Prediction the Survivability of Male Breast Cancer Respondents in Detroit Michigan Area by Machine Learning Methods and Compare with Bayesian Cox Proportional Regression Model

by

Roungu Ahmmad

PhD Candidate of Biostatistics and DataScience

Department of Data Science

John D Bower School of Population Health

The University Mississippi Medical Center MS 39216 USA

April 30, 2021

Abstract

Background: The death rate in Michigan is high compare with all the state in USA. Especially Detroit is one of the area where the death rate highest among male and female in Michigan. Several factors affecting this death rate but we are considering the breast cancer outcomes among male respondents. This is very rear event but the risk of death is very much considerable and take into account in the study. The prediction of survivability of breast cancer male is very complex manner because of diagnosis systems and a lot of factors affecting this event. Surveillance, Epidemiology, and End Results Program (SEER) is a large institute to keep the record of this high volume of respondents and all records of breast cancer but we have very few numbers of respondents of male breast cancer events.

Methods: Machine learning methods and Bayesian Cox proportional methods is applied for prediction the survivability of male breast cancer patients. Random survival methods is applied for conditionally prediction the survivability of male breast cancer patients and Bayesian Cox model is applied for the same modeling structure and finally AUC, prediction errors, Brier score and OOB survival rate score applied for comparison which methods provided better prediction.

Results: Machine learning methods proposed that age is the most potential factors for death of male breast cancer patients. Primary tumor size is less than 21mm, then the median survival time is higher compare with larger tumor size and primary tumor diameter is an another important features that, with increasing of diameter, the change of survivability decreasing and with increasing distance of tumor from prime site, the median survival time decreasing. For Cox PH model

and Bayesian Cox PH model provided approximately same outcomes but some predictors in Cox PH model shown insignificant. By the Bayesian Cox model, for single year increase of age, approximately 5% chance increase of death (HR: 0.05, SD = 0.006). Similarly, for T1-4a stage cancer patients have approximately 30% lower chance of death compare with T0-X stages. Total tumor around prime site is negatively related with survival probability of male breast cancer patients. Another important things is, White breast cancer patients have lower rate of death compare with black breast cancer patients. For Weibull, Log-logistic and log-normal survival prior, we get approximately same outcomes but Weibull prior fit best for this data set considering log-PMLL, DIC, WAIC and Posterior inference frailty variance. Considering all the comparison of OOB survival, OOB Brier score, prediction error, AUC, machine learning methods provided better prediction compare with Bayesian Cox Proportional model and Cox PH regression model.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background of Study

Cancer is a disease in which cells in the body grow out of control and when some cell in breast is grown abnormally, it causes of breast cancer. So breast cancer can be begin in different parts of the breast among male and female. Different kinds of breast cancer found depend on which cell in the breast turn into cancer symptoms. According to CDC report except skin cancer, breast cancer is the most common cancer for women in the United States. According to Kopans (2008) over the last decade, the rate of getting breast cancer has not changed for overall women in US, but the rate has increased dramatically in Detroit, Michigan area among different race and age groups. American cancer society reported each year in the United States, about 245 thousand cases of breast cancer are diagnosed and out this around 41 hundred women die each year because of breast cancer but very few recode found in case of male respondents. Some researchers mentioned that, breast cancer of male is a part of death rate in USA.

A large number of literature have been shown that the risk factors for breast cancer is a combination of factors but the main factors that influence the risk include being getting older. Most breast cancers are found in 50 years or older. Some other researches shown that other factors such as smoking, being exposed to chemicals that can cause cancer, and changes in other hormones due to night shift working also may increase the risk of breast cancer. According the report of Detroit Atla, about 11% of all new cases of breast cancer in the United States are found in younger ages but very few researches founded on male breast cancer events and consequence predictors. In this study we will consider more than 30 potential factors that affecting the breast cancer and will find more important factors for the events and predict the survivability of them with machine learning methods and Bayesian cox regression model as well as comparison among them.

1.2 Methods

Most of the studies on breast cancer used statistical survival methods. A key aspect of survival analysis is the presence of censored data, indicating that the event of interest has not occurred during the study period. The presence of censored data requires the use of specialised techniques. Traditionally, the Cox proportional hazards model has been the most widely used technique for analysing censored data, but the Cox model was designed for small data sets and does not scale well to high dimensions. Machine learning techniques that inherently handle high-dimensional data have been adapted to handle censored data, allowing machine learning to offer more flexible alternatives for analysing high-dimensional, right-censored, heterogeneous data.

When the data is high-dimensional, usually, cox or any statistically methods for survival analysis are not educated for implementation. According to James et al. (2013) some methods of analysis become infeasible as the number of coefficients to be estimated exceeds the number of observations from which to estimate them, and so a unique solution cannot be found. Rathore et al. (2017) shown that different sources of clinical data can provide complementary information about the event and that the integration of multiple sources of data leads to better prediction of cognitive decline than the use of a single source. However, both models have considerable limitations. Although both models have been validated with large cohort data, their discriminatory ability, area under the ROC (receiver operating characteristics) curve. Both models make implicit assumptions that risk factors relate to cancer development in a linear way and are mostly independent from other risk factors. Thus, both models likely oversimplify complex relationships and non-linear interactions in numerous risk factors.

2 Literature Review

2.1 Machine Learning Review

Several literature found on machine learning approaches usually used in breast cancer detection. Very recently Jain and Kumar (2020), applied machine learning approaches for prediction the breast cancer and founded 97.89% accurately predict the breast cancer. Agarap (2018) used Machine learning method on Wisconsin Diagnostic dataset for detection of breast cancer and predict breast cancer events. Some advanced machine leaning method such as support vector network and artificial neural network methods used for cancer diagnosis and prognosis (Sharma et al. (2017) , Amrane et al.

(2018) and Sharma et al. (2018)). Gayathri and Sumathi (2016) provided a comparative study of relevance vector machine with various machine learning methods for detection of breast cancer. Jerez et al. (2010) used multi-layer perception techniques, self-organization maps methods and K-nearest neighbour methods for predicting the missing data of real breast cancer events. According to Obermeyer and Emanuel (2016) machine learning methods offers an alternative approach to standard prediction modeling that may address current limitations of survival methods and improve the accuracy of breast cancer prediction tools. Vanneschi et al. (2011) mentioned that this methods has been used in models related to cancer prognosis and survival and produced better accuracy and reliability estimates. Still today, very few studies applied machine methods for personalized breast cancer survival prediction or compared the predictive accuracy and reliability with models commonly used in clinic practice, but Heidari et al. (2018) mentioned that, machine learning approaches provided more accurate prediction on short term breast cancer risk.

2.2 Variables Importance

Variable importance evaluation can be separated into two groups, those that use the model information and those that do not. The advantage of using a model-based approach is that is more closely tied to the model performance and that it may be able to incorporate the correlation structure between the predictors into the importance calculation. Regardless of how the importance is calculated for most classification models, each predictor will have a separate variable importance for each class (the exceptions are classification trees, bagged trees and boosted trees). All measures of importance are scaled to have a maximum value of 100, unless the scale argument. In random survival forest for each tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The difference between the two accuracy's are then averaged over all trees, and normalized by the standard error. For regression, the MSE is computed on the out-of-bag data for each tree, and then the same computed after permuting a variable. The differences are averaged and normalized by the standard error. If the standard error is equal to 0 for a variable, the division is not done.

There are several approaches have for measuring the importance score. Gini score is more popular for measuring the importance which give the means Gini score produced by overall trees. Gini Permutation importance measure the mean decrease in classification accuracy after permuting over all trees. Let t be the tree size then the Variable importance score will be

$$VI^{(t)}(x_j) = \frac{\sum_{i\epsilon\hat{\sigma}^{(t)}} I\left(y_i = \hat{y}_i^{(t)}\right)}{|\hat{\sigma}^{(t)}|} - \frac{\sum_{i\epsilon\hat{\sigma}^{(t)}} I\left(y_i = \hat{y}_{i,\pi_j}^{(t)}\right)}{|\hat{\sigma}^{(t)}|} \tag{1}$$

Where, $\hat{y}_i^{(}t) = f^t(x_i)$ = predicted class before permuting

$\hat{y}_{i,\pi_j}^{(t)} = f^t(x_{i,\pi_j})$ = predicted class after permuting of X_j

And $X_{i,\pi_j} = (x_{i,1}, x_{i,2}, \ldots, x_{i,j-1}, X_{\pi_j(i),j}, \ldots, X_{\pi_j(p),j}$

Note that, $VI^{(t)}(X_j) = 0$ by definition if X_j is not in the tree t.

For all over tree the raw importance is

$$VI^{(t)}(x_j) = \frac{\sum_{t=1}^{ntree} VI^{(t)}(X_j)}{ntree} \tag{2}$$

To give a better intuition, features that are selected at the top of the trees are in general more important than features that are selected at the end nodes of the trees, as generally the top splits lead to bigger information gains. In this project we are using Gini impurity score for selecting the significant important variables for random survival tree. For selecting the significant criterion, we are applying level of significance equal 5% and Jackknife criterion.

2.3 Machine Learning Model

The usual analyses for survival rely on parametric or semi-parametric methods, for which non-linear and interaction effects need to be explicitly modelled. Classification and regression trees (CART) have proved a valuable resource in modelling complex and non-linear effects for categorical and continuous outcomes, which has motivated a range of attempts to adapt tree methodology to right-censored survival outcomes. Ishwaran et al. (2008) proposed a method for analysis survival data that very helpful for fitting and implementation. Any tree method relies on the recursive binary partitioning of a given features space into increasingly smaller regions, containing observations with similar responses, until a certain stopping criterion is reached. The regions are referred to as nodes, and the final regions created on termination of the growth of a tree are known as terminal nodes, or leafs, while the initial node is commonly referred to as the root node. There are two necessary features to any tree algorithm: a node splitting rule, which informs how a partition is best split and a stopping rule, according to Zhou and McArdle (2015), which provides a criterion for termination the growth of a tree . For a given feature space with x possible variables, and c possible split values, variable x* and split point c* are chosen in a way that maximizes the difference between the two daughter nodes.

While a tree has great potential for modelling non-linear and interactive effects, as will be detailed later, the very way a tree is grown makes it highly variable and very unstable. Each split of a node is dependent on previous partitioning [Strobl et al. (2009)]. Consider a situation in which several samples are taken from a larger dataset and used to grow individual trees. Even a slight variation in the data may result in a selection of a different variable x* or splitting point c* for one of the early nodes, or even the root node, giving rise to vastly different trees. The sensitivity of a single tree to minor training data variations is likely to result in poor generalization to new data. Breiman (1996) mentioned that, a way to combat the over fitting reduce the variability - is by introducing a measure of randomness in the way of bootstrapping. In this procedure individual trees are grown for multiple bootstrap samples drawn from the data, and subsequently aggregated over, producing a single ensemble tree. This procedure of bootstrap aggregation, is commonly known as bagging.

Random forests and also,perhaps to a lesser extent, bagged forests are capable of modelling complex associations, even in the situation of a high number of predictors and have been successfully applied in the context of genetic and epidemiological data. Unlike single trees, forests are a both stable and powerful tool for data analysis. A single tree, however, can easily be interpreted, but an averaged ensemble decidedly less so. As such, while forests resolve variability, this comes at the cost of interpret-ability. Various measures have been proposed to quantify variable contribution, with the most common method relying on in-variable permutation or random node assignment.

The survival difference can be quantified using the log-rank statistic or the log-rank score statistic. The log-rank test has been shown to be a valid test for splitting survival trees given both proportional and non-proportional hazards. For a split on a continuous predictor defined as $x < c$ and $x > c$ let $t_1, \ldots, t_m$ be the event times in the parent node h. Let $d_{k,l}$ and $y_{k,l}$ be the number of events and subjects at risk at time k in the left daughter node, respectively, and let $d_{k,r}$ and $Y_{k,r}$ be the same for the right daughter node, then $Y_{k,l}$ will be all the subject i at risk time k with covariates values $x_i \leq c$

and $Y_{k,r}$ the same for subjects with $x_i \geq c$.

$$\begin{aligned} Y_{k,l} &= \#T_i \geq t_k, x_i \leq c \\ Y_{k,r} &= \#T_i \geq t_k, x_i \geq c \end{aligned} \tag{3}$$

Let Y_k and d_k be the total number of subjects at risk time k, and total event at time k, then the log-rank statistic for the split point on the variable x is

$$L(c,x) = \frac{\sum_{k=1}^{m} \left(d_{k,l} - Y_{k,l} \frac{d_k}{Y_K} \right)}{\sqrt{\sum_{k=1}^{m} \frac{Y_{k,l}}{Y_k} \left(1 - \frac{Y_{k,l}}{Y_k}\right) \left(\frac{Y_k - d_k}{Y_k - 1}\right) d_k}} \tag{4}$$

The large the value of $|L(c,x)|$ the greater difference between the survival curves and the greater the node separation. The objective is to find a best variable x* wiht an optimal split c* such that $|L(c*, x8)| \geq |L(c,x)|$ for all variables x and split point c. An alternative splitting rule is given by the log-rank score statistic, which differs from the log-rank statistic in the assumption that the variable x is ordered. This approach was first described by Hothorn and Lausen in 2003. Ishwaran provides a concise formal notation in line with the one given here for the log-rank statistic. Additionally, Ishwaran describes a log-rank based randomized approach for splitting, where a single random split point c is selected for each variable x, the log-rank statistic computed, and the variable and split point with the largest statistic selected for splitting. The particular splitting rule used, the number of variables considered at each split along with the number of split points must all be considered when growing a tree.

