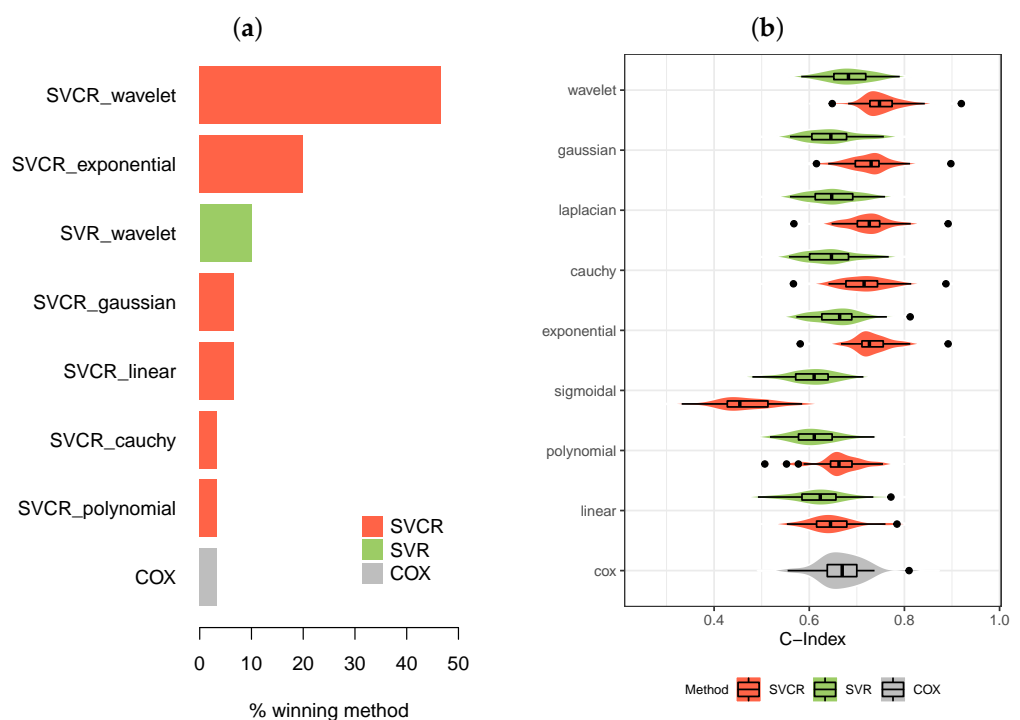


Figure 2. (a) C-index winning method on CGD dataset. (b) Boxplots of the C-index by CGD dataset.

Table 5. Variables and statistical descriptive for the ages (in years) of patients with congenital and coronary heart disease followed up in both protocols.

Age	Congenital		Coronary	
	Conventional Care	Fast-Track Care	Conventional Care	Fast-Track Care
Total	20	50	37	38
Mean	8.5	12.2	60.5	58.4
Standard deviation	13.4	11.1	12.9	8.8
Minimum	0.3	0.8	18.0	38.0
Median	4.0	10.0	63.0	58.0
Maximum	49.0	38.0	81.0	79.0

The parameters used in the grid-search for the tuning process were similar to previous applications varying the parameters $C \in \{0.25; 0.5; 1; 2; 4\}$, γ and $A \in \{0.125; 0.25; 0.5; 1; 2; 4; 8\}$. The concordance was investigated by repeated holdout validation with 100 repetitions and a split ratio of 70–30% of training–test, respectively. Figure 3 displays the grid search results of the wavelet SVCR for the 36 different possibilities, where it is observed that the highest values of C-index are obtained by increasing the values of A . Table 6 presents the average values obtained from 100 repetitions, revealing the superior performance of the wavelet kernel for both SVR and SVCR approaches. Notably, the results show less variation in the wavelet kernel across tests with the SVR approach, which is expected due to the low censoring rate. This trend is further illustrated in Figure 4, where the wavelet and Gaussian kernel models exhibit the best performance.

