

**A peer-reviewed version of this preprint was published in PeerJ on 11 June 2015.**

[View the peer-reviewed version](https://doi.org/10.7717/peerj.1009) (peerj.com/articles/1009), which is the preferred citable publication unless you specifically need to cite this preprint.

Zhou L, Xu Q, Wang H. 2015. Rotation survival forest for right censored data. PeerJ 3:e1009 <https://doi.org/10.7717/peerj.1009>

# Rotation survival forest for right censored data

Lifeng Zhou, Qingsong Xu, Hong Wang

Recently, survival ensembles have found more and more applications in biological and medical research when censored time-to-event data are often confronted. In this research, we investigate the plausibility of extending rotation forest, originally proposed for classification purpose, to survival analysis. Supported by the proper statistical analysis, we show that rotation survival forests are able to outperform the state-of-art survival ensembles on right censored data. We also provide a C-index based variable importance measure for evaluating covariates in censored survival data.

# Rotation Survival Forest for Right Censored Data

Lifeng Zhou<sup>a</sup>, Qingsong Xu<sup>a</sup>, Hong Wang<sup>a,\*</sup>

<sup>a</sup> School of Mathematics and Statistics, Central South University, China

---

---

## 1. Introduction

In biological and medical research, time-to-event data are often confronted. Survival analysis focuses on studying the relationship between covariates and the time until an event of interest occurs. The analysis has become complicated when the data are censored due to various reasons such as some patients are uncooperative and withdraw from a clinical trial or some patients do not experience the event (death or occurrence of a disease) when a clinical trial ends. Many parametric or semi-parametric models such as Cox-proportional hazards model and its extensions (David, 1972; Cox and Oakes, 1984) are developed to investigate such relationship in censored data. However, when the underlying assumptions are not satisfied, these models may not lead to faithful conclusions. Therefore, non-parametric models such as survival trees (LeBlanc and Crowley, 1995; Bou-Hamad et al., 2011) and neural networks (Faraggi and Simon, 1995) are evolved to relax or remove the restrictive assumptions.

Recently, ensemble based approaches which combine one of the previous parametric and non-parametric models with state-of-the-art ensemble learning techniques are applied to create accurate and diverse base learners. Bagging, one of the most simple but ingenious ensemble techniques, was first applied with survival trees to censored data in (Dannegger, 2000) and (Benner, 2002). A general bagging method for arbitrary tree growing algorithms with a conditional survival function was proposed in (Hothorn et al., 2004). The popular random forest (Breiman, 2001) method was also

---

\*Corresponding Author. School of Mathematics and Statistics, Central South University

Email addresses: lfzhou@csu.edu.cn (Lifeng Zhou), qsxu@csu.edu.cn (Qingsong Xu), wh@csu.edu.cn (Hong Wang)

24 extended to survival analysis scenario by Hothorn et al. (2006) and later by  
 25 Ishwaran et al. (2008). In (Hothorn et al., 2006), estimated inverse prob-  
 26 ability of censoring weights were used as sampling weights in constructing  
 27 the bootstrapping samples and the final ensemble prediction is a weighted  
 28 average of log-survival time predictions from all survival trees. In (Ishwaran  
 29 et al., 2008), Nelson-Aalen estimates of cumulative hazard functions were  
 30 obtained and averaged from all nodes, and four different splitting criteria  
 31 including log-rank statistics and conservation-of-event principle were provid-  
 32 ed for constructing the so-called random survival forest (RSF). The uniform  
 33 consistency for RSF was further proved in (Ishwaran and Kogalur, 2010).  
 34 Boosting techniques using different type of base learners such as regression  
 35 trees (Hothorn et al., 2006; Chen et al., 2013) and smoothing splines (Lu  
 36 and Li, 2008) were also studied. A Bayesian ensemble using Cox propor-  
 37 tional hazard model, Weibull regression and accelerated failure time model  
 38 for high dimensional survival data was presented in (Bonato et al., 2011).  
 39 All these survival ensembles have been proved to be more effective than pre-  
 40 vious monolithic models (Hothorn et al., 2004, 2006; Ishwaran et al., 2008;  
 41 Ishwaran and Kogalur, 2010; Bonato et al., 2011)

42 In this article, we introduce rotation survival forests, a novel tree ensemble  
 43 for analyzing survival data. The proposed rotation survival forest (RotSF)  
 44 methodology extends the original rotation forest(RotF) approach (Rodriguez  
 45 et al., 2006) from classification to survival analysis. In RotF, accurate and  
 46 diverse classifiers are obtained through variable reconstruction by assembling  
 47 information extracted from variable subsets. The randomness and diversity is  
 48 introduced in two forms. First, a bootstrap sample of the training dataset is  
 49 randomly generated. Second, variables are randomly divided into a number of  
 50 disjoint subsets and principal component analysis (PCA) is applied in turn to  
 51 each variable data subset to obtain the so-called rotation matrix. Multiplying  
 52 the original dataset by the rotation matrix will lead to a new rotated dataset.  
 53 These rotated datasets will be used to train tree base learners within the  
 54 ensemble.