The survival time, and censoring status for individuals $i = 1, \dots, n$ can then be written as $(T_{1,h}, \delta_{1,h}), \dots, (T_{n(h),h}, \delta_{n(h),h})$ where, $\delta_{i,h} = 0$ denotes a subject right censored at time $T_{i,h}$ otherwise the event occurs. Define the n(h) distinct event times as $t_{1,h}, t_{2,h}, \dots, t_{n(h),h}$. At time, $t_{1,h}$, $d_{1,h}$ is the number of death and $Y_{1,h}$ is the subjects at risk, then, Cumulative hazard function (CHF) for a terminal event can be estimated as

$$\hat{H}_h(t) = \sum_{t_{l,h} \leq t}^{n} \frac{d_{l,h}}{Y_{l,h}} \tag{5}$$

So, At time, $t_{1,h}$, $d_{1,h}$ is the number of death and $Y_{1,h}$ is the subjects at risk, then, survival function for a terminal event can be estimated as

$$S_h(t) = e^{-\hat{H}_h(t)} \tag{6}$$

To obtain the ensemble cumulative hazard function, the average over the survival trees is taken. Defining the hazard for a tree grown from a bootstrap sample b as $H_b^*(t|x)$, then the bootstrap ensemble $H_e^*(t|x_i)$ is given by

$$H_e^*(t|x_i) = \frac{1}{B} \sum_{b=1}^{B} H_b^*(t|x) \tag{7}$$

Where, B is the number of survival trees. For a subject i with covariates vectors x_i all survival trees are used. This approach gives a CHF estimate which, given a large enough number of trees, is comparable to one that would be obtained with a leave-one-out cross-validation. Such a CHF estimate for a subject i is valid, as it has been obtained from prediction using only those trees in which i was not included as an observation. Then, the indicators, $I_{i,b}$ is used to select the correct trees and will equal to 1 if the observation in OOB and 0 if it lies in-bag. Then the survival probability will be

$$S_e(t) = exp\left(-(H_e^{**}(t|x_i)\right) = exp\left\{-\frac{\sum_{b=1}^{B} I_{i,b} H_b^{*}(t|x)}{\sum_{b=1}^{B} I_{i,b}}\right\} \tag{8}$$

In the final step of the algorithm OOB data is used to calculate the prediction error for the ensemble CHF. The concordance index for the Cox proportional hazards model is obtained using the prognostic index and observed survival times and status. In a random survival forest context, the prognostic index is replaced by a predicted outcome: the ensemble mortality. Note that while OOB data can be used to obtain the OOB ensemble mortality and by extension the OOB prediction error, a valid prediction error can also be obtained in a cross-validation procedure. When choosing the latter option, the ensemble mortality scores for the subjects of each test set are obtained by dropping test observations down the trees of a survival random forest grown from the training set, and calculating their ensemble CHF and from the ensemble CHF the ensemble mortality. Ensemble mortality is based on the conservation of events principle. The conservation of events for a given terminal node $h\epsilon T$ in a given tree can be written as

$$\sum_{i=1}^{n(h)} \hat{H}_h(T_{i,h}) = \sum_{i=1}^{n(h)} \delta_{i,h} \tag{9}$$

For each terminal nodes $h\epsilon T$, which shows that the total number of deaths is conserved within h. Summing the estimated CHF over all observed survival times over all terminal nodes h amounts to the total number of deaths

$$\sum_{i=1}^{n} H(T_i|x_i) = \sum_{h\epsilon T} \sum_{i=1}^{n(h)} \hat{H}_h(T_{i,h}) = \sum_{h\epsilon T} \sum_{i=1}^{n(h)} \delta_{i,h} = \sum_{i=1}^{n} \delta_i \tag{10}$$

Then, mortality for a given individual i is defined as the expected value of the CHF, summed over T_i, conditioned on x_i, which is the number of death that would be expected under the conditional of the survival behaviors

$$M_i = E_i \sum_{j=1}^{J} H(T_j|x_j) = E_i \sum_{j=1}^{J} \sum_{h\epsilon T} \sum_{i=1}^{n(h)} \hat{H}_h(T_{i,h}) \tag{11}$$

In a survival tree this null hypothesis is naturally enforced in the terminal nodes, as the growth of a tree ensures that all individuals in a terminal have the same CHF. Then, the ensemble mortality can be defined as

$$\hat{M}^*_{e,i} = \sum_{j=1}^{J} H^*_e(T_j|x_i) = \sum_{j=1}^{J} \sum_{h\epsilon T} \sum_{i=1}^{n(h)} \hat{H}_h(T_{i,h}) \tag{12}$$

In the random survival forest model, usually predict the median survival times and predict the survival probability for the observed survival time T.

For classification criterion, the model predict the survival probability each treat consider the CHF of the each criterion and predict the mortality or survival probability each trees.

3 Bayesian Cox Regression Model

3.1 Model Formulation

The Cox proportional hazards (PH) model is ubiquitous in Statistics. However, it often fails to fit, at least without some form of tinkering. For example, when a treatment effect is negative at the beginning of a study and positive by the end, namely when treatment interacts with time, the PH assumption fails. It is then necessary to either fit an interaction with time or to stratify on treatment. Other semiparametric models such as the accelerated failure time (AFT) model or the proportional odds (PO) model will also fail under such circumstances since they also constrain in such a way that survival curves are not allowed to cross. Sometime we have a lot of predictors that affecting the outcomes and we can not use cox model for educated inferencing and interpretation because of lot of interaction effect to consider.

In what follows, Bayesian non parametric approach to survival regression that allows curves to cross, or not. Full inferences regarding the treatment effect, straightforward or not, are easily obtained using a Markov chain Monte Carlo implementation of the Bayesian model. Bayesian non parametric and semiparametric models in survival analysis have become popular recently due to the advances in computing technology and the development of efficient computational algorithms. The Dirichlet process (DP) is probably the most frequently used tool in Bayesian non parametric inference. Some common parametric regression model consider the prior information of the estimators and apply MCMC methods for estimating the more efficient and relevant inferences.

Suppose that a sample of n individuals have possible censored survival times $Y_1 \leq, \ldots, \leq Y_n$. Let $\delta_i = 1$ if the ith subject are death or event occurs and 0 otherwise. We assume d-dimensional predictors X_i for n individuals and the triple outcomes, (Y_i, δ_i, X_i). The most basic statistical model was proposed

$$S_X(t) = P_X(Y > t) = exp\left(-H_X(t)\right) = exp\left(-\int_o^t h_X(dy)\right) \tag{13}$$

The hazard function is proportional over times and for d-dimensional vectors of parameter β, so the hazard function will be

$$\begin{gathered} H_X(dy) = e^{X\beta} h(dy) \\ H_X(t) = e^{X\beta} H(t) \end{gathered} \tag{14}$$

So, the likelihood of the observed data (Y_i, δ_i, X_i) will be

$$L(\beta) = \left(\prod_{\delta_i=0}^{n} P_{X_i}(Y > Y_i)\right)\left(\prod_{\delta_i=1}^{n} P_{X_i}(Y = Y_i)\right) \tag{15}$$

Many inferential methods in statistics are based on finding the parameters that are the most likely for known data, in the sense of those parameters that have the largest value of L. To derive explicit formula, choose number $\Delta_j > 0$, such that

$$L_\Delta = \prod_{i=1}^{n} P_{X_i}(Y > Y_j + \Delta_j); \delta_i = 0 \tag{16}$$

$$L_\Delta = \prod_{i=1}^{n} P_{X_i}(Y_j - \Delta_j < Y < Y_j + \Delta_j); \delta_i = 1 \tag{17}$$

So, we can write the above two equation in survival function by using the above relationship mentioned. Let define $Z_j = \int_{Y_{j-1}+\Delta_{j-1}}^{Y_j-\Delta_j} h(dy)$ and $Z_{j0} = \int_{Y_j-\Delta_j}^{Y_j+\Delta_j} h(dy)$, then,

$$L_\Delta = Exp\left(-\sum_{j=1}^{m}\left\{Z_j R_j(\beta) + Z_{j0} R_j^0(\beta) + S_j(Z_{j0},\beta)\right\}\right) \tag{18}$$

where

$$R_j(\beta) = \sum_{k=j}^{m}\left(\sum_{Y_i=Y_k} e^{X\beta}\right) = \sum_{Y_i>Y_k} e^{X\beta} \tag{19}$$

$$R_j^0(\beta) = \sum_{Y_i=Y_j,\delta_i=0} e^{X_i\beta} + \sum_{Y_i>Y_j} e^{X_i\beta} \tag{20}$$

and

$$S_j(Z_{j0},\beta) = \sum_{Y_i=Y_j,\delta_i=1} log\left(1 - exp(-Z_{j0}e^{X_i\beta})\right) \tag{21}$$

There are some common frailty model can be consider for the survival and density function. Commonly for semiparametric frailty model used Accelerated Failure Time (AFT), Cox Proportional Hazard (PH) and Proportional Odds (PO) model. For the AFT model has the survival and density functions

$$S_{X_{ij}}(t) = S_0(e^{X_{ij}^T\beta+v_i}t) \tag{22}$$

and

$$f_{X_{ij}}(t) = e^{X_{ij}^T\beta+v_i}f_0(e^{X_{ij}^T\beta+v_i}t) \tag{23}$$

Similarly for Proportional hazard model,

$$S_{X_{ij}}(t) = \{S_0(t)\}^{(e^{X_{ij}^T\beta+v_i}t)} \tag{24}$$

and

$$f_{X_{ij}}(t) = e^{X_{ij}^T\beta+v_i}\{S_0(t)\}^{(e^{X_{ij}^T\beta+v_i}-1)}f_0(t) \tag{25}$$

For Proportional odds model, the survival and density functions

$$S_{X_{ij}}(t) = \frac{e^{-X_{ij}^T\beta-v_i}S_0(t)}{1+(e^{-X_{ij}^T\beta-v_i}-1)S_0(t)} \tag{26}$$

and

$$f_{X_{ij}}(t) = \frac{e^{-X_{ij}^T\beta-v_i}f_0(t)}{[1+(e^{-X_{ij}^T\beta-v_i}-1)S_0(t)]^2} \tag{27}$$

3.2 Prior Distribution

Let us consider the baseline cumulative hazard function $H(t) = \int_0^t h(dy)$. A useful way to estimate properties of the baseline hazard density $h(dy)$ is to assume a parametric model and then estimate the parameters involved. Parametric probability distribution for the set of increasing functions $H(t)$ for $t \geq 0$ is a gamma process $Z(t)$. This is a stochastic process with independent increments whose increments have the gamma distribution

$$Z(t) - Z(s) \approx G\left\{\theta\left[\alpha(t) - \alpha(s)\right], \lambda\right\} \tag{28}$$

where, $\alpha(t)$ is some strictly increasing function that is continuously differential for $t > 0$, and $Z \approx G(\theta, \lambda)$ means that Z is a random variables with the gamma density with pdf

$$f(Z; \theta, \lambda) = \frac{\lambda^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\lambda Z} \tag{29}$$

Where, $0 \leq Z \leq \infty$, from the above $\alpha(t) = t$ or $\alpha(t) = t^\sigma$ for some values of $\sigma > 0$. So the mean and variance will be

$$E\left\{Z(t) - Z(s)\right\} = \left\{\theta[\alpha(t) - \alpha(s)]\right\}/\lambda = \mu\left\{\alpha(t) - \alpha(s)\right\} \tag{30}$$

$$Var\left\{Z(t) - Z(s)\right\} = \left\{\theta[\alpha(t) - \alpha(s)]\right\}/\lambda^2 = \mu\left\{\alpha(t) - \alpha(s)\right\}/\lambda \tag{31}$$

For, $\mu = \theta/\lambda$. If $\alpha(t) = t, E(Z(t) = \mu t$ so that, $\alpha(t) = t$ corresponds to noisy exponential baseline survival times. Similarly, if $\alpha(t) = t^{\sigma}$, then $E(Z(t) = \mu t^{\sigma}$ corresponds to Weibull survival distribution. If the function $\alpha(t)$ is assumed to fixed, and θ, λ are parameters to be estimated and the parameter will be $\mu = \theta/\lambda, 1/\lambda$.

The sample path of the gamma process $Z(t)$ with independent increments, the differences all follows gamma distribution. If Z_j and Z_{j0} are considered parameters or hidden variables in the observed data (Y_i, δ_i, X_i) with the probability density gamma, then the parameter θ, λ are considered hyper parameters. In a Bayesian framework, the hyper parameter themselves are given probability density such as prior distribution. In this case we assume, gamma prior distribution $\theta, \lambda \approx G(\epsilon, \epsilon)$ for some small $\epsilon > 0; 0.001$ and an uniformative normal prior for each components β_j of $\beta \epsilon R^d$, specifically, all the coefficients follow normal density with mean zero and standard deviation $1/\epsilon$ according to the definition of Ibrahim et al. (2001).

More specifically let, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a vectors of regression coefficients, v_i is an unobserved frailty associated with $S_i(t)$ and $S_o(t)$ is the baseline survival with density $f_o(t)$ corresponding to $x_{ij} = o$ and $v_o = 0$. Let $\Gamma(a, b)$ denotes a gamma distribution with mean a/b and $N_p(\mu, \Sigma)$ a p-variate normal distribution.

$$
\begin{gathered}
\beta \sim N_p(\beta_0, S_o) \\
S_0(.)|\alpha, \theta \sim TBP_L(\alpha, S_\theta(.)), \alpha \sim \Gamma(a_0, b_0), \theta \sim N_2(\theta_0, V_0) \\
(v_1, \ldots, v_n)^T|\tau \sim ICAR(\tau^2), \tau^{-2} \sim \Gamma(a_\tau, b_\tau), or \\
(v_1, \ldots, v_n)^T|\tau, \psi \sim GRF(\tau^2, \psi), \tau^{-2} \sim \Gamma(a_\tau, b_\tau), \psi \sim \Gamma(a_\psi, b_\psi), or \\
(v_1, \ldots, v_n)^T|\tau \sim IID(\tau^2), \tau^{-2} \sim \Gamma(a_\tau, b_\tau)
\end{gathered} \tag{32}
$$

where, TBP refer to Transformed Bernstein Polynomial Zhou et al. (2017) prior for baseline survival probability and relevant parameters, Intrinsic Conditionally auto-regressive prior Besag (1974), which is the time dependent and spatial data analysis framework and finally IID refers to Independent and identical Gaussian prior distrbution. In our cases, IID parior consider for the predictors coefficients.

3.3 Likelihood Function

Under the above consider, the full combined likelihood function of the observed data, including the prior distribution for Z_j, Z_{j0} and θ, λ, β corresponding to the above likelihood function is

$$
\begin{aligned}
L_\Delta = {} & \frac{\epsilon^\epsilon}{\Gamma(\epsilon)}\theta^{\epsilon-1}e^{-\theta\epsilon}\frac{\epsilon^\epsilon}{\Gamma(\epsilon)}\lambda^{\epsilon-1}e^{-\lambda\epsilon} \\
& \times \prod_{\alpha=1}^{d}\left[\frac{1}{\sqrt{2\pi}}exp(-\epsilon^2\beta_\alpha^2/2\right] \\
& \times \prod_{j=1}^{m}\left[\frac{\lambda^{\theta W_j^\Delta}}{\Gamma(\theta W_{j0}^\Delta)}Z_j^{\theta W_j^\Delta-1}e^{-\lambda Z_j}e^{Z_jR_j(\beta)}\right] \\
& \times \prod_{j=1}^{m}\left[\frac{\lambda^{\theta W_{j0}^\Delta}}{\Gamma(\theta W_{j0}^\Delta)}Z_{j0}^{\theta W_{j0}^\Delta-1}e^{-\lambda Z_{j0}}e^{Z_{j0}R_j^0(\beta)}\right] \\
& \times \prod_{j=1}^{m}\left[\prod_{Y_i=Y_j,\delta=1}\left(1-exp(-Z_{j0}e^{X_i\beta})\right)\right]
\end{aligned}
\tag{33}
$$

As each $\Delta_j \to 0$, then $W_j^\Delta \to W_j = \alpha(Y_j) - \alpha(Y_{j-1}) > 0$. If $d_j = 0$, the jth factors has a delta -function singularity at $Z_{j0} = 0$ as $\Delta_j \to 0$, so L_Δ does not need to re-scaled. Thus, ignoring the constant that depend on Δ_j for $d_j > 0$, the limit of the likelihood is the limiting full function is

$$
\begin{aligned}
L = {} & C \times \lambda^{\epsilon-1}e^{-\lambda\epsilon}\left(\theta^{r+\epsilon-1}e^{-\theta\epsilon}\right)exp\left(-\epsilon^2\sum_{\alpha=1}^{d}\beta_\alpha^2/2\right) \\
& \times \prod_{j=1}^{m}\left[\frac{\lambda^{\theta W_j}}{\Gamma(\theta W_j)}Z_j^{\theta W_j-1}exp\left\{-Z_j(\lambda+R_j(\beta))\right\}\right] \\
& \times \prod_{d_j\geq 1}^{m} exp\left\{-Z_{j0}(\lambda+R_j(\beta))\right\}\left[\frac{\prod_{Y_i=Y_j,\delta=1}(1-exp(-Z_{j0}e^{X_i\beta}))}{Z_{j0}}\right]
\end{aligned}
\tag{34}
$$

In the above equation, C depends on ϵ and $\alpha(Y_i)$ and r is the number of distinct time with $d_j > 1$. The inference depends on which parameter values are relatively more likely and relatively larger values of L for the observed data (Y_i, δ_i, X_i).