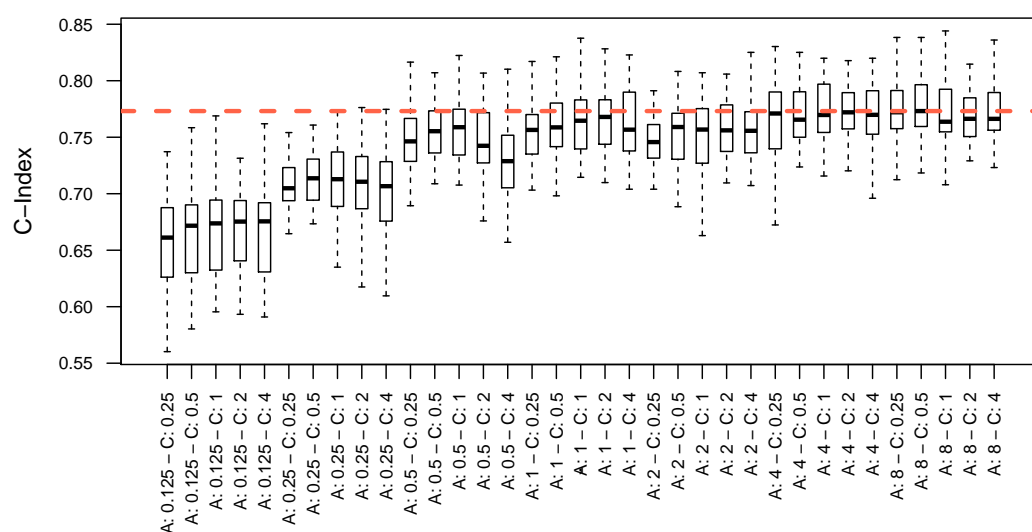


Figure 3. C-index distribution for the wavelet SVCR tuning process. Dotted line represents the chosen values by the median, $A = 8$ and $C = 0.5$.

Table 6. Mean and standard deviation of C-index for the biomedical real data.

COX									
Mean	0.757								
SD	0.030								
SVR									
	linear	polynomial	sigmoidal	Gaussian	exponential	Cauchy	Laplacian	wavelet	
Mean	0.768	0.766	0.770	0.798	0.798	0.796	0.794	0.800	
SD	0.031	0.042	0.029	0.029	0.029	0.030	0.030	0.028	
SVCR									
	linear	polynomial	sigmoidal	Gaussian	exponential	Cauchy	Laplacian	wavelet	
Mean	0.770	0.762	0.771	0.801	0.797	0.797	0.793	0.810	
SD	0.031	0.040	0.030	0.029	0.027	0.029	0.028	0.028	

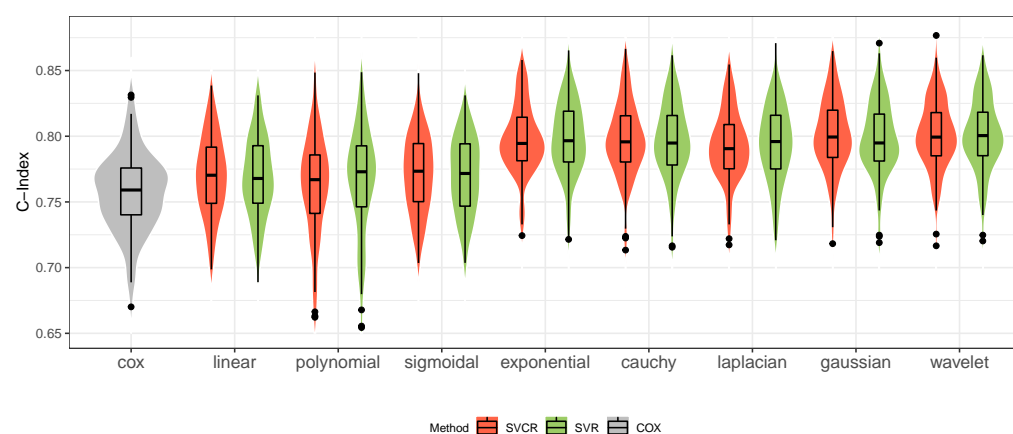


Figure 4. C-index distribution to the biomedical real data.