55 RotF is unique among all ensemble learning algorithms in that all the  
 56 information in the original dataset is preserved, though in different forms  
 57 of representation. Various experiments results and theoretical analysis have  
 58 shown that RotF is a very competitive ensemble learning method. Howev-  
 59 er, applications of RotF have focused primarily on classification problems so  
 60 far. It is of a great value to generalize rotation forest to right censored sur-  
 61 vival data. Similar to RotF, our proposed rotation survival forest approach

random split the  $p$  covariates into  $k$  variable subset and then some feature extraction method is executed on each variable subset. Different from RotF, coefficients of covariates which are not chosen in  $k$  subsets of variable are set to 0 in the rotation matrix. To further diverse the data, we employ a double bagging approach in preparing the data for rotation which in turn improves the prediction accuracy.

A byproduct of our proposed approach is that importance of variables can be calculated easily using the so-called out-of-bag (OOB) data (Breiman, 2001). The proposed importance measure based on Harrell's C-index (Harrell et al., 1996) is very useful in survival analysis, as researchers usually have strong interest in determining the most significant covariates that affecting the survival probability.

In order to carry out the empirical comparisons, we establish a similar experimental framework to that in (Ishwaran et al., 2008); we test on the famous public survival datasets available from public repositories. We consider an extended version of classification and regression tree (CART)(Breiman et al., 1984) as the base classifier for our rotation ensemble in that it has been the most commonly used non-parametric method in analyzing survival data (Bou-Hamad et al., 2011). To estimate prediction errors of various survival models, we use C-index as the evaluation criterion as suggested by Ishwaran et al. (2008). The results obtained in the comparisons are further validated by some proper statistical tests.

The main contributions of this paper can be summarized as follows:

- We extend the original rotation forest approach from classification to survival analysis of right censored data;
- We provide a C-index based variable importance measure for evaluating covariates in censored survival data.

## 2. Methods

Survival analysis is the study of relationship between survival time  $\tau$  and a set of covariates  $\mathbf{X} = (X_1, X_2, \dots, X_p)$ . Here,  $\tau$  is not fully observed, i.e.  $\tau = \min(U, C)$  is composed of both true survival time  $U$  and censored time  $C$  and an indicator variable  $\delta = I(U \leq C)$  takes 1 for true event time  $U$  and 0 otherwise. The major goal of survival analysis is to estimate survival experiences of different patient groups via the so-called cumulative survival function, which captures the probability that the event does not occur until

a given time. In this study, we propose to model the survival function using a novel non-parametric learning ensemble called Rotation Survival Forest (RotSF).

Assume  $\mathbf{X}$  to be a variable set  $\mathbf{V}$  of  $p$  covariates and  $D$  be the dataset containing the training samples in a form of  $n \times (p + 2)$  matrix, namely  $D = (\tau_q, \delta_q, \mathbf{X}_q), q = 1, 2, \dots, n$ .

In the following, we will give a high-level description of how RotSF trains a base survival learning algorithm  $S_i$ :

1. Generate a bootstrap sample  $D'$  of size  $n$  from  $D$ .
2. Split  $\mathbf{V}$  randomly into  $k$  disjoint subsets  $V_{i,j} (j = 1, \dots, k)$  such that each variable subset contains  $M$  covariates. If  $p$  is not divisible by  $M$ , there would be some variables not included in any subset and denote these remaining variables by  $RV$ .
3. In the  $j$ -th iteration, generate a bootstrap sample  $D''$  of size  $n$  from  $D'$ . Let  $X_{i,j}$  be the subset of  $D''$  with variable set  $V_{i,j}$ .
4. Run PCA on  $X_{i,j}$  and obtain the variable loadings matrix (rotation matrix)  $M_{i,j}$  for  $V_{i,j}$ .
5. Repeat above steps 3 and 4 for all  $j = 1, 2, \dots, k$ , and obtain a group of  $M_{i,j}$ 's ( $j = 1, 2, \dots, k$ ) for all  $V_{i,j}$ 's.
6. For  $RV$  covariates, set the their rotations (loadings) to 0, and hence all covariates have corresponding rotation values. Then rearrange these values according to the covariates order in  $V$  and we get the rearranged rotation matrix  $R_i^a$ .
7. Use the newly formed data  $(\tau_i, \delta_i, X R_i^a)$  as the training set for the base survival tree algorithm, we will get a base survival tree learner  $S_i$ .