3.4 Parameters Estimation by MCMC

By using Markov Chain Monte Carlo methods we estimate the parameters and hidden variables $(\theta, \lambda, Z_j, Z_{j0}, \beta)$. Let define a Markov Chain Q_n that takes its values in the space of possible parameter vectors $(\theta, \lambda, Z_j, Z_{j0}, \beta)$ and which has s stationary or asymptotic distribution that is proportional to the above likelihood function, that means that Q_n depends most of its time where the likelihood is the largest. Mean or median values of components or function of the components of Q_n can be used to provided estimates of the parameters affecting the true survival times T_i. The Markov chain proceeds by changing or updating each of the components of the vectors in the turn in a way that depends on the conditional probability distribution of that parameters given the data and all the other parameters.

3.4.1 Updating θ Values

Ignoring multiplicative constant and alos ignoring the factors of the likelihood, that donot depend on θ, the conditional density of θ given the data and other parameters are

$$\theta^{r+\epsilon-1}e^{\theta\epsilon}\lambda^{\theta W}\prod_{j=1}^{m}\left[\frac{Z_j^{\theta W_j}}{\Gamma(\theta W_j)}\right] \tag{35}$$

where, $W = \sum_{j=1}^{m} W_j$. The density of the above equation is asymptotic to $C\theta^{r+m+\epsilon-1}$ as $\theta \to 0$ and decays faster than exponentially at infinity, and can be updated efficiently by one steps of a Metropolis random walk sampler. The density of above function is log-concave function of θ so it can be updated by Gibbs sampler steps that samples directly from the distribution. In general, a function $f(\theta)$ is decreasing in θ, then the log-concave of the above equation 35, follows from the identity,

$$\frac{d^2}{d\theta^2}log\Gamma(\theta) = Var\left\{logG(\theta,1)\right\} \tag{36}$$

Where, $G(\theta, 1)$ represent gamma-distribution of the parameters.

3.4.2 Updating λ Values

Ignoring the multiplicative constants and factors that don't dependent on λ, the conditional density of for the data and other parameters is

$$\lambda^{\theta W+\epsilon-}exp\left\{-\lambda(\epsilon + \sum_{j=1}^{m}(Z_j + Z_{j0}))\right\} \tag{37}$$

Where, $Z_{jo} = 0$ if $d_j = 0$, this can be updated by a Gibbs sampler step by sampling from the gamma distribution

$$\lambda \approx G\left(\epsilon + \theta W, \epsilon + \sum_{j=1}^{m}(Z_j + Z_{j0})\right) \tag{38}$$

3.4.3 Updating Z_j Values

Ignoring term and factors that do not involve on Z_j , then the conditional distrbution will be given by the observed data and other parameters

$$Z_j^{\theta W_j - 1} e^{Z_j(\lambda + R_j(\beta))} \tag{39}$$

Thus, Z_j can be updated by Gibbs sampling from the gamma distribution

$$Z_j \approx G\left\{\theta W_j, \lambda + R_j(\beta)\right\} \tag{40}$$

3.4.4 Updating Z_{j0} for $d_j \geq 1$ Values

Similar way we can update for ignoring the independent part from the likelihood and the conditional density of Z_{jo} given the observed data and other parameters is

$$e^{-Z_{j0}(\lambda + R_j^0(\beta))} \left\{ \frac{\prod_{Y_i = Y_j, \delta_i \geq 1}\left(1 - exp(-Z_{j0}e^{X_i\beta})\right)}{Z_{j0}} \right\} \tag{41}$$

The above density is normalizable in Z_{j0} and can be updated by one steps of a Metropolis random walk.

3.4.5 Updating β Values

Ignoring multiplicative constants and factors that do not depend on β, the conditional density of β given the data and other parameters is

$$exp\left\{-\sum_{j=1}^{m}\big(Z_jR_j(\beta)+Z_{j0}R_j^0(\beta)-S_j(Z_{j0},\beta)\big)-\frac{1}{2}\epsilon^2\beta_a^2\right\} \tag{42}$$

Where,

$$S_j(Z_{j0},\beta)=\sum_{Y_i=Y_j,\delta_i=1} log\left(1-exp(-Z_{j0}e^{X_i\beta})\right) \tag{43}$$

If there is not observed death at time $Y_i = Y_j$ then the conditional survival will be zero and the second term do not appear. So, baring linear dependencies, among sample covariates the conditional likelihood in realizable in each beta so that it can be updated by a single step of metropolis random walk. The advantages of Markov chain Monte Carlo process is that it gives the information about the conditional distribution of baseline hazard functions

$$\begin{aligned} Z_j &\approx H(Y_j-)-H(Y_{j-1}) \\ Z_{j0} &\approx d(H(Y_j))=h(dY_j) \end{aligned} \tag{44}$$

Given the observed data. We are primarily interested on the parameters (θ,λ,β) and not in the baseline hazard, the Z_j, Z_{j0} can be integrated out of the likelihood function to obtain a marginal density of the interested parameters. So evaluating the integrals $\int LdZ_j$ then we will get the likelihood functions

$$L = C\lambda^{\epsilon-1}e^{-\epsilon\lambda}\prod_{j=1}^{m}\left(\frac{\lambda}{\lambda+R_j(\theta)}\right)^{W_j\theta}$$

$$\times\, exp\left(-\epsilon^2\sum_{a=1}^{d}\beta_a^2\right) \tag{45}$$

$$\times \prod_{d_j\geq 1} exp\left(-Z_{j0}(\lambda+R_j^0(\beta))\right)\left(\frac{\prod_{Y_i=Y_j,\delta_i=1}(1-exp(Z_{j0}e^{X_i\beta}))}{Z_{j0}}\right)$$

While, λ no longer has a single gamma update, the parameter θ now has a gamma update, specifically,

$$\theta \approx G\left(r+1, \frac{\sum_{j=1}^{m} W_j log\left\{\lambda+R_j(\beta)\right\}}{\lambda}\right) \tag{46}$$

The parameters, Z_{j0} can be integrated by using the identity and $d_j = 1$ the jth factors in compact in a simple form. In particularly if $d_j \leq 1$, for all j, there are no ties observed death times then the integral leads to

$$L = L(\theta,\lambda,\beta) = C\lambda^{\epsilon-1}e^{\lambda\epsilon}\theta^r\prod_{j=1}^{m}\left\{\left(\frac{\lambda}{\lambda+R_j(\beta)}\right)^{W_j\theta} log\left(\frac{\lambda+R_j(\beta)}{\lambda+R_j^0(\beta)}\right)\right\} \tag{47}$$

If $d_j = 0$, then $R_j^0(\beta) = R_j(\beta)$ and the logarithmic factors does not appear. According to Kalbfleisch (1978), the likelihood of the above equation there is no information about Z's although the conditional density if beta is known precisely.

The posterior density of hazard will be the density function of the hidden parameter given by the data and the other parameters. For baseline hazard we consider the gamma prior and after simplification we get it also follow gamma density. For simulating the sample, we first choose a sample size n, the number of covariate d. Chosen any arbitrary values θ, λ and the risk parameters $\beta \epsilon R^d$ and also choose strictly-increasing continuous function $\alpha(t)$ then we can simulate

$$H(t) = Z(t) \approx G(\theta\alpha(t), \lambda) \approx \frac{1}{\lambda} G(\theta\alpha(t), 1) \tag{48}$$

The observation time Y can be cansored and the path of z(t) are right continuous jumps every time interval. so $P(Z(Y_i)) > s = exp(-e^{X_i\beta})$ where $s = Z(t)$. Let Z_i be independent exponential distributed random varaibles with mean $e^{-X_i\beta}$ then

$$Z_i \approx e^{-X_i\beta}\left(-log(U_i)\right) \tag{49}$$

So, to simulate, Y_i we need a approximate sample path of z(t). The independent gamma random variables is the path of the sample generation candidates $Q_j \approx G(\theta\Delta(j,m), 1)$ then $Z(k/m) \approx G(\theta, \alpha(k/m), \lambda) \approx \frac{1}{\lambda}\sum_{j=1}^{k} Q_j$. Thus, we can simulate Y_i as

$$Y_i = min\left\{\frac{k}{m}, \frac{1}{\lambda}\sum_{j=1}^{m} Q_k \geq Z_i\right\} \tag{50}$$

Which is equivalently

$$Y_i = \frac{1}{m} min \left\{ k, \sum_{j=1}^{m} Q_k \geq \lambda Z_i \right\} \tag{51}$$

We have the density of Z_i which is $Z_i \approx \lambda exp(-X_i\beta)(-log(U_i)) \approx \lambda Z_i$. To include censoring, we define censoring times,

$$Y_i^c = \frac{1}{m} min \left\{ k, \sum_{j=1}^{m} Q_k \geq \lambda Z_i^c \right\} \tag{52}$$

For $Z_i^c \approx \mu e^{-X_i\beta}(-log(U_i)$. In the same way for constant, μ and censoring indicators we can simulated the samples. $\mu > 0$ in same way for some constant. Define $\delta_i = 1$ if $Y_i < Y_i^c$ and $\delta_i 0$ if Y_i^c. The last observed time are

$$Y_i^o = min\{Y_i, Y_i^c\} = \frac{1}{m} min\{k, \sum_{j=1}^{m} Q_j \geq Z_i^o\} \tag{53}$$

where $Z_i^o = min\{Z_i, Z_i^o\}$. In general if all the predictors are independent then all are follow gamma distributions and coefficient can be considered as independent gamma variables.

4 Results and Discussions

4.1 Data Descriptions

For this study, we are selecting SEER male breast cancer dataset. In this dataset have 60 more predictors but only few listed variable are more related on male breast cancer outcomes. For prediction the survivability of male breast cancer patients, we are selecting 22 predictors and removing all the predictors and values that are not listed properly and missing values. Finally we have 689 complete dataset. From the table 1 we see that the overall age male breast cancer patients is 66.91 years but alive respondent have little bit smaller average age than death patients. Tumor size for death respondents is 55.59 mm^2 but for alive cases 38.58mm^2.

Table 1: Data description Table for considering all the predictors and classified to status of respondents

Variable	Total	Alive	Death
Age	66.91 (12.73)	63.49 (12.25)	69.26 (12.54)
Tumor Size	48.89 (21.25)	38.58 (23.01)	55.99 (16.59)
Tumor Extension Area (mm^2)	196.99 (158.05)	166.54 (136.33)	217.96 (168.40)
Lymph Nodes Size (mm)	153.79 (191.27)	132.84 (181.99)	168.22 (196.33)
Tumor Away from Prime site	15.63 (14.74)	13.22 (9.75)	17.30 (17.18)
primary tumor diameter	18.31 (13.86)	14.49 (11.04)	20.95 (14.96)
Total Tumor on Prime site			
1	450 (65%)	179 (64%)	271 (66%)
2	178 (26%)	78 (28%)	100 (25%)
3	45 (7%)	18 (6%)	27 (7%)
4	10 (1%)	6 (2%)	4 (1%)
5	6 (1%)	0 (0%)	6 (1%)
Race			
Black	183 (27%)	63 (22%)	120 (29%)
White	506 (73%)	218 (78%)	288 (71%)
Marital Status			
Unmarried	125 (18%)	55 (20%)	70 (17%)
Married	564 (82%)	226 (80%)	338 (83%)
AJCC Cancer Stage			
I	204 (30%)	92 (33%)	112 (27%)
II-IIIA	285 (41%)	111 (40%)	174 (43%)
II-IIIB	127 (18%)	48 (17%)	79 (19%)
IIIC	13 (2%)	11 (4%)	2 (0%)
UNK	60 (9%)	19 (7%)	41 (10%)
AJCC Cancer Stage T			
T0-TX	182 (26%)	100 (36%)	82 (20%)
T1-T4a	168 (24%)	32 (11%)	136 (33%)
T1-T4b	110 (16%)	47 (17%)	63 (15%)
T1-T4c	229 (33%)	102 (36%)	127 (31%)
AJCC Cancer Stage N			
N0	229 (33%)	126 (45%)	103 (25%)
N1-N3a	299 (43%)	110 (39%)	189 (46%)
N2-N3c	161 (23%)	45 (16%)	116 (28%)
AJCC Cancer Stage M			
M0	337 (49%)	193 (69%)	144 (35%)
M1	318 (46%)	83 (30%)	235 (58%)
MX	34 (5%)	5 (2%)	29 (7%)
Treatment of Prime Site			
No Pathelogic Specimen	270 (39%)	117 (42%)	153 (38%)
Resection	419 (61%)	164 (58%)	255 (63%)
Treatment not Surgary			
Autopsy	575 (83%)	246 (88%)	329 (81%)
Not Primary Site	114 (17%)	35 (12%)	79 (19%)
Sugrary On Prime Site			
No	278 (40%)	116 (41%)	162 (40%)
Yes	411 (60%)	165 (59%)	246 (60%)
Regonal Node Evaluation			
No Autopsy Used	302 (44%)	98 (35%)	204 (50%)
Pre-surgary Treat or Radiation	387 (56%)	183 (65%)	204 (50%)
Mets DX Evaluation			
Carsinomatosis	31 (4%)	9 (3%)	22 (5%)
Distant Lymph Nodes	130 (19%)	35 (12%)	95 (23%)
Distant Metastage	528 (77%)	237 (84%)	291 (71%)
Tumor Marker			
Boarderline	279 (43%)	120 (43%)	177 (43%)
Positive	392 (57%)	161 (57%)	231 (57%)
Breast Subtype M			
M0	616 (89%)	266 (95%)	350 (86%)
M1	56 (8%)	14 (5%)	42 (10%)
MX	17 (2%)	1 (0%)	16 (4%)
Breast Subtype T			
T0-TX	373 (54%)	129 (46%)	244 (60%)
T1-T4a	24 (3%)	10 (4%)	14 (3%)
T1-T4b	79 (11%)	41 (15%)	38 (9%)
T1-T4c	213 (31%)	101 (36%)	112 (27%)
Breast Subtype N			
N0	326 (47%)	170 (60%)	156 (38%)
N1	175 (25%)	64 (23%)	111 (27%)
N2	53 (8%)	15 (5%)	38 (9%)
N3	58 (8%)	20 (7%)	38 (9%)
NX	77 (11%)	12 (4%)	65 (16%)
Sample Size	689	281	408

From the table 1 we see that, 42% of the respondent alive for no pathological specimen treatment applied on the other hand, 38% patients died for this treatment. Similarly for, resection treatment, 63% died an only 58% alive. After sugary on prime site of cancer, approximately same proportion of patients died and alive, but treatment other than sugary, for example, autopsy only its 88% patients alive. For the regional node, not used autopsy 50% patients died but only 35% alive, on the other hand, pre-sugary or radiation therapy, 65% patients alive and 50% died. Tumor lymph node size is an important factors for a patient died or alive. From the table 1, we see that tumor lymph size for death patients is higher compare with alive respondents. Similar outcomes for tumor diameter on the prime sites. For larger diameter of tumor size, the proportion of patients died is higher. Another important thing is take into account that, how much scatter the tumor around the breast. From the table 1 we see that for death respondents have higher scatteredness compare with alive patients.