6. Final Comments

This paper proposes the WSVRC approach and compares different methods and kernel functions applied to survival data. Three different approaches are discussed: (i) the standard survival analysis approach, i.e., the Cox proportional hazards model; (ii) the support vector machine approach for regression models, i.e., the SVR; and (iii) the support vector approach for censored data, i.e., the SVCR, with the novel WSVCR model. The

simulation results and applications show that the WSVCR approach is either equivalent to or outperforms the other methods. Hence, it is not advisable to persist with a model that requires a strong assumption to be validated. Instead, the WSVCR approach, which is more flexible and general, may be employed.

Furthermore, this article presents an extensive study on different kernel functions and demonstrates that kernels such as exponential, cauchy, and Laplacian show good performance and are not often considered in many papers, as most of them only employ the most common kernels (i.e., linear, polynomial, and Gaussian).

Author Contributions: Conceptualization, A.A.; methodology, A.A. and M.M.; software, M.M., J.S.P. and A.A.; validation, A.A. and R.O.; investigation, M.M., R.O. and A.A.; data curation, M.M. and J.S.P.; writing—original draft preparation, M.M. and J.S.P.; writing—review and editing, M.M., A.A. and R.O.; supervision, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Council for Scientific and Technological Development (CNPq) through the grant 305305/2019-0 (R.O.), Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES), from the Brazilian government and Science Foundation Ireland Career Development Award grant number 17/CDA/4695 and SFI research centre award 12/RC/2289P2 (M.M.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This Appendix refers to more general summary of the simulation results displayed in Section 4. The Figures A1–A4 are related to the median, minimum, maximum and standard deviation over the 100 holdout repetitions. The values are showed in traffic light scale.

<i>n</i>	30			100			300				Mean
<i>r</i>	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.95	
COX	0.630	0.623	0.667	0.672	0.687	0.696	0.653	0.660	0.684	0.609	0.658
SVR											
linear	0.641	0.604	0.690	0.663	0.670	0.690	0.647	0.653	0.677	0.655	0.659
polynomial	0.667	0.667	0.717	0.595	0.587	0.578	0.649	0.656	0.684	0.689	0.649
sigmoidal	0.704	0.714	0.828	0.668	0.691	0.720	0.627	0.641	0.643	0.748	0.699
gaussian	0.806	0.852	0.886	0.732	0.740	0.768	0.720	0.731	0.746	0.777	0.776
exponential	0.799	0.828	0.846	0.730	0.750	0.770	0.717	0.729	0.744	0.740	0.765
cauchy	0.793	0.833	0.838	0.729	0.739	0.759	0.714	0.729	0.736	0.713	0.758
laplacian	0.789	0.822	0.841	0.725	0.741	0.748	0.707	0.721	0.733	0.721	0.755
wavelet	0.789	0.847	0.875	0.722	0.734	0.786	0.716	0.733	0.743	0.832	0.778
SVCR											
linear	0.655	0.649	0.667	0.660	0.690	0.734	0.645	0.657	0.684	0.698	0.674
polynomial	0.667	0.667	0.706	0.605	0.608	0.625	0.648	0.652	0.688	0.666	0.653
sigmoidal	0.714	0.714	0.800	0.670	0.698	0.735	0.623	0.635	0.592	0.491	0.667
gaussian	0.806	0.841	0.875	0.726	0.750	0.776	0.721	0.735	0.756	0.835	0.782
exponential	0.806	0.830	0.848	0.733	0.751	0.778	0.720	0.740	0.754	0.829	0.779
cauchy	0.786	0.821	0.833	0.726	0.742	0.769	0.720	0.732	0.751	0.831	0.771
laplacian	0.778	0.827	0.833	0.726	0.740	0.755	0.712	0.730	0.739	0.807	0.765
wavelet	0.798	0.827	0.889	0.718	0.750	0.786	0.715	0.733	0.752	0.829	0.780

Figure A1. Median of C-index distribution in the simulation study.

