A Comparative Study for Modelling the Survival of Breast Cancer Patients in the West of Iran

Mozhgan Safe¹, Hossein Mahjub¹,² & Javad Faradmal¹,³

¹ School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran
² Research Center for Health Sciences, Hamadan University of Medical Sciences, Hamadan, Iran
³ Modeling of Noncommunicable Disease Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

Correspondence: Javad Faradmal, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran. E-mail: javad.faradmal@umsha.ac.ir

Received: April 7, 2016   Accepted: June 25, 2016   Online Published: July 6, 2016
doi:10.5539/gjhs.v9n2p215          URL: http://dx.doi.org/10.5539/gjhs.v9n2p215

Abstract

Background: Breast cancer is the main cause of women cancer mortality. Therefore, precise prediction of patients’ risk level is the major concern in therapeutic strategies. Although statistical learning algorithms are high quality risk prediction methods, but usually their better prediction quality leads to more loss of interpretability. Therefore, the aim of this study is to compare ‘Model-Based Recursive Partitioning’ and ‘Random Survival Forest’; whether the partitioning, as the more interpretable learning technique, could be a suitable successor for forests.

Patients and Methods: The applied dataset for this retrospective cohort study includes the information of 539 Iranian females with breast cancer. To model the patients’ survival, various learning algorithms were fitted and their accuracy measures were statistically compared by means of several precision criteria.

Results: This study verified the stability of ‘Model-based Recursive Partitioning’, further to ‘Random Survival Forest’ deficiency to present a unique pervasive model. Moreover, except ‘Log-Logistic-Based Recursive Partitioning’, none of the methods significantly outperformed ‘Exponential- Based Recursive Partitioning’.

Conclusions: Briefly, it was concluded that the loss of interpretability due to the use of over complex models, may not always counterbalanced by the amount of prediction improvements.

Keywords: breast cancer, model-based recursive partitioning, random survival forest

1. Background

Breast cancer as the most common cancer among Iranian females (Aboutorabi, Hadian, Ghaderi, Salehi, & Ghasipour, 2015), is word widely the 18% proportional share of all cancers in women (Cvetković & Nenadović, 2016). Certainly, the prerequisites for the disease prevention and proficient treatment findings are the precise detection of affective risk factors in diseases formation and progression (Mert, Kılıç, Bilgili, & Akan, 2015). For a long time ago, conventional survival models have been used to achieve this aim and to provide necessary requisites for breast cancer prognostication (Ahmadinejad, Movahedinia, Movahedinia, Naieni, & Nedjat, 2013; Azarkish, Najmabadi, Roudsari, & Shandiz, 2015; Sadoughi, Afshar, Olfatbakhsh, & Mehrdad, 2016). Although the prominent interpretability of these models has made them pioneer for simple medical explanations, but the lower bias and accurate prediction of recently introduced learning algorithms, has stimulated the statistical focus on machine learning methods. The superior performance of these state-of-the-art learning techniques has been confirmed previously in many medical studies (Chao, Yu, Cheng, & Kuo, 2014; Dezfuly & Sajedi, 2015; Habibi, Ahmadi, & Alizadeh, 2015).

‘Model-Based Recursive Partitioning’ (MoBRP) could be referred as an exemplary interpretable machine learning technique (Zeileis, Hothorn, & Hornik, 2008a). Actually, this partitioning is a hybrid tree such that combines the traditional parametric survival models with newly introduced recursive partitioning methods (Zeileis et al., 2008a). By this way, MoBRP gains the profits of both classical and modern analysis; the prominent interpretability for recognized affective risk factors, in addition to its accurate survival time prediction
Further to MoBRP, ‘Random Survival Forest’ (RSF) could be referred as another outstanding learning technique (Ishwaran, Kogalur, Blackstone, & Lauer, 2008). In fact, this more intricate and time-consuming repetitious algorithm is an ensemble of survival trees and therefore would be so expert to find high order effects and interactions (Hastie, Tibshirani, Friedman, & Franklin, 2005).

Although the MoBRP ability to recognize nonlinearity and high order interactions, has made it superior to ordinary parametric survival models, but its detection ability for more complicated structures is hierarchically inferior to RSF. Briefly, more precise prognostication requires more costly computational analysis which usually provides less interpretable results; whereas, many publications have certified that the loss of interpretability due to the use of over complex algorithms may not always counterbalanced by the amount of prediction improvements (Haibe-Kains, Desmedt, Sotiriou, & Bontempi, 2008).