4.2 Variable Selections

As mentioned above, when we have a lot of predictors and very few number of respondents, usually statistical modeling framework is not appropriate for inferences and interpretations for our cases. For these reason we are using machine learning approaches for prediction the survivability for breast cancer respondents and inferences. Most important things are which predictors have greater impact on survivability of patients. Machine learning methods proposed some predictors are more important than others with significant levels and some other criterion.

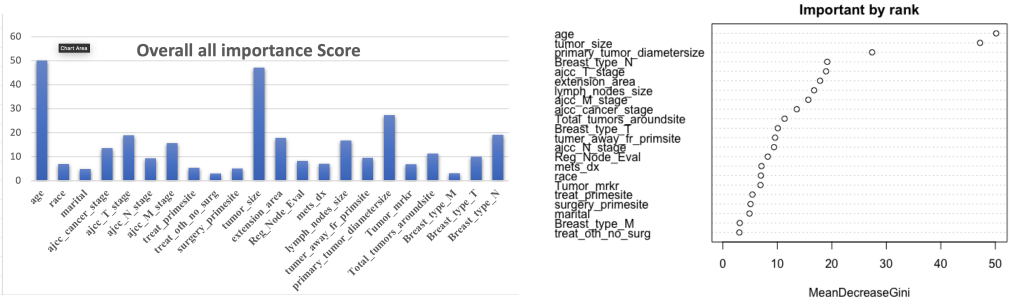


Figure 1: Variable important according to the rank and the histogram corresponding to the important scores

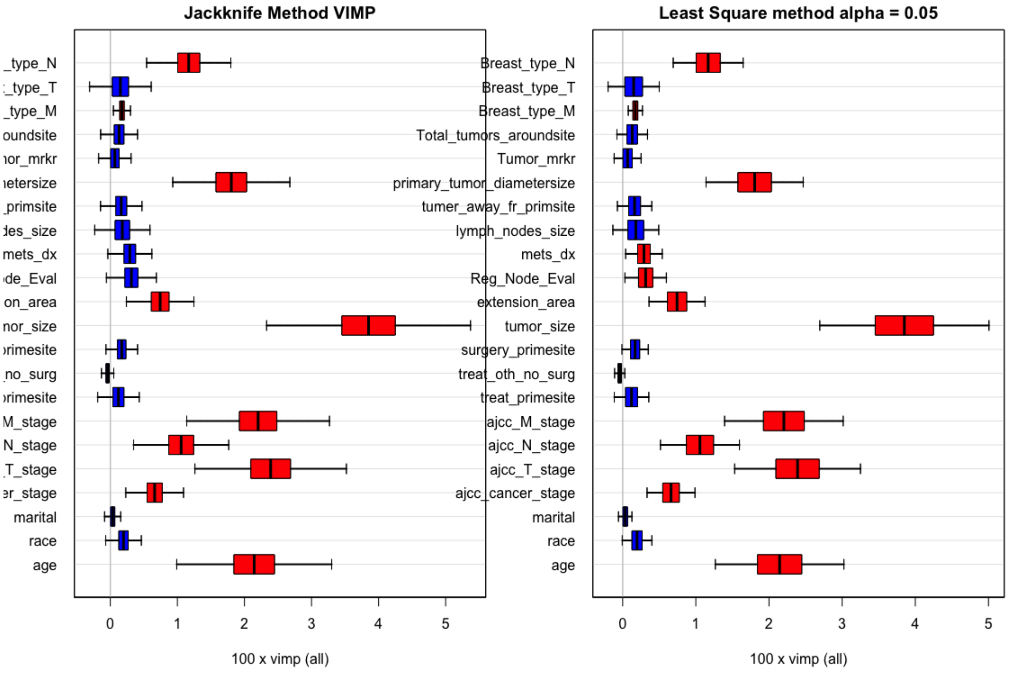


Figure 2: Variables Important by Jackknife and 95% Confidence Level selection Criterion

From the figure 1 and 2 we see that very few predictors are important for predictions of survivability of male breast cancer patients. By random survival forest methods, and using 5% level of significance and Jackknife selection criterion, we are using only the significant predictors for random survival forest trees and Bayesian cox regression model. From the figure 1 we are getting the rank of importance of the predictors but from the figure 2 we are getting the statistical selection criterion which variables are significantly important. The important score calculated by Ginni mean impurity score.

4.3 Random Survival Forest Trees

In survival analysis many different regression modeling strategies can be applied to predict the survivability of future events. Often, however, the default choice of analysis relies on Cox regression modeling due to its convenience. Extensions of the random survival forest approach proposed by Breiman (1996) to survival analysis provide an alternative way to build a prediction model. Usually from the random survival forest tree predict conditionally the median survival time with consider the censoring indicators. In the tree consider the variable number of tree in each classification and node size. From the figure 3 we are presenting four differ trees with node sizes 1000 and numbers of tree is 50, 100, 150 and 500 respectively. Figure 3, from first plot we see that cancer tumor away from prime site is the most important predictors and predict that more than 50 mm away from prime site patients have median survival time is approximately 209 months. Similarly we can predict that tumor less than 55mm away and age less than 70 years, for state cancer stage IIA-IIIA and IIIC patients have median survival time 168 months.

From the figure 3, in second plot we see that tumor less than 65mm away and age greater than 76 years, for breast type NX adjusted patients only survive 244 months, on the other hand other breast type patients survive 191 months. Similarly, we can predict that, tumor away less than 65mm, age less than 76 years, for tumor lymph node size greater than 740mm, the patients only survive 188 months. When increase the number of tree in each node, from the figure 3 third plot, ntree = 150, the predicted survival median time is 105.8 months for tumor size is less than 21mm, on the other hand, tumor size is greater than 21mm, for age less than 66 years, II-IIIA and IIIC stage breast cancer patients only 173 months survive but stage I, II-IIIB and UNK adjusted stage breast cancer patients and breast type N0, N2 only 77.92 months but breast type N1, N3 and NX adjusted have 118.87 months.

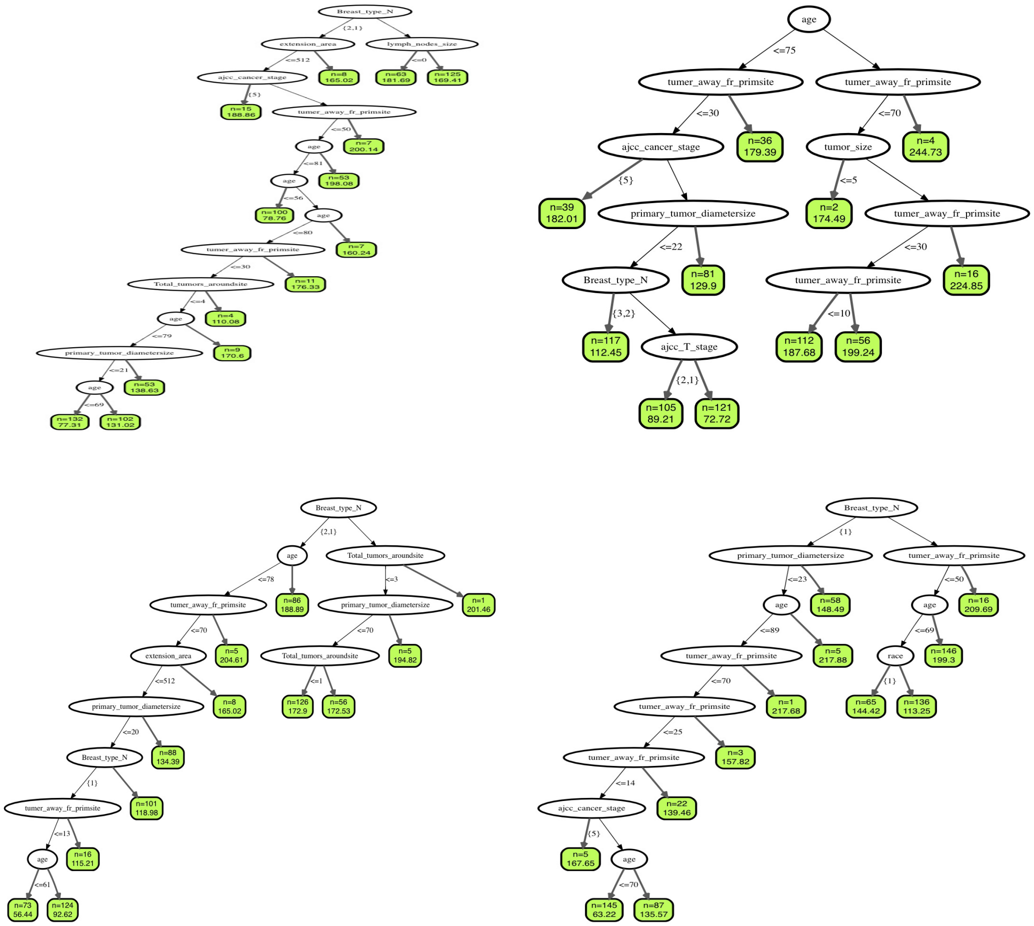


Figure 3: Random Survival Forest Tree for (1): ntree = 50, (2): ntree = 100, (3): ntree = 150 and (4) ntree = 500

The forest tree also predict the survival probabilities by considering the given condition of the predictors. From the figure 4, when ntree is 50, we find that, for breast type N0, and tumor size less than 35mm, the patients survival probability approximately 0.71 but the tumor size is greater than 35mm, and cancer stages are T0-TX adjusted, survival probability only 0.39 and for stages T1-4a, T1-4c has 0.3. Similarly we can interpret that, for breast type not N0, and tumor soze less than 42mm, the patients survival probability approximately 60% but for greater than 42mm, and tumor diameters is less than 21mm, the patient survive 32% but for greater than 21mm diameters, only 18% survive. However, when, tree size for 500, the patients have 29% chance for survivability of breast type N2 and NX adjusted, but for breast type N0, N1 and N3 and age less than 62 years, for tumor size less than 30mm patients have 70% change survivability but tumor size greater than 30mm, only 37% chance.

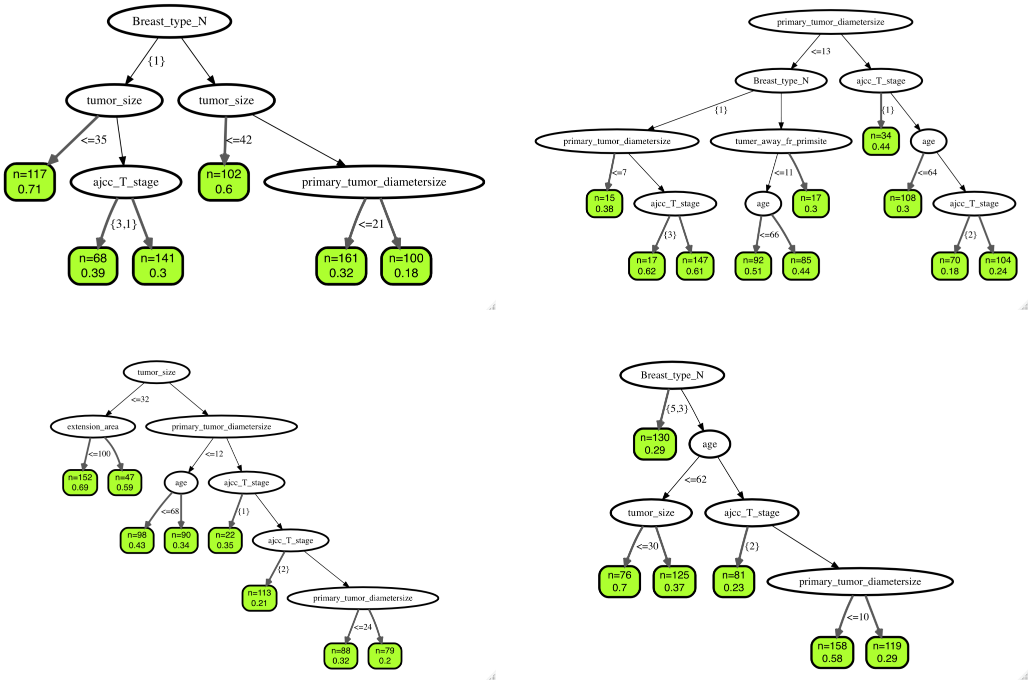


Figure 4: Random Forest Tree for predicted survival probability for (1) ntree = 50, (2) ntree = 100, (3) ntree = 150, and (4) ntree = 500

We can predict the overall survival probability of male breast cancer patients by using random survival forest trees. From the figure 4.3, we see that, for age less than 69 years and white male breast cancer have higher survival probability compare with black for different conditions. For age less than 69 years of black male breast cancer patients, if the tumor extension are is less than $200mm^2$, N0 breast type patients have higher survival probability compare with N1, N2, N3 and NX adjusted breast type patients. However, the extension area grater than $200mm^2$, the patients have lower survival probability. From the figure 4.3, age greater than 69 years, NX adjusted breast cancer patents have lower survival probability compare with others type of breast. For breast not NX adjusted and prime site tumor diameters is greater than 21mm, the survival probability decreasing very sharply with increasing of times on the other hand, tumor diameter greater than 21mm, and for different cancer stages the survival probability decreasing slowly. We can also compare that for ajcc cancer stage, I,II-IIIA, and IIIC patients have higher survival compare with II-IIIB and UNK adjusted stage breast cancer patients.

