International Journal of Advanced Health Science and Technology
Multidisciplinary : Rapid Review : Open Access Journal e-ISSN:2808-6422; p-ISSN:2829-3037

RESEARCH ARTICLE OPEN ACCESS

Manuscript received April 18, 2022; revised April 31, 2022; accepted May 31, 2022; date of publication June 20, 2022;
Digital Object Identifier (DOI): https://doi.org/10.35882/ijahst.v2i3.9

Development and Modeling of Decision Tree for Survival Data with Multiple Events Using Deviance and Cox-Snell Residuals within Node Homogeneity Technique

Kazeem Adesina Dauda
Department of Mathematics and Statistics, Kwara State University, Malete.

Corresponding author: Kazeem Adesina Dauda (e-mail: Kazeem.dauda@kwasu.edu.ng; qdauda70@gmail.com).

ABSTRACT It is very common in medical studies for a patient to experience more than one event rather than one of interest. This led to exposing an individual to multiple risks and medical practitioners need to account for these risks concerning some prognostic factors. There are many methods of dealing with multiple events in survival data classically, however, these methods break down when considering the top-down effect of the prognostic factors concurrently and when the risks of events are correlated (competing risks). This study aimed to develop a decision tree using a within-node homogeneity procedure in survival analysis with multiple events to classify individual risks for the competing risks. Since the CART methodology involves recursive portioning of covariates into different subgroups, this study considers the use of Deviance and Modified Cox-Snell residuals as a measure of impurity in the Classification Regression Tree (CART) during the process of partitioning. The flexibility and predictive accuracy of our learning algorithm would then be compared with other existing methods through simulation and the freely available online real-life data. The results of the simulation revealed that: using Deviance and Cox-Snell residuals as a response within the node homogeneity classification tree performs better than using other residuals irrespective of performance indices. Results from empirical studies of the two real-life data that the proposed model with Cox-Snell residual (Deviance=16.6498) performs better than both the Martingale residual (deviance=160.3592) and Deviance residual (Deviance=556.8822). Conclusively, using Cox-Snell residual (Mean Square Error (MSE)=0.01783563) as a measure of impurity in CART revealed improved performance than using any other residual methods (MSE=0.1853148, 0.8043366). This implies that the proposed methods have the capability of accounting for individual effects based on the prognostic biomarkers.

INDEX TERMS Martingale residual, cox-snell residual, deviance residual, CART, and within-node homogeneity.

I. INTRODUCTION

Survival data is very common in the field of epidemiology, public health, and biomedical sciences. A typical example of survival data is the study of patients diagnosed with covid-19 time until the patient was discharged alive or dead from the isolation centers [1]. Here only one event is of interest to the researcher, however, when considering another effect such as death due to other risks factors in the isolation center, then this is referred to as survival data with multiple events or simply competing risks [2]. In another word, when other events prevent the occurrence of events of interest, then, we say competing events (risk) exist [3]. Researchers such [4], [5], [6], and so on have applied the classical concept of analyzing survival data in the realm of nonparametric and semi-parametric methods. Most of these studies focus on the prediction of clinical outcomes and the identification of risk prognostic factors in multiple events survival analysis, ignoring the concept of machine learning (CART) that can efficiently predict the prognostic risk factors using a top-down tree procedure and identify possible interaction effects among

the biomarkers [7,8]. Therefore, the concept and fundamentals of machine learning techniques then become useful in the field of survival analysis for both low- and high-dimensional survival data [9,10,11].

Classification and Regression Tree (CART) is the most commonly and popularly used tool for exploring large data building. Moreover, it is one of the most stretchy, instinctive, and powerful information analytics and can be applied in the area of determining prognostic subgroups with the same results within each subgroup but different outcomes between the subgroup. It is a categorized, successive classification structure that recursively partitions the set of progressive observations into smaller subgroups based on the binary division of the covariates [12,13,14]. CART was first brought together by Breiman et al. [15] and it gaining more attention recently because of its predictive ability in model building (tree-building) for difficult data mining problems [16,17]. There are other numbers techniques that can handle this difficulty in machine learning, they include Neural Network [18,19], Deep Neural Network [20,21], Multivariate Adaptive Regressions Spline[12], Support Vector Machine [22], and so on. However, all these studies were not directly applicable to the survival data with competing risks and their results were rely on the case of a single event. Some of the previous studies considered the use of mean absolute error (MAE) and the sum of square error (SSE) for the numerical response, which cannot hold in survival data sets.

In this research, we focus on a reliable decision tree (within-node homogeneity), which will help in categorizing new observations into groups. The purpose of this study is that the existing classical statistical tools such as the sub-distribution hazard function, cause-specific hazard function, and so on; are inappropriate in this scenario, or of limited utility in addressing these types of classification problems. There are several reasons for this drawback. Firstly, the difficulties in variable selection, this is due to the predictors themselves. Secondly, different degree of variation or variance of the predictors occurs as a result of the predictors been failed to follow a normal distribution. Thirdly, the existence of complex interactions and patterns. Lastly, the results of traditional methods may be difficult to interpret in this setting. All these difficulties make classification and regression trees robust over the entire classical methods and it is Supervised machine learning and also known as a non-parametric.