The RSF performance has been evaluated several times previously. For instance comparative studies, using both of RSF and Cox proportional hazard, for modeling the survival of patients with different cancers as breast (Kurt Omurlu, Ture, & Tokatli, 2009), prostate (Gerds, Kattan, Schumacher, & Yu, 2013), head and neck (Datema et al., 2012), as well as patients with systolic heart failure (Hsieh, Gorodeski, Blackstone, Ishwaran, & Lauer, 2011). Forests were also compared with variety of learning techniques (Mirmohammadi, Shishehgar, & Ghapanchi, 2014; Pang, Datta, & Zhao, 2010) and survival trees, as the forest elements (Yosefian, Mosa Farkhani, & Baneshi, 2015); but as far as we know, the RSF has never been compared with MoBRP. Therefore, the aim of this study is to compare the accuracy of predictions obtained with MoBRP and RSF; as if MoBRP which is more interpretable technique, could be a suitable successor for computationally expensive RSF.

2. Patients and Methods

2.1 Patients

In this retrospective cohort study, the information of 539 eligible women with breast cancer was gathered. All the involved patients have undergone, at least, one surgery for tumor extraction, from 1995 to 2013. The surgeries were under the supervision of the Diagnostic Center of Darolaytam-e Mahdieh of Hamedan, as the referential center in the west of Iran.

The interested event was death of breast cancer and the survival time was measured in days from surgery to death. Almost, 63% of patients were censored and had never experienced the event of interested, in the follow up duration.

2.2 Model-Based Recursive Partitioning Algorithm

Simply, MoBRP is a classification tree that is capable for parametric model fitting in each node of the tree (Zeileis et al., 2008a). Therefore, participied variables in MoBRP could be considered for two distinct objectives: (i) partitioning variables which are used for grow trees and forming the terminal nodes and (ii) model variables, which are used for explain the survival time in each node. It is worth noting that; these two types of variables could partially or totally be the same.

In order to grow this special tree, each node would be partitioned if the instability of the fitted model associated to it, is statistically significant through some partitioning variables; more precisely, for all terminal nodes and through each partitioning variable, the stabilities of models are assessed and the variable which is responsible for the most instability is selected as partitioning variable for its associated terminal node. Additionally, the cut point selection would be in favor of some objective functions to globally optimize the models of terminal nodes (Zeileis et al., 2008a)

2.3 Random Survival Forests Algorithm

The conclusion of a RSF algorithm is simply the average of its constituent trees. To grow a forest including B trees, the following steps are repeated B times to produce each tree (Hastie et al., 2005; Ishwaran et al., 2008). A random sample with replacement, and of size N, is drowning from the original observations. At each node of the tree, p covariates are randomly selected and the partitioning would be based on the covariate which can provide the largest survival differences between generated nodes. Note that the covariate and its cut point are selected by a heuristic search through the all terminal nodes and for each possible split point along. Furthermore, different criterion could be used for this difference assessment (Hastie et al., 2005; Ishwaran et al., 2008).

The growing would continue until a tree reaches its constrain about the least permissible number of observations at terminal nodes.
2.4 Performance Assessment

In order to compare the efficiency of the aforementioned risk prediction methods, the five most common accuracy measures, for nonparametric survival algorithms, were used: the time-dependent receiver operating characteristic (ROC) curve (Heagerty, Lumley, & Pepe, 2000), the sensitivity and specificity, the concordance index (Harrell, Lee, & Mark, 1996), the Brier score (Brier, 1950; Graf, Schmoor, Sauerbrei, & Schumacher, 1999) and the hazard ratio (HR) from Cox’s proportional hazards model (Cox, 1972).

2.5 Analysis Framework

In this study, the MoBRP was grown based on the four most common parametric survival models (Corbiere & Joly, 2007); as: Exponential, Weibull, Log-Logistic and Log-Normal.

The applied growing rules for RSF were Log-Rank and Log-Rank Score, that is, node splitting through maximization of log-rank and its standardized statistics, respectively (Ishwaran et al., 2008).

To derive MoBRP and RSF risk group prediction, the patients with the lower 33% of the predicted risk score were considered as low risk; and the remains as high risk groups. It should be mentioned; this experimental proportion has been certified in various breast cancer prognostications (Buyse et al., 2006; Haibe-Kains et al., 2008).