From the figure **??** we see that tumor size less than 47mm, the patients survival probability approximately 70%, on the other hand, the tumor size greater than 47mm and age less than 66 years, for breast type N0 patients will approximately 30% higher chance of survival compare with breast type N1 N2 N3 and NX adjusted. The patients age greater than 66 years and tumor size also greater than 47mm, T1-4a, T1-4c stages breast cancer patients have only 10% chance of survivability on the other hand for T0-X, T1-4b stages breast cancer patients have 35% chance. So approximately 25% lower change of survive for the patients with T1-4a and T1-4c stage breast cancer and the age greater than 66 years and tumor size is greater than 47mm. For prediction the survival probability of male breast cancer patients, random forest tree is very powerful for conditional interpretation. From the figure 4.3, we see that breast type N0 patients have higher survival probability compare with other type of breast such as N1, N2, N3 and NX adjusted considering some potential conditions. When tumor 70mm away from prime site and extension of tumor around prime site is less than 200mm^2, N0 breast type patients have higher survival probability compare with other types however, the prime tumor grater than 70mm away from prime site, N3 breast type patients have higher survival probability. When tumor greater than 70mm away from prime sites, N1 breast type patients have very low survivability among all other breast. Tumor extension area greater than 200 mm^2, NX adjusted breast patients have higher survival compare with less than 200 mm^2 extension area.

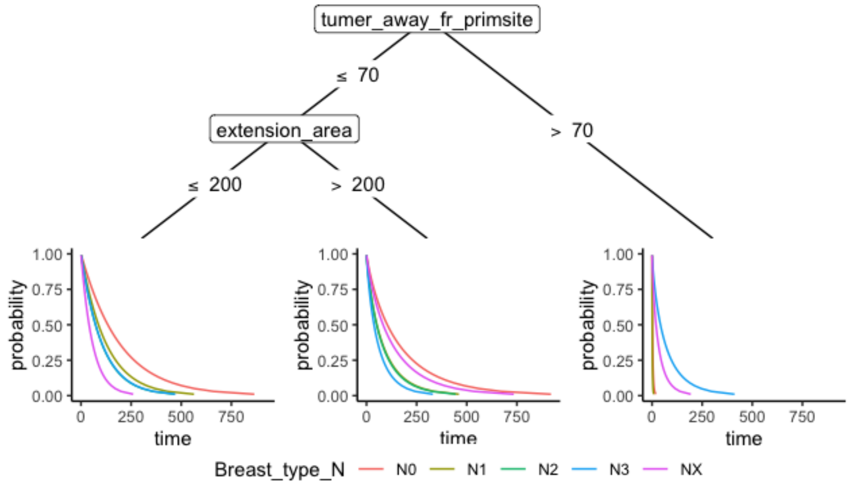


Figure 5: Random survival forest Tree for Full model prediction the overall survival Probability

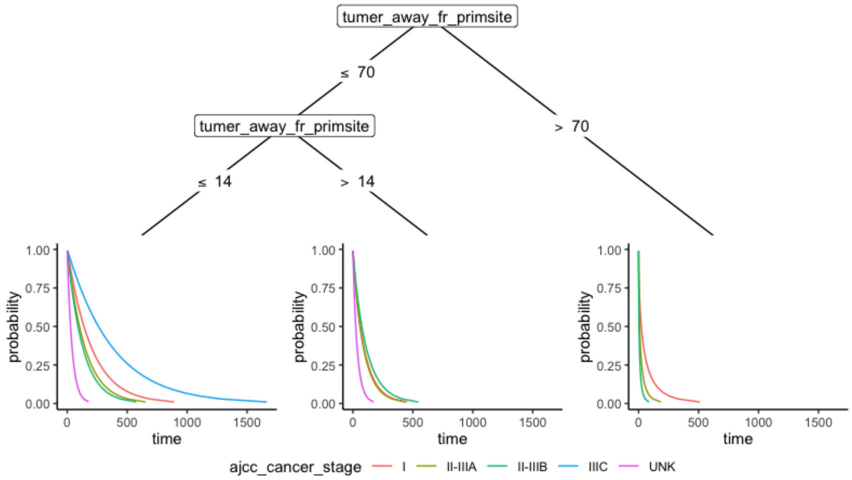


Figure 6: Random forest Tree for prediction the survivability of male breast cancer patients

Similar prediction can be done from the figure 5 we see that, tumor greater than 70mm away from prime site, cancer stage I patients have higher survival compare with stage II-IIIA, II-IIIB, IIIC and NK unadjusted. On the other hand, for less than 14mm away from prime site, IIIC stage patients have higher survival compare with all others but tumor between 14mm and 70mm, II-IIIB stage breast cancer patients have higher survivability. For tumor away from prime site less than 14mm, IIIC type cancer patients have higher survival rate on the other have, greater than 70mm, have lower survival for stage IIIC patients. Cancer stage is an important factors for patients survivability. Usually cox model or other statistically model predict the survival for consideration the linear or categories of the distance but in machine learning methods consider the predictors as continuous and classified significant levels and predict the survival probability base on the tree.

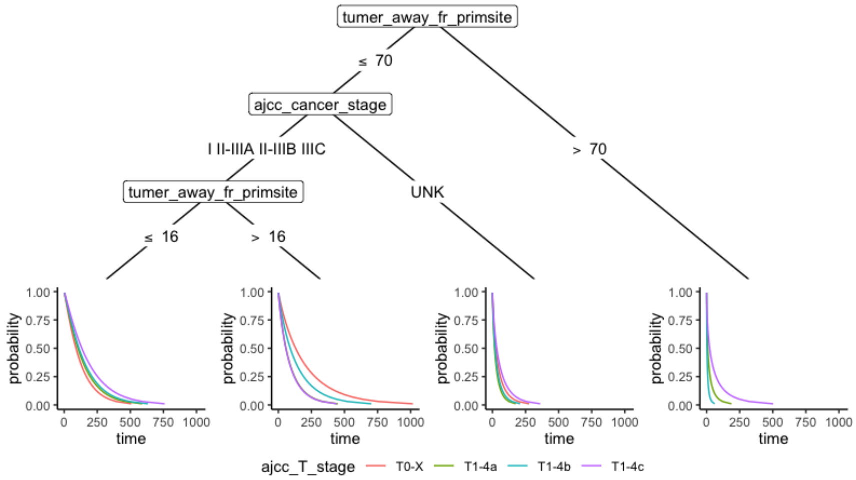


Figure 7: Random survival forest Tree for

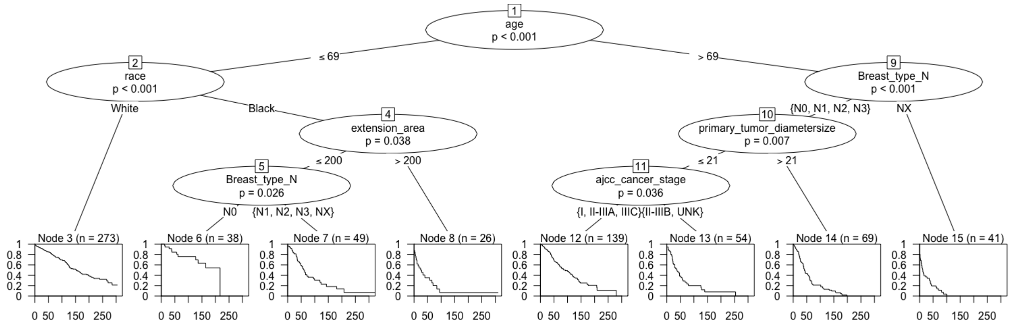


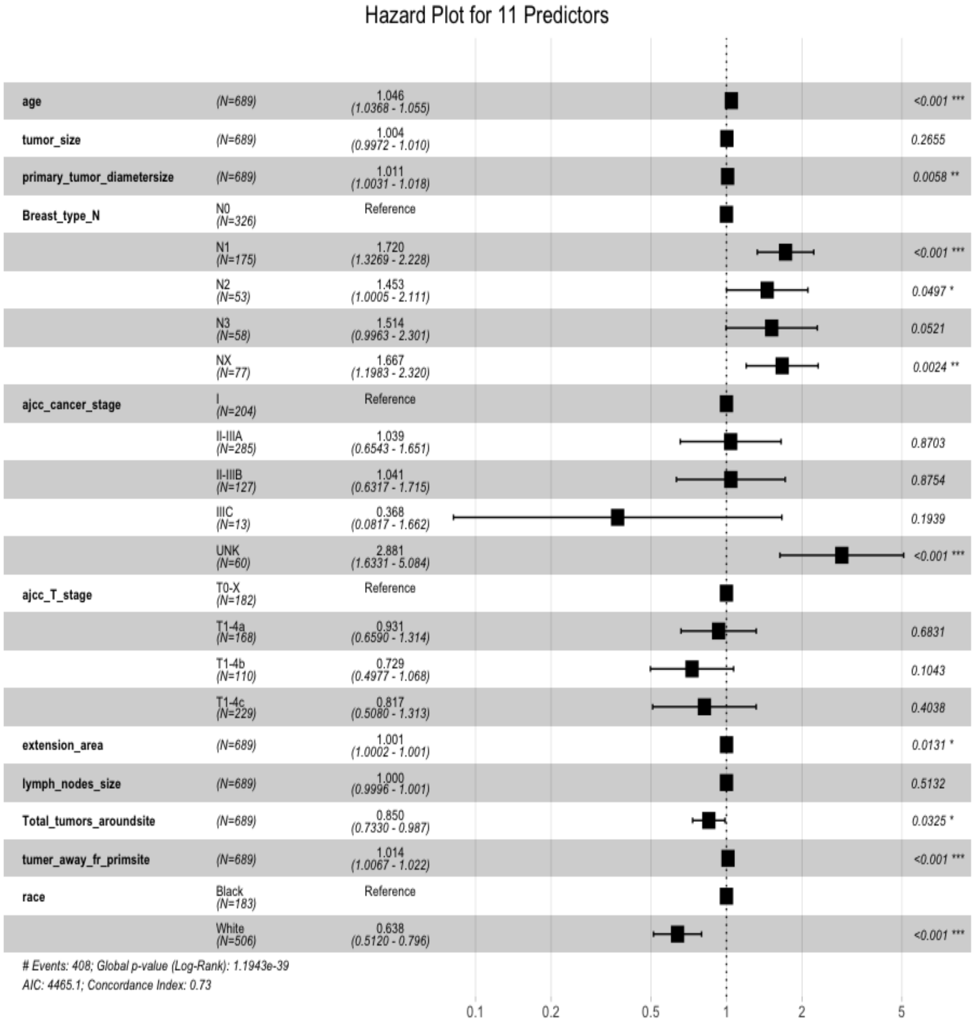
Figure 8: Random survival forest Tree for prediction the overall survival probability of male breast cancer patients

We can also predict the survival probability of different T stage breast cancer by considering all the predictors as same way mentioned in the above. From the figure 7 we see for cancer stage I, II-IIIA, II-IIIB, IIIC and tumor less than 16mm away from prime site, T1-4c stage breast cancer patients have higher survival probability compare with T0-X, T1-4a, and T1-4b stages. On the other hand, for tumor between 16mm and 70mm, T0-X stage breast cancer patients have higher survival probability. Very important finding here, the two outcomes reciprocal for the given condition. Usually for the given conditions, the T1-4b stage breast cancer patients pity stable but for other conditions, tumor great than 70mm away from prime site, the patients in this stage are very unstable. The survival probability for T1-4b stage breast cancer patients is the lowest for tumor away from greater than 70mm from the prime site. In this condition, the patients in stage T1-4c, have highest survivability among all the stages of the breast cancer.

4.4 Bayesian Cox Model Outcomes

Bayesian cox model is more applicable when the sample sizes are small and the prior information is available for the predictors coefficients. For our study, we have small dataset and a lot of predictors involves for the outcomes. Some bayesian approaches will be more applicable. By fitting cox regression we find that from the figure 9 we see that, most of the predictors are insignificant. Only few are statistically significant which will give us inefficient interpretation and inferences.

Figure 9: Cox Proportional Hazard Model for the same predictors



So by using Bayesian cox model with different prior we get better fit compare with usual cox model. From the table 2 we fitted the data with different prior of survival and the coefficients and getting the outcomes. From the table 2, we also presented the relevant information of model selection criterion such as Log Psuedo Marginal Log-likelihood value, Deviance Information Criterion and Wannabe-Akaike information criterion and posterior frailty inference variance with median and standard errors. All the prior pity same outcomes but for our data set Weibull prior fitted best compare with Log-logistics and log Gaussian prior. We are also selecting for survival prior Identical and independent Gaussian and frailty density Proportional hazard model.

Table 2: Fitting Bayesian Cox Proportional Hazard Regression Model with Weibull Prior, Log-logistic Prior and Log-normal prior with Log Psuedo Marginal Log-likelihood value, Deviance Infromation Criterion and Watanabe-Akaike information criterion

	Weibull Prior			**Loglogistic Prior**			**Lognormal Prior**		
Predictors	**Mean**	**Median**	**Std.Dev.**	**Mean**	**Median**	**Std. Dev.**	**Mean**	**Median**	**Std.Dev.**
age (Con)	0.052	0.052	0.006	0.052	0.052	0.006	0.047	0.047	0.005
Primary Tumor Diameter	0.014	0.014	0.005	0.014	0.014	0.005	0.012	0.012	0.004
Breast Type (Ref:N0)									
N1	0.656	0.652	0.168	0.656	0.652	0.168	0.577	0.575	0.135
N2	0.55	0.541	0.237	0.550	0.541	0.237	0.493	0.492	0.197
N3	0.616	0.598	0.244	0.616	0.598	0.244	0.489	0.488	0.219
NX Adjusted	1.147	1.132	0.253	1.147	1.132	0.253	0.907	0.903	0.185
Cancer Stage (Ref:T0-TX)									
T1-4a	-0.316	-0.317	0.193	-0.316	-0.317	0.193	-0.249	-0.24	0.159
T1-4b	-0.479	-0.466	0.222	-0.479	-0.466	0.222	-0.461	-0.461	0.186
T1-4c	-0.515	-0.519	0.185	-0.515	-0.519	0.185	-0.433	-0.424	0.157
Prime Tumor Extension Area	0.002	0.002	0.001	0.003	0.002	0.001	0.002	0.001	0.001
Total Tumors Primesite	-0.161	-0.16	0.092.	-0.161	-0.160	0.092	-0.143	-0.142	0.079
Tumor Away From Primsite	0.017	0.017	0.005	0.017	0.017	0.005	0.015	0.014	0.004
Race (Ref:Black)									
White	-0.537	-0.531	0.153	-0.537	-0.531	0.153	-0.511	-0.507	0.13
Log Pseudo Marginal LL	-2269.5			-2292.3			-2296.6		
DIC	4468.8			4555.1			4584.3		
WAIC	4494.7			4562.3			4586.8		
Posterior Frailty Var	1.416	1.273	0.698	0.601	0.551	0.365	0.179	0.124	0.171

From the fitted survival cure we see that in figure 10, the survivability for different cancer stage, different breast type and race have almost same as random survival forest trees. From the figure 10 we find, IIIC stage breast cancer have highest survival probability compare with other stages on the other hand, stages unadjusted NK stages have lowest survival. Similarly for the second plot of figure 10 we find that, N0 breast type have highest survival probability on the other hand, adjusted NX breast type have lowest survival rate. From the third plot of the same figure we see that, T1-4c stages have higher survival rate compare with T0-X,T1-4a, T1-4b and T1-4c stage breast cancer. All the outcomes is approximately same but some cases its not supported. From the figure 10 in forth plot we see that, white male breast cancer patients have higher survival probability compare with black respondents.