As with most ensemble learning methods, ensemble size  $L$  needs to be set beforehand. Similar to RotF, a fixed value  $M = 2$  is given to the size of variable subset. The base survival tree algorithm adopted here is the CART algorithm extended by LeBlanc and Crowley (1992).

The pseudo-code of the proposed RotSF algorithm is presented in Algorithm 1:

---

**Algorithm 1** The Rotation Survival Forest Algorithm

---

```

1: INPUT:
2:   Training data:  $D = (\tau_q, \delta_q, \mathbf{X}_q), q = 1, \dots, n$ , where  $\mathbf{X}$  is a variable
   set  $\mathbf{V}$  of  $p$  covariates
3:    $M$ : Number of variables within a subset
4:    $L$ : Ensemble size
5:    $S_i$ : Survival tree base learning algorithm
6: OUTPUT:
7:   Survival ensemble  $S$ 
8: procedure ROTSF( $D, M, L$ )
9:   while  $i$  in  $1 : L$  do
10:     Preparing the rotation matrix  $R_i^a$ :
11:     Generate a bootstrap sample  $D'$  from  $D$ 
12:     Randomly split variables  $\mathbf{V}$  into  $k = p/M$  equal size subsets
        $V_{i,j}, j = 1, \dots, k$ 
13:     Denote the variables not chosen (remaining variables) as  $RV$ 
14:     while  $j$  in  $1 : k$  do
15:       Generate a bootstrap sample  $D''$  from  $D'$ 
16:       Let  $X_{i,j}$  be subset of  $D''$  with variable set  $V_{i,j}$ 
17:       Apply PCA on  $X_{i,j}$  to get the rotation matrix  $M_{i,j}$ 
18:     end while
19:     If there are any  $RV$  variables, set the corresponding rotations to 0
20:     Arrange all  $M_{i,j}$ s to match variable order in  $V$  and obtain rotation
       matrix  $R_i^a$ 
21:     Train a survival tree base learner  $S_i$  using  $(\tau_i, \delta_i, X R_i^a)$  as the
       training set
22:   end while
23:   return The survival ensemble  $S$ 
24: end procedure
25: In prediction, the average hazard rate  $r$  for new data  $\mathcal{X}$ :

```

$$r = \frac{1}{L} \sum_{S_i \in S} S_i(\mathcal{X} R_i^a)$$

---

128 From Algorithm 1, we know that similar to random survival forest, RotS-  
129 F is also a parallel algorithm and the first "while" part can be executed  
130 concurrently to save time in case of large survival data.

### 131 3. Results & Discussion

132 In this section, we investigate the performance of the proposed RotSF  
133 model and compare with three popular survival models for censored survival  
134 data.

#### 135 3.1. Datasets

136 We evaluate the performance of our algorithm on three well-known sur-  
137 vival benchmark datasets which have been extensively analyzed in the sta-  
138 tistical literature. A brief introduction and summary of the used datasets  
139 are given below and in Table 1.

##### 140 3.1.1. Primary Biliary Cirrhosis(PBC) Dataset

141 This dataset is from the Mayo Clinic trial in primary biliary cirrhosis  
142 (PBC) of the liver conducted between 1974 and 1984. There are 418 patients  
143 in this study, 257 of whom have censored data. The currently used dataset  
144 is taken from Appendix D of Fleming and Harrington (1991).

##### 145 3.1.2. Chronic Myelogenous Leukemia(CML) Dataset

146 This dataset contains survival informatin in a randomised trial comparing  
147 three treatments for Chronic Myelogeneous Leukemia(CML). In this dataset,  
148 507 samples, 108 of which are censored, are simulated according to structure  
149 of the data by the German CML Study Group used in Hehlmann et al. (1994).

##### 150 3.1.3. Veterans' Administration Lung Cancer(Veteran) Dataset

151 This dataset includes survival data for 137 patients from Veteran's Ad-  
152 ministration Lung Cancer Trial and was first made public by Kalbfleisch and  
153 Prentice (1980). There are 9 censored observations.