II. METHODOLOGY

The single or univariate event decision tree type refers to whether or not the device used to create each split discriminates based on a single event. In competing risks survival analysis, under the univariate event, a tree will be built by focusing on the event of interest and adjusting for other risks. We proposed to use within-node homogeneity by using Deviance, Cox-Snell, and Martingale residuals in the sub-distribution hazard function developed by Fine and Gray [23] i.e. hazard function associated with Cumulative Incidence Function (CIF). The sub-distribution hazard function $\tilde{\lambda}(t)$ is assumed to take the form:

$$\tilde{\lambda}(t; Z_i) = \tilde{\lambda}_{j0}(t)\exp\left[\beta^j I(Z_i < c)\right] \quad (1)$$

Where $\tilde{\lambda}_{j0}(t)$ is an unspecified baseline sub-distribution and β^j is an unknown parameter corresponding to cut-point c.

A. GENERAL ALGORITHM TO GROW TREE

The impurity functions for every possible binary split of the predictor space Z will be considered. The split could be of several forms: splits on a single covariate splits on linear combinations of predictors, and Boolean combinations of splits. The simplest form, in which each split relates to only one covariate, can be described as follows:

1. If a covariate Z is continuous or ordinal, then we look at all the possible cut points c that divide the sample into two groups, $Z \leq c$ and $Z > c$.
2. If a covariate is nominal, then we consider all the possible ways of forming the two groups.

We repeat this process for all subsequent branches of the tree until we reach one of the predefined stopping criteria. The usual criteria are that the sample size in each terminal node is very small, the subgroups are homogeneous with respect to the covariates, or F1 cannot be estimated (e.g., the subgroup contains no observations of the event of interest).

B. BEST SPLIT

The "best split" is defined to be the one corresponding to the minimum statistic (e.g Gray, deviance, etc). Subsequently, the data are divided into two groups according to the best split. Apply this splitting scheme recursively to the sample until the predictor space is partitioned into many regions. There will be no further partition to a node when any of the following occurs:

1. The node contains less than, say 10 or 20, observations.
2. All the observed times in the subset are censored.
3. All the observations have identical covariate vectors or the node has only complete observations with identical survival times.

This procedure results in a large tree (ω), which could be used for data structure exploration.

C. ALGORITHM TO PRUNE TREE

The idea of pruning is to iteratively cut off branches of the initial tree, ω, in order to locate a limited number of candidate sub-trees from which an optimally sized tree is selected. For the proposed method, we adopt a pruning algorithm that exerts little computational burden. The steps for adopting this algorithm are as follows:

a. Initially growing a large tree.
b. To each of the internal nodes in the original tree, assign the maximal splitting statistics contained in the corresponding branch. This statistic reflects the strength of linking the branch to the tree.
c. Among all these internal nodes, find the one with the smallest statistic. That is, find the branch that has the weakest link and then prune off this branch from the tree.

d. The second pruned tree can be obtained similarly by applying the above two steps to the first pruned tree.
e. Repeating this process until the pruned tree contains only the root node, finally, a sequence of nested trees is obtained. The desired tree can be obtained by plotting the size of these trees against their weakest linking statistics. The tree corresponding to the "kink" point in the curve is chosen as the best one.

D. IMPURITY MEASURES

Two different impurity measures were considered in the study, including sum-of-square and absolute impurities measures.

a. **Sum-of-squares impurity function** We propose to utilize the Cox-Snell (Cs) and Deviance (D) residuals as a result measure for occasion 1 risk.

For individual i in group k, the Cox-Snell (Cs) for the event of interest (j = 1) is defined as follows:

$$\hat{C}s_{ik}^{1} = I_{\{\delta_{ik=1}\}} - \widehat{\Lambda}_0^1(T_{ik}). \quad (2)$$

Where $\widehat{\Lambda}_0^1(.)$ Is the estimated cause-specific cumulative hazard function. This function can be calculated by $\widehat{\Lambda}_0^1(T_{ik}) = \int_0^{T_{ik}} \frac{dN_{1k}(S)}{Y_k(s)}$

Where $N_{1k}(t)$ and $Y_k(t)$ are the number of event 1 and number of risks at time t for group k, respectively.

For a node h, we proposed using an impurity function $\phi_{ss}(h)$ based on the sum of the square cox-Snell variations,

$$\phi_{ss}(h) = \sum_{i\epsilon h}(\hat{C}s_{ih}^{1} - \overline{Cs}_h^1)^2, \quad (3)$$

where $\overline{Cs}_h^1 = \frac{\sum_{i\epsilon h}\hat{C}s_{ih}^{1}}{N_h}$.

Let s be a possible split for the parent node P , and let L and R be the corresponding left child node and right child node, respectively, for this split. To obtain the split that maximizes the within-node homogeneity, we need to find an s that has the largest reduction in within node impurity from the parent node P. That is, we need to find an s that maximizes the impurity function. $\Phi_{SS}(s,P) = \phi_{ss}(P) - \phi_{ss}(L) - \phi_{ss}(R)$. The process can be simplified as

$$\Phi_{SS}(s,P) = \frac{N_L N_R}{N_P}(\overline{Cs}_L^1 - \overline{Cs}_R^1)^2 \quad (4)$$

This process is repeated until we research the stopping criteria stated above.

