A Predictive Perspective on Machine Learning Survival Models: Inventory, Evaluation Metrics, and Empirical Analysis.

WORKING PAPER[1]

Diego Vallarino Navarro
diego.vallarino@gmail.com
Mayo 2023

Abstract

This study examines the use of survival machine learning models in healthcare for predicting outcomes. A total of 20 survival models, including more than 10 survival machine learning algorithms, were investigated to identify their main strengths and weaknesses. In the first part of the study, we discuss the various models, their data types, and evaluation methods, comparing similarities and differences between the most popular models. We also highlight the necessary rules that all machine learning strategies for survival analysis must follow. In the second part, four machine learning algorithms representing each family (trees, multi-task, kernel, and deep network) were applied to analyze a breast cancer dataset and two other simulated datasets using the R coxed library. The results show how machine learning models can be used to recommend treatments and improve population health by analyzing survival. We also identify the best algorithms to apply based on more than twelve restrictions, such as censored data,

MSC 2000 *subject classifications*. 62N01, 62P10, 68T05
Key words and phrases. survival, extinction, machine learning models

---

1. Introduction

Survival analysis is crucial for predicting patients' time-to-event outcomes and supporting healthcare practitioners in making the best treatment decisions (Ping, W., et al. 2017), not just in illness analysis or monitoring procedures but also in assisting in the quantitative and qualitative improvement of preventative medicine (lifestyle interventions, vaccine efficacy, screening programs, among others). In recent years, machine learning algorithms have matured into outstanding instruments for survival analysis, providing precise and trustworthy forecasts (Balan, T. A., & Putter, H. 2020; Balan, T. A. 2018; Alyass, A. et al. 2015).

Considering the relevance of the use of machine learning models, the primary goal of this study is to highlight the most widely used machine learning approaches for predictive survival analysis. This article provides a thorough examination of machine learning models for survival analysis, which may aid healthcare practitioners and researchers in selecting the best model for their datasets.

The paper is structured as follows: In the first part, a discussion of the models is made based on the availability of the data that contains the experiment that has been carried out. In turn, the different ways of evaluating machine learning models will be presented through objective statistics and qualitative indicators. Weaknesses common to all machine learning models for survival analysis will also be analyzed. as well as the weaknesses and strengths of the main models.

In the second part, the application of four machine learning algorithms for survival analysis will be studied in depth. A comparison is shown between a dataset that contained the R survival library and a dataset on breast cancer. The document ends with the conclusions drawn from this analysis and recommendations for new lines of research.

2. Machine Learning Survival Models

We conduct a full review of several methods used for survival analysis depending on data availability in this study (see Ping, W., et al 2017). Furthermore, we emphasize the common weakness of these algorithms and provide a full study of the many metrics used to assess their performance. Finally, we discuss the benefits and drawbacks of these machine learning models for survival analysis, which may be used to choose the best strategy for different datasets.

To carry out the above-mentioned analysis, the following conceptual framework will be followed:

Figure 1: conceptual structure of the methodological review
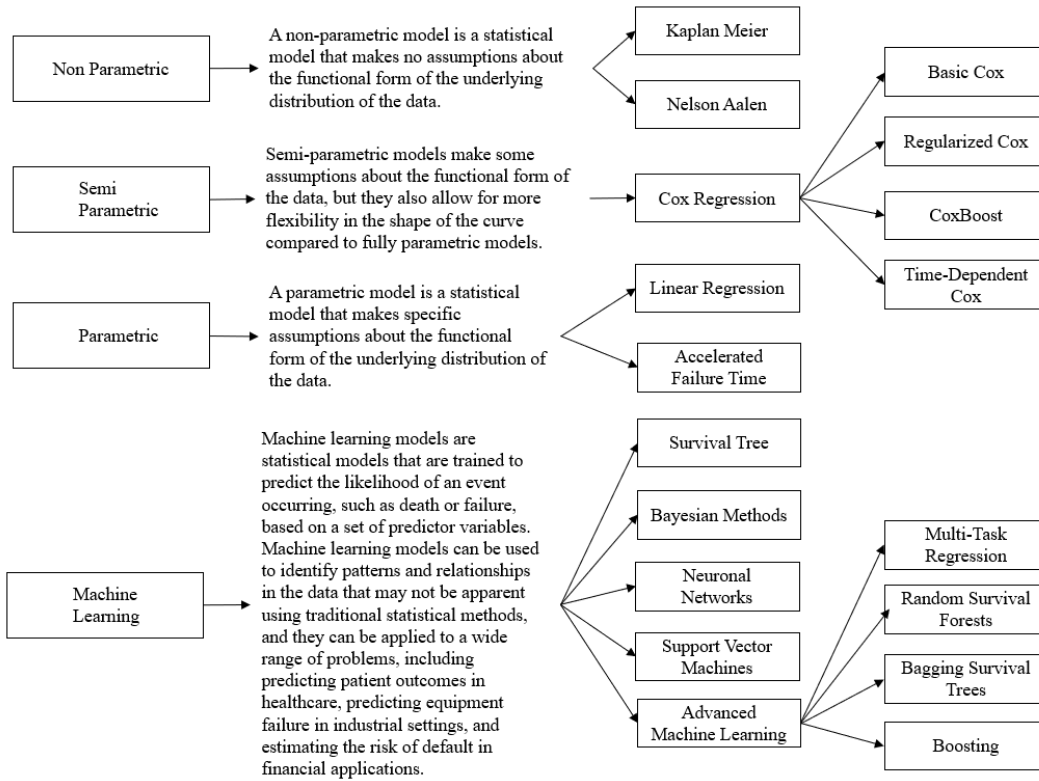
Source: own development based on Ping, W., et al. 2017.

2.1 Discussion of algorithms based on the availability of some type of data.

Survival algorithms may forecast a patient's survival following a diagnosis, but the efficacy of these algorithms is heavily reliant on the quality and availability of patient data. There are several sorts of data that may influence how these algorithms are used. For example, such as basic patient data that includes demographic and clinical information as illness stage, comorbidities, and therapy (Balan, T. A., & Putter, H. 2020; Balan, T. A. 2018; Alyass, A. et al. 2015).

We can also have problems when we use datasets with censored data, or competitive risk data, and even with data that presents pacients longitudinal information (Alyass, A. et al. 2015). Survival algorithms are applicable, even though with certain limitations (Jin, P., et al 2021; Cuperlovic-Culf, M. 2018).

In the following paragraphs, we will go through each of the consequences that this sort of data has for survival analysis.