Table 1: Summary of three benchmark datasets used in the paper

Dataset	Samples	Covariates	Censored data	Censoring rate
PBC	418	17	257	61.48%
CML	507	5	108	21.30%
Veteran	137	6	9	6.5%

### 3.2. Evaluation metric & statistical tests

In medical decision making, researchers and doctors are usually concerned with the relative risks between patients with various covariates. To evaluate the accuracy of such relative risks, Harrells concordance index (C-index) measure was proposed in (Harrell et al., 1996). Currently, it is a widely adopted statistic in evaluating different survival models and will also be the evaluation metric in our later experiments.

To compare the performance of various survival models, the non-parametric Friedman test (Demšar, 2006) is applied. Friedman's test statistic is based on the average ranked performance of the algorithms on each run of the datasets and can be calculated according to the following formula:

$$FT = \frac{12}{nm(m+1)} \sum_{j=1}^m \left( \sum_{i=1}^n r_i^j \right)^2 - 3n(m+1) \quad (1)$$

where  $m$  denotes the number of survival models,  $n$  the number of runs, and  $r_i^j$  the rank of survival models  $j$  on the  $i$ -th run. If the value of  $FT$  is large enough, the null hypothesis that there is no significant difference among the different survival models can be rejected and a Nemenyi post-hoc test can be adopted to find where the difference lies.

For two survival models  $C_1$  and  $C_2$ , the Nemenyi statistic  $z$  is calculated as follows:

$$z = \frac{R_{j1} - R_{j2}}{\sqrt{\frac{m(m+1)}{6n}}} \quad (2)$$

where  $R_j$  denotes the mean rank of survival models  $C_j$  on all runs of the dataset, namely,  $R_j = \frac{1}{n} \sum_{i=1}^n r_i^j$ . The performance of two survival models is significantly different if the  $z$  value is larger than a certain critical value (Demšar, 2006).

### 3.3. Evaluating covariate importance

Covariate(variable) importance plays an importance role in the interpretation of a survival model. In this study, we introduce a new variable important measure, mean C-index decrease measure. It is based on C-index values difference on the OOB data before and after permuting the values of the variable in consideration. The variable importance for variable  $i$  in terms of mean C-index decrease is defined by:

$$VI_i^c = \frac{100}{L} \sum_{j=1}^L (C_i - C_{\bar{i}}) \quad (3)$$

183 where  $C_i$  and  $C_{\bar{i}}$  denotes C-index values on the current OOB data before and  
184 after the permutation of variable  $i$ .

185 If the variable in question is not associated with the survival outcome,  
186 value permutation will have no influence on the prediction and hence no in-  
187 fluence on the C-index. On the contrary, if outcomes and variables are indeed  
188 associated, permuting variable values results in a worse prediction power and  
189 will lead to a decrease in the C-index value. The C-index difference before  
190 and after randomly permuting the variable will reflect the importance level of  
191 the current variable. Averaging the decreased values across all survival tree  
192 base models, we will get a list of mean decrease C-index values. The higher  
193 a mean decrease C-index value is, the more important a variable would be.

### 194 3.4. Experiment Results

195 We conduct our experiments on a system with a Pentium Dual-Core  
196 3.20GHz CPU and 4G RAM. The proposed RotSF algorithm is implemented  
197 in the R programming language. In the experiments, all training and test-  
198 ing sets are from a random 80% and 20% split of the benchmark datasets.  
199 Hereafter, the reported performance results are based on 1000 random runs  
200 of RotSF and other methods.

#### 201 3.4.1. Covariate importance result

202 For illustration purpose, we calculate covariates' importance in the P-  
203 BC dataset using the above mean C-index measure. A bar plot of top 10  
204 important covariates are shown in Figure 1.

205 According the obtained result, covariates such as "bili" (serum bilirun-  
206 bin), "age" are very important clinical indicators for predicting survival of  
207 patients with primary biliary cirrhosis, as permutation of these variables can  
208 cause a decrease in C-index of more than 2 percent on average. Compared  
209 with the results in (Bou-Hamad et al., 2011), one may find that there is a  
210 significant overlap in the most important covariates found by RotSF, omit-  
211 one-covariate approach and rotation survival forest. In addition, the top two  
212 covariates "bill" and "age" are the same and in the same order.

