Random Survival Forests

Hemant Ishwaran
Min Lu
and Udaya B. Kogalur

August 4, 2021

Introduction

Early applications of random forests (RF) focused on regression and classification problems. Random survival forests (Ishwaran et al., 2008) (RSF) was introduced to extend RF to the setting of right-censored survival data. Implementation of RSF follows the same general principles as RF: (a) Survival trees are grown using bootstrapped data; (b) Random feature selection is used when splitting tree nodes; (c) Trees are generally grown deeply, and (d) The survival forest ensemble is calculated by averaging terminal node statistics (TNS).

The presence of censoring is a unique feature of survival data that complicates certain aspects of implementing RSF compared to RF for regression and classification. In right-censored survival data, the observed data is (T, δ) where T is time and δ is the censoring indicator. The observed time T is defined as the minimum of the true (potentially unobserved) survival event time T^o and the true (potentially unobserved) censoring time C^o; thus $T = \min(T^o, C^o)$ and the actual event time might not be observed. The censoring indicator is defined as $\delta = 1\{T^o \leq C^o\}$. When $\delta = 1$, an event has occurred (i.e., death has occurred) and we observe the true event time, $T = T^o$. Otherwise when $\delta = 0$, the observation is censored and we only observe the censoring time $T = C^o$: thus we know that the subject has survived to time C^o, but not when the subject actually dies.

Hereafter we denote the data by $(T_1, \mathbf{X}_1, \delta_1), \ldots, (T_n, \mathbf{X}_n, \delta_n)$ where $\mathbf{X}_i$ is the feature vector (covariate) for individual i and T_i and δ_i are the observed time and censoring indicators for i. RSF trees just like RF trees are grown using resampling (for example by using the bootstrap; the default is to use .632 sampling without replacement). However for notational simplicity, we will sometimes ignore this distinction.

RSF splitting rules

The true event time being subject to censoring must be dealt with when growing a RSF tree. In particular, the splitting rule for growing the tree must specifically account for censoring. Thus, the goal is to split the tree node into left and right daughters with dissimilar event history (survival) behavior.

Log-rank splitting

The default splitting rule used by the package is the log-rank test statistic and is specified by `splitrule="logrank"`. The log-rank test has traditionally been used for two-sample testing with survival data, but it can be used for survival splitting as a means for maximizing

between-node survival differences (Ciampi et al., 1988; Segal, 1988, 1995; LeBlanc and Crowley, 1992, 1993).

To explain log-rank splitting, consider a specific tree node to be split. Without loss of generality let us assume this is the root node (top of the tree). For simplicity assume the data is not bootstrapped, thus the root node data is $(T_1, \mathbf{X}_1, \delta_1), \ldots, (T_n, \mathbf{X}_n, \delta_n)$. Let X denote a specific variable (i.e., one of the coordinates of the feature vector). A proposed split using X is of the form $X \leq c$ and $X > c$ (for simplicity we assume X is nominal) and splits the node into left and right daughters, $L = \{X_i \leq c\}$ and $R = \{X_i > c\}$, respectively. Let

$$t_1 < t_2 < \cdots < t_m$$

be the distinct death times and let $d_{j,L}, d_{j,R}$ and $Y_{j,L}, Y_{j,R}$ equal the number of deaths and individuals at risk at time t_j in daughter nodes L, R. At risk means the number of individuals in a daughter who are alive at time t_j, or who have an event (death) at time t_j:

$$Y_{j,L} = \#\{T_i \geq t_j,\ X_i \leq c\}, \qquad Y_{j,R} = \#\{T_i \geq t_j,\ X_i > c\}.$$

Define

$$Y_j = Y_{j,L} + Y_{j,R}, \qquad d_j = d_{j,L} + d_{j,R}.$$

The log-rank split-statistic value for the split is

$$L(X,c) = \frac{\displaystyle\sum_{j=1}^{m} \left(d_{j,L} - Y_{j,L} \frac{d_j}{Y_j} \right)}{\sqrt{\displaystyle\sum_{j=1}^{m} \frac{Y_{j,L}}{Y_j} \left(1 - \frac{Y_{j,L}}{Y_j}\right) \left(\frac{Y_j - d_j}{Y_j - 1}\right) d_j}}.$$

The value $|L(X,c)|$ is a measure of node separation. The larger the value, the greater the survival difference between L and R, and the better the split is. The best split is determined by finding the feature X^* and split-value c^* such that $|L(X^*,c^*)| \geq |L(X,c)|$ for all X and c.