2.1.1   Baseline Patient Data

Basic patient data is a crucial element in developing a survival algorithm for healthcare professionals. Demographic information such as age, sex, race, and ethnicity, along with clinical data such as disease stage, comorbidities, and treatment history, can significantly impact a patient's survival rate. Developing a survival algorithm requires extensive knowledge and consideration of various factors that can affect a patient's outcome (Jin, P., et al 2021; Cuperlovic-Culf, M. 2018). Healthcare professionals can utilize different algorithms such as trees, forest, neuronal networks, deep learning, multi-tasking, boosting, and "others" to generate survival predictions (Allison, P.D., 2003; Allison, P.D., 2002; Schafer, J. L., 1997). However, it is essential to consider the limitations and constraints of these algorithms while making predictions to avoid errors and biases (Azodi, C. B. at al 2020). So, basic patient data is essential in developing accurate survival algorithms for effective patient care, but many times it is incomplete to find a good performance in machine learning models.

2.1.2 Censored Data

A distinctive feature of survival data is the concept of censoring data. If the event of interest is death, then the event time is censored for those subjects who are still alive at the end of the study. This means that the statistical analysis must proceed without knowing when the subject will die (Allison, P.D., 2003; Allison, P.D., 2002; Schafer, J. L., 1997).

All that is known about the time of his death is that it is after the date the study ended. More generally, in a follow-up study, subjects who drop out are censored, because they are usually lost to follow-up and the time of their event is unknown (see Raghunathan, T. E., 2004). The date of the event is not observed, although it is not a missing piece of data either, since these two types of unobserved data have completely different characteristics and empirical explanations (Yuan, H., et al, 2022).

For right-censored subjects, the only thing that is known is that their event occurred after the time they were censored. An implicit concept in the definition of censorship is that, if the study had been prolonged (or if the subjects had not dropped out), eventually the outcome of interest would have been observed to occur for all subjects. Conventional statistical methods for analysis of survival data make the important assumption of independent or uninformative censoring (see Khan, F. M., & Zubek, V. B., 2008).

This means that, at a given moment, the subjects who remain in follow-up have the same future risk for the occurrence of the event as those subjects who are no longer followed up (either due to censorship or abandonment of the study), as if the losses to follow-up were random and therefore not informative (see Basak, P., et al 2022).

What the current literature clearly shows is that the treatment of the censored data is fundamental in order to have an exact vision of the survival analysis experiment that will be carried out (Jiang, Y., et al, 2022). Therefore, the line of research of this work will seek to understand which are the best methods of incorporating censored data, both from the right (the most common in analysis models), and the data censored from the left. The latter have not been thoroughly analysed in the literature and can provide a number of clues in time-to-event statistical analyses (Liu, T., Zhang, L., Jin, G., & Pan, Z., 2022).

When dealing with survival data, it is common to encounter censored data, which occurs when the exact time of an event is unknown, but it is known that the event has not occurred before or at a particular time. Right-censored data, interval censored data, and left-censored data are the three types of censored data. When it comes to handling large volumes of censored data, there are several algorithms available that are effective (Liu, T., Zhang, L., Jin, G., & Pan, Z., 2022).

Survival Random Forest is one such algorithm that can handle censored data well. It is a machine learning algorithm that constructs multiple decision trees and combines their outputs to make predictions (see Imani, F., et al 2019, Jin, Z, et al, 2020). Multi-Tasking Linear Regression (MTLR) is another algorithm that can handle censored data efficiently. It uses a Bayesian approach to estimate the distribution of survival times and is useful when dealing with multiple outcomes (Wang, L., Li, Y., Zhou, J., Zhu, D., & Ye, J., 2017). XGboost is also a popular algorithm that can handle large volumes of censored data and can handle both continuous and categorical variables (see Ilie, S., 2020).

We will present evidence for claims in the next section of this paper.

2.1.3 Competitive Events/Risk Data

A competing risk is an incident that prevents the principal event of interest from occurring (see Barrett, JK, Siannis, F., & Farewell, V.T., 2011). For example, in a research where the main outcome was time to death from a cardiovascular cause, death from a non-cardiovascular cause was included as a competing event. That is, a person who dies of cancer is no longer at danger of

dying from cardiovascular reasons. Regardless of whether the time of patient follow-up is prolonged, a subject will not be seen to die of cardiovascular reasons after dying of cancer.

When determining which kind of incident happened first, competing risks may be present in studies in which distinct categories of occurrences involve non-fatal consequences. Thus, a research with three categories of events is possible: heart disease diagnosis, cancer diagnosis, and death. Each kind of occurrence acts as a competing risk since a cancer diagnosis preceded by a heart illness or death prevents the latter two events from happening first (see Gorfine, M. and Nevo, D., 2020; Kin, J. et al, 2020).

Conventional statistical approaches for analysing survival data presuppose the absence of competing hazards. If information regarding a subject's risk of experiencing one kind of event does not convey information about a subject's risk of experiencing the other type of event, the risks are said to be independent (Kin, J. et al, 2020). The strategies discussed in this article are equally applicable in situations where competing risks are independent of one another and in contexts where competing risks are not independent of one another. In biomedical applications, biology often reveals at least some correlation between conflicting dangers, which may be extremely severe in many circumstances.

As a result, separate competing hazards may be uncommon in biological applications. Analysts commonly filter participants when a competing event occurs while analysing survival data with conflicting hazards. Thus, if the outcome is time to death from cardiovascular causes, an analyst may consider a subject censored if that subject dies from non-cardiovascular causes. Censoring people at the moment of death from non-cardiovascular reasons, on the other hand, might be difficult.

The analytical goal of competitive event data, like ordinary survival analysis, is to estimate the likelihood of an event among many potential occurrences over time, enabling individuals to fail competitive events.

As a result, understanding that competing events will always emerge when we longitudinally analyse various patients and compute the time to event will be crucial in the line of study provided in this work proposal (Haneuse, S and Lee, K., 2016). The main issue is to be able to create some sort of statistical approach capable of analysing the significance of various competing occurrences and comprehending the worth of the information they give to the survival analysis (Peng, M., & Xiang, L., 2019). Some occurrences will yield a lot of information, such as the example of the cancer patient in a cardiovascular disease study, while other competitive events that are not thought

significant a priori may be so owing to their consequences in later temporal sequences, inside or outside of time of investigation (Xu, Y., Scharfstein, D., Müller, P., & Daniels, M., 2022).

The following are some strategies for dealing with the event risk issue. Competitive Survival Analysis (CSA) is a statistical approach for analyzing survival data where numerous events have the potential to impact the result of interest. The Cox proportional hazards model, along with competition data, is used in CSA to simulate the hazard rates of both the event of interest and the competing event. This enables for a more accurate forecast of the desired result while accounting for the possible influence of the competing event (see Hong, C., Yi, F., & Huang, Z., 2022).