213 Different approaches giving similar results provides further confidence  
214 that the covariates found with C-index measure can be used in clinical deci-  
215 sion making to evaluate the survival risks of patients.

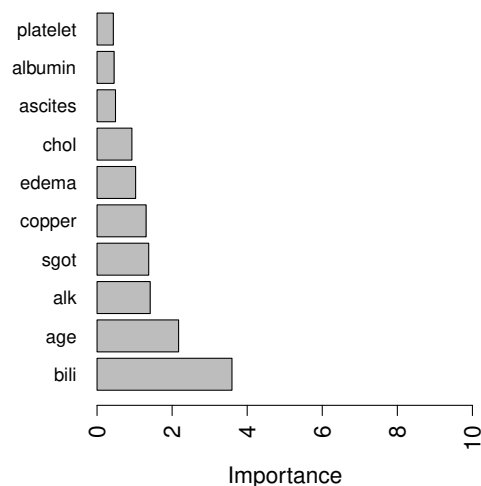


Figure 1: Top ten important variables by RotSF

### 3.4.2. The effect of double bagging

Next, we will test RotSF's performance with different bagging schemes: RotSF with double bagging(RotSF, the proposed algorithm); RotSF with single bagging(RotSFsb). For both two approaches, the same settings, i.e.  $L = 1000$  and  $M = 2$  are applied. For illustration and simplicity, we only report the results with the PBC dataset in the following Table 2.

Table 2: RotSF's performance with different bagging schemes

Statistic	RotSF	RotSFsb
Min	0.7000	0.7032
1st Quintile	0.8072	0.8046
Median	0.8370	0.8318
Mean	0.8347	0.8309
3rd Quintile	0.8650	0.8614
Max	0.9473	0.9402

To decide whether the two bagging schemes are significantly different, a Wilcoxon signed-rank test is applied. As the  $p$ -value turns out to be less

than  $2.2e-16$ , we can reject the null hypothesis that these two schemes have the same predictive power. Together with the results shown in Table 2, we can come to the conclusion that double bagging outperforms single bagging under the C-index metric. This confirms our assumption that double bagging scheme can result in a more diversified training data and this does improve the algorithm's performance.

### 3.4.3. Performance comparison result

Here, we will compare the proposed method with three state-of-the-art survival models. The first method is Cox proportional hazard (Cox) model (David, 1972); the second method is random survival forest (RSF) and the third method is gradient boosted model (GBM)(Ridgeway, 2004). Comparisons with these models are conducted with corresponding "survival", "randomForestSRC", and "gbm" packages in R. For the ease of notation, survival models RotSF, RSF, Cox and GBM are denoted by A, B, C, D, respectively when necessary. In the experiments, we want all classifiers to have the same opportunities to achieve the best results, thus the default settings are adopted. For ensemble methods, i.e. RotSF, RSF and GBM, 1000 trees are built.

The following Figure 2 reports the performance of RotSF, RSF, LLR and CART algorithms in term of C-index on 1000 runs of the experiments.

The Friedman rank sum test statistics on PBC, CML and Veteran datasets are 800.9539, 2498.443 and 403.2192, respectively and all are significant as all three  $p$ -values are less than  $2.2e-16$ .

Thus, to find out which pairs of algorithms are significantly different, we can compute the Nemenyi test statistics for different pairs of survival models, i.e.,  $z_{BA}$ ,  $z_{CA}$  and  $z_{DA}$ . The post hoc Nemenyi test results are shown in the following Table 3.

Table 3: Friedman test and Nemenyi test results on all datasets

Dataset	Nemenyi $z_{BA}$	Nemenyi $z_{CA}$	Nemenyi $z_{DA}$
PBC	4.849742	14.98224	26.24057
CML	2.182384	39.33487	25.82488
Veteran	4.78046	14.53191	15.39793