Log-rank score splitting

The package also implements splitting using the log-rank score test (Hothorn and Lausen, 2003). This is specified by the option `splitrule="logrankscore"`. To describe this rule, assume the variable X has been ordered so that $X_1 \leq X_2 \leq \cdots \leq X_n$ where for simplicity we assume there are n unique values for X (no ties). Now compute the "ranks" for each survival time T_j,

$$a_j = \delta_j - \sum_{k=1}^{\Gamma_j} \frac{\delta_k}{n - \Gamma_k + 1}$$

where $\Gamma_k = \#\{t : T_t \leq T_k\}$. The log-rank score test is defined as

$$S(x,c) = \frac{\sum_{X_j \leq c} a_j - n_L \bar{a}}{\sqrt{n_L \left(1 - \frac{n_L}{n}\right) s_a^2}}$$

where $\bar{a}$ and s_a^2 are the sample mean and sample variance of $\{a_j : l = 1, \ldots, n\}$ and $n = n_L + n_R$ where n_L is the sample size of the left daughter node. Log-rank score splitting defines the measure of node separation by $|S(X, c)|$. Maximizing this value over X and c yields the best split.

Randomized splitting

All models in the package including RSF allow the use of randomized splitting specified by the option `nsplit`. Rather than splitting the node by considering all possible split-values for a variable, instead a fixed number of randomly selected split-points $c_1, \ldots, c_{\texttt{nsplit}}$ are chosen (Ishwaran et al., 2008, 2010; Ishwaran, 2015). For example, the best randomized split using log-rank splitting is the maximal value of

$$|L(X, c_1)|, \ldots, |L(X, c_{\texttt{nsplit}})|.$$

For each variable X, this reduces n split-statistic evaluations (worst case scenario) to `nsplit` evaluations. Not only does randomized splitting greatly reduce computations, it also mitigates the well known tree bias of favoring splits on variables with a large number of split-points, such as continuous variables or factors with a large number of categorical labels (Loh and Shih, 1997). Related work includes Geurts et al. (2006) who investigated extremely randomized trees. Here a single random split-point is chosen for each variable (i.e., `nsplit` $= 1$). Traditional deterministic splitting (all split values considered) is specified by `nsplit = 0`.

Terminal node statistics (TNS)

RSF estimates the survival function, $S(t|\mathbf{X}) = \mathbb{P}\{T^o > t|\mathbf{X}\}$, and the cumulative hazard function (CHF),

$$H(t|\mathbf{X}) = \int_{(0,t]} \frac{F(du|\mathbf{X})}{S(u|\mathbf{X})}, \qquad F(u|\mathbf{X}) = \mathbb{P}\{T^o \leq u|\mathbf{X}\}.$$

Below we describe how these two quantities are estimated using a survival tree.

In-bag (IB) estimator

Once the survival tree is grown, the ends of the tree are called the terminal nodes. The survival tree predictor is defined in terms of the predictor within each terminal nodel. Let h be a terminal node of the tree and let

$$t_{1,h} < t_{2,h} < \cdots < t_{m(h),h}$$

be the unique death times in h and let $d_{j,h}$ and $Y_{j,h}$ equal the number of deaths and individuals at risk at time $t_{j,h}$. The CHF and survival functions for h are estimated using the bootstrapped Nelson-Aalen and Kaplan-Meier estimators

$$H_h(t) = \sum_{t_{j,h} \leq t} \frac{d_{j,h}}{Y_{j,h}}, \qquad S_h(t) = \prod_{t_{j,h} \leq t} \left(1 - \frac{d_{j,h}}{Y_{j,h}}\right).$$

The survival tree predictor is defined by assigning all cases within h the same CHF and survival estimate. This is because the purpose of the survival tree is to partition the data into homogeneous groups (i.e., terminal nodes) of individuals with similar survival behavior. To estimate $H(t|\mathbf{X})$ and $S(t|\mathbf{X})$ for a given feature $\mathbf{X}$, drop $\mathbf{X}$ down the tree. Because of the binary nature of a tree, $\mathbf{X}$ will fall into a unique terminal node h. The CHF and survival estimator for $\mathbf{X}$ equals the Nelson-Aalen and Kaplan-Meier estimator for $\mathbf{X}$'s terminal node:

$$H^{\text{IB}}(t|\mathbf{X}) = H_h(t), \quad S^{\text{IB}}(t|\mathbf{X}) = S_h(t), \text{ if } \mathbf{X} \in h.$$

Note that we use the notation IB because the above estimators are based on the training data, which is the IB data.