The Cumulative Incidence Function (CIF) is a statistical tool for estimating the likelihood of an event of interest occurring in the context of competing hazards. The CIF calculates the marginal probability of encountering the event of interest at a given time point while accounting for competing occurrences. This allows for a more accurate forecast of the likelihood of the event of interest when competing occurrences are considered (see Zhang, M. J., Zhang, X. U., & Scheike, T. H., 2008; Lambert, P. C., 2017).

Lunn-McNeil (LM) is a statistical approach used to predict the likelihood of encountering the event of interest vs a competing event in the setting of conflicting risks. It entails developing a dummy variable for the competing event, which is then included into the Cox proportional hazards model. This provides for a more precise calculation of the event's hazard rate while adjusting for the influence of the competing event (see Huszti, E., Abrahamowicz, M., Alioum, A., & Quantin, C., 2011).

2.1.4 Longitudinal patient Data

Longitudinal patient data is information gathered over time on a patient's health status or other circumstances that may influence their health. This data can be very varied, ranging from changes in the person's income to changes in the employment situation. This sort of information may give important insights into the patient's general health and how it develops over time. Covariates are factors that may be linked to the result under study but are not of main concern, and they may be included in the analysis to assist account for their effects (see Thenmozhi, M., Jeyaseelan, V., Jeyaseelan, L., Isaac, R., & Vedantam, R., 2019).

Longitudinal patient data may be utilized in survival analysis to increase prediction accuracy and allow for a more complete investigation of the variables influencing the outcome of interest. Information about a patient's medication usage, nutrition, or exercise habits, for example, might

be gathered over time and included as variables in the study. These variables give extra information that may have an indirect impact on the patient's survival and incorporating them in the study may increase prediction accuracy (see Allison, P. D., 2014).

Furthermore, longitudinal data is often gathered on a regular basis and provides information on how a patient's health evolves over time. This data may be utilized to simulate the time-varying impacts of variables and give a more complete study of how these factors influence the desired result.

As we have been able to identify, the more data we have in our dataset, the more information we can get from them. This implies that as we develop survival analysis experiments with a greater number and diversity of data points, the better performance the machine learning algorithms will have and the greater predictive power they will have.

Obviously, it is important to note that it is relevant to understand what the best techniques for data management with certain characteristics and particularities are such as NAs, missing data, censored data, competitive events, among other types of data.

2.3 Common weaknesses for survival machine learning algorithms and some solutions.

Survival analysis is being undertaken using machine learning algorithms, although these approaches have common limitations and risks (Jin, P., et al 2021; Cuperlovic-Culf, M. 2018). In this section, we will explore these limitations, as well as some potential solutions.

The lack of interpretability is the most basic limitation of using machine learning models in survival analysis. Understanding "black box" machine learning models is challenging (Azodi, C. B. et al 2020, Guidotti, R., et al. 2018, Angermueller, C., et al. 2016, Alyass, A., et al. 2015). They are often used to generate forecasts, but they do not reveal the underlying relationships between the variables utilized to generate such predictions (Miller. T., 2019). This is a critical requirement to keep in mind when using machine learning models to survival analysis, since knowledge of the underlying corelations between the variables may affect treatment and other intervention strategies (Lipton, Z.C., 2018).

In addition to interpretability, another problem of using machine learning models to survival analysis is overfitting (Libbrecht, M.W. and Noble, W.S. 2015). When a model captures too much of the noise in the data and does not generalize well to new data, this is known as overfitting. This may result in inaccurate predictions because the model cannot effectively represent the underlying

data connections. Overfitting in survival analysis may lead to erroneous predictions of the time until an event happens (Tarca, A.L., et al. 2007).

Although this is a general weakness in all machine learning analyses, it becomes very evident in survival analysis algorithms, since one more complexity will be added, which is the introduction of the time variable. This results in a higher complexity of the overfitting problem (see Yin, Q., et al., 2022).

Thirdly, applying machine learning models in survival analysis for prediction is limited by the necessity for huge volume of data in each dataset (Azodi, C. B. at al 2020). Large volumes of data re usually necessary to train machine learning algorithms, which may be challenging to gather in the medical and social sciences.

Moreover, datasets typically include suppressed information, which may result in biased outcomes, . Censored data refers to observations in which the result of interest, such as death or recurrence of an illness, is missing (Allison, P.D., 2003; Allison, P.D., 2002; Schafer, J. L., 1997). Although it is not an exclusive weakness of machine learning algorithms, when using large volumes of data, the probability of finding suppressed information is greater, so in these cases it must be considered as a relevant weakness in relative terms.

Using visualization to analyze machine learning model predictions may also give insight into a variable's underlying connections. Lastly, the use of data augmentation methods, like synthetic data (see Haradal, S. et al., 2018; Perez, J., et, al., 2023) may aid in alleviating the problem of insufficient data.

Using data augmentation methods, it is possible to produce extra data points that may be utilized to train the model. If we need to reduce the time-effect-in-covariances weaknesses one common approach is to use stratification, which involves dividing the analysis into distinct time intervals. By doing so, the impact of time on the outcome can be evaluated more precisely. Another technique is to use time-dependent covariates, which enable us to examine how the relationship between a particular variable and the outcome changes over time (Kelly, C.J., et al, 2019; Moradian, H., et al, 2022).

2.4 Specific *pros* and *cons* for survival machine learning algorithms.

Despite the limits of machine learning algorithms in survival analysis, there are possible solutions to these issues. One way is to create predictions using ensembles of machine learning models. Ensembles integrate the forecasts of many models to obtain a more accurate forecast (Azodi, C.

B. 2020; Ping, W., et al 2017). This may lessen the danger of overfitting and increase the accuracy of the forecast.

Most machine learning models can learn from new data. This is a differential characteristic when we compare machine learning models with traditional survival models. Algorithms can "learn" as they access new data, and this has a distribution similar to the training data. If the fresh data behaves similarly to the training data, there is no need to retrain the algorithm (see Alyass, A., Turcotte, M., Meyre, D., 2015).

The matrix below compares the performance of various machine learning models for survival analysis based on several characteristics such as censored data, missing data, small and large number of observations, number of variables, overfitting, interpretability, covariate independence, computational time, hyper-parameter robustness competitive events, and non-linear relationships.

Each column in the matrix indicates a single model can handle certain feature. The values range from 1 (poor performance) to 5 (good performance).