<i>n</i>	30			100			300				Mean
<i>r</i>	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.95	
COX	0.333	0.150	0.000	0.501	0.497	0.431	0.579	0.567	0.545	0.000	0.360
SVR											
linear	0.306	0.105	0.000	0.496	0.469	0.344	0.544	0.547	0.537	0.273	0.362
polynomial	0.333	0.407	0.000	0.399	0.441	0.361	0.548	0.520	0.501	0.272	0.378
sigmoidal	0.393	0.414	0.412	0.520	0.539	0.462	0.533	0.532	0.531	0.400	0.473
gaussian	0.536	0.273	0.375	0.587	0.568	0.597	0.632	0.652	0.584	0.397	0.520
exponential	0.500	0.364	0.333	0.582	0.565	0.550	0.626	0.643	0.600	0.331	0.509
cauchy	0.500	0.273	0.313	0.572	0.548	0.525	0.621	0.640	0.577	0.313	0.488
laplacian	0.464	0.308	0.333	0.556	0.558	0.542	0.616	0.617	0.581	0.331	0.491
wavelet	0.607	0.524	0.611	0.592	0.600	0.573	0.626	0.648	0.610	0.455	0.585
SVCR											
linear	0.321	0.095	0.000	0.496	0.502	0.469	0.541	0.537	0.538	0.000	0.350
polynomial	0.250	0.286	0.000	0.482	0.399	0.350	0.566	0.521	0.528	0.000	0.338
sigmoidal	0.357	0.450	0.474	0.503	0.515	0.468	0.523	0.513	0.483	0.000	0.429
gaussian	0.536	0.273	0.333	0.597	0.571	0.571	0.626	0.658	0.655	0.441	0.526
exponential	0.500	0.273	0.375	0.582	0.558	0.549	0.621	0.658	0.655	0.429	0.520
cauchy	0.464	0.273	0.333	0.592	0.552	0.563	0.620	0.654	0.644	0.419	0.511
laplacian	0.464	0.273	0.333	0.579	0.558	0.542	0.619	0.640	0.651	0.426	0.509
wavelet	0.611	0.571	0.667	0.587	0.623	0.549	0.627	0.658	0.664	0.602	0.616

Figure A2. Minimum of C-index distribution in the simulation study.

<i>n</i>	30			100			300				Mean
<i>r</i>	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.95	
COX	0.944	0.889	1.000	0.789	0.836	0.904	0.721	0.740	0.825	1.000	0.865
SVR											
linear	0.893	0.880	1.000	0.762	0.854	0.894	0.739	0.743	0.771	1.000	0.854
polynomial	0.929	0.944	1.000	0.749	0.793	0.810	0.715	0.757	0.808	1.000	0.851
sigmoidal	0.944	1.000	1.000	0.775	0.851	0.886	0.701	0.735	0.760	1.000	0.865
gaussian	0.972	1.000	1.000	0.838	0.848	0.976	0.782	0.803	0.819	1.000	0.904
exponential	1.000	1.000	1.000	0.811	0.863	0.976	0.774	0.798	0.816	1.000	0.904
cauchy	0.944	1.000	1.000	0.816	0.887	0.976	0.776	0.793	0.807	1.000	0.900
laplacian	1.000	1.000	1.000	0.826	0.866	0.965	0.770	0.777	0.808	1.000	0.901
wavelet	0.982	1.000	1.000	0.814	0.839	0.976	0.778	0.801	0.821	1.000	0.901
SVCR											
linear	0.893	0.882	1.000	0.765	0.858	0.896	0.752	0.765	0.849	1.000	0.866
polynomial	0.929	0.938	1.000	0.747	0.756	0.826	0.715	0.739	0.835	0.946	0.843
sigmoidal	0.897	0.952	1.000	0.767	0.839	0.896	0.707	0.747	0.709	1.000	0.851
gaussian	0.944	1.000	1.000	0.828	0.859	0.976	0.786	0.817	0.829	1.000	0.904
exponential	1.000	1.000	1.000	0.824	0.887	0.965	0.779	0.808	0.825	1.000	0.909
cauchy	0.944	1.000	1.000	0.816	0.891	0.965	0.778	0.814	0.830	1.000	0.904
laplacian	1.000	1.000	1.000	0.824	0.891	0.965	0.771	0.803	0.817	1.000	0.907
wavelet	0.982	1.000	1.000	0.815	0.866	0.941	0.781	0.807	0.851	1.000	0.904