Since exponential is the simplest parametric model, it was selected as the benchmark for MoBRP, whether the more complexity of other methods can provide them superior prognostications. All the comparisons were statistically tested except the values of sensitivity and specificity; due to the lack of any proven statistical test.

Lastly, although the train and test sets were randomly selected from the dataset, but in order to demystify the style of subsets selection, all the analyses were repeated substituting the sets (Michiels, Koscielny, & Hill, 2005).

3. Results

The patients’ median lifetime was 8.85 years and the 5-year survival rate was 68% (95%CI: 64%-72%). The longest observed follow up duration was approximately 19 years. The variable’s importance, provided by both of forests, demonstrated the ‘Progesterone Receptor Status’ (PR) and ‘Number of Involved Lymph Nodes’, respectively as the two most relevant variables to explain the survival duration. These variables were also significant in all MoBRPs.

Table 1 represents the sensitivities and specificities for risk group predictions. Excepting Exponential-BRP, almost all sensitivities were more than 0.7. Although the most sensitivity were associated to RSFs, but they were able to provide more sensitivities only for train set and MoBRP methods were pioneer in the test set.

Furthermore, the specificities of the forests were lowest for test set, in spite of their moderate specificities for train set. Also, Exponential-BRP showed the most specificity despite its lowest sensitivities.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sensitivity</th>
<th></th>
<th>Specificity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train</td>
<td>test</td>
<td>train</td>
<td>test</td>
</tr>
<tr>
<td>Exponential-BRP</td>
<td>0.597</td>
<td>0.510</td>
<td>0.624</td>
<td>0.702</td>
</tr>
<tr>
<td>Weibull-BRP</td>
<td>0.747</td>
<td>0.779</td>
<td>0.461</td>
<td>0.496</td>
</tr>
<tr>
<td>Log-Logistic-BRP</td>
<td>0.819</td>
<td>0.789</td>
<td>0.417</td>
<td>0.494</td>
</tr>
<tr>
<td>Log-Normal-BRP</td>
<td>0.787</td>
<td>0.791</td>
<td>0.431</td>
<td>0.438</td>
</tr>
<tr>
<td>RSF Log-Rank</td>
<td>0.933</td>
<td>0.678</td>
<td>0.465</td>
<td>0.339</td>
</tr>
<tr>
<td>RSF Log-Rank Score</td>
<td>0.945</td>
<td>0.716</td>
<td>0.458</td>
<td>0.354</td>
</tr>
</tbody>
</table>

M-BRP: Model-Based Recursive Partitioning.

Addition to C-index and IBS, which are common for risk score and group assessments, IAUC and HR have been included in Tables 2 and 3, respectively. The results of risk score prediction were confirmed by the results of risk group prediction, through the signification of statistical tests involving C-indices and IBSs. Excluding IAUCs, almost none of the MoBRPs were significantly more accurate than Exponential-BRP, for each of the measures or each of the train or test sets. Moreover, RSFs were in agreement and their superiorities were declared, unanimously by all the measures; however, their excellence was specific to train set and they failed to significantly outperform in the test set. This paradox could also be seen for all the prediction methods, according to IAUCs.
Table 2. Performance for risk score prediction for each of the train and test set

<table>
<thead>
<tr>
<th>Model</th>
<th>C-index</th>
<th>IAUC</th>
<th>IBSC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train</td>
<td>test</td>
<td>train</td>
</tr>
<tr>
<td>Exponential-BRP</td>
<td>0.653</td>
<td>0.625</td>
<td>0.667</td>
</tr>
<tr>
<td>Weibull-BRP</td>
<td>0.672</td>
<td>0.615</td>
<td>0.658</td>
</tr>
<tr>
<td>Log-Logistic-BRP</td>
<td>0.681</td>
<td>0.609</td>
<td>0.669</td>
</tr>
<tr>
<td>Log-Normal-BRP</td>
<td>0.686</td>
<td>0.615</td>
<td>0.686</td>
</tr>
<tr>
<td>RSF Log-Rank</td>
<td>0.761**</td>
<td>0.556</td>
<td>0.773**</td>
</tr>
<tr>
<td>RSF Log-Rank Score</td>
<td>0.755**</td>
<td>0.556</td>
<td>0.763**</td>
</tr>
</tbody>
</table>

M-BRP: Model-Based Recursive Partitioning; *Significant at 5%; **Significant at 1%; All the tests were designed as if the accuracy of every method is more than Exponential-BRP.