Similarly, for individual i in group k, the Deviance (D) for the event of interest (j = 1) is defined as follows:

$$\widehat{D}_{ik}^{1} = I_{\{\delta_{ik=1}\}} - \widehat{\Lambda}_0^1(T_{ik}). \quad (5)$$

For a node h, we proposed using an impurity function $\phi_{ss}(h)$ based on the sum of the square Deviance residual,

$\phi_{ss}(h) = \sum_{i\epsilon h}(\widehat{D}_{ih}^{1} - \overline{D}_h^1)^2$, where $\overline{D}_h^1 = \frac{\sum_{i\epsilon h}\widehat{D}_{ih}^{1}}{N_h}$.

To maximizes the impurity function,

$\Phi_{SS}(s,P) = \phi_{ss}(P) - \phi_{ss}(L) - \phi_{ss}(R)$. The process can be simplified as

$$\Phi_{SS}(s,P) = \frac{N_L N_R}{N_P}(\overline{D}_L^1 - \overline{D}_R^1)^2. \quad (6)$$

b. **Absolute value impurity function**.

Consequently, we propose using the following impurity function based on the absolute value function using Cox-Snell (CABS) and Deviance (DABS) residuals. Then, the absolute sum of the impurity function is generally defined for both residuals as

$\Phi_{ABS}(s,P) = \phi_{ABS}(P) - \phi_{ABS}(L) - \phi_{ABS}(R)$, for each split when growing the tree.

In all, four methods were proposed these are deviance sum of square (DSS), deviance absolute sum (DABS), the cox-snell sum of square (CSS), and cox-snell absolute sum (CABS) impurity measures respectively. The existing methods include the martingale sum of square and martingale absolute sum impurity measure proposed in [14].

E. SIMULATION PROCEDURE

We investigate the ability of our proposed models to detect the structure in data with competing risks. In TABLE 1, models 1 and 2 were spawned from exponential distributions, and models 3 and 4 were also spawned from Lognormal distributions. Two covariates Z_1 and Z_2 were related to survival times, and four other covariates Z_3 *to* Z_6 were not related to survival times. The covariates Z_1, Z_3, and Z_4 were dichotomous variables generated from a binomial distribution with parameter $p = 0.5$. While Z_2, Z_5, and Z_6 were generated

from a uniform distribution with the following parameters: (a, b) for Z_2, (a, b) for Z_5, and (a, b) for Z_6, respectively. 0, 0.25, 0.5, and 0.75 and 1. We consider a 50% censoring rate for all the simulations on both additive and interaction between Z_1 and Z_2 models. 100 trees were constructed with 500 sample sizes for each simulation. To measure the performance of our model, both the number of terminal nodes in each tree and the predictive ability of the tree were considered. We generated validation data sets with 500 observations without censoring for each simulation and calculate the mean absolute difference between the event 1 failure time for each observation in a validation data set and the median of event 1 failure time predicted by the tree. Similarly, this is also repeated in the second event, and below is the computation procedure for event 1 and event 2 times respectively;

$$MAE_1 = \frac{1}{N_1}\sum_{h=1}^{|T^t|}\sum_{i=1}^{N_{ih}}|T_{i1h} - \hat{\tau}_{1h}|, \quad (7)$$

$$MAE_2 = \frac{1}{N_2}\sum_{h=1}^{|T^t|}\sum_{i=1}^{N_{ih}}|T_{i2h} - \hat{\tau}_{2h}| \quad (8)$$

Where N_{1h} is the number of type 1 events in node h, N_1 is the total number of type 1 events, T_{i1h} is the ith failure time for event 1 in terminal node h, and $\hat{\tau}_{1h}$ is the median event 1 failure time based on the training data for terminal node h. This interpretation is also applicable to event 2 when 1 is chosen as the complexity penalty function. 20 numbers of minimum observations in each terminal node are allowed and a 10-fold Cross Validation method was used to select the final tree. All these simulations were similar to Fiona's [14].

TABLE 1
Description of models used for simulation in data structure performance

Model	Cumulative incidence function $F_1(t/Z_1, Z_2)$	Expected terminal nodes $\|T^t\|$
1	**Exponential** $\lambda = 0.1 + 0.3\mathrm{I}\{(Z_1 > 0.5) \cap (Z_2 > 0.5)\}$	3
2	**Exponential** $\lambda = 0.05 + 0.2\mathrm{I}(Z_1 > 0.5) + (Z_2 > 0.5)$	4 or 2
3	**Log normal** $\mu = 2 + 1.5\mathrm{I}\{(Z_1 > 0.5) \cap (Z_2 > 0)\}, \sigma = 1$	3
4	**Log normal** $\mu = 2 - 0.85\mathrm{I}(Z_1 > 0.5) + 0.85Z_2, \sigma = 1$	4 or 2

III. NUMERICAL RESULTS

A. RESULTS OF SIMULATION STUDIES

Simulation studies were conducted on models presented in TABLE 1 to assess the performance of the proposed model and whether its suits the data in terms of data structure. From TABLE 2 and 3, when mean and median were used to evaluate the performance of the proposed model, we observed that the proposed model has a value that is much closer to the cut point estimate than the existing (MSS & MABS) techniques irrespective of the distributions (i.e. whether the distribution is exponential or log-normal) and the censoring rates.