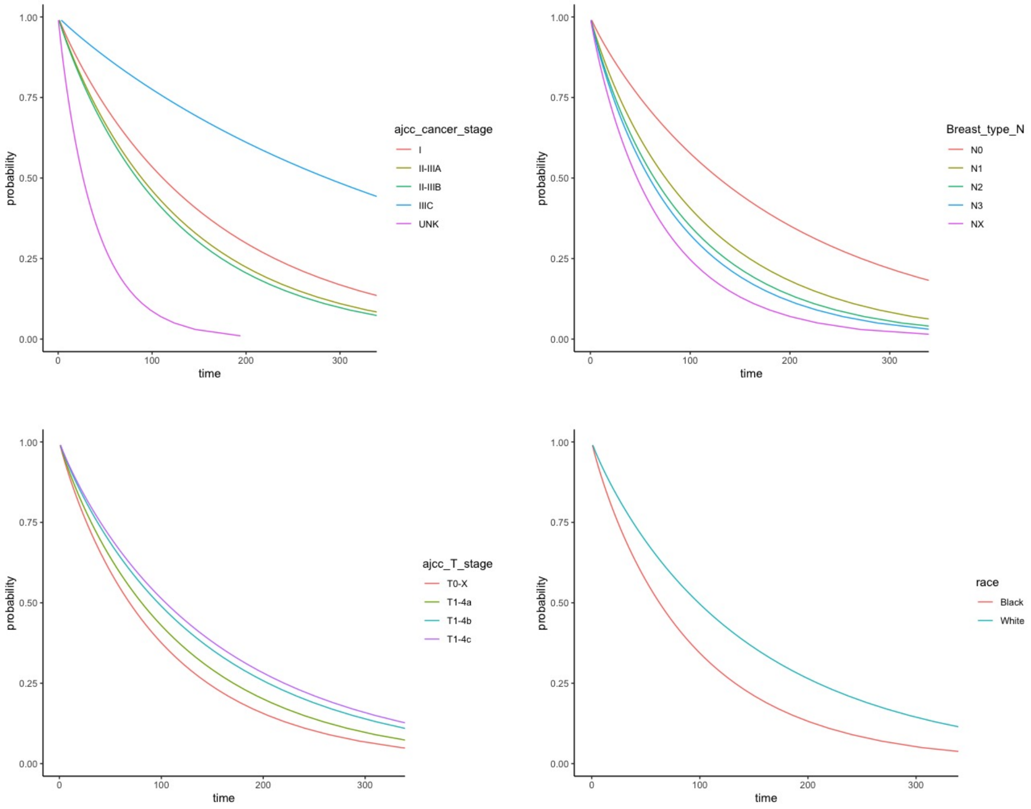


Figure 10: Survival Curve for ajcc cancer stage, breast type, Cancer stage T and race

By using Bayesian cox proportional hazard model we found all the predictors convergence and stable for estimation. From the figure 11 we see the the trace plots all are pity consistent and stable. In our model we select gamma density is the prior for all the predictors coefficients and Weibull as the priors for survival density. The trace plot here is shown in figure 11 for Weibull prior and independent Gaussian prior consider for the Bayesian joint frailty model . The density plot for the coefficients in figure 4.5 all the coefficient consistent and convergences.

4.5 MCMC Trace Plot

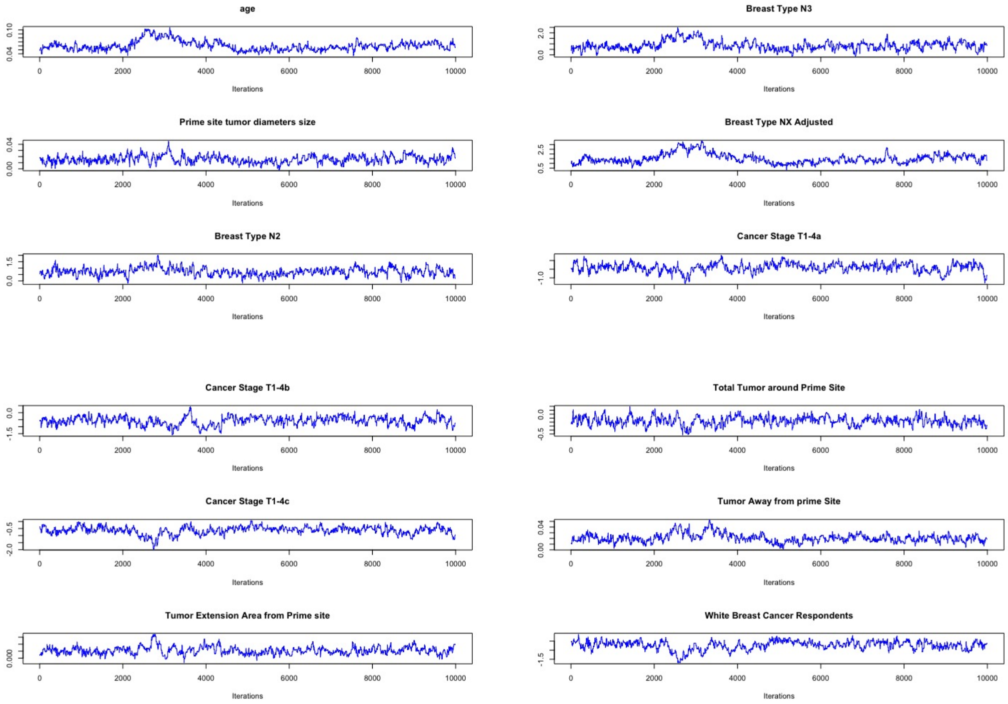


Figure 11: Trace Plot for all the coefficients

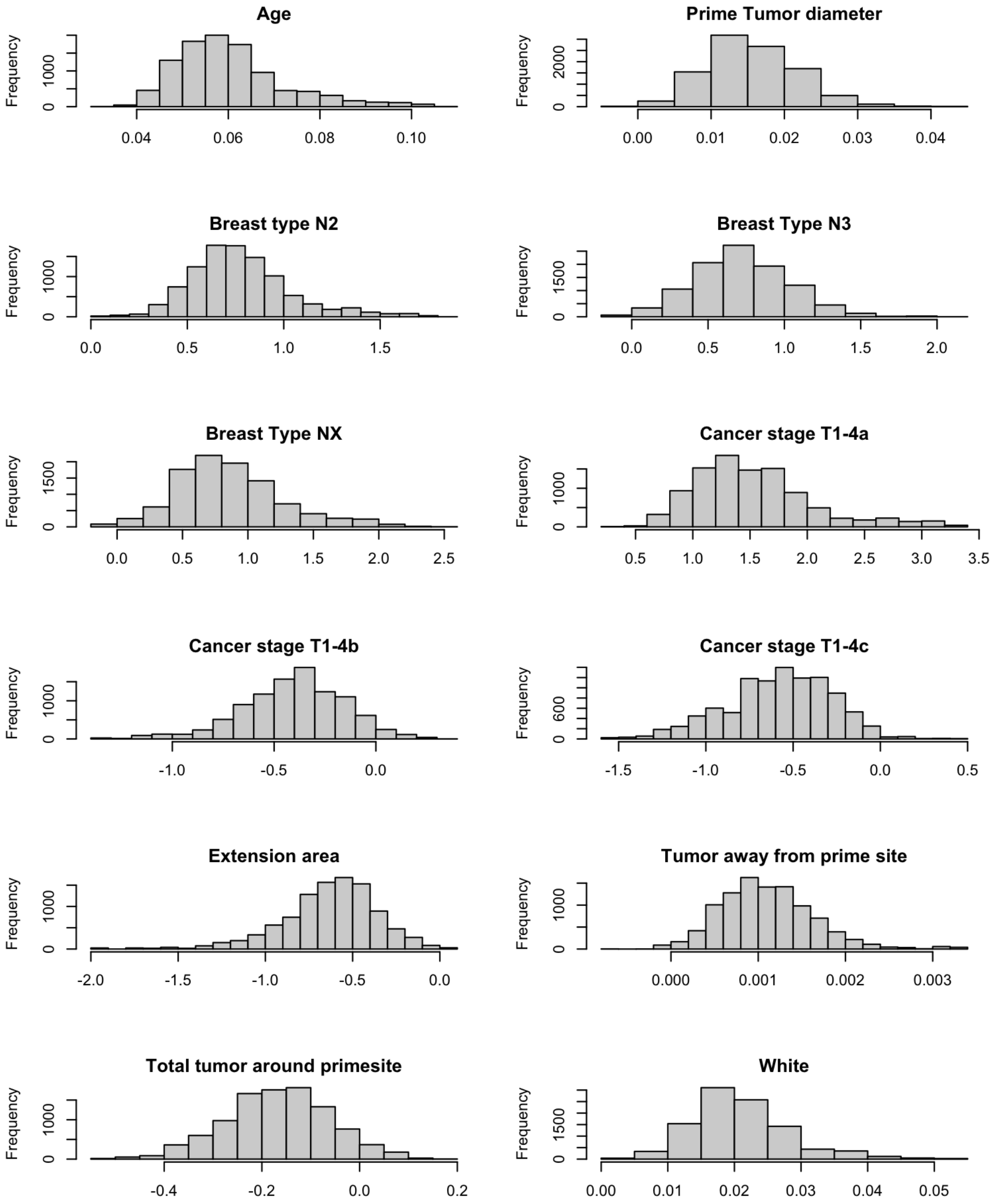


Figure 12: Coefficient Density plot

5 Comparison and Validations

For comparison the prediction performance of different model different methods are available. Usually prediction errors, mean square errors, AIC, BIC, sensitivity and specificity analysis, ROC curve, AUC and so on measurements are used. For comparison of random survival forest model and cox model, OOB survival, OOB mortality curve, AUC, Brier score and predictors errors are more popular. From the figures 5 we see that the Concordance index for different model with same predictors. From this plots 5 we see that C-index is higher for Random survival models compare with Bayesian Cox PH model and usual cox model. We also see that, from the figure 5 Bayesian Cox model have better performance compare with usual cox model.

We can also see that from the figure 5 in the second plots, the prediction error for random survival forest methods have lower error compare with all other models. The reference is the no predictors model and consider the time and censoring indicator are are only associated each others. From the figure 5 we aslo conclude that, Bayesian cox model have lower prediction error compare with cox model. So, for our dataset we can conclude that the machine learning approaches give better prediction with lower prediction errors and higher Concordance.

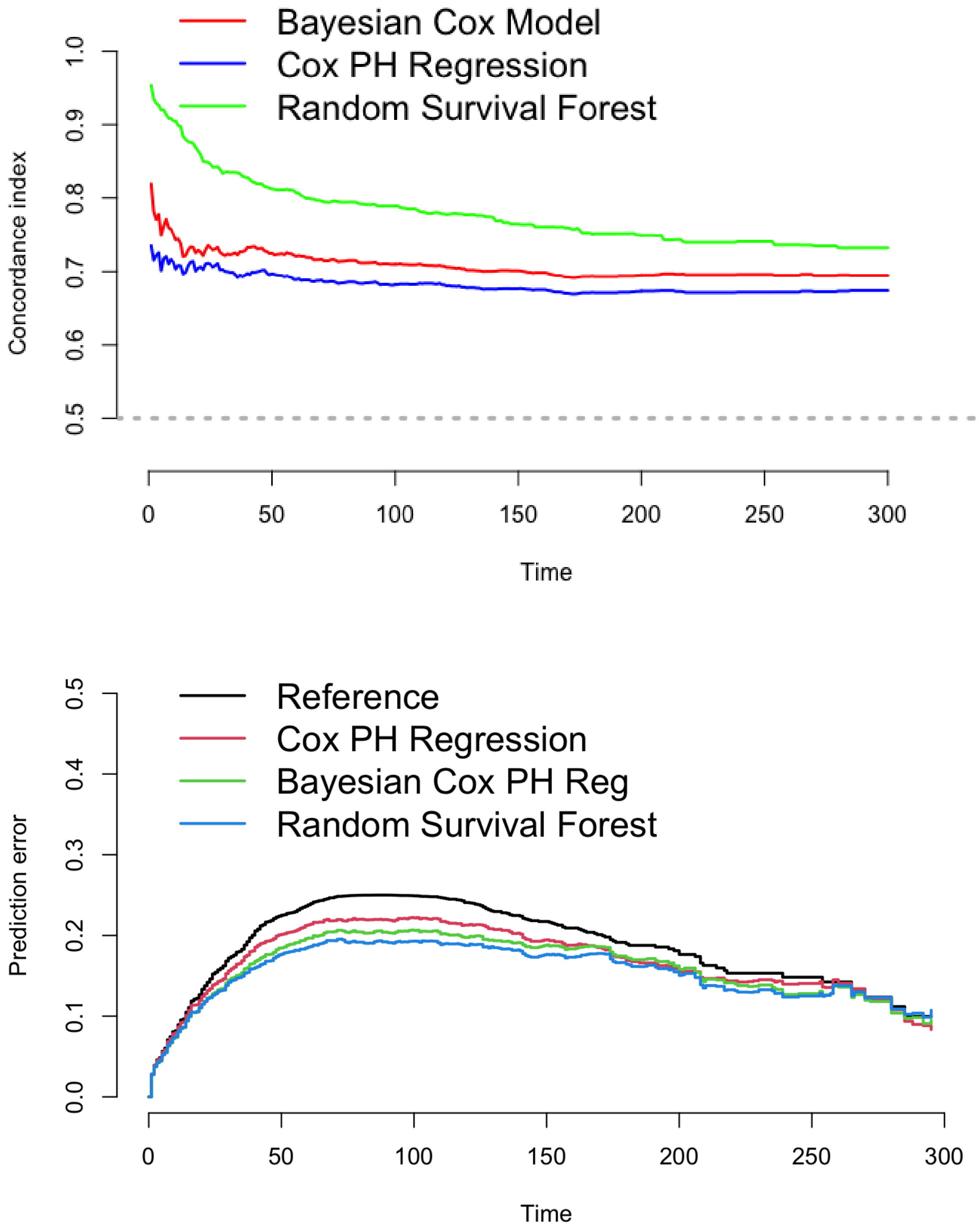


Figure 13: Fig 1: Concordance plot for Bayesian Cox model, Cox PH model and Random survival forest model and Fig 2: Prediction error Plot for 1: Reference Cox model 2: Cox PH model 3: Bayesian Cox Model and 4: Random survival forest model

We can also compare the prediction performance by using AUC time dependent ROC curve of Cox model and random survival forest model. In random survival forest model we can calculate the out of bag survival score and out of bag brier score. The approximate relation between AUC and OOB survival rate is $OOB-Survival-rate \approx 1-AUC$. So for compare with this two outcome from machine learning and Cox model, we can conclude that the out of bag survival rate is low compare with 1- AUC rate of cox model for different time points. Also for this figure 15 we can say that the survival forest model is pity consistent because the OOB brier score decrease after time 50. For the figure 15 we also find that OOB CRPS and OOB mortality plots consistent and follows decreasing trends. So, Machine learning methods are more relevant in this studies.

