University of Kent

Kent Academic Repository

Sossi, M. and Freitas, Alex A. (2026) *Three Data Bipartitioning Methods to Improve Prediction of Time to Death*. In: Proceedings of the 2026 IEEE International Conference on Healthcare Informatics (ICHI 2026). IEEE . pp. 466-471. IEEE Computer Society – Conference Publishing Services E-ISBN 979-8-3315-6426-1.

Downloaded from
https://kar.kent.ac.uk/115521/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.1109/ICHI69079.2026.000

This document version
Author's Accepted Manuscript

DOI for this version

Licence for this version
CC BY (Attribution)

Additional information
For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Three Data Bipartitioning Methods to Improve Prediction of Time to Death

Michael Sossi
School of Computing
University of Kent
Canterbury, UK
ms2533@kent.ac.uk

Alex A. Freitas
School of Computing
University of Kent
Canterbury, UK
A.A.Freitas@kent.ac.uk

***Abstract*—Accurate individual mortality risk prediction is a central challenge in survival analysis, especially in heterogeneous populations, where severe hazards can obscure subtler risks. We introduce a bipartition of survival data to improve prediction. Using data from the English Longitudinal Study of Ageing, we generate two relatively homogeneous subgroups—those at imminent risk and those at eventual risk of death—using three methods: disjunctive binary splits, an inverse oblique tree, and supervised principal components. For each method, a separate model is trained for each part, after which results are combined. Findings demonstrate improved prediction accuracy compared to overall models without partitioning.**

***Index Terms*—survival analysis, English Longitudinal Study of Ageing, partitioned modelling.**

I. Introduction

Identifying factors that influence the time until death in at-risk populations is a key to crafting interventions that could potentially increase longevity. During the past 50 years, the effort to improve predictive accuracy in the prediction of time until death and the factors that drive it [1] has led to improvements in existing—and a proliferation of new—statistical and machine learning algorithms. In that time, hundreds of clinical studies have addressed this challenge.

Modelling time until death involves not just training instances where the time until death is known, but also training instances for which the time is not known. Some participants may drop out of the study, researchers may lose contact with them, or the study may end before their death. This is known as *right censoring*, in that the time of death, if it occurs at all before the end of the study, cannot be determined. Survival analysis was developed to address the challenges posed by censoring in predicting the time until an event [2, 3].

The results of a survival analysis can be expressed as an estimated survival curve, which shows the predicted probability of survival over time. This is compared with the empirical Kaplan-Meier survival curve [4] built with the raw data, accounting for censoring. The performance of a survival analysis model can be assessed using the Brier score, a measure of predictive accuracy of the fit survival curve at different points in time [5]. The integrated Brier score (IBS) [5, 6] is a summary of Brier scores that measures model fit.

Despite algorithmic improvements, many predictive models of time-to-death appear to be driven not only by a strong signal among the elderly but also by a strong signal among those who are very sick at the start of the study. This is reflected in a common shape of the Brier score curves, where there is high predictive accuracy at the start and at the end of the study, but a fall-off between [e.g. 7, 8, 9, 10, 11, 12]. The weaker predictions typical in the middle of the studies present an opportunity to improve overall prediction. Also, strong signals at the beginning and end of the studies may mask weaker signals there. In particular, a strong signal among the very sick at the start of a study may prevent an accurate prediction of early death among the not obviously symptomatic.

If a study population could be partitioned into two groups, one with a strong signal near the beginning and the other with a strong signal near the end, the masked weaker signals would be exposed and overall prediction could be improved.

In many studies [e.g. 13, 14, 15], the presence of severe conditions strongly associated with imminent death in some fraction of the study groups may have masked the factors predicting death for those without these dire symptoms, and this presence may have lowered the overall predictive accuracy.

The primary goal of the current study is to improve the prediction of early death at the individual level, by first separating individuals into imminent and eventual risk partitions (an $\mathcal{I}$-group and an $\mathcal{E}$-group). This separation is achieved by maximising the difference in expected survival times between the two groups, measured as the difference in the area under the Kaplan-Meier survival curves.

The main contribution of this work is to propose three bipartitioning methods that, with their results combined after being modelled separately, improve prediction relative to an omnibus model. Each method is an adaptation of one used in other learning tasks, here applied to uncovering $\mathcal{I}$ and $\mathcal{E}$ groups in a survival analysis. Two of the methods use decision tree-inspired approaches to find the best split, essentially inverting the usual tree process so that the split determines group membership instead of group membership determining the split. The third method uses a supervised principal component technique that assumes the partition is driven by latent features. Each method showed promise in its ability to separate the $\mathcal{I}$ and $\mathcal{E}$ groups. The two groups from each method are modelled separately using three survival analysis modelling methods—Cox proportional hazards (PH),

accelerated failure time (AFT), and random survival forests (RSF)—with combined results compared to models built before separation. Results show stronger performance using data partitioning across partitioning methods and model types.

While this approach may seem to be a form of decision tree, decision trees are typically 'deep', in that nodes are split until the leaf nodes can no longer support splitting. In contrast, in this study, we are only interested in a single split into two groups. So, this is not so much a decision tree approach as a segmented modelling approach, which is not typical in survival analysis. A similar approach has been used successfully for decades in a very different modelling domain—credit risk scoring [16, 17]—where it is known as a segmented scorecard or hybrid segmentation-based model approach.