TABLE 2
Examining tree structure with one event-of-interest, through exponential distribution. N=500

	Percent terminal nodes in the final tree							Average	
	1	2	3	4	5	6	≥ 7	MAE1	MAE2
Model 1: Exponential $\lambda = 0.1 + 0.3\mathrm{I}\{(Z_1 > 0.5) \cap (Z_2 > 0.5)\}$									
MSS	13	3	4	4	3	13	3	1.5336	1.5098
MABS	60	14	13	3	3	4	3	1.5340	1.5190
DSS(Proposed)	59	13	13	4	4	4	3	1.5310	**1.5095**
DABS(Proposed)	60	15	**14**	3	3	3	2	1.5318	1.5160
CSS(Proposed)	64	7	4	7	5	6	7	1.5273	1.5185
CABS(Proposed)	86	4	1	1	4	4	0	**1.5246**	1.5172
	Percent terminal nodes in the final tree							Average	
	1	2	3	4	5	6	≥ 7	MAE1	MAE2
Model 2: Exponential $\lambda = 0.05 + 0.2\mathrm{I}(Z_1 > 0.5) + (Z_2 > 0.5)$									
MSS	60	12	11	5	4	4	4	1.5143	1.5265
MABS	79	6	5	3	2	3	2	1.5134	1.5224
DSS(Proposed)	60	**13**	10	5	4	4	4	**1.5025**	1.5268
DABS(Proposed)	70	10	8	4	2	3	3	1.5144	1.5235
CSS(Proposed)	66	6	6	5	6	6	5	1.5131	1.5237
CABS(Proposed)	89	2	2	3	0	2	2	1.5112	**1.5205**

TABLE 3
Examining tree structure with one event-of-interest, through Log-normal distribution. N=500

	Percent terminal nodes in the final tree							Average	
	1	2	3	4	5	6	≥ 7	MAE1	MAE2
Model 3: Log normal $\mu = 2 + 1.5\mathrm{I}\{(Z_1 > 0.5) \cap (Z_2 > 0)\}, \sigma = 1$									
MSS	60	13	13	3	4	3	4	1.6826	2.0103
MABS	60	14	**14**	3	3	2	4	1.6841	2.0121
DSS(Proposed)	60	13	13	3	4	3	4	**1.6826**	**2.0075**
DABS(Proposed)	59	14	**14**	3	4	3	3	1.6829	2.0102
CSS(Proposed)	62	13	7	5	4	5	4	1.7326	2.0208
CABS(Proposed)	78	15	5	0	0.5	0.5	1	1.7581	2.0127
	Percent terminal nodes in the final tree							Average	
	1	2	3	4	5	6	≥ 7	MAE1	MAE2
Model 4: Log normal $\mu = 2 - 0.85\mathrm{I}(Z_1 > 0.5) + 0.85Z_2, \sigma = 1$									
MSS	59	13	13	3	4	5	3	1.7128	1.9956
MABS	60	14	13	2	4	4	3	1.7123	**1.9867**
DSS(Proposed)	60	12	12	3	5	5	3	1.7123	2.0007
DABS(Proposed)	60	13	13	3	4	4	3	**1.7097**	1.9904
CSS(Proposed)	63	12	6	**5**	5	5	4	1.7596	1.9988
CABS(Proposed)	78	14	3	1	1	2	1	1.7624	1.9946

Additionally, when the measure of spread and error were taken into consideration, the proposed model still out-fit the existing model except when the censoring rates is high in the Log-normal distribution. The results revealed the impact of censoring rates in competing risks and address how to choose the appropriate cut point in the regression tree. Interestingly, FIGURE 1 to 3 also justifies the efficacy of the proposed models based on both roots mean square error and mean absolute error. Finally, the boxplot virtual display distinguishes between the proposed model and the existing model and it also revealed the performance of the proposed model to detect cut points more perfectly than the existing model. In some cases, we also observed that the existing techniques out-fit the proposed model only when the censoring rates are low with a smaller cut point estimate as was shown above.

The bold-faced in TABLE 2 and 3 represent the ideal number of nodes. For the multiplicative hazard model, three-terminal nodes are expected and in the additive hazard model, 4 terminal nodes are of interest. The average Mean Absolute Error was considered to check the performance of our grown tree. Results from this simulation revealed that the proposed models fit the data structure more than the existing methods based on the rate at which they detect the terminal nodes. Additionally, the Mean Absolute Error of the proposed models both on exponential and log-normal distributions was less than the existing model

B. REAL-LIFE DATA

To further investigate the performance of the proposed models, the study wishes to apply the proposed model to real-life examples. We consider two real-life data both in bone marrow transplant [24] and prostate cancer [25].

TABLE 4
Results of bone marrow transplant through cause 1 (relapse)

Models	Deviance	Number of terminal nodes	The first covariate to be split	Mean of Squared Residuals
Proposed model 1	556.8822	4	Age	1.2703
Proposed model 2	**16.6498**	7	Age	**0.0376**
Existing Model	160.3592	4	Age	0.3669

In TABLE 4 the results from the various methods were presented based on the bone marrow transplant when relapse was considered an event of interest. The deviance results were used as a measure of model selection and models with the lowest Deviance values are generally preferred over other models. It was observed that when using the proposed model 2 (cox-snell) residual as a common response in univariate within node homogeneity provides a deviance value that is lower than other existing methods. Interestingly, when the mean squared error was considered a measure of model selection, the proposed model 2 was also found to be the most effective method in classification and regression tree.