For  $\alpha = 0.05$ , the critical value of Nemenyi's test is 2.5742. It can be

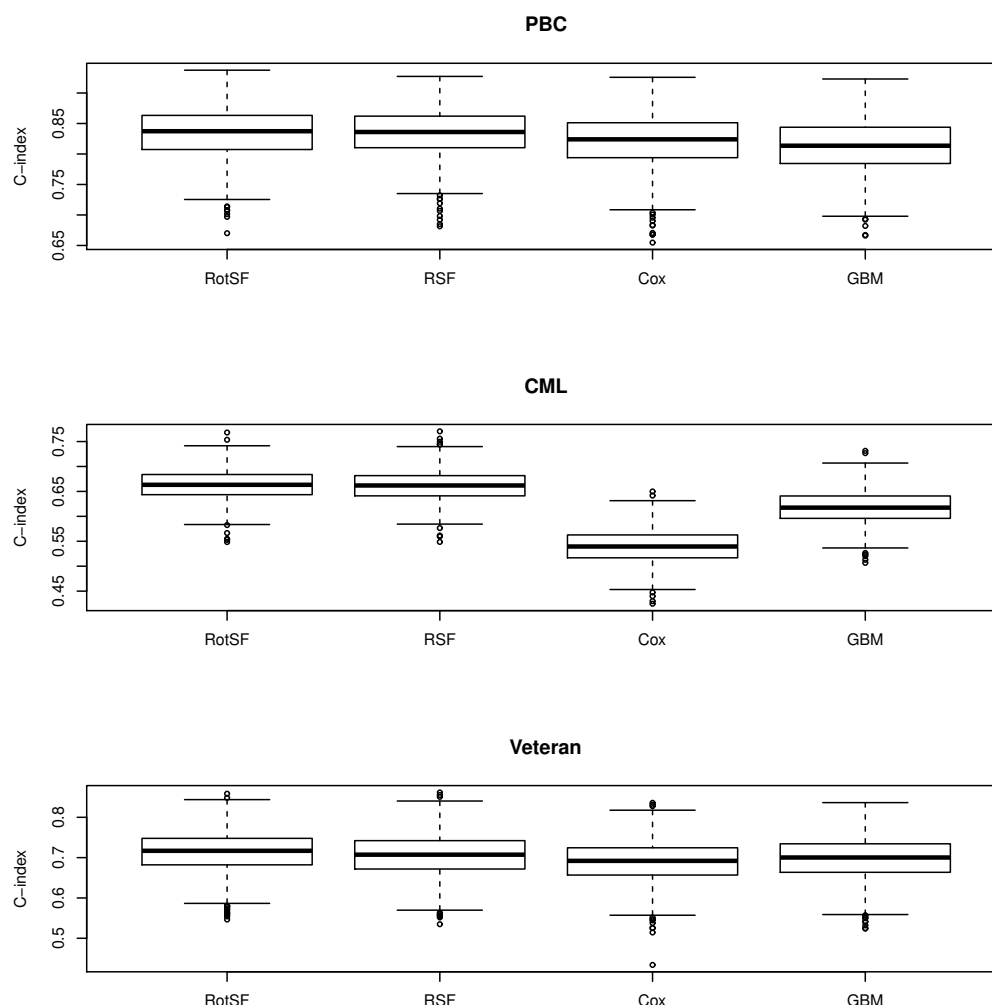


Figure 2: Boxplots of performance in terms of C-index

252 seen that all Nemenyi statistics except the  $z_{BA}$  value on CML dataset exceed  
 253 2.5742. Thus, in terms of C-index, there exists significant differences between  
 254 RotSF and the other three algorithms on PBC and Veteran datasets and also  
 255 significant differences between RotSF and Cox & GBM on CML dataset. In  
 256 other words, on PBC and Veteran datasets, RotSF is significantly better than  
 257 RSF, Cox and GBM; on CML dataset, RotSF is also significantly better than

Cox and GBM.

Tough RotSF beats RSF in 543 out 1000 runs on CML dataset, the difference between RotSF and RSF is not significantly different at this time according to the Nemenyi statistic. However, if we repeat the experiments more times, for example, 2000 times, the Nemenyi test statistic becomes 3.612997 and is above the critical value 2.5672. Thus, RotSF is also significantly better than RSF on CML dataset.

#### 4. Conclusion

In this study, we have developed a new ensemble learning algorithm, rotation survival forest, for survival analysis. By studying the well-known benchmark datasets, we have found that RotSF generally outperforms state-of-the-art survival models such as rotation survival forest, Cox proportional hazard and generalized boosted model in terms of C-index metric. As a non-parametric approach, RotSF does not impose parametric assumptions on hazard functions, and it extends the well-known rotation forest methodology to survival analysis.

This study also provides a mean C-index decrease measure to evaluate variable importance. The important covariates identified by RotSF agrees strongly with results reported in previous studies and may provide useful clues for clinical decision making. It is clear that other methods (ensembles and not) are available but the goal here is to illustrate some key features of the proposed method and not to provide an exhaustive comparison across methods.