Table 1: main characteristics of machine learning survival models

	Surv. Tree	Random Forest	MTLR	SVM	XGBoosting	DeepSurv	DeepHit
Censored Data	2	5	4	3	4	3	3
Missing Data	2	3	2	2	2	3	3
Small number Observations	4	4	5	4	4	4	4
Big number of Observations	2	5	3	3	4	5	5
Number Variables	3	4	4	3	4	5	5
Overfitting	5	4	4	4	4	2	2
Interpretability	5	3	3	2	2	2	2
Independency	4	3	2	2	4	5	5
Computational time	4	4	4	2	3	3	3
Hyper-param Sensitive	4	4	4	4	3	2	2
Competitive Events	2	5	4	3	4	3	3
Non-linear relationships	3	4	4	3	3	4	4
Processing time (+10 var x +40.000 obs.)	4	4	4	2	4	3	3
Punctuation	**44**	**52**	**47**	**37**	**43**	**45**	**45**

Source: own elaboration
Note: a valuation of 1 = low ML performance and 5 = high ML performance

According to the matrix, no one model performs the best across all attributes. Rather, each model has advantages and disadvantages, and the model chosen should be based on the unique features of the data and the research topic being addressed.

If the dataset includes missing or censored data, decision tree-based models like Survival Tree and Survival Random Forest may be appropriate Basak, P., Linero, A., Sinha, D., & Lipsitz, S., 2022; Bertsimas, D., et al., 2022; Jin, Z., et al., 2020). Deep learning-based models such as DeepSurv and DeepHit may be more appropriate if the dataset contains a high number of observations and variables owing to their capacity to capture complex patterns in the data (Ohlsson, A., 2018; Katzman, J. L. et al., 2018; Gutierrez, J. A., & Ramirez, G., 2020; Mao, J., et al, 2019; Huang, S., et al. 2018). MTLR may be a preferable alternative for datasets with a limited number of observations since it can manage the short sample size and minimize overfitting (see Li, Y., et al., 2018; Li, Y., et al., 2019)

When it comes to interpretability, the decision tree-based model (Survival Tree) is the most interpretable of the models in the matrix (Basak, P., Linero, A., Sinha, D., & Lipsitz, S., 2022). It is worth noting, however, that the majority of the models in the matrix have poor interpretability, which is a frequent shortcoming of machine learning models.

Decision tree-based models (Survival Tree and Random Forest) and MTLR are the quickest in terms of computing time, whereas deep learning-based models (DeepSurv and DeepHit) are slower owing to their complicated design (see Basak, P., Linero, A., Sinha, D., & Lipsitz, S., 2022; Cartocci, A., Cevenini, G., & Barbini, P., 2021; Fotso, S. 2018; Brard, C., et al., 2017).

As a result, the machine learning model used should be dependent on the properties of the dataset and the research issue being addressed. It is critical to consider the advantages and disadvantages of each model and choose the one that best meets the objectives of the analysis.

3. Case Studies

We present the performance of 4 machine learning models on three different datasets: SimulatedONE, SimulatedTWO, and NKI Breast Cancer Data (Lum, P., Singh, G., Lehman, A. et al, 2013). The SimulatedONE is a data frame with 2000 observations generated using coxed R library. It contains 10 variables, with 30% of censored data. For SimulatedTWO the same library has been used, with the same characteristics. The only change that was introduced was that 80% of censored data was considered.

The NKI Breast Cancer Data (a variation of this dataset is in the survival R library) includes survival data of 272 breast cancer patients[2], which is the most extensive dataset we analyzed.

---

[2] NKI Breast Cancer Data: https://data.world/deviramanan2016/nki-breast-cancer-data

Our empirical analysis provides insights into the performance of different machine learning models for survival analysis and their suitability for various types of datasets. Next, we will present the results that have been achieved.

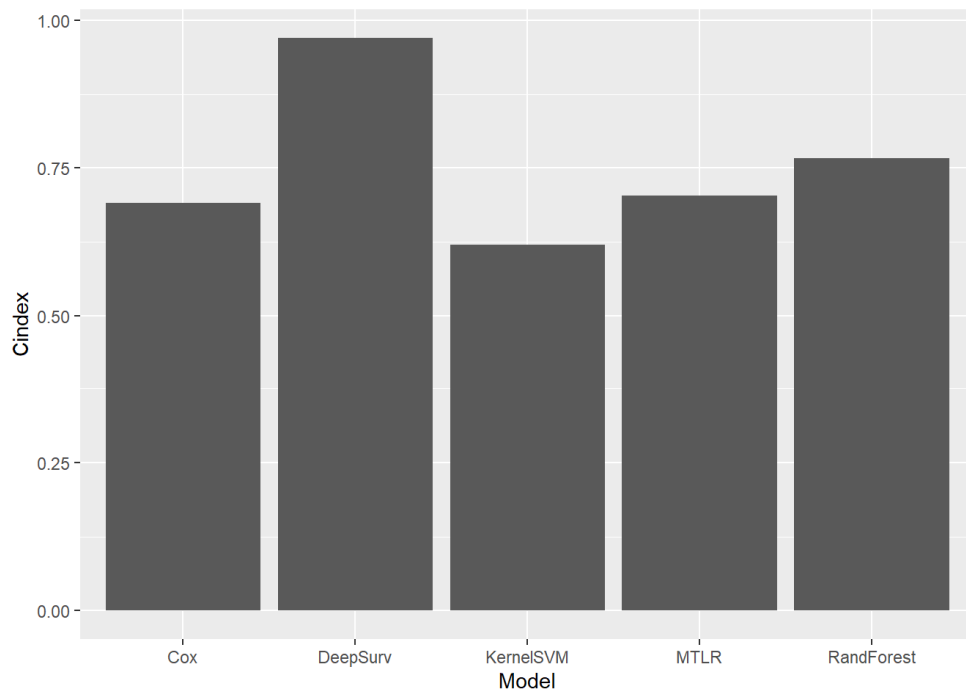**Figure 2: results from different machine learning models on SimulatedONE dataset.**
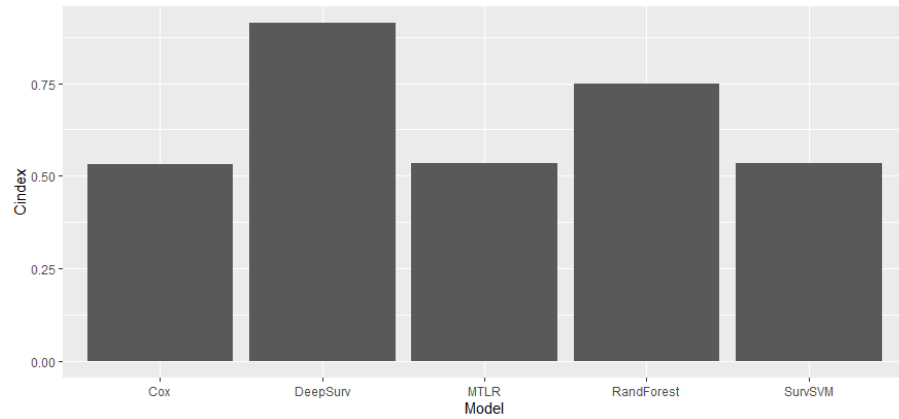
Source: own elaboration