Moreover, the virtual display in FIGURES 1, 3, and 5 above revealed that the explanatory variable age plays a key role in constructed classification trees along with other covariates therein irrespective of the models. This suggests that the survival wood of patients with bone marrow transplants rests on the age of that patient while other covariates (factors) follow.

Another important statistical tool to be considered is called variable importance. The variable importance assigns the highest values to variables with the greatest discrepancy between original prediction performance and prediction performance after permutation. The resulting variable importance values have no meaning on an absolute scale, but their relative sizes can be useful for comparing across different predictor variables (for full details see [26] and [27]). Results right of FIGURE 2, 4, and 6 revealed that age is an important factor in bone marrow transplant when relapse is of interest since it has a longer bar than other factors such as platelets and tcells. Additionally, the left-hand sides of FIGURE 2, 4, and 6 represent the possible number of trees grown before the tree converges (stabilized). The conversion was observed to be at around 200 trees at was shown and this authenticates the 200 number trees grown in the simulation studies.

TABLE 5
Results of bone marrow transplant through cause 2 (death in remission)

Models	Deviance	Number of terminal nodes	The first covariate to be split	Mean of Squared Residuals
Proposed model 1	372.4276	3	Age	0.9120
Proposed model 2	**49.5625**	3	Age	**0.1175**
Existing Model	86.5535	4	Age	0.2120

Next, we consider the death in remission and the corresponding results were shown in TABLE 5 above. Equivalently, deviance and mean squared residuals were used to measure the performance of the proposed models and the existing model. However, it was also observed that the proposed model 2 (cox-snell) performs better than other empirical methods based on the two selected measurements mentioned earlier.

Besides, the tree display in FIGURES 7, 9, and 11 revealed that the covariate age plays a key role in constructing classification trees when remission is the variable of interest in bone marrow transplant while other explanatory variables follow. These results are similar to the results generated when relapse is of interest.

Finally, FIGURE 8, 10, and 12 show how the dependent variables (age, platelet, and Tcells) are important to the outcome of death in remission based on a certain period. Consistently, age was found to be the most important covariate in determining the outcome of death in remission

in bone marrow transplants. This is also similar to results of relapse when variable importance was considered.

IV. DISCUSSION

We have considered two proposed methods in classification and regression tree and compare our results with an existing method. But not only covered the survival analysis but also considered the competing risk event with just two events of interest. Moreover, we have discussed and compared different measures of impurity used for both proposed and existing methods through the cumulative incidence function (CIF) in the within-node homogeneity tree. In particular, we have compared the use of deviance and cox-snell residuals as a common response in the classification tree with the martingale residual.

Based on the simulation results in TABLE 2 and 3, it was observed that the results showcase the tree selection. The proposed models perform best in detecting the data structure in exponential distribution (see TABLE 2) and the existing techniques only outshine the proposed models in terms of lognormal distribution based on the percentage of a terminal node in the additive model and also behave poorly in the multiplicative model as it was shown in TABLE 3.

Results from empirical studies from the bone marrow transplant data in TABLE 4 showed that the proposed model with Cox-Snell residual (Deviance=16.6498) performs better than both the Martingale residual (deviance=160.3592) and Deviance residual (Deviance=556.8822) when relapse is considered as a variable of interest. Inconsistently, when death in remission is considered, the proposed model outperformed the existing model, which tends to have high variability (see TABLE 5).

Additionally, results from Prostate cancer in TABLE 4 and 5 also reveal the better performance of the proposed model over the existing one in both causes. When Cox-Snell residual was considered, the results generated MSE=0.01783563 and deviance=14.9732200, of which both measures are less than the measures produced by Deviance and Martingale residua. This implies that the proposed model is performing better at the expense of real-life data. Moreover, these results validate those obtained from the Monte-Carlo studies.

The major implication of this study is that it has the capability of predicting the survival of patients from the disease with respect to the prognostic markers. However, the study only focuses on low-dimensional survival data with multiple events. Therefore, there is a need to extend the study to high dimensional survival data with competing risks

V. CONCLUSION

The purpose of this study is to introduce CART into survival analysis with multiple events with the aim of helping the medical practitioners to arrange patients into several groups with comparable risks. The study successfully developed a tree for Univariate competing risk by using within-node homogeneity through the deviance and cox-snell residuals as a common response in a classification tree. The methods are particularly intended for information with contending risks, where more than one risk is of interest.

All the proposed models seem to fit the data structure better than the existing model in the exponential and lognormal distributions through the simulation schemes. Consequently, when the real-life data were considered, many preferences were given to the proposed models than the existing ones.

In general, when it comes to data structure performance, no methods appeared to be able to select covariate or number of terminal nodes than the proposed methods. In that case, the Cox-Snell residual in the within-node homogeneity classification tree performed better than any existing methods.

VI. ACKNOWLEDGMENT

The author like to thank the anonymous reviewers for their valuable comments that significantly improve the study.

REFERENCES:

[1] K. E. Cevasco, A. A. Roess, H. M. North, S. A. Zeitoun, R. N. Wofford, G. A. Matulis, A. F. Gregory, M. H. Hassan, A. D. Abdo, and M. E. von Fricken, "Survival analysis of factors affecting the timing of COVID-19 non-pharmaceutical interventions by U.S. universities". BMC Public Health **21,** 1985 2021. https://doi.org/10.1186/s12889-021-12035-6

[2] Y. Jeon and W. K. Lee, "Competing Risk Model in Survival Analysis," Cardiovasc Prev Pharmacother. 2(3):77-84, 2020.