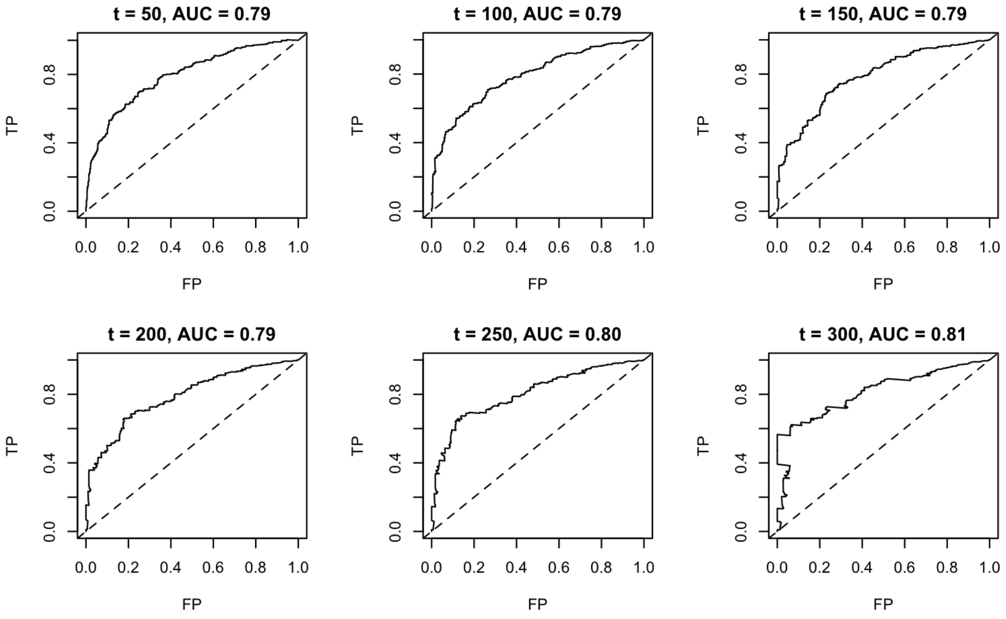


Figure 14: Bayesian Cox Model Fitted AUC curve for different time points

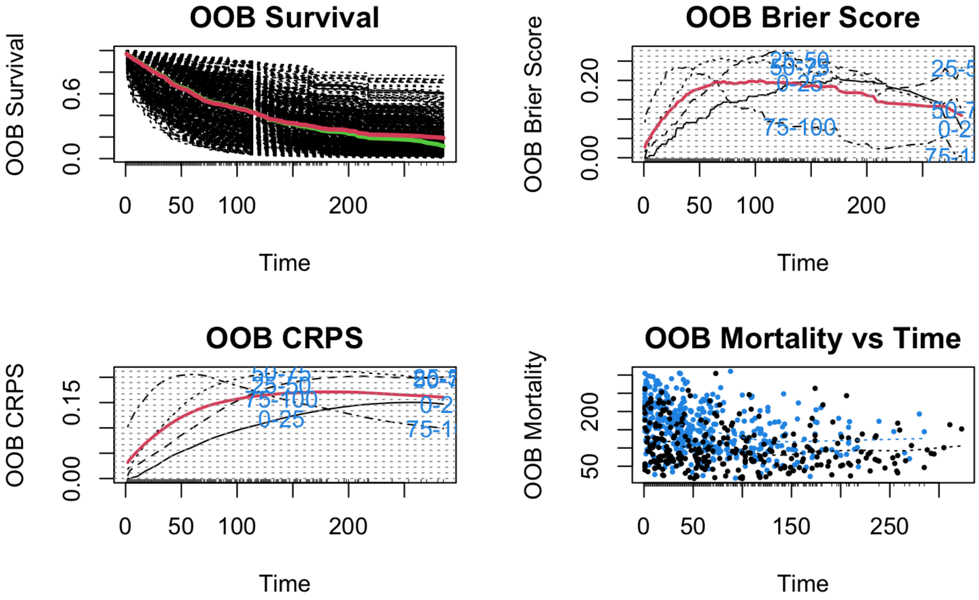


Figure 15: Random Survival Forest Error Curve for different Time Points

From the figure 5 we see that the fitted plot all predictors in the random survival forest methods. From this figures we see that the mortality rate decreasing from the starting age but after age 60 year, the rate increase very sharply. Similarly outcome find from the figure 5 in tumor size, that with increasing the tumor size, the mortality rate is also increasing. Similar finding have been found for primary tumor diameters size. For breast type, generally N0 have lower rate of mortality but NX adjusted have higher rate of mortality. From the figure also shown that , for tumor extension area from prime site not linear relation with mortality but for lymph node size have increasing trend with mortality. Most important factors tumor away from prime site is highly positive relation with mortality. From the figure 5, we see that white male breast cancer patients have lower rate of mortality compare with black respondents.

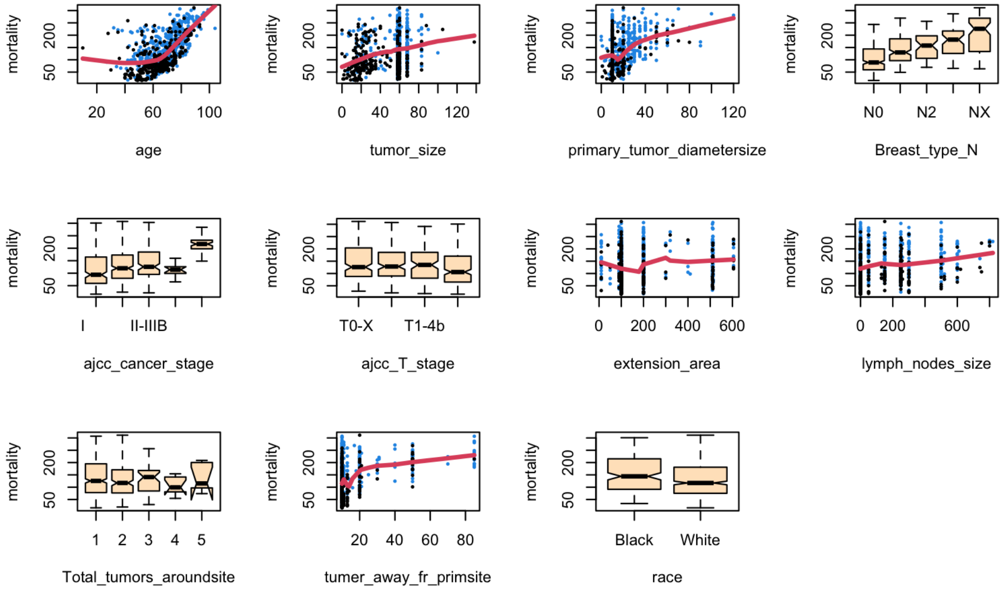


Figure 16: Random survival Model fitted plot for all the predictors curve

6 Conclusions

Overall, considering machine learning approaches may be more powerful tools for predicting the survivability of male breast cancer patients. So, the main purpose in this study was considering regression tree methods for survival data and predicted the survivability of male breast cancer in Detroit Michigan by this methods and compare with Bayesian cox model outcomes. By the random survival forest methods, we find that age is the most potential factors for death of male breast cancer patients. Random Survival Forest methods also suggest primary tumor size is less than 21mm, then the median survival time is higher compare with larger tumor size. Primary tumor diameter is an another important features that, with increasing of diameter, the change of survivability decreasing. With increasing distance of tumor from prime site, the median survival time decreasing. finally, breast type have a complex interaction among other features.

For Cox PH model and Bayesian Cox PH model we find approximately same outcomes but some predictors in Cox PH model shown insignificant. From Bayesian Cox model, for single year increase of age, approximately 5% chance increase of death (HR: 0.05, SD = 0.006). Similarly, for T1-4a stage cancer patients have approximately 30% lower chance of death compare with T0-X stages. Total tumor around prime site is negatively related with survival probability of male breast cancer patients. Another important things is, White breast cancer patients have lower rate of death compare with black breast cancer patients. For Weibull, Log-logistic and log-normal survival prior, we get approximately same outcomes but Weibull prior fit best for this data set considering log-PMLL, DIC, WAIC and Posterior inference frailty variance. Bayesian Cox model is more power full because it consider the MCMC method for generating random from well know prior distributions. In this study, we have very few numbers of respondents, so MCMC methods will be more applicable for generating samples and finally, we will compare the prediction power with machine learning outcomes.

References

Agarap, A. F. M. (2018). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing*, pages 5–9.

Amrane, M., Oukid, S., Gagaoua, I., and Ensari, T. (2018). Breast cancer classification using machine learning. In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pages 1–4. IEEE.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Gayathri, B. and Sumathi, C. (2016). Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–5. IEEE.

Heidari, M., Khuzani, A. Z., Hollingsworth, A. B., Danala, G., Mirniaharikandehei, S., Qiu, Y., Liu, H., and Zheng, B. (2018). Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. *Physics in Medicine & Biology*, 63(3):035020.

Ibrahim, J., Chen, M., and Sinha, D. (2001). Bayesian survival analysis springer series in statistics. *New York, NY: Springer. doi*, 10:978–1.

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

Jain, S. and Kumar, P. (2020). Prediction of breast cancer using machine learning. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 13(5):901–908.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., and Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial intelligence in medicine*, 50(2):105–115.

Kalbfleisch, J. D. (1978). Non-parametric bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2):214–221.

Kopans, D. B. (2008). Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology*, 246(2):348–353.

Obermeyer, Z. and Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216.

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548.

Sharma, A., Kulshrestha, S., and Daniel, S. (2017). Machine learning approaches for breast cancer diagnosis and prognosis. In *2017 International conference on soft computing and its engineering applications (icSoftComp)*, pages 1–5. IEEE.

Sharma, S., Aggarwal, A., and Choudhury, T. (2018). Breast cancer detection using machine learning algorithms. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pages 114–118. IEEE.

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14(4):323.

Vanneschi, L., Farinaccio, A., Mauri, G., Antoniotti, M., Provero, P., and Giacobini, M. (2011). A comparison of machine learning techniques for survival prediction in breast cancer. *BioData mining*, 4(1):1–13.

Zhou, H., Hanson, T., and Zhang, J. (2017). spbayessurv: fitting bayesian spatial survival models using r. *arXiv preprint arXiv:1705.04584*.

Zhou, Y. and McArdle, J. J. (2015). Rationale and applications of survival tree and survival ensemble methods. *Psychometrika*, 80(3):811–833.

7 Appendex

Relevant Code

```
# Require Package

‘‘‘{r, message=FALSE, warning=FALSE}
library(tidyverse)
library(readxl)
library(Publish)
library(survival)
library(survminer)
library(ggparty)
library(model4you)
library(plyr)
library(randomForestSRC)
library(randomForest)
library(survivalROC)
library(ggparty)
library(plyr)
library(prodlim)
library(pec)
library(spBayesSurv)
library(coda)
library("fields")
library("BayesX")
library("R2BayesX")
library(forestmangr)
library(rms)
‘‘‘

# Data input and Data cleaning
‘‘‘{r, message=FALSE, warning=FALSE}
male_BC<- read_excel("male_BC_no_missing.xlsx")
male_BC = male_BC %>% type_convert() %>% mutate(status = if_else(death ==

male_BC[sapply(male_BC, is.character)] <- lapply(male_BC[sapply(male_BC,

colnames(male_BC)

male_BC = male_BC %>%
  mutate(
    time = as.integer(time),
    age = as.integer(age),
    treat_primesite = as.integer(treat_primesite),
    treat_oth_no_surg = as.integer(treat_oth_no_surg),
    tumor_size = as.integer(tumor_size),
    extension_area = as.integer(extension_area),
    Reg_Node_Eval = as.integer(Reg_Node_Eval),
    lymph_nodes_size = as.integer(lymph_nodes_size),
    tumer_away_fr_primsite = as.integer(tumer_away_fr_primsite),
    primary_tumor_diametersize = as.integer(primary_tumor_diametersize),
    Total_tumors_aroundsite = as.integer(Total_tumors_aroundsite)
  )
```

```
male_BC = as.data.frame(male_BC)
male_BC
‘‘‘


# Variable imprtance selection
‘‘‘{r}
male_BC$status = as.factor(male_BC$status)
male_BC = as.data.frame(male_BC)
o <- rfsrc(status ~ ., data = male_BC %>% select(-time))
oo <- subsample(o)
‘‘‘
‘‘‘{r, fig.height=8, fig.width=12}
par(mfrow = c(1,2))
plot.subsample(oo, jknife = TRUE, main = "Jackknife␣Method␣VIMP")
plot.subsample(oo, alpha = 0.05, main = "Least␣Square␣method␣alpha␣=␣0.05
‘‘‘


‘‘‘{r}
fit = randomForest(status ~. , data = male_BC %>% select(-time))
library(caret)
impd = varImp(fit)
varImpPlot(fit, type = 2, main = "Important␣by␣rank")
‘‘‘
‘‘‘{r}
impd = varImp(fit)
impd = data.frame(impd)
df <- cbind(newColName = rownames(impd), impd)
rownames(df) = 1:nrow(df)

write.csv(df, "vipdata.csv")
# using excell plot bar diagram

‘‘‘

# Random survival tree
‘‘‘{r}
male_BC_treeData = male_BC %>% select(
  time,
```

~~~
  status,
  age,
  tumor_size,
  primary_tumor_diametersize,
  Breast_type_N,
  ajcc_cancer_stage,
  ajcc_T_stage,
  extension_area,
  lymph_nodes_size,
  Total_tumors_aroundsite,
  tumer_away_fr_primsite,
  race
)
male_BC_treeData$status = as.integer(male_BC$status)
male_BC_treeData
```

# Random Survival Forest tree
```{r}
vd.obj <- rfsrc(Surv(time, status)~ .,
data = male_BC_treeData,
ntree = 500, nodesize = 50)
par(mfrow = c(2,3))
p1 = plot(get.tree(vd.obj, 50))

p2 = plot(get.tree(vd.obj, 100))

p3 = plot(get.tree(vd.obj, 150))

p4 = plot(get.tree(vd.obj, 500))
p1;p2;p3;p4
```



```{r}
## classification
male_BC_class = male_BC_treeData %>% select(-time)
male_BC_class$status = as.factor(male_BC_class$status)
mod_c = rfsrc(status ~ ., data = male_BC_class, ntree = 500, nodesize = 5
~~~

~~~~
p1 = plot(get.tree(mod_c, 50, target = "0"))
p2 = plot(get.tree(mod_c, 100, target = "0"))
p3 = plot(get.tree(mod_c, 150, target = "0"))
p4 = plot(get.tree(mod_c, 500, target = "0"))
p1;p2;p3;p4
```


Party plots
```{r, fig.height=5, fig.width=16}
library(party)
stree = ctree(Surv(time, status) ~ . , data = male_BC_treeData)
plot(stree)
```

```{r, fig.height=6, fig.width=14}
dt = male_BC_treeData  %>% mutate(status = factor(status)) %>% select(-ti

stree = ctree( status ~ ., data = dt)
plot(stree, type = "simple")
plot(stree)
```

```{r, fig.height=4, fig.width=8}
dt = male_BC_treeData %>% select(time, status, tumor_size,extension_area,
## model
bmod <- survreg(Surv(time, status) ~ Breast_type_N, data = dt, model = TR
tree <- pmtree(bmod)