Figure 3: results from different machine learning models on SimulatedTWO dataset.
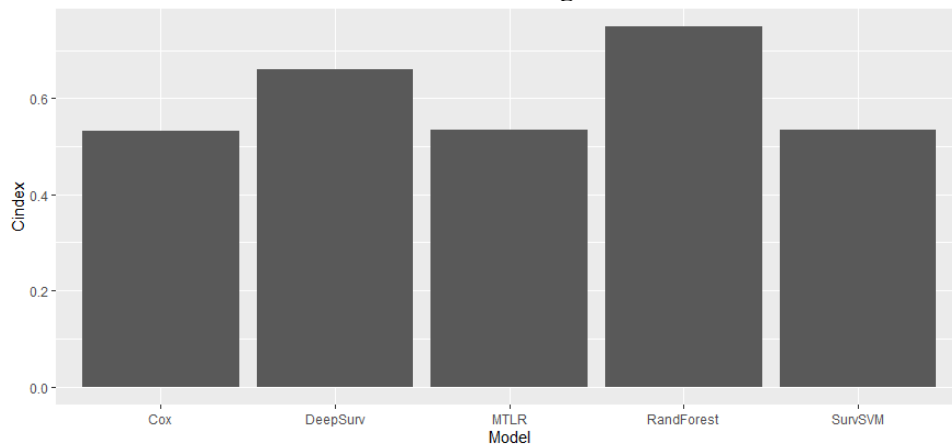
Source: own elaboration

Figure 4: results from different machine learning models on NKI Breast Cancer Data dataset.

Source: own elaboration

For the evaluation and comparison of models we have used the Cindex, which allows a comparative evaluation of the predictive capacity of the models and the codes that are presented below.

Table 2: Code for the main survival machine learning models

Algorithm	R code
Survival Tree	LTRCART.obj <- LTRCART(Surv(time, End, status) ~ karno + age + trt, data=data.train_tree) LTRCIT.obj <- LTRCIT(Surv(time, End, status) ~ karno + age + trt, data=data.train_tree)
MTLR Model	mtlr(Surv(time, status)~., data = data.train, nintervals = 9)
XGBoost	xgboost(data, label = output_vector, max.depth = 4, eta = 1, nthread = 2, nrounds = 10,objective = "binary:logistic")
Random Forest	rfsrc(Surv(time, status) ~ ., data.train)
Survival Kernel SVM	survivalsvm(Surv(time, status) ~ ., data = data.train, type = "regression", gamma.mu = 1, opt.meth = "quadprog", kernel = "lin_kernel")
DeepSurv Model	deepsurv(data = data.train, frac = 0.3, activation = "relu", num_nodes = c(4L, 8L, 4L, 2L), dropout = 0.5, early_stopping = TRUE, batch_size = 32L, epochs = 100L)
DeepHit Model	deephit(data = data.train, frac = 0.3, activation = "relu", num_nodes = c(4L, 8L, 4L, 2L), dropout = 0.1, early_stopping = TRUE, epochs = 100L, batch_size = 32L)

Source: own elaboration

The evidenced results imply that the DeepSurv model performed the best in two cases, with a Cindex of 0.84512 in the first test (figure 2). This indicates that the model's predicted probabilities

of the event occurring are well-calibrated and the model can effectively differentiate between patients who will experience the event and those who will not and the time when it happens.

The other models, including Cox, MTLR, Random Forest, and Kernel SVM, all performed reasonably well but not as well as DeepSurv. The differences in performance could be due to various factors, such as differences in the model's architecture, the type of loss function used, or the hyperparameters chosen during training.

Something happens when we change the percentage of censored data. We have noticed that the performance of the models, with the same characteristics (same hyperparameters) have a different absolute and relative performance. It is evident what happens when a dataset similar characteristic, but the only differential is the percentage of censored data.

As we noted, in the case of SumulatedONE and SimulatedTWO the relative performance of the DeepSurv algorithm has decreased notably. We can conclude that this algorithm was affected by the volume of censored data, in line with what we have presented in the previous section.

It is good to note at this moment that the standard parameters have been used to develop each of the models, as they are presented in Table 2. It would be desirable, in future investigations, to deepen the analysis that certain changes could generate in some of the models hyperparameters.

4. Conclusions and Next Steps.

To summarize, although machine learning algorithms have made tremendous progress in resolving some of the shortcomings of classic survival analysis approaches, there is still room for advancement. Overfitting, interpretability, data amount and quality, and temporal effects in variables are among shortcomings shared by all machine learning models. Although ensemble machine learning systems have been created to mitigate these flaws, the most complicated models remain hyperparameter sensitive.

Furthermore, since access to longitudinal or supplemental data is restricted, many assumptions cannot be checked ex post, posing a challenge to the validity of the findings. Nonetheless, by changing parameters as they learn from fresh data, machine learning algorithms provide huge productivity increases.

To address these constraints, future research should concentrate on building more resilient machine learning algorithms that can manage vast amounts of data while maintaining accuracy over time, as well as enhancing access to complementary and longitudinal data. Furthermore, the

problem of interpretability must be addressed, ensuring that models can be understood and confirmed by domain experts.

Overall, machine learning algorithms are an important tool for survival analysis since they provide great prediction skills and allow researchers to identify patterns and insights in big and complicated datasets. With continued investment in research and development, the potential of machine learning for survival analysis will expand, advancing our knowledge of illness and improving patient outcomes.

There are numerous crucial next steps that researchers should examine in order to progress the area of machine learning in survival analysis. One critical step is to test multiple models with simulated data to evaluate their resilience to data quality, including prediction result repeatability and hyperparameter sensitivity. Researchers can find which algorithms are most useful for certain use cases by comparing the predicted performance of various models.

Another critical next step is to investigate methods to address the present machine learning algorithms' shortcomings in survival analysis, such as overfitting, interpretability, and the time impact of covariates. By resolving these flaws, researchers will be able to create more trustworthy and accurate models that may be used to make therapeutic choices.

Finally, it is critical to improve the interpretability of particular algorithms so that domain experts can understand and evaluate them. Clinicians and doctors will be able to make better judgments and enhance patient outcomes as a result of this.

References

Survival Tree

Basak, P., Linero, A., Sinha, D., & Lipsitz, S. (2022). Semiparametric analysis of clustered interval-censored survival data using soft Bayesian additive regression trees (SBART). *Biometrics, 78(3), 880-893.*

Bertsimas, D., Dunn, J., Gibson, E., & Orfanoudaki, A. (2022). Optimal survival trees. *Machine Learning, 1-73.*

Hao, L., Kim, J., Kwon, S., & Ha, I. D. (2021). Deep learning-based survival analysis for high-dimensional survival data. *Mathematics, 9(11), 1244.*

Härdle, W., Klinke, S., & Müller, M. (2004). *Applied nonparametric regression.* Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. *Springer Science & Business Media.*

Breiman, L. (1984). *Classification and regression trees.* CRC press.