[3] P. Macek, M. Biskup, M. Terek-Derszniak, M. Manczuk, H. Krol, E. Naszydlowska, J. Smok-Kalwat, S. Gozdz and M. Zak," Competing Risks of Cancer and Non-Cancer Mortality When Accompanied by Lifestyle-Related Factors—A Prospective Cohort Study in Middle-Aged and Older Adults," Frontiers in Oncology , 10, 2020.

[4] V. Zuccaro, C. Celsa, M. Sambo, S. Battaglia, P. Sacchi, S. Biscarini, P. Valsecchi, T. C. Pieri, I. Gallazzi, M. Colaneri, M. Sachs, S. Roda, E. Asperges, M. Lupi, A. Di Filippo, E. Seminari, A. Di Matteo, S. Novati, L. Maiocchi, M. Enea, M. Attanasio, C. Cammà, and R. Bruno, "Competing-risk analysis of coronavirus disease 2019 in-hospital mortality in a Northern Italian centre from SMAtteo COvid19 REgistry (SMACORE)," Sci Rep. 2021 Jan 13;11(1):1137. doi: 10.1038/s41598-020-80679-2. PMID: 33441892; PMCID: PMC7806993.

[5] G. Nijman, M. Wientjes, J. Ramjith, N. Janssen, J. Hoogerwerf, E. Abbink, M. Blaauw, T. Dofferhoff, M. van Apeldoorn, K. Veerman, Q. de Mast, J. Ten Oever, W. Hoefsloot, M. H. Reijers, R. van Crevel, and J. S. van de Maat , "Risk factors for in-hospital mortality in laboratory-confirmed COVID-19 patients in the Netherlands: A competing risk survival analysis," PLoS One. 2021 Mar 26;16(3):e0249231. doi: 10.1371/journal.pone.0249231. PMID: 33770140; PMCID: PMC7997038.

[6] M. Kojiro,"Introduction to Survival Analysis in the Presence of Competing Risks," Annals of Clinical Epidemiology 2021;3(4):97–100

[7] J. J. Liao, and G. F. Liu, "A flexible parametric survival model for fitting time to event data in clinical trials," Pharm Stat 2019;18(5):555–567.

[8] G. F. Liu, and J. J. Liao, "Analysis of time-to-event data using a flexible mixture model under a constraint of proportional hazards," J Biopharm Stat 2020;30(5):783–796.

[9] J. J. Liao, M. Z. Farooqui, P. Marinello, J. Hartzel, K. Anderson, J. Ma, C. K. Gause, "Using artificial intelligence tools in answering important clinical questions: the keynote-183 multiple myeloma experience,"Contemp Clin Trials 2020;p106179

[10] Y. Tseng, H. Wang, T. Lin, J. Lu, C. Hsieh, and C. Liao, "Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer,"*JAMA Netw*

Open. 2020;3(8):e2011768. doi:10.1001/jamanetworkopen.2020.11768.
[11] S. Bussy, A. Guilloux, S. Gaïffas, A-S. Jannot, "C-mix: a high-dimensional mixture model for censored durations, with applications to genetic data," Stat Methods Med Res 2019; 28(5):1523–1539.
[12] K. A. Dauda, W. B. Yahya, and A. W. Banjoko, "Survival Analysis With Multivariate Adaptive Regression Splines Using Cox-Snell Residual," Journal of Annals. Computer Science Series. 2015;13(2): 25-41.
[13] K. A. Dauda, B. Pradhan, B. U. Shankar, and S. Mitra, "Decision tree for modeling survival data with competing risks", Biocybernetics and Biomedical Engineering, 2019;39(3):697-708. https://doi.org/10.1016/j.bbe.2019.05.001
[14] M. C. Fiona, "Classification Trees For Survival Data With Competing Risks," Department of Biostatistics, University of Pittsburgh; 2008.
[15] L. Breiman., J. Friedman., R. Olshen., and C. Stone, "Classification and Regression Trees," Wadsworth, Belmont California, 1984.
[16] A Triantafyllidis, H. Kondylakis, D. Katehakis, A. Kouroubali, L. Koumakis, K. Marias, A. Alexiadis, K. Votis, and D. Tzovaras, "Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer," Systematic Review JMIR Mhealth Uhealth 2022;10(4):e32344. doi: 10.2196/32344PMID: 35377325
[17] M. Z. Alam, M. S. Rahman, and M. S. Rahman," A Random Forest-based predictor for medical data classification using feature ranking." *Inform Med Unlocked*. 2019; 15:1–12. doi: 10.1016/j.imu.2019.100180
[18] K. A. Dauda, K. O. Olorede, and, S. A. Aderoju, "A novel hybrid dimension reduction technique for efficient selection of bio-marker genes and prediction of heart failure status of patients," Scientific African, Volume 12, 2021,e00778, ISSN 2468-2276, https://doi.org/10.1016/j.sciaf.2021.e00778.
[19] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," *Stroke Vasc Neurol*. 2017;2(4):230–43. 10.1136/svn-2017-000101
[20] S. Barbieri, S. Mehta, B. Wu, C. Bharat, K. Poppe, L. Jorm, and R. Jackson, "Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach", *International Journal of Epidemiology*, 2021;, dyab258, https://doi.org/10.1093/ije/dyab258
[21] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM, Sensors", 2021, (21)2852. https://doi.org/10.3390/s21082852
[22] S. Piri, D. Delen, and T. Liu, "A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets", Decis. Support Syst., 106 (2018), 15–29. https://doi.org/10.1016/j.dss.2017.11.006
[23] J. P. Fine, and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," J Am Stat Assoc. 1999;**94**:496–509. doi: 10.1080/01621459.1999.10474144.
[24] J. P. Klein, and M. L. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data", 2005.
[25] G. L. Lu-Yao, P. C. Albertsen,, D. F. Moore, W. Shih, Y. Lin, R. S. DiPaola, M. J. Barry, A. Zietman, M. O'Leary, E. Walker-Corkery, S. L. Yao, "Outcomes of localized prostate cancer following conservative management," Journal of the American Medical Association, 2009 302, 1202 - 1209.
[26] R. Olshen, "Remembering leo breiman," The Annals of Applied Statistics, 2010, 4(4):1644–1648.
[27] L. Breiman, "Statistical modeling: The two cultures (with rejoinder),"Statistical Science, 2001b, 16(3):199–231.