Out-of-bag (OOB) estimators

To define the OOB estimator, let $I_i \in \{0, 1\}$ indicate whether case i is IB or OOB. Let $I_i = 1$ if and only if i is OOB. Drop i down the tree and let h denote i's terminal node. The OOB tree estimators for i is

$$H^{\text{OOB}}(t|\mathbf{X}_i) = H_h(t), \quad S^{\text{OOB}}(t|\mathbf{X}_i) = S_h(t), \text{ if } \mathbf{X}_i \in h \text{ and } I_i = 1.$$

Observe these are NULL unless i is OOB for the tree.

Ensemble CHF and Survival Function

The ensemble CHF and survival function are determined by averaging the tree estimator. Let $H_b^{\text{IB}}(t|\mathbf{X})$ and $S_b^{\text{IB}}(t|\mathbf{X})$ be the IB CHF and survival estimator for the bth survival tree. The IB ensemble estimators are

$$\bar{H}^{\text{IB}}(t|\mathbf{X}) = \frac{1}{\texttt{ntree}} \sum_{b=1}^{\texttt{ntree}} H_b(t|\mathbf{X}), \quad \bar{S}^{\text{IB}}(t|\mathbf{X}) = \frac{1}{\texttt{ntree}} \sum_{b=1}^{\texttt{ntree}} S_b(t|\mathbf{X}).$$

Let O_i record trees where case i is OOB. The OOB ensemble estimators for individual i are

$$\bar{H}_i^{\text{OOB}}(t) = \frac{1}{|O_i|} \sum_{b \in O_i} H_b^{\text{IB}}(t|\mathbf{X}_i), \quad \bar{S}_i^{\text{OOB}}(t) = \frac{1}{|O_i|} \sum_{b \in O_i} S_b^{\text{IB}}(t|\mathbf{X}_i), \qquad i = 1, \ldots, n.$$

An important distinction between the two sets of estimators is that OOB estimators are used for inference on the training data and for estimating prediction error and only apply to the features $\mathbf{X} = \mathbf{X}_i$ in the training data. In-bag estimators on the other hand are used for prediction and can be used for any feature $\mathbf{X}$.

Why are two estimates provided?

Why does RSF estimate both the CHF and the survival function? Classically, we know the two are related by

$$H(t) = -\log(S(t)).$$

The problem is that even if it were true that $H_b = -\log(S_b)$ for every tree $b = 1, \ldots, \texttt{ntree}$, the above identity will not hold for the ensemble. Let $\bar{S}$ and $\bar{H}$ denote the ensemble survival

and CHF and let $\mathbb{E}_b$ denote ensemble expectation (i.e., $\bar{S} = \mathbb{E}_b[S_b]$ and $\bar{H} = \mathbb{E}_b[H_b]$). Then by Jensen's inequality for convex functions,

$$-\log(\bar{S}) = -\log(\mathbb{E}_b[S_b]) \leq \mathbb{E}_b[-\log(S_b)] = \mathbb{E}_b[H_b] = \bar{H}.$$

In other words, $-\log$ of the survival ensemble does not necessarily equal the ensemble of $-\log$ of the tree survival function. The inequality above also shows that taking $-\log$ of the ensemble survival will generally be smaller than the true ensemble CHF. This is why RSF provides the two values separately.

Prediction error

Prediction error for survival models is measured by $1-$C, where C is Harrell's concordance index (Harrell Jr et al., 1982). Prediction error is between 0 and 1, and measures how well the predictor correctly ranks two random individuals in terms of survival. Unlike other measures of survival performance, Harrell's C-index does not depend on choosing a fixed time for evaluation of the model and specifically takes into account censoring of individuals. The method is a popular means for assessing prediction performance in survival settings since it is easy to understand and interpret.