```{r, fig.height=4, fig.width=8}
dt = male_BC_treeData %>% select(time, status, tumor_size,extension_area,
## model
bmod <- survreg(Surv(time, status) ~ Breast_type_N, data = dt, model = TR
tree <- pmtree(bmod)


# get data for geom_node_plot's gglist
obs_nodes <- predict(tree, type = "node")
get_plot_data <- function(i, data) {
~~~~

```
    dat <- subset(data, obs_nodes == i)
    imod <- update(bmod, data = dat)
    gg <- survreg_plot(imod, data = dat)
    cbind(gg$data, id = i)
}
survplot_data <- ldply(unique(obs_nodes), .fun = get_plot_data, data = dt

# plot
p <- ggparty(tree) +
    geom_edge() +
    geom_edge_label() +
    geom_node_label(aes(label = splitvar),
        ids = "inner")

p + geom_node_plot(gglist = list(geom_line(data = survplot_data,
        mapping = aes(x = pr,
            y = probability,
            colour = Breast_type_N)
    ), xlab("time"),
        theme_classic()))
‘‘‘

Final to take this condition
‘‘‘{r, fig.height=5, fig.width=10}
dt = male_BC_treeData %>%
  select(time, status, tumor_size,extension_area, tumer_away_fr_primsite,
## model
bmod <- survreg(Surv(time, status) ~ ajcc_cancer_stage, data = dt, model
tree <- pmtree(bmod)


# get data for geom_node_plot's gglist
obs_nodes <- predict(tree, type = "node")
get_plot_data <- function(i, data) {
    dat <- subset(data, obs_nodes == i)
    imod <- update(bmod, data = dat)
    gg <- survreg_plot(imod, data = dat)
    cbind(gg$data, id = i)
}
survplot_data <- ldply(unique(obs_nodes), .fun = get_plot_data, data = dt
# plot
p <- ggparty(tree) +
    geom_edge() +
    geom_edge_label() +
    geom_node_label(aes(label = splitvar),
        ids = "inner")

p + geom_node_plot(gglist = list(geom_line(data = survplot_data,
        mapping = aes(x = pr,
            y = probability,
            colour = ajcc_cancer_stage)
    ), xlab("time"),
        theme_classic()))
‘‘‘

Final to take this condition
‘‘‘{r, fig.height=5, fig.width=10}
dt = male_BC_treeData %>%
  select(time, status, tumor_size,extension_area, tumer_away_fr_primsite,
## model
bmod <- survreg(Surv(time, status) ~ ajcc_T_stage, data = dt, model = TRU
tree <- pmtree(bmod)


# get data for geom_node_plot's gglist
obs_nodes <- predict(tree, type = "node")
get_plot_data <- function(i, data) {
    dat <- subset(data, obs_nodes == i)
    imod <- update(bmod, data = dat)
    gg <- survreg_plot(imod, data = dat)
    cbind(gg$data, id = i)
}
survplot_data <- ldply(unique(obs_nodes), .fun = get_plot_data, data = dt

# plot
p <- ggparty(tree) +
    geom_edge() +
    geom_edge_label() +
    geom_node_label(aes(label = splitvar),
        ids = "inner")

p + geom_node_plot(gglist = list(geom_line(data = survplot_data,
        mapping = aes(x = pr,
            y = probability,
            colour = ajcc_T_stage)
    ), xlab("time"),
        theme_classic()))
```

# Comparison

```{r}
mod = coxph(Surv(time, status) ~ ., data = male_BC_treeData)
male_BC$lp.mod <- predict(mod, type = "lp")
fun.survivalROC <- function(lp, t) {
    res <- with(male_BC,
                survivalROC(Stime        = time,
                            status       = status,
                            marker       = get(lp),
                            predict.time = t,
                            method       = "KM"))
    with(res, plot(TP ~ FP, type = "l", main = sprintf("t␣=␣%.0f,␣AUC␣=␣%
    abline(a = 0, b = 1, lty = 2)
    res
}

## 2 x 5 layout
layout(matrix(1:6, byrow = T, ncol = 3))

## Model with age and sex
res.survivalROC <- lapply(1:6 * 50, function(t) {
    fun.survivalROC(lp = "lp.mod", t) })
```

Plot for diagnosiss
```{r}
```

~~~
plot.survival(rfsrc(Surv(time, status)~ . , male_BC_treeData, cens.model
```


```{r}
v.obj <- rfsrc(Surv(time,status)~ ., data = male_BC_treeData, ntree = 100
plot.variable(v.obj, plots.per.page = 4)
```



Bayesian Cox model
```{r}
library(ggparty)
library(survival)
library(model4you)
library(plyr)
bmod = survreg(Surv(time, status) ~  race, data = male_BC_treeData, model
survreg_plot(bmod)
bmod = survreg(Surv(time, status) ~ Breast_type_N , data = male_BC_treeDa
survreg_plot(bmod)
bmod = survreg(Surv(time, status) ~ ajcc_cancer_stage, data = male_BC_tre
survreg_plot(bmod)
bmod = survreg(Surv(time, status) ~ ajcc_T_stage, data = male_BC_treeData
survreg_plot(bmod)
```



```{r}
# simulate data based on Weibull regression
library(prodlim)
library(pec)
library(survival)
library(randomForestSRC)
set.seed(13)
dat <- SimSurv(100)
dat = male_bayeData %>% select(-district)
# fit three different Cox models and a random survival forest
# note: low number of trees for the purpose of illustration
~~~

```
set.seed(111)
mcmc=list(nburn= 1000, nsave= 10000, nskip= 2, ndisplay=1000)
prior = list(maxL = 15)
cox12 <- survregbayes(formula = Surv(time, status) ~.
                        + frailtyprior("iid", district),
                     data = male_bayeData,
                     survmodel = "PH",
                     dist = "weibull",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E)
cox2 <- coxph(Surv(time,status)~ ., data=dat, x=TRUE, y=TRUE)
rsf1 <- rfsrc(Surv(time,status)~., data=dat, ntree=500, forest=TRUE)
#
# compute the apparent estimate of the C-index at different time points
#
A1 <- pec::cindex(list("Cox␣Model"=cox12, "Random␣Surival␣Model"=rsf1),
                  formula=Surv(time,status)~.,
                  data=dat,
                  eval.times=10)

ApparrentCindex <- pec::cindex(list("Bayesian␣Cox␣Model"=cox12,
                                    "Cox␣PH␣Regression"=cox2,
                                    "Random␣Survival␣Forest"=rsf1),
                               formula=Surv(time,status)~.,
                               data=dat,
                               eval.times=seq(1,300,1))
print(ApparrentCindex)
plot(ApparrentCindex, col = c("red", "blue", "green"), ylim = c(0.45, 1))
‘‘‘

dat = male_bayeData %>% select(- district)
# fit some candidate Cox models and compute the Kaplan-Meier estimate
Models <- list("Cox␣PH␣Regression"=coxph(Surv(time,status)~. ,
                              data=dat,x=TRUE,y=TRUE),
               "Bayesian␣Cox␣PH␣Reg"=survregbayes(formula = Surv(time, st
                        + frailtyprior("iid", district),
                     data = male_bayeData,
                     survmodel = "PH",
                     dist = "weibull",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E),
               "Random␣Survival␣Forest" =
                 coxph(Surv(time,status)˜., data=dat,x=TRUE,y=TRUE))

#rsf1 <- rfsrc(Surv(time,status)˜., data=dat,ntree=15,forest=TRUE)
# compute the apparent prediction error
PredError <- pec(object=Models, formula=Surv(time,status)˜.,
                 data=dat, exact=TRUE,
                 cens.model="marginal",
                 splitMethod="none",
                 B=0,
                 verbose=TRUE)
print(PredError,times=seq(5,30,5))

summary(PredError)
plot(PredError,xlim=c(0,300), ylim = c(0,.5))

‘‘‘

cox model
‘‘‘{r, fig.height=8, fig.width=10}
fit.coxph <- coxph(Surv(time, status) ˜., data = male_BC_treeData)
ggforest(fit.coxph, data = male_BC_treeData, main = "Hazard␣Plot␣for␣11␣P
‘‘‘


‘‘‘{R}
set.seed(111)
mcmc=list(nburn= 1000, nsave= 10000, nskip= 2, ndisplay=1000)
prior = list(maxL = 15)
res1 <- survregbayes(formula = Surv(time, status) ˜
                       age +
                       primary_tumor_diametersize +
                       Breast_type_N +
                       ajcc_T_stage +
                       extension_area +
```

~~~~
                       Total_tumors_aroundsite +
                       tumer_away_fr_primsite +
                       race +
                       frailtyprior("iid", district),
                     data = male_bayeData,
                     survmodel = "PH",
                     dist = "weibull",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E)
summary(res1)
sumar1 = summary(res1)
sumar1 = round_df(data.frame(sumar1$coeff), 3)
write.csv(sumar1, "summary1.csv")
```

```{r}
set.seed(111)
mcmc=list(nburn= 1000, nsave= 10000, nskip= 2, ndisplay=1000)
prior = list(maxL = 15)
res2 <- survregbayes(formula = Surv(time, status) ~
                       age +
                       primary_tumor_diametersize +
                       Breast_type_N +
                       ajcc_T_stage +
                       extension_area +
                       Total_tumors_aroundsite +
                       tumer_away_fr_primsite +
                       race +
                       frailtyprior("iid", district),
                     data = male_bayeData,
                     survmodel = "PH",
                     dist = "loglogistic",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E)
summary(res2)
sumar2 = summary(res2)
sumar2 = round_df(data.frame(sumar2$coeff), 3)
write.csv(sumar2, "summar2.csv")
~~~~

```
‘‘‘

‘‘‘{r}
set.seed(111)
mcmc=list(nburn= 1000, nsave= 10000, nskip= 2, ndisplay=1000)
prior = list(maxL = 15)
res3 <- survregbayes(formula = Surv(time, status) ~
                       age +
                       primary_tumor_diametersize +
                       Breast_type_N +
                       ajcc_T_stage +
                       extension_area +
                       Total_tumors_aroundsite +
                       tumer_away_fr_primsite +
                       race +
                       frailtyprior("iid", district),
                     data = male_bayeData,
                     survmodel = "PH",
                     dist = "lognormal",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E)
summary(res3)
sumar3 = summary(res3)
sumar3 = round_df(data.frame(sumar3$coeff), 3)
write.csv(sumar3, "summary3.csv")
‘‘‘


‘‘‘{r}
par(mfrow = c(3,1))
traceplot(mcmc(res1$beta[1,]),
          main = "age", col = "blue")
traceplot(mcmc(res1$beta[2,]),
          main = "Prime␣site␣tumor␣diameters␣size", col = "blue")
traceplot(mcmc(res1$beta[4,]),
          main = "Breast␣Type␣N2", col = "blue")
traceplot(mcmc(res1$beta[5,]),
          main = "Breast␣Type␣N3", col = "blue")
traceplot(mcmc(res1$beta[6,]),
          main = "Breast␣Type␣NX␣Adjusted", col = "blue")
traceplot(mcmc(res1$beta[7,]),
          main = "Cancer␣Stage␣T1-4a", col = "blue")
traceplot(mcmc(res1$beta[8,]),
          main = "Cancer␣Stage␣T1-4b", col = "blue")
traceplot(mcmc(res1$beta[9,]),
          main = "Cancer␣Stage␣T1-4c", col = "blue")
traceplot(mcmc(res1$beta[10,]),
          main = "Tumor␣Extension␣Area␣from␣Prime␣site", col = "blue")
traceplot(mcmc(res1$beta[11,]),
          main = "Total␣Tumor␣around␣Prime␣Site", col = "blue")
traceplot(mcmc(res1$beta[12,]),
          main = "Tumor␣Away␣from␣prime␣Site", col = "blue")
traceplot(mcmc(res1$beta[13,]),
          main = "White␣Breast␣Cancer␣Respondents", col = "blue")
‘‘‘

‘‘‘{r}
par(mfrow = c(3,2))
hist(mcmc(res1$beta[1,]), main = "Age", xlab = "")
hist(mcmc(res1$beta[2,]), main = "Prime␣Tumor␣diameter", xlab = "")
hist(mcmc(res1$beta[3,]), main = "Breast␣type␣N2", xlab = "")
hist(mcmc(res1$beta[4,]), main = "Breast␣Type␣N3", xlab = "")
hist(mcmc(res1$beta[5,]), main = "Breast␣Type␣NX", xlab = "")
hist(mcmc(res1$beta[6,]), main = "Cancer␣stage␣T1-4a", xlab = "")
hist(mcmc(res1$beta[7,]), main = "Cancer␣stage␣T1-4b", xlab = "")
hist(mcmc(res1$beta[8,]), main = "Cancer␣stage␣T1-4c", xlab = "")
hist(mcmc(res1$beta[9,]), main = "Extension␣area", xlab = "")
hist(mcmc(res1$beta[10,]), main = "Tumor␣away␣from␣prime␣site", xlab = ""
hist(mcmc(res1$beta[11,]), main = "Total␣tumor␣around␣primesite", xlab =
hist(mcmc(res1$beta[12,]), main = "White", xlab = "")
‘‘‘


Comparison Perfomence Measure

‘‘‘{r}
# Comparison of Weibull model and Cox model

male_BC = male_bayeData %>% mutate(status = if_else(status == 0,2,1))
```

```
f1 <- survregbayes(formula = Surv(time, status) ~ age +
            primary_tumor_diametersize +
            Breast_type_N +
            ajcc_T_stage +
            extension_area +
            Total_tumors_aroundsite+
            tumer_away_fr_primsite+
            race +
            frailtyprior("iid", district),
                     data = male_BC,
                     survmodel = "PH",
                     dist = "weibull",
                     mcmc = mcmc,
                     prior = prior,
                     Proximity = E)

f2 <- coxph(Surv(time,status)~ age +
            primary_tumor_diametersize +
            Breast_type_N +
            ajcc_T_stage +
            extension_area +
            Total_tumors_aroundsite+
            tumer_away_fr_primsite+
            race,
          data=male_BC,x=TRUE,y=TRUE)

f3 <- rfsrc(Surv(time,status)~
                age +
                primary_tumor_diametersize +
                Breast_type_N +
                ajcc_T_stage +
                extension_area +
                Total_tumors_aroundsite +
                tumer_away_fr_primsite +
                race,
              data = male_BC, ntree=500, forest=TRUE)

error <- pec(list("Bayesian␣Cox␣PH␣Reg"=f1,
                  "Cox␣PH␣Reg"=f2,
                  "Random␣Surival␣Forest"=f3),
             data=male_BC,
             formula=Surv(time,status!=0)~1)
plot(error, xlim = c(0,250),
     ylim = c(0, 0.8),
     col = c("black", "red","green","blue"))
```
```

... The End ...
```