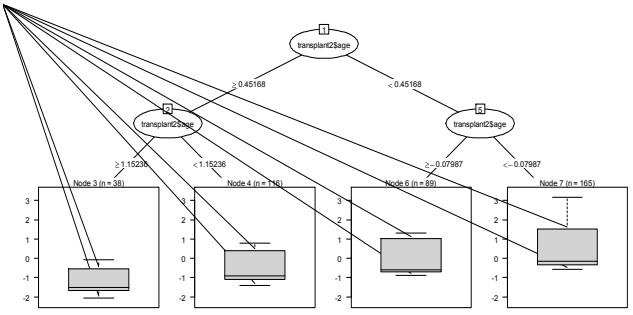


FIGURE 1. Within node tree for data from bone marrow transplant through cause 1 (relapse) via martingale residual

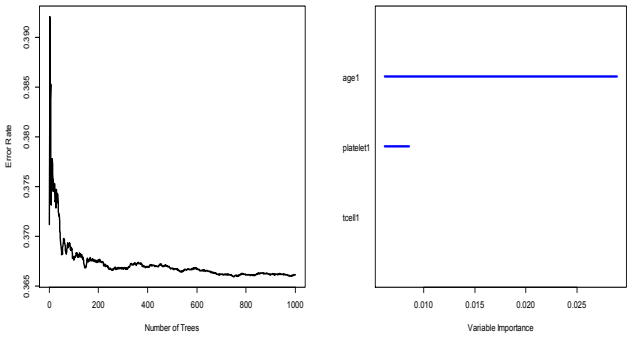


FIGURE 2. Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residua

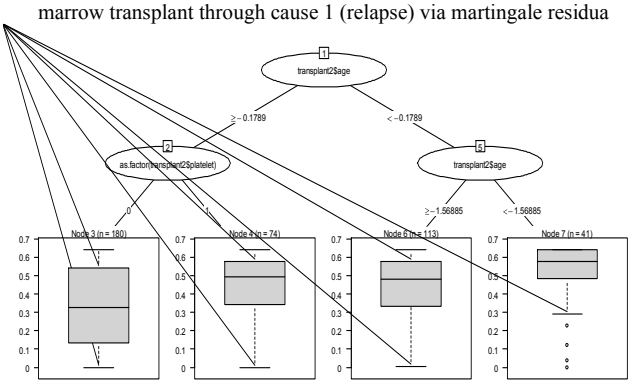


FIGURE 3. Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual

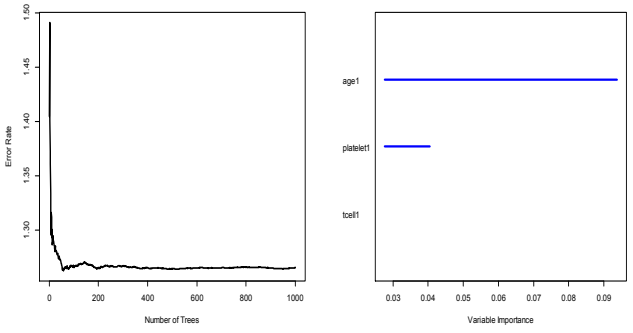


FIGURE 4. Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residual.

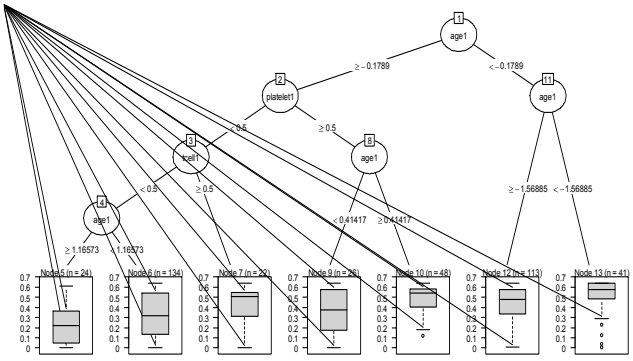


FIGURE 5. Within node tree for data from bone marrow transplant through cause 1 (relapse) via cox-snell residual