Mortality (the predicted value used by RSF)

To compute the concordance index we must define what constitutes a worse predicted outcome. For survival models this is defined by the concept of *mortality* which is the predicted value used by RSF. Let $t_1 < \cdots < t_m$ denote the entire set of unique event times for the learning data. The IB ensemble mortality for a feature $\mathbf{X}$ is defined as

$$\bar{M}^{\text{IB}}(\mathbf{X}) = \sum_{j=1}^{m} \bar{H}^{\text{IB}}(t_j|\mathbf{X}).$$

This estimates the number of deaths expected if all cases were similar to $\mathbf{X}$. OOB ensemble mortality which is used for the C-index calculation is defined by

$$\bar{M}_i^{\text{OOB}} = \sum_{j=1}^{m} \bar{H}_i^{\text{OOB}}(t_j), \qquad i = 1, \ldots, n.$$

Individual i is said to have a worse outcome than individual j if

$$\bar{M}_i^{\text{OOB}} > \bar{M}_j^{\text{OOB}}.$$

C-index calculation

The C-index is calculated using the following steps:

PUT YOUR CONCORDANCE ALGORITHM HERE: REPLACE YOUR M_i with $\bar{M}_i^{\text{OOB}}$

Brier score

PUT YOUR BRIER SCORE SECTION HERE

Illustration

To illustrate, we use the survival data from [Hsich] consisting of 2231 adult patients with systolic heart failure. All patients underwent cardiopulmonary stress testing. During a mean follow-up of 5 years (maximum for survivors, 11 years), 742 patients died. The outcome is all-cause mortality and a total of $p = 39$ covariates were measured for each patient including demographic, cardiac and noncardiac comorbidity, and stress testing information.

PUT YOUR ILLUSTRATION/FIGURE 1 HERE

PUT YOUR BRIER SCORE/FIGURE 2 HERE. GIVE SOME EXPLANATIONS

Variable Importance

PUT YOUR VARIABLE IMPORTANCE SECTION HERE

Partial Plots

Another useful tool for interpreting the results from a RSF analysis is the partial dependence plot (Friedman, 2001). Figure [link] displays the partial dependence plot for the most important variable, peak VO_2, from our previous analysis (i.e., the variable with the largest VIMP from Figure [link]. The figure displays 5 year OOB survival as a function of peak VO_2 and in particular we see that survival depends strongly on this value and increases with increasing peak VO_2 capacity.

An important feature of the partial dependence plot is that it displays the dependence of survival on the target variable while adjusting for all other variables. This is accomplished by integrating out the effect of the other variables. Specifically, let $\bar{S}_X^{\text{OOB}}(t|X_i = x)$ be the OOB ensemble survival function for case i where X represents peak VO_2 and the observed peak VO_2 value X_i is replaced by some prechosen value x. In other words, patient i's variables are set to their observed values except peak VO_2 which is fixed at x. The OOB partial predicted survival function for peak VO_2 at x equals

$$\bar{S}_X^{\text{OOB}}(t|x) = \frac{1}{n}\sum_{i=1}^{n} \bar{S}_X^{\text{OOB}}(t|X_i = x).$$

The value $\bar{S}_X^{\text{OOB}}(t|x)$ is what is displayed on the vertical axis of Figure [link] for $t = 5$ years as x is varied.

PUT YOUR PARTIAL PLOT EXAMPLE HERE

Partial dependence plots can be defined for more than one variable. For example, if the target variables are $X^{(k)}$ and $X^{(l)}$, the OOB partial predicted survival function at $X^{(k)} = a$ and $X^{(l)} = b$ equals

$$\bar{S}^{\text{OOB}}_{X^{(k)},X^{(l)}}(t|a,b) = \frac{1}{n}\sum_{i=1}^{n} \bar{S}^{\text{OOB}}_{X^{(k)},X^{(l)}}(t|X_i^{(k)} = a, X_i^{(l)} = b).$$

Figure [link] displays the partial dependence plot of peak VO_2 and BUN which are the top two variables from our previous analysis. The contour plot shows how 5 year OOB survival depends jointly on these two variables. In particular, low peak VO_2 combined with high

BUN yields poor survival, wheras high peak VO_2 combined with low BUN yields improved survival.

PUT YOUR P-PLOTS HERE

Competing Risks

Here we briefly outline the extension of RSF to competing risks developed in Ishwaran et al. (2014). In competing risks, unlike survival where there is only one event type, the individual is subject to $J > 1$ competing risks. As in survival data, a complication is that the individual can be right-censored. Formally, let T^o be the true event time and let $\delta^o \in \{1, \ldots, J\}$ record the event type. Let C^o denote the true censoring time. Under the presence of right-censoring we only observe $T = \min(T^o, C^o)$ and the censoring indicator $\delta = \delta^o \cdot I\{T^o \leq C^o\}$. Thus for each individual one either observes the time an event occurs $T = T^o$ and the type of event which occured $\delta = \delta^o \in \{1, \ldots, J\}$. Otherwise if the individual is right-censored, we observe the censoring time $T = C^o$ and the censoring indicator is $\delta = 0$.