The R code and the supplementary material are available at url: <https://github.com/whcsu/rotsf> and we are working hard to provide an R package for the proposed RotSF algorithm as soon as possible. The proposed algorithm still has room for improvement. First, RotSF has an extra parameter which controls the variables within a subset and we just set it to 2 in our experiments. We can further test its sensitivity or use cross-validation to tune this parameter. Second, as double bagging is applied, RotSF is computationally more intensive than other ensemble methods. Fortunately, RotSF is easily parallelizable, which could help in dealing with big data. Third, same as rotation forest, PCA is chosen for feature extraction in our approach. One may try other feature extraction methods as well.

## References

- Benner, A., 2002. Application of aggregated classifiers in survival time studies, in: *Compstat*, Springer. pp. 171–176.
- Bonato, V., Baladandayuthapani, V., Broom, B.M., Sulman, E.P., Aldape, K.D., Do, K.A., 2011. Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics* 27, 359–367.
- Bou-Hamad, I., Larocque, D., Ben-Ameur, H., 2011. A review of survival trees. *Statistics surveys* 5, 44–71.
- Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and regression trees*. CRC press.
- Chen, Y., Jia, Z., Mercola, D., Xie, X., 2013. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Computational and mathematical methods in medicine* 2013, 873595.
- Cox, D.R., Oakes, D., 1984. *Analysis of survival data*. volume 21. CRC Press.
- Dannegger, F., 2000. Tree stability diagnostics and some remedies for instability. *Statistics in medicine* 19, 475–491.
- David, C.R., 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* 34, 187–220.
- Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 7, 1–30.
- Faraggi, D., Simon, R., 1995. A neural network model for survival data. *Statistics in medicine* 14, 73–82.
- Fleming, T.R., Harrington, D.P., 1991. *Counting processes and survival analysis*. John Wiley&Sons, Hoboken, NJ, USA.
- Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* 15, 361–387.

- 322 Hehlmann, R., Heimpel, H., Hasford, J.t., Kolb, H., Pralle, H., Hossfeld, D.,  
323 Queisser, W., Loffler, H., Hochhaus, A., Heinze, B., 1994. Randomized  
324 comparison of interferon-alpha with busulfan and hydroxyurea in chronic  
325 myelogenous leukemia-the german cml study group. *Blood* 84, 4064–4077.
- 326 Hothorn, T., Bhlmann, P., Dudoit, S., Molinaro, A., Van Der Laan, M.J.,  
327 2006. Survival ensembles. *Biostatistics* 7, 355–373.
- 328 Hothorn, T., Lausen, B., Benner, A., RadespielTrger, M., 2004. Bagging  
329 survival trees. *Statistics in medicine* 23, 77–91.
- 330 Ishwaran, H., Kogalur, U.B., 2010. Consistency of random survival forests.  
331 *Statistics & probability letters* 80, 1056–1064.
- 332 Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random  
333 survival forests. *The Annals of Applied Statistics* 2, 841–860.
- 334 Kalbfleisch, J.D., Prentice, R.L., 1980. The statistical analysis of failure time  
335 data. John Wiley & Sons, New York, USA.
- 336 LeBlanc, M., Crowley, J., 1992. Relative risk trees for censored survival data.  
337 *Biometrics* 77, 411–425.
- 338 LeBlanc, M., Crowley, J., 1995. A review of tree-based prognostic models,  
339 in: *Recent Advances in Clinical Trial Design and Analysis*, Springer. pp.  
340 113–124.
- 341 Lu, W., Li, L., 2008. Boosting method for nonlinear transformation models  
342 with censored survival data. *Biostatistics* 9, 658–667.
- 343 Ridgeway, G., 2004. The gbm package. R Foundation for Statistical Com-  
344 puting, Vienna, Austria .
- 345 Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation forest: A new  
346 classifier ensemble method. *Pattern Analysis and Machine Intelligence*,  
347 *IEEE Transactions on* 28, 1619–1630.