Bayesian Methods

Brard, C., Le Teuff, G., Le Deley, M. C., & Hampson, L. V. (2017). Bayesian survival analysis in clinical trials: What methods are used in practice? *Clinical Trials, 14(1), 78-87.*

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *In Bayesian inference and decision techniques (pp. 383–402). Springer, Berlin, Heidelberg.*

ANN

Chi, C. L., Street, W. N., & Wolberg, W. H. (2007). Application of artificial neural network-based survival analysis on two breast cancer datasets. *In AMIA annual symposium proceedings (Vol. 2007, p. 130). American Medical Informatics Association.*

Ching, T., Zhu, X., & Garmire, L. X. (2018). Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology, 14(4), e1006076.*

Fotso, S. (2018). Deep neural networks for survival analysis based on a multi-task framework. *arXiv preprint arXiv:1801.05512.*

Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer letters, 471, 61-71.*

Joshi, R., & Reeves, C. (2006). Beyond the Cox model: artificial neural networks for survival analysis part II. *In Proceedings of the eighteenth international conference on systems engineering (pp. 179-184).*

Pölsterl, S., Navab, N., & Katouzian, A. (2016). An efficient training algorithm for kernel survival support vector machines. *arXiv preprint arXiv:1611.07054.*

Van Belle, V., Pelckmans, K., Van Huffel, S., & Suykens, J. A. (2011). Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artificial intelligence in medicine*, 53(2), 107-118.

DeepHit

Mao, J., Wang, Y., & Hu, Y. (2019). *DeepHit:* A Deep Learning Approach to Survival Analysis With Competing Risks. *IEEE transactions on biomedical engineering, 66(6), 1747-1756.*

Gutierrez, J. A., & Ramirez, G. (2020). Deep learning in medical decision making: A systematic review and future directions. *Medical Decision Making, 40(3), 416-429.*

Lee, C., Zame, W., Yoon, J., & Van Der Schaar, M. (2018). Deephit: A deep learning approach to survival analysis with competing risks. *In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).*

Ropcke, S., & Lieder, F. (2019). Deep learning in medical decision making: A systematic review and future directions. *Artificial Intelligence in Medicine, 97, 1-9.*

Deo, R. C., & Arora, S. (2020). Deep learning for medical decision making: A review and future directions. *Frontiers in artificial intelligence, 3, 48.*

DeepSurv

Katzman, J. L. et al. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *Digital Medicine, 1(1).*

Mayer-Baumgartner, S. (2018). DeepSurv: a deep learning approach to survival analysis. *BMC Medical Research Methodology, 18(1).*

Ohlsson, A. (2018). An Introduction to Survival Analysis Using Deep Learning. *Journal of Computational Science, 24, pp.1-7. https://doi.org/10.1016/j.jocs.2018.02.007*

SVM

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research, 7, 1-30.*

Fouodo, C. J., König, I. R., Weihs, C., Ziegler, A., & Wright, M. N. (2018). Support Vector Machines for Survival Analysis with R. *R Journal, 10(1).*

Khan, F. M., & Zubek, V. B. (2008, December). Support vector regression for censored data (SVRc): a novel tool for survival analysis. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 863-868). IEEE.

Pölsterl, S., Navab, N., & Katouzian, A. (2016). An efficient training algorithm for kernel survival support vector machines. *arXiv preprint arXiv:1611.07054.*

Van Belle, V., Pelckmans, K., Suykens, J. A., & Van Huffel, S. (2007, July). Support vector machines for survival analysis. *In Proceedings of the third international conference on computational intelligence in medicine and healthcare (cimed2007) (pp. 1-8).*

Rendon Forest

Imani, F., Chen, R., Tucker, C., & Yang, H. (2019, August). Random forest modeling for survival analysis of cancer recurrences. *In 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE) (pp. 399-404). IEEE.*

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008*). Random survival forests. The annals of applied statistics, 2(3), 841-860.*

Ishwaran, H., & Kogalur, U. B. (2007). Random survival forests for R. *R news, 7(2), 25-31.*

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020, October). RFRSF: Employee turnover prediction based on random forests and survival analysis. *In International Conference on Web Information Systems Engineering (pp. 503-515). Springer, Cham.*

Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine, 47(1), 31-39.*

XGBoosting

Barnwal, A., Cho, H., & Hocking, T. (2022). Survival regression with accelerated failure time model in XGBoost. *Journal of Computational and Graphical Statistics, 1-11.*

Li, Y., Li, Y., & Zhou, H. (2018). A survey of survival analysis based on machine learning techniques. *IEEE Access, 6, 74204-74214.*

Mangasarian, O. L., & Wild, M. (1997). Data mining via XGBoosting. *In Proceedings of the 14th international conference on Machine learning (pp. 337-345).*

Vieira, D., Gimenez, G., Marmorela, G., & Estima, V. (2021). XGBoost Survival Embeddings: improving Statistical Properties of XGBoost Survival Analysis Implementation. *Loft. Python.*

Zhang, Y., & Yang, J. (2017). *XGBoost:* A scalable tree boosting system. *In Proceedings of the 22nd acm signed international conference on knowledge discovery and data mining (pp. 785-794). ACM.*

Zhang, X., & Yang, J. (2015). Practical machine learning with XGBoost. *In Proceedings of the 20th acm signed international conference on knowledge discovery and data mining (pp. 829-837). ACM.*

Bagging Survival Trees (BST)

Bou-Hamad, I., Larocque, D., & Ben-Ameur, H. (2011). A review of survival trees. *Statistics surveys, 5, 44-71.*

Lee, C., Zame, W., Alaa, A., & Schaar, M. (2019, April). Temporal quilting for survival analysis. *In The 22nd international conference on artificial intelligence and statistics (pp. 596-605). PMLR.*

Van Wieringen, W. N., Kun, D., Hampel, R., & Boulesteix, A. L. (2009). Survival prediction using gene expression data: a review and comparison. *Computational statistics & data analysis, 53(5), 1590-1603.*

Boosting

Bursa, M., & Todorov, V. (2019). Boosting and survival analysis: A review. *Expert Systems with Applications, 117, 336-346.*

Ilie, S. (2020). Boosting for survival analysis. *Advanced Analytics, 5(2), 9-17.*

Li, Y., Li, Y., Li, M., & Zhang, B. (2020). Boosting regression trees for survival analysis. *Computational and Mathematical Methods in Medicine, 2020.*

MTL

Alaa, A. M., & van der Schaar, M. (2017, December). Deep multi-task gaussian processes for survival analysis with competing risks. *In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 2326-2334).*

Li, Y., Wang, L., Zhou, J., & Ye, J. (2019). Multi-task learning based survival analysis for multi-source block-wise missing data. *Neurocomputing, 364, 95-107.*

Li, Y., Yang, T., Zhou, J., & Ye, J. (2018, May). Multi-task learning based survival analysis for predicting alzheimer's disease progression with multi-source block-wise missing data. *In Proceedings of the 2018 SIAM international conference on data mining (pp. 288-296). Society for Industrial and Applied Mathematics.*

Wang, L., Li, Y., Zhou, J., Zhu, D., & Ye, J. (2017, November). Multi-task survival analysis. *In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 485-494). IEEE.*

General References.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, *112*(4), 545.