Competing risk splitting rules

There are three splitting rules used by the package to grow a competing risk tree:

(1) Generalized log-rank test, specified by `splitrule = "logrank"`. This tests for equality of the event-specific hazard functions and is most appropriate when the analysis focuses on determining risk factors for an event-specific hazard. The generalized log-rank test is based on the weighted difference of the Nelson-Aalen event-specific CHF estimates in the daughter nodes.

(2) Gray's test, specified by `splitrule = "logrankCR"`. This is a modification of Gray's test (Gray, 1988) and tests for the equality of the cause-specific cumulative incidence functions (CIF). This is most appropriate when the goal is long term probability prediction.

(3) Composite (weighted) splitting. This is specified using `cause` and is an integer value between 1 and J indicating the event of interest for splitting a node, where splitting is either based on the generalized log-rank test or Gray's test specified by `splitrule` as described above. If not specified, the default is to use a composite splitting rule that averages over all event types. Can also be a vector of non-negative weights of length J specifying weights for each event (for example, passing a vector of ones reverts to the default composite split-statistic).

Event-specific TNS

Let $(T_i, \mathbf{X}_i, \delta_i)_{1 \leq i \leq n}$ denote the data where T_i is the observed time, $\delta_i \in \{0, 1, \ldots, J\}$ is the observed censoring indicator, and $\mathbf{X}_i$ is the feature. Let $n_{i,b}$ be the number of times case i occurs in bootstrap sample of the bth tree. Let $h_b(\mathbf{X})$ be the terminal node of the bth tree containing $\mathbf{X}$. Denote IB node-specific event counts by

$$N_{j,b}^{\text{IB}}(t|\mathbf{X}) = \sum_{i \in h_b(\mathbf{X})} n_{i,b} I\{T_i \leq t, \delta_i = j\}, \quad j = 1, \ldots, J$$

and IB number at risk by

$$Y_b^{\text{IB}}(t|\mathbf{X}) = \sum_{i \in h_b(\mathbf{X})} n_{i,b} I\{T_i \geq t\}.$$

The IB tree estimator for the event-specific CIF, $F_j(t|\mathbf{X}) = \mathbb{P}\{T^o \leq t, \delta^o = j|\mathbf{X}\}$, is the Aalen-Johansen estimator (Aalen and Johansen, 1978):

$$F_{j,b}^{\text{IB}}(t|\mathbf{X}) = \int_{(0,t]} S_b^{\text{IB}}(u - |\mathbf{X}) Y_b^{\text{IB}}(u|\mathbf{X})^{-1} N_{j,b}^{\text{IB}}(du|\mathbf{X}), \quad j = 1, \ldots, J,$$

where $S_b^{\text{IB}}(t|\mathbf{X}) = \prod_{u \leq t} \left[1 - \sum_{j=1}^J N_{j,b}^{\text{IB}}(du|\mathbf{X}) / Y_b^{\text{IB}}(u|\mathbf{X})\right]$ is the Kaplan-Meier estimate of event-free survival.

Event-specific ensembles

Tree-averaging $F_{j,b}^{\text{IB}}(t|\mathbf{X})$ yields the IB ensemble estimate for the event-specific CIF

$$\bar{F}_j^{\text{IB}}(t|\mathbf{X}) = \frac{1}{\texttt{ntree}} \sum_{b=1}^{\texttt{ntree}} F_{j,b}^{\text{IB}}(t|\mathbf{X}), \quad j = 1, \ldots, J.$$

Let O_i record trees where case i is OOB. The OOB ensemble estimator for case i is

$$\bar{F}_{i,j}^{\text{OOB}}(t) = \frac{1}{|O_i|} \sum_{b \in O_i} F_{j,b}^{\text{IB}}(t|\mathbf{X}_i), \quad j = 1, \ldots, J.$$

Expected number of life years lost (cause-j mortality)