Figure A3. Maximum of C-index distribution in the simulation study.

<i>n</i>	30			100			300				Mean
<i>r</i>	0.05	0.25	0.50	0.05	0.25	0.50	0.05	0.25	0.50	0.95	
COX	0.116	0.151	0.204	0.061	0.073	0.097	0.032	0.036	0.050	0.245	0.107
SVR											
linear	0.112	0.144	0.213	0.066	0.080	0.117	0.034	0.036	0.049	0.164	0.102
polynomial	0.137	0.129	0.207	0.072	0.071	0.104	0.037	0.045	0.062	0.139	0.100
sigmoidal	0.098	0.114	0.139	0.058	0.071	0.092	0.030	0.035	0.047	0.140	0.082
gaussian	0.087	0.135	0.156	0.049	0.059	0.075	0.030	0.032	0.042	0.137	0.080
exponential	0.091	0.134	0.154	0.052	0.064	0.087	0.026	0.029	0.041	0.136	0.081
cauchy	0.089	0.148	0.178	0.051	0.066	0.088	0.029	0.031	0.042	0.162	0.088
laplacian	0.095	0.146	0.161	0.054	0.064	0.083	0.027	0.030	0.043	0.147	0.085
wavelet	0.072	0.095	0.098	0.046	0.053	0.078	0.030	0.032	0.038	0.116	0.066
SVCR											
linear	0.118	0.142	0.185	0.062	0.075	0.105	0.038	0.041	0.053	0.184	0.100
polynomial	0.141	0.153	0.207	0.060	0.072	0.096	0.034	0.046	0.054	0.181	0.104
sigmoidal	0.108	0.098	0.146	0.061	0.071	0.090	0.034	0.034	0.053	0.185	0.088
gaussian	0.087	0.120	0.151	0.044	0.052	0.076	0.032	0.033	0.036	0.138	0.077
exponential	0.093	0.135	0.144	0.049	0.056	0.080	0.030	0.032	0.034	0.144	0.080
cauchy	0.091	0.137	0.167	0.047	0.060	0.076	0.031	0.033	0.036	0.147	0.082
laplacian	0.097	0.147	0.156	0.049	0.059	0.078	0.029	0.032	0.036	0.145	0.083
wavelet	0.074	0.088	0.097	0.044	0.048	0.080	0.031	0.033	0.036	0.115	0.065

Figure A4. Standard deviation of C-index distribution in the simulation study.