## **Table 1**(on next page)

Table 1: Summary of three benchmark datasets used in the paper

Table 1: Summary of three benchmark datasets used in the paper

Dataset	Samples	Covariates	Censored data	Censoring rate
PBC	418	17	257	61.48%
CML	507	5	108	21.30%
Veteran	137	6	9	6.5%

## **Table 2**(on next page)

Table 2: RotSF's performance with different bagging schemes

Table 2: RotSF's performance with different bagging schemes

Statistic	RotSF	RotSFsb
Min	0.7000	0.7032
1st Quintile	0.8072	0.8046
Median	0.8370	0.8318
Mean	0.8347	0.8309
3rd Quintile	0.8650	0.8614
Max	0.9473	0.9402

### **Table 3**(on next page)

Table 3: Friedman test and Nemenyi test results on all datasets

Table 3: Friedman test and Nemenyi test results on all datasets

Dataset	Nemenyi	Nemenyi	Nemenyi
PBC	4.849742	14.98224	26.24057
CML	2.182384	39.33487	25.82488
Veteran	4.78046	14.53191	15.39793

Figure 1: Top ten important variables by RotSF

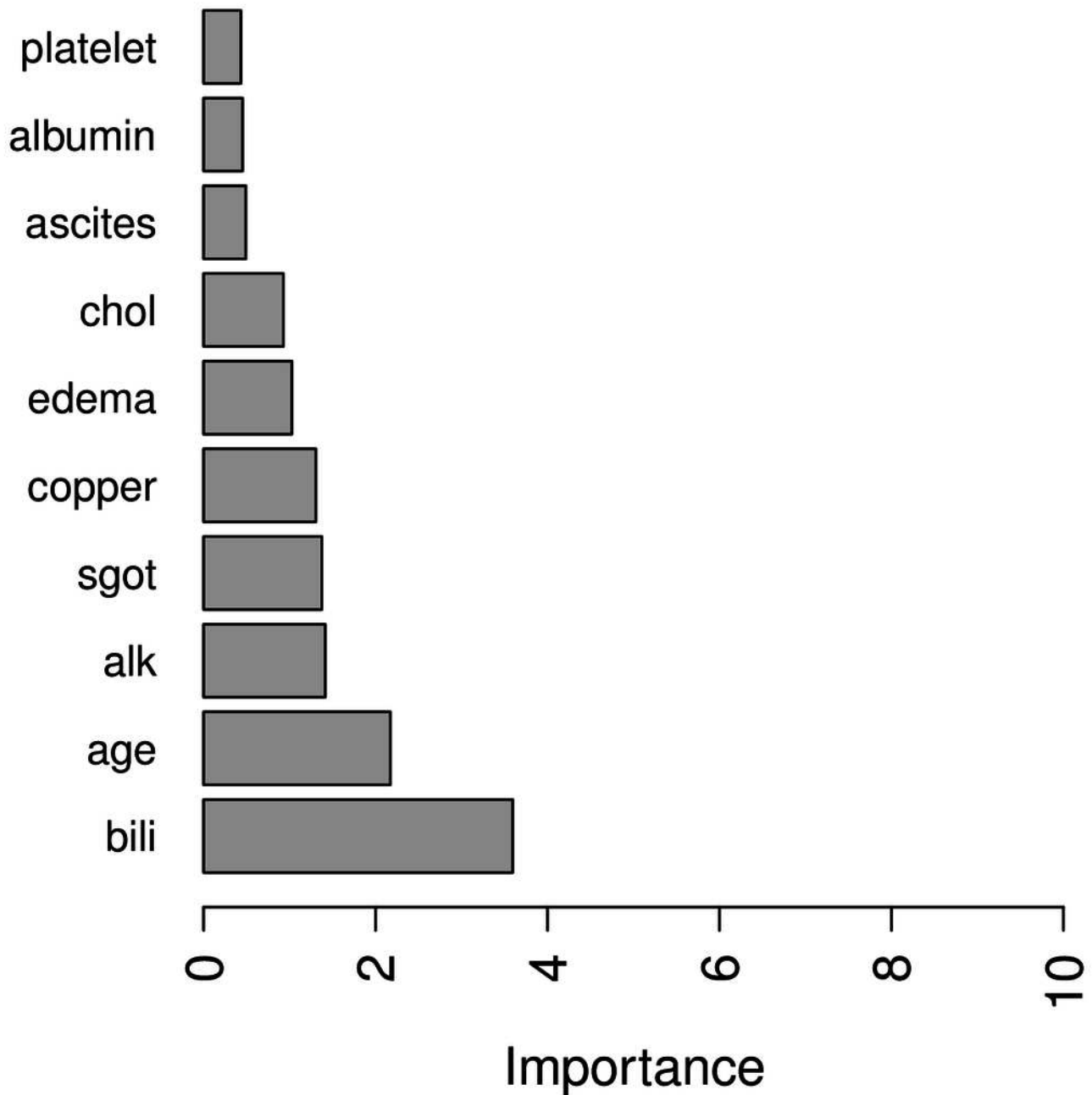


Figure 2: Boxplots of performance in terms of C-index