Alyass, A., Turcotte, M., Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC medical genomics*, *8*(1), 1-12.

Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular systems biology*, *12*(7), 878.

Athey, S., Bickel, P. J., Chen, A., Imbens, G., & Pollmann, M. (2021). Semiparametric Estimation of Treatment Effects in Randomized Experiments (No. w29242). *National Bureau of Economic Research*.

Axelrod, R., & Nevo, D. (2022). A Sensitivity analysis approach for the causal hazard ratio in randomized and observational studies. *arXiv preprint arXiv:2202.12420*.

Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*(6), 601-609.

Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends in genetics*, *36*(6), 442-455

Balan, T. A. (2018). *Advances in frailty models* (Doctoral dissertation, Leiden University).

Balan, T. A., & Putter, H. (2020). A tutorial on frailty models. *Statistical methods in medical research*, *29*(11), 3424-3454.

Barrett, J. K., Siannis, F., & Farewell, V. T. (2011). A semi-competing risks model for data with interval-censoring and informative observation: an application to the MRC cognitive function and ageing study. *Statistics in medicine*, *30*(1), 1-10.

Cartocci, A., Cevenini, G., & Barbini, P. (2021). A compartment modeling approach to reconstruct and analyze gender and age-grouped CoViD-19 Italian data for decision-making strategies. *Journal of Biomedical Informatics*, 118, 103793.

Chai, H., Zhou, X., Zhang, Z., Rao, J., Zhao, H., & Yang, Y. (2021). Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. *Computers in biology and medicine*, *134*, 104481.

Chai, H., Zhang, Z., Wang, Y., & Yang, Y. (2021). Predicting bladder cancer prognosis by integrating multi-omics data through a transfer learning-based Cox proportional hazards network. *CCF Transactions on High-Performance Computing*, *3*(3), 311-319.

Chen, L., Shao, K., Long, X., & Wang, L. (2020). Multi-task regression learning for survival analysis via prior information-guided transductive matrix completion. *Frontiers of Computer Science*, *14*(5), 1-14.

Chicco, D. (2017) Ten quick tips for machine learning in computational biology. *BioData* Min. 10, 35

Comment, L., Mealli, F., Haneuse, S., & Zigler, C. (2019). Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks. *arXiv preprint arXiv:1902.09304*.

Cortiñas Abrahantes, J., Sotto, C., Molenberghs, G., Vromman, G., & Bierinckx, B. (2011). A comparison of various software tools for dealing with missing data via imputation. *Journal of Statistical Computation and Simulation*, 81(11), 1653-1675.

Cui, P., Shen, Z., Li, S., Yao, L., Li, Y., Chu, Z., & Gao, J. (2020, August). Causal inference meets machine learning. In Proceedings of the 26th ACM SIGKDD *International Conference on Knowledge Discovery & Data Mining* (pp. 3527-3528).

Cuperlovic-Culf, M. (2018) Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8, 4

Gao, F., Xia, F., & Chan, K. C. G. (2020). Defining and Estimating Subgroup Mediation Effects with Semi-Competing Risks Data. *arXiv preprint arXiv:2010.00061*.

Gorfine, M., & Zucker, D. M. (2022). Shared Frailty Methods for Complex Survival Data: A Review of Recent Advances. *arXiv preprint arXiv:2205.05322*.

Guidotti, R. et al. (2018) A survey of methods for explaining black box models. *ACM Comput*. Surv. 51, 1–42

Haneuse, S., & Lee, K. H. (2016). Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal. *Circulation: Cardiovascular Quality and Outcomes*, 9(3), 322-331.

Haradal, S., Hayashi, H., & Uchida, S. (2018, July). Biosignal data augmentation based on generative adversarial networks. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 368-371). IEEE.

Horton, N.J. and Kleinman, K.P., 2007. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61, 79–90

Hong, C., Yi, F., & Huang, Z. (2022). Deep-CSA: Deep Contrastive Learning for Dynamic Survival Analysis With Competing Risks. *IEEE Journal of Biomedical and Health Informatics*, 26(8), 4248-4257.

Huszti, E., Abrahamowicz, M., Alioum, A., & Quantin, C. (2011). Comparison of selected methods for modeling of multi-state disease progression processes: a simulation study. *Communications in Statistics-Simulation and Computation*, 40(9), 1402-1421.

Jin, P., Haider, H., Greiner, R., Wei, S., & Häubl, G. (2021). Using survival prediction techniques to learn consumer-specific reservation price distributions. *Plos one*, 16(4), e0249182.

Jiang, R. (2022). A novel parameter estimation method for the Weibull distribution on heavily censored data. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(2), 307-316.

Jiang, Y., Jin, C., Yu, H., Wu, J., Chen, C., Yuan, Q., ... & Li, R. (2021). Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. *Annals of surgery*, 274(6), e1153-e1161.

Jin, C., Yu, H., Ke, J., Ding, P., Yi, Y., Jiang, X., ... & Li, R. (2021). Predicting treatment response from longitudinal images using multi-task deep learning. *Nature communications*, 12(1), 1-11.

Jing, B., Deng, Y., Zhang, T., Hou, D., Li, B., Qiang, M., ... & Li, C. (2020). Deep learning for risk prediction in patients with nasopharyngeal carcinoma using multi-parametric MRIs. *Computer Methods and Programs in Biomedicine*, 197, 105684.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1), 1-12.

Keret, N., & Gorfine, M. (2020). Optimal Cox Regression Subsampling Procedure with Rare Events. *arXiv preprint arXiv:2012.02122*.

Kim, J. (2020). Developing statistical models and constructing clinical systems for analysing semi-competing risks data produced from medicine, public heath, and epidemiology. *The Korean Journal of Applied Statistics*, 33(4), 379-393.

Kneib, T., & Hennerfeind, A. (2008). Bayesian semi parametric multi-state models. *Statistical Modelling*, 8(2), 169-198.

Lambert, P. C. (2017). The estimation and modeling of cause-specific cumulative incidence functions using time-dependent weights. The Stata Journal, 17(1), 181-207.