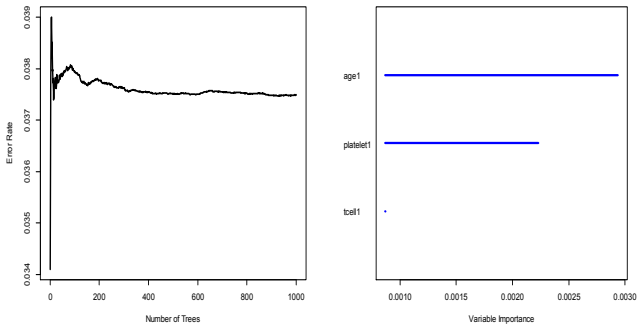


FIGURE 6. Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via cox-snell residual

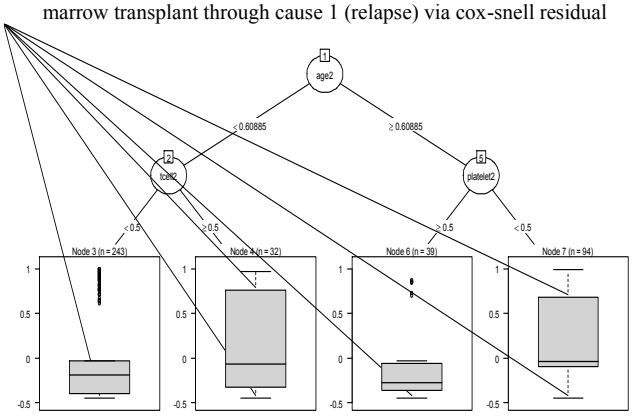
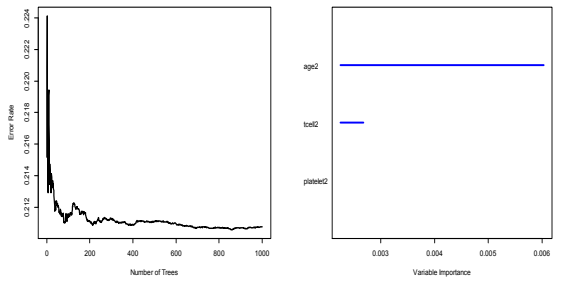


FIGURE 7. Within node tree for data from bone marrow transplant through cause 2 (death in remission) via martingale residual



FIFURE 8. Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residua

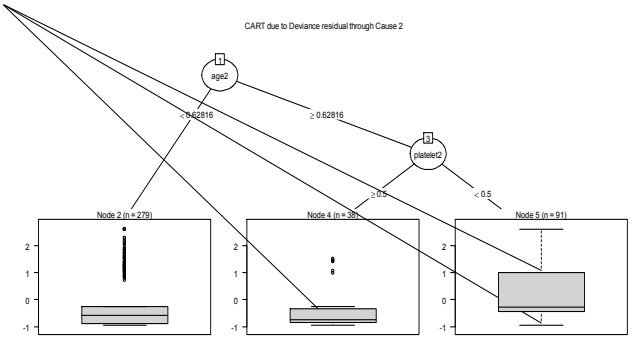


FIGURE 9. Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual

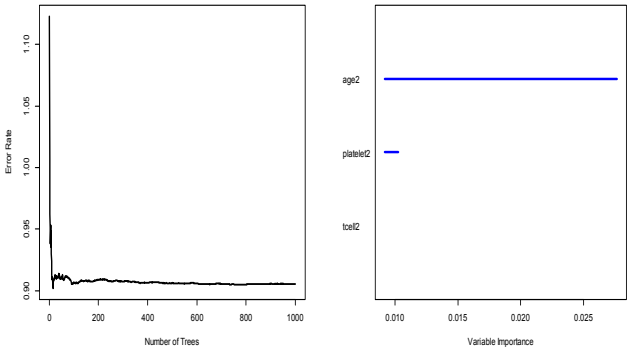


FIGURE 10. Graphical representation of the variable important in bone marrow transplant through cause 2 (death in remission) via martingale residual.

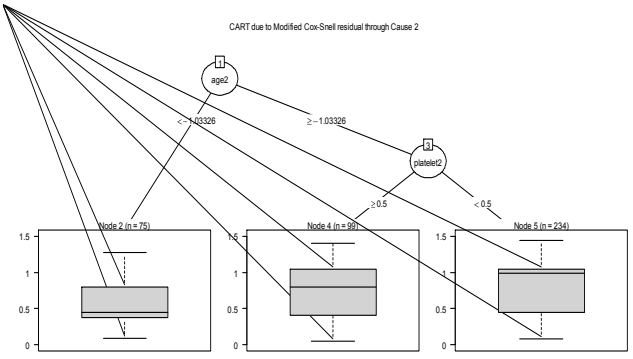


FIGURE 11. Within node tree for data from bone marrow transplant through cause 2 (death in remission) via cox-snell residual

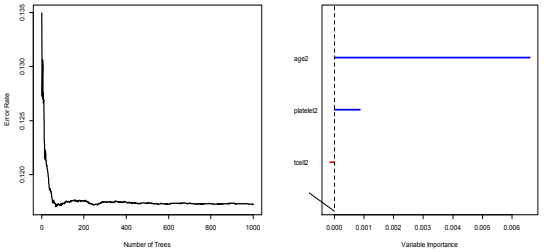


FIGURE 12. Graphical representation of the variable important in bone marrow transplant through cause 2 (death in remission) via cox-snell residual