The predicted value for RSF competing risks is the one-dimensional summary of the cumulative incidence refered to as the expected number of life years lost due to cause j (Andersen, 2013). In right-censored data it is not feasible to get a reliable estimate of the expected lifetime. Therefore for a fixed time point τ we consider the restricted mean lifetime conditional on $\mathbf{X}$

$$R_{\text{MLT}}(\mathbf{X}) = \int_0^\tau S(t|\mathbf{X})\, dt.$$

In practice, the package sets τ to the maximum observed follow-up time, t_m. We extend the notation ofAndersen (2013) to the case with covariates and note the relation $S(t|\mathbf{X}) + \sum_{j=1}^J F_j(t|\mathbf{X}) = 1$. The expected number of years lost before time τ is

$$\tau - R_{\text{MLT}}(\mathbf{X}) = \tau - \int_0^\tau S(t|\mathbf{X})\, dt = \int_0^\tau \sum_{j=1}^J F_j(t|\mathbf{X})\, dt.$$

Our summary value is $M_j(\tau|\mathbf{X}) = \int_0^\tau F_j(t|\mathbf{X})\, dt$, which the above shows equals the expected number of life years lost due to cause j before time τ. We shall also call $M_j(\tau|\mathbf{X})$ the cause-j mortality. This value is estimated by replacing F_j with the IB ensemble estimator $\bar{F}_j^{\text{IB}}(t|\mathbf{X})$. For the purposes of prediction error estimation it is replaced with the OOB estimator, yielding OOB expected number of life years lost for an individual i

$$\bar{M}_{i,j}^{\text{OOB}}(\tau) = \int_0^\tau \bar{F}_{i,j}^{\text{OOB})}(t)\, dt, \quad j = 1, \ldots, J.$$

Illustration

To illustrate, we use the follicular cell lymphoma data fromPintilie (2006). The subset of 541 patients includes all patients identified as having follicular type lymphoma. Patients were treated with radiation alone or with radiation and chemotherapy. The two types of events are relapse and death. Figure [link] displays the averaged OOB ensemble CIF for the two events from a RSF competing risk analysis using the composite Gray splitting rule. Figure [link] displays VIMP where prediction error was measured using the truncated Harrell's C-index (Ishwaran et al., 2014). Here VIMP was calculated using the generalized log-rank splitting rule, in which separate RSF analyses were run for each event type. This type of analysis is most appropriate where the goal is to identify risk factors specific to an event. We see that age of the individual is highly predictive of death but not the competing risk of relapse.

PUT CR figure from our StatsRef paper here

References

O. O. Aalen and S. Johansen. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, pages 141–150, 1978.

P. K. Andersen. Decomposition of number of life years lost according to causes of death. *Statistics in Medicine*, 32(30):5278–5285, 2013.

A. Ciampi, S. A. Hogg, S. McKinney, and J. Thiffault. RECPAM: a computer program for recursive partition and amalgamation for censored survival data. *Comp. Methods Programs Biomed.*, 26(3):239–256, 1988.

J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63 (1):3–42, 2006.

R. J. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, pages 1141–1154, 1988.

F. E. Harrell Jr, R. M. Califf, D. B. Pryor, K. L. Lee, R. A. Rosati, et al. Evaluating the yield of medical tests. *JAMA*, 247(18):2543–2546, 1982.

T. Hothorn and B. Lausen. On the exact distribution of maximally selected rank statistics. *Computational Statistics & Data Analysis*, 43(2):121–137, 2003.

H. Ishwaran. The effect of splitting on random forests. *Machine Learning*, 99(1):75–118, 2015.

H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860, 2008.

H. Ishwaran, U. B. Kogalur, E. Z. Gorodeski, A. J. Minn, and M. S. Lauer. High-dimensional variable selection for survival data. *Journal of the American Statistical Association*, 105(489):205–217, 2010.

H. Ishwaran, T. A. Gerds, U. B. Kogalur, R. D. Moore, S. J. Gange, and B. M. Lau. Random survival forests for competing risks. *Biostatistics*, 15(4):757–773, 2014.

M. LeBlanc and J. Crowley. Relative risk trees for censored survival data. *Biometrics*, pages 411–425, 1992.

M. LeBlanc and J. Crowley. Survival trees by goodness of split. *Journal of the American Statistical Association*, 88(422):457–467, 1993.

W.-Y. Loh and Y.-S. Shih. Split selection methods for classification trees. *Statistica Sinica*, pages 815–840, 1997.

M. Pintilie. *Competing risks: a practical perspective*, volume 58. John Wiley & Sons, 2006.

M. R. Segal. Regression trees for censored data. *Biometrics*, pages 35–47, 1988.

M. R. Segal. Extending the elements of tree-structured regression. *Statistical Methods in Medical Research*, 4(3):219–236, 1995.