Lee, C., Zame, W., Yoon, J., & Van Der Schaar, M. (2018, April). Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

Lee, E. T., & Wang, J. (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Li, L., Greene, T., & Hu, B. (2018). A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Statistical methods in medical research*, *27*(8), 2264-2278.

Li, S., Deng, Y. Q., Zhu, Z. L., Hua, H. L., & Tao, Z. Z. (2021). A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging. *Diagnostics*, *11*(9), 1523.

Li, Y., Yang, T., Zhou, J., & Ye, J. (2018, May). Multi-task learning based survival analysis for predicting alzheimer's disease progression with multi-source block-wise missing data. In *Proceedings of the 2018 SIAM international conference on data mining* (pp. 288-296). Society for Industrial and Applied Mathematics.

Li, Y., Wang, L., Wang, J., Ye, J., & Reddy, C. K. (2016, December). Transfer learning for survival analysis via efficient l2, 1-norm regularized cox regression. In *2016 IEEE 16th International Conference on Data Mining (ICDM)* (pp. 231-240). IEEE.

Libbrecht, M.W. and Noble, W.S. (2015) Machine learning applications in genetics and genomics. *Nat. Rev. Genet*. 16, 321–332

Liu, T., Zhang, L., Jin, G., & Pan, Z. (2022). Reliability Assessment of Heavily Censored Data Based on E-Bayesian Estimation. *Mathematics*, *10*(22), 4216.

Lipton, Z.C. (2018) The mythos of model interpretability. *Comm. ACM* 16, 10

Luck, M., Sylvain, T., Cardinal, H., Lodi, A., & Bengio, Y. (2017). Deep learning for patient-specific kidney graft survival analysis. *arXiv preprint arXiv:1705.10245*.

Ma, C. et al. (2014) Machine learning for big data analytics in plants. *Trends Plant* Sci. 19, 798–808

Marx, V. (2013) Biology: the big challenges of big data. *Nature* 498, 255–260

Meng, X.-L., 2000. Missing data: dial M for ???. *Journal of the American Statistical Association*, 95,1325-1330.

Miller, T. (2019) Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell*. 267, 1–38

Montavon, G. et al. (2018) Methods for interpreting and understanding deep neural networks. *Digit. Signal Process*. 73, 1–15

Molnar, C. (2019) Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. (1st edn), leanpub.com

Nevo, D., & Gorfine, M. (2020). Causal inference for semi-competing risks data. *arXiv preprint arXiv:2010.04485*.

Olden, J.D. and Jackson, D.A. (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model*. 154, 135–150

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., & Zheng, Y. (2019). Recent progress on generative adversarial networks (GANs): A survey. *IEEE access*, *7*, 36322-36333.

Peng, M., & Xiang, L. (2019). Joint regression analysis for survival data in the presence of two sets of semi-competing risks. *Biometrical Journal*, *61*(6), 1402-1416.

Perez, J., Arroba, P., & Moya, J. M. (2023). Data augmentation through multivariate scenario forecasting in Data Centers using Generative Adversarial Networks. *Applied Intelligence*, *53*(2), 1469-1486.

Raghunathan, T. E., (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health*, 25, 99–117.

Siannis, F., Farewell, V. T., & Head, J. (2007). A multi-state model for joint modeling of the terminal and non-terminal events with application to Whitehall II. *Statistics in medicine*, *26*(2), 426-442.

Schrider, D.R. and Kern, A.D. (2018) Supervised machine learning for population genetics: a new paradigm. *Trends Genet*. 34, 301–312

Stephens, Z.D. et al. (2015) Big data: astronomical or genomical? *PLoS Biol*. 13

Tada, T., Kumada, T., Toyoda, H., Kiriyama, S., Tanikawa, M., Hisanaga, Y., ... & Tanaka, J. (2016). HBcrAg predicts hepatocellular carcinoma development: An analysis using time-dependent receiver operating characteristics. *Journal of hepatology*, *65*(1), 48-56.

Tarca, A. L., Carey, V. J., Chen, X. W., Romero, R., & Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS computational biology*, *3*(6), e116.

Thenmozhi, M., Jeyaseelan, V., Jeyaseelan, L., Isaac, R., & Vedantam, R. (2019). Survival analysis in longitudinal studies for recurrent events: applications and challenges. *Clinical Epidemiology and Global Health*, *7*(2), 253-260.

Titman, A. C., & Sharples, L. D. (2010). Model diagnostics for multi-state models. *Statistical Methods in Medical Research*, *19*(6), 621-651.

Xu, Y., Scharfstein, D., Müller, P., & Daniels, M. (2022). A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks. *Biostatistics*, *23*(1), 34-49.

Yin, Q., Chen, W., Wu, R., & Wei, Z. (2022, December). Cox-ResNet: A Survival Analysis Model Based on Residual Neural Networks for Gene Expression Data. In *2022 IEEE International Conference on Networking, Sensing and Control (ICNSC)* (pp. 1-6). IEEE.

Yu, H., Zhang, X., Song, L., Jiang, L., Huang, X., Chen, W., ... & Fu, G. (2021). Large-scale gastric cancer screening and localization using multi-task deep neural network. *Neurocomputing*, *448*, 290-300.

Yuan, H., Xie, F., Ong, M. E. H., Ning, Y., Chee, M. L., Saffari, S. E., ... & Liu, N. (2022). AutoScore-Imbalance: An interpretable machine learning tool for the development of clinical scores with rare events data. *Journal of Biomedical Informatics*, 129, 104072.

Zhang, M. J., Zhang, X. U., & Scheike, T. H. (2008). Modeling cumulative incidence function for competing risks data. *Expert review of clinical pharmacology*, *1*(3), 391-400.

Zhang, J. L., & Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, *28*(4), 353-368.

Zhou, C., Zhou, L., Liu, F., Chen, W., Wang, Q., Liang, K., ... & Zhou, L. (2021). A Novel Stacking Heterogeneous Ensemble Model with Hybrid Wrapper-Based Feature Selection for Reservoir Productivity Predictions. *Complexity*, *2021*.

Books

Allison, P. D. (2014). *Event history and survival analysis: Regression for longitudinal event data* (Vol. 46). SAGE publications.

Allison, P. D., 2002. *Missing Data*. SAGE University Papers.

Ibrahim, J. G., Chen, M. H., Sinha, D., Ibrahim, J. G., & Chen, M. H. (2001). *Bayesian survival analysis* (Vol. 2). New York: Springer.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: a self-learning text* (Vol. 3). New York: Springer.

Koop, G. (2003). *Bayesian econometrics*. John Wiley & Sons.

Kuss, M., & Dawson, R. (2007). *Introduction to Bayesian statistics*, 2nd ed. John Wiley & Sons.

Little, R. J. A. and Rubin, D. B., 2002. *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley.

Schafer, J. L., 1997. *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.