Table 3. Performance for risk group prediction for each of the train and test set

<table>
<thead>
<tr>
<th>Model</th>
<th>C-index</th>
<th>HR</th>
<th>IBSC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train</td>
<td>test</td>
<td>train</td>
</tr>
<tr>
<td>Exponential-BRP</td>
<td>0.680</td>
<td>0.657</td>
<td>1.96</td>
</tr>
<tr>
<td>Weibull-BRP</td>
<td>0.707</td>
<td>0.706</td>
<td>2.05</td>
</tr>
<tr>
<td>Log-Logistic-BRP</td>
<td>0.762</td>
<td>0.705</td>
<td>2.47</td>
</tr>
<tr>
<td>Log-Normal-BRP</td>
<td>0.752</td>
<td>0.682</td>
<td>2.37</td>
</tr>
<tr>
<td>RSF Log-Rank</td>
<td>0.869**</td>
<td>0.527</td>
<td>3.87*</td>
</tr>
<tr>
<td>RSF Log-Rank Score</td>
<td>0.850**</td>
<td>0.587</td>
<td>3.19*</td>
</tr>
</tbody>
</table>

M-BRP: Model-Based Recursive Partitioning; *Significant at 5%; **Significant at 1%; All the tests were designed as if the accuracy of every method is more than Exponential-BRP.

Log-Logistic-BRP was the sole method which its risk score predictions outperformed, in both of test and train sets; though, this supremacy seems uncertain as only one measure, i.e. IBS, affirmed that. However, Log-Logistic-BRP has performed leastwise similar to other methods; therefore, another generic Log-Logistic-BRP was fitted by participation of all the observations in order to provide a pervasive infrastructure for physicians. In this implementation, a four-terminal-node tree was formed. The percent of high-risk patients according to risk group prediction were respectively, 17%, 34%, 67% and 85%, associated to each terminal node; and the median lifetimes were 108, 100, 98 and 89 months, correspondingly.

The stability of conclusions was confirmed by achieving the similar results after the replacement of training and test sets. As before, Exponential-BRP had the least sensitivity; however, it demonstrated most specificity in risk prediction. Although all the assessment criteria statistically certified more accuracy of RSF for this new training set, but none of them could discover any more precise for new test set.

4. Discussion

Since for none of the assessment measures, RSF significantly outperformed the Exponential-BRP in both of train and test sets; this investigation certified that the probable accuracy improvement caused by costly computational RSF does not compensate the loss of interpretability. The superiority of RSF only for its training set, testifies the RSF over-fitting and its deficiency to provide a generic pervasive model. Consistent with this conclusion, the RSF over-fitting has been cited as its disadvantage by many previous medical studies (Mirmohammadi et al., 2014). Whereas, other studies (Ishwaran, 2007) which have compared RF with Classification and Regression Trees, have claimed the RF more capability for over-fit controlling. The idea of seldom over fitting with random forest classification is also affirmed by other documentations (Hastie et al., 2005).

Considering Log-Logistic-BRP, the present order for the percentage of high-risk patients and the median lifetime related to terminal nodes clearly attests the ability of tree to divide the population to homogeneous subsets with the same risk levels. Therefore, in addition to MoBRP regression-structure, its excellent classification-structure should also be highlighted. Meanwhile, it is cited that the results of RSF would not be easy to interpret for clinicians, due to RSF inability for any classification (Walschaerts, Leconte, & Besse, 2012).

There are many studies which have estimated the risk factor importance of breast cancer by means of RSF. In line with our findings Ishwaran et al. and Kurt Omurlu et al. reported the importance of ‘PR’ and ‘Number of
Nodes’; but in contrast with us, they found ‘Number of Nodes’ as a more important risk factor (Ishwaran et al., 2008; Kurt Omurlu et al., 2009). In spite of these diagnosed factors, it has been claimed that the predicted importance by RSF is bias due to its bias of variable selection (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). To explain more, note that other than MoBRP which is a special unbiased tree, there are numerous studies that have discussed the bias of variable selection in the algorithm of usual tree models (Hothorn, Hornik, & Zeileis, 2006; Kim & Loh, 2001). Clearly, a forest composed of bias trees would lead to bias predictions of variable importance.

Addition to the mentioned factors, all the MoBRPs recognized the significant effects for ‘Tumor Size’; since the adverse effect of this factor was previously certified in many clinical researches (Faradmal, Soltanian, Roshanaei, Khodabakhshi, & Kasaeian, 2014); therefore, MoBRPs seems to be more able for risk factor recognition.