References

- Kaplan, E.L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [\[CrossRef\]](#)
- Cox, D.R. Regression Models and Life-Tables. *J. R. Stat. Soc.* **1972**, *34*, 187–220. [\[CrossRef\]](#)
- Fouodo, C.J.; König, I.R.; Weihs, C.; Ziegler, A.; Wright, M.N. Support Vector Machines for Survival Analysis with R. *R J.* **2018**, *10*, 412–423. [\[CrossRef\]](#)
- Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 412–414.
- Cervantes, J.; Li, X.; Yu, W.; Li, K. Support vector machine classification for large data sets via minimum enclosing ball clustering. *Neurocomputing* **2008**, *71*, 611–619. [\[CrossRef\]](#)
- Zhang, L.; Zhou, W.; Jiao, L. Wavelet support vector machine. *IEEE Trans. Syst. Man Cybern. Part (Cybernetics)* **2004**, *34*, 34–39. [\[CrossRef\]](#)
- Van, M.; Hoang, D.T.; Kang, H.J. Bearing fault diagnosis using a particle swarm optimization-least squares wavelet support vector machine classifier. *Sensors* **2020**, *20*, 3422. [\[CrossRef\]](#)
- Fan, Y.; Sun, Z. CPI big data prediction based on wavelet twin support vector machine. *Int. J. Pattern Recognit. Artif. Intell.* **2021**, *35*, 2159013. [\[CrossRef\]](#)
- Hazarika, B.B.; Gupta, D.; Natarajan, N. Wavelet kernel least square twin support vector regression for wind speed prediction. *Environ. Sci. Pollut. Res.* **2022**, *22*, 86320–86336. [\[CrossRef\]](#)
- Sun, W.; Xu, C. Carbon price prediction based on modified wavelet least square support vector machine. *Sci. Total Environ.* **2021**, *754*, 142052. [\[CrossRef\]](#)
- Yahia, S.; Said, S.; Zaied, M. Wavelet extreme learning machine and deep learning for data classification. *Neurocomputing* **2022**, *470*, 280–289. [\[CrossRef\]](#)
- Mo, Z.; Zhang, Z.; Tsui, K.L. The variational kernel-based 1-D convolutional neural network for machinery fault diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [\[CrossRef\]](#)
- Szu, H.H.; Telfer, B.A.; Kadambe, S.L. Neural network adaptive wavelets for signal representation and classification. *Opt. Eng.* **1992**, *31*, 1907–1916. [\[CrossRef\]](#)
- Liu, J.W.; Zuo, F.L.; Guo, Y.X.; Li, T.Y.; Chen, J.M. Research on improved wavelet convolutional wavelet neural networks. *Appl. Intell.* **2021**, *51*, 4106–4126. [\[CrossRef\]](#)
- Zubek, V.B.; Khan, F.M. Support Vector Regression for Censored Data (SVRc): A Vowel Tool for Survival Analysis. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; Volume 1.
- Chen, Y.; Jia, Z.; Mercola, D.; Xie, X. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Comput. Math. Methods Med.* **2013**, *2013*, 873595. [\[CrossRef\]](#)
- Shivaswamy, P.K.; Chu, W.; Jansche, M. A support vector approach to censored targets. In Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, USA, 28–31 October 2007; pp. 655–660.
- Van Belle, V.; Pelckmans, K.; Van Huffel, S.; Suykens, J.A. Support vector methods for survival analysis: A comparison between ranking and regression approaches. *Artif. Intell. Med.* **2011**, *53*, 107–118. [\[CrossRef\]](#)
- Therneau, T.M.; Lumley, T. Package ‘survival’. *R Top. Doc.* **2015**, *128*, 28–33.
- Wang, J.; Williams, M.; Karafili, E. *Apply Machine Learning Approaches to Survival Data*; Imperial College: London, UK, 2018.