In partial agreement with our study, D’Eredita cited ‘Lymph Status’, ‘Tumor Size’ and ‘Histological Grade’ as the most informative medical factors (D’Eredita, Giardina, Martellotta, Natale, & Ferrarese, 2001). Also, the study of Delen et al. (Delen, Walker, & Kadam, 2005) affirmed the effectiveness of ‘Tumor Size’ through the sensitivity analysis of Artificial Neural Network. In agreement to our results, they found ‘Number of Involved Lymph Nodes’ as another important factor but in contrast to ours, they reported the ‘Stage of Disease’ as a more informative index. In their study decision tree showed the best prediction following with Artificial Neural Network and Logistic Regression. They also suggested that modern pattern recognition tools should be used as the complementary for traditional statistical modelling. It is worth noting that, their idea could be considered as a reference for MoBRP which combines the algorithm of both methods.

The prediction quality of RSF has been appraised many times. Bou-Hamad et al. assessed the RSF performance for predicting the survival of patients with primary biliary cirrhosis of the liver. The resulted IBS from a 10-fold cross validation certified the best implementation for RSF following with bagging (Bou-Hamad, Larocque, & Ben-Ameur, 2011). In another real data application to model the survival time of Iranian females with breast cancer, random forest showed the highest level of accuracy among other learning techniques (Montazeri, Montazeri, Montazeri, & Beigzadeh, 2015). In spite of mentioned studies in the context of learning algorithms, there are also comparisons between RSF and Cox, as the most widely used method for modeling the censored data (Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006). Through some of these comparative studies Cox has shown, not only the same (Hsieh et al., 2011), but also better performance than all diversity of forests (Datema et al., 2012). Additionally, the supremacy of Cox has been confirmed by other simulation studies (Kurt Omurlu et al., 2009). Evidently and in agreement to our findings, random forest is so prominent among learning algorithms; however, its computational complexity may not always guarantee its superiority over traditional models.

Other than MoBRP, random forest has been also compared with survival trees with so simpler structure. In a study which was designed to model the survival of Iranian patients suffering from acute myocardial infarction, both of IBS and C-index ascertained the RSF more precise results; furthermore, the difference between indices in training and test sets, evidenced the more generalizability of the forest (Yosefian et al., 2015). It should be added that the RSF less prediction error rates, has been also proven in another study in the field of breast cancer survival modeling (Walschaerts et al., 2012). Maroco et al. compared the performance of seven data mining classifiers in addition to three traditional models including logistic regression. They concluded that the overall accuracy of random forest out performed three different types of classification trees and logistic regression (Maroco et al., 2011). Eventually, although forest outperforms each of its constituent trees but our study statistically certified that MoBRP, as a less complex algorithm, can provide sufficient accuracy and less over-optimization.

For none of the aforementioned studies, statistical test was used to compare the accuracy measures. In actuality, the investigation of Haibe-Kains et al. is referring as the first study that has statistically compared the performance of learning algorithms for breast cancer prognostication from gene expression data (Haibe-Kains et al., 2008). In accordance with the present study, they claimed that the loss of interpretability as the consequences of complex models does not equilibrate the provided excess prices in breast cancer prognostication (Haibe-Kains et al., 2008).

It should be noted that the retrospective design of this study could be referred as its main limitation. Although, it was preferred to conduct a study including all Iranian females with breast cancer, but the target population of this survey only includes breast cancer patients from the west of Iran. Finally, regarding to the high observed censoring rate and the high possibility of being cured, variety of cure models seem to be a proper proposal for survival modelling and more investigations.
5. Conclusion

Briefly, more precise prognostication is in the wake of more costly computational analysis which usually provides less interpretable recognitions. This study certified MoBRP as a compromise between a high quality prediction and the ease of interpretation for clinicians; therefore, a good analytical model, in medical fields.

Acknowledgments

This investigation which is a part of Ph.D. thesis; is resulted by the fund of Vice Chancellor for Research and Technology of Hamadan University of Medical Sciences. We appreciate them and also the Diagnostic Center of Hamedan Mahdieh Darolaytam for helping us in collecting data.

Competing Interests Statement

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

References


Mirmohammadi, S. N., Shishehgar, M., & Ghapanchi, F. (2014). Applications of ANNs, SVM, MDR and FR


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).