21. Brentnall, A.R.; Cuzick, J. Use of the concordance index for predictors of censored survival data. *Stat. Methods Med. Res.* **2018**, *27*, 2359–2373. [[CrossRef](#)]
22. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
23. Wilcoxon, F. Probability tables for individual comparisons by ranking methods. *Biometrics* **1947**, *3*, 119–122. [[CrossRef](#)]
24. Steck, H.; Krishnapuram, B.; Dehing-Oberije, C.; Lambin, P.; Raykar, V.C. On ranking in survival analysis: Bounds on the concordance index. *Adv. Neural Inf. Process. Syst.* **2007**, *20*.
25. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
26. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1997**, *28*, 155–161.
27. Zang, E.; Afifi, A. Support vector machines (SVMs) with universal kernels. *Appl. Artif. Intell.* **2011**, *25*, 575–589. [[CrossRef](#)]
28. Courant, R.; Hilbert, D. *Methods of Mathematical Physics*; Interscience Publ. Inc.: New York, NY, USA, 1953; Volume 1, p. 106.
29. Jebara, T. Multi-task feature and kernel selection for SVMs. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AL, Canada, 4–8 July 2004; p. 55.
30. Chapelle, O.; Vapnik, V. Model selection for support vector machines. In Proceedings of the Advances in Neural Information Processing Systems, Denver, CO, USA, 29 November–4 December 1999; pp. 230–236.
31. Yger, F.; Rakotomamonjy, A. Wavelet kernel learning. *Pattern Recognit.* **2011**, *44*, 2614–2629. [[CrossRef](#)]
32. Gholizadeh, S.; Salajegheh, E.; Torkzadeh, P. Structural optimization with frequency constraints by genetic algorithm using wavelet radial basis function neural network. *J. Sound Vib.* **2008**, *312*, 316–331. [[CrossRef](#)]
33. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
34. Frank, M.; Wolfe, P. An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **1956**, *3*, 95–110. [[CrossRef](#)]
35. Platt, J. Sequential Minimal Optimization: A fast Algorithm for Training Support Vector Machines. *Adv. Kernel-Methods-Support Vector Learn.* **1998**, 208.
36. Bender, R.; Augustin, T.; Blettner, M. Generating survival times to simulate Cox proportional hazards models. *Stat. Med.* **2005**, *24*, 1713–1723. [[CrossRef](#)]
37. Carroll, K.J. On the use and utility of the Weibull model in the analysis of survival data. *Control Clin. Trials* **2003**, *24*, 682–701. [[CrossRef](#)]
38. Crowther, M.J.; Lambert, P.C. Simulating biologically plausible complex survival data. *Stat. Med.* **2013**, *32*, 4118–4134. [[CrossRef](#)]
39. Tantithamthavorn, C.; McIntosh, S.; Hassan, A.E.; Matsumoto, K. An empirical comparison of model validation techniques for defect prediction models. *IEEE Trans. Softw. Eng.* **2016**, *43*, 1–18. [[CrossRef](#)]
40. Louzada, F.; Ara, A.; Fernandes, G.B. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surv. Oper. Res. Manag. Sci.* **2016**, *21*, 117–134. [[CrossRef](#)]
41. Cawley, G.C.; Talbot, N.L.; Janacek, G.J.; Peck, M.W. Sparse bayesian kernel survival analysis for modeling the growth domain of microbial pathogens. *IEEE Trans. Neural Netw.* **2006**, *17*, 471–481. [[CrossRef](#)] [[PubMed](#)]
42. Fernandes, A.M.d.S.; Mansur, A.J.; Canêo, L.F.; Lourenço, D.D.; Piccioni, M.A.; Franchi, S.M.; Afiune, C.M.C.; Gadioli, J.W.; Oliveira, S.d.A.; Ramires, J.A.F. Redução do período de internação e de despesas no atendimento de portadores de cardiopatias congênitas submetidos à intervenção cirúrgica cardíaca no protocolo da via rápida. *Arq. Bras. Cardiol.* **2004**, *83*, 18–26. [[CrossRef](#)] [[PubMed](#)]
43. Cavalcante, T.; Ospina, R.; Leiva, V.; Cabezas, X.; Martin-Barreiro, C. Weibull Regression and Machine Learning Survival Models: Methodology, Comparison, and Application to Biomedical Data Related to Cardiac Surgery. *Biology* **2023**, *12*, 442. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.