The structure of the remainder of this paper is as follows: In section II we provide an overview of our data source and a brief description of the creation of our dataset for survival analysis, we discuss our three proposed partitioning methods in section III, section IV describes results from our initial experiments with partitioning, and section V summarises our conclusions and provides a brief roadmap for continuing work.

II. Dataset

Our analyses used data from the English Longitudinal Study of Ageing (ELSA) [18], a panel study of adults aged 50 and over living in private households in England. ELSA began in 2002, collecting data every two years, with the 10th wave completed in March 2023. Participants provide demographic, economic, social, psychological, cognitive, health, biological, and genetic information through interviews and questionnaires, while periodic nurse visits every four years yield physical and biomarker assessments. For a sample of respondents known to have died, end-of-life interviews are conducted with proxies. Of the original 19,905 recruits, 11,391 completed Wave 1. Data are linkable across waves via unique respondent identifiers and were obtained from the UK Data Service (Dataset 5050) [19].

Of the original recruited sample, we used data from 6,132 respondents with completed nurse visit interviews between July 2004 and August 2005 from the end of Wave 2. The month and year of each participant's nurse visit were used as the start time for each instance. Information about the time of death of participants in the ELSA study was made available through linkage with the National Health Service Central Register, but merges with that database were only undertaken by the ELSA team through the end of 2012, halfway through the Wave 6 interviews. Therefore, mortality information for respondents is only usable through the first five waves (completed in June 2011) [20].

Figure 1 shows the Kaplan-Meier survival curve for our ELSA dataset. The curve is concave, which is a typical shape for a healthy study population. Convex curves typify study groups who share a grave or critical condition, such as cancer or advanced heart disease.

We used the information from the participant interviews and questionnaires, along with the measures taken during the nurse visits, to construct 422 features. These were used both for the partitioning and the fitting of survival models.

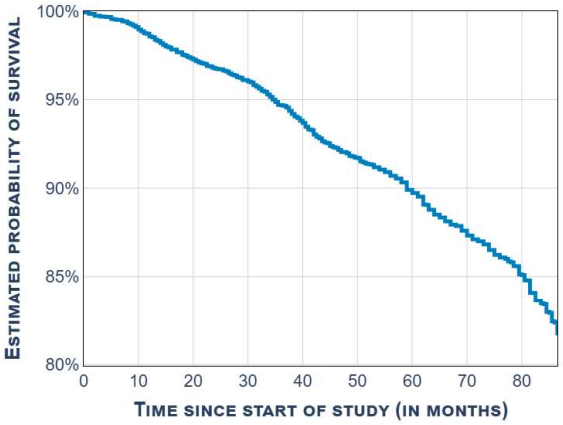


Fig. 1: *Kaplan-Meier survival curve for the consolidated ELSA dataset.*

III. Proposed Partitioning Methods

Building a model to perform the partition of the dataset into an $\mathcal{I}$-group and an $\mathcal{E}$-group required development of novel methods for finding the maximum difference between the survival functions of the two unlabelled groups. The most common method of determining the difference between two survival functions is the log-rank test [21]. However, the log-rank test is tied to the proportional hazards assumption, insofar as the log-rank is a poor estimator of the difference between survival functions if this assumption does not hold [22, p. 223]. Because the area under the survival curve is the expected value of survival time [23], the difference in areas is the difference in expected survival times. This relation is non-parametric and assumption-free [24], and arises from the simple relation between the survival function and the cumulative probability function. Maximising the difference in expected survival time is exactly what we are looking for to partition the $\mathcal{I}$-group and the $\mathcal{E}$-group.

Note that the partitioning for each method is built on the entire dataset. This approach—as opposed to a validation approach on smaller samples (such as split-half validation)—is the one used in credit risk hybrid segmentation-based models, where the validation occurs at the model-building not the segmentation stage, and the value of a partitioning is measured by the improvement of the partitioned model over the unpartitioned model for a given model type.

The separation that maximises the *area between the survival curves* (ABSC) for the two groups requires constructing what is essentially a classifier of the two unlabelled groups. The measure of the ABSC we use is the difference in the *restricted area under* two estimated *survival curves* (rAUSC) [22, 24],

with the rAUSC for each survival curve calculated by

$$\sum_{t_i=t_0}^{t_n} S_K(t_i)(t_{i+1} - t_i),$$

where $t_0, \ldots, t_{n+1}$ are the discrete event times from the beginning of the study through the time of the last known death in either group, and $S_K(t_i)$ is the Kaplan-Meier estimated survival probability at time t_i.

In order to avoid partitions with one of the groups being trivially small, we impose a penalty factor on the rAUSC calculation during the search for a partition:

$$\left(\sum_{i=0}^{n} S_K(t_i)(t_{i+1} - t_i)\right)(p(1-p))^{\lambda}, \tag{1}$$

where p and $(1-p)$ are the proportions of the instances in the dataset for the two groups and the non-negative value λ is a hyperparameter of the constraint. High values of λ lead to more balanced partitions, while low values only avoid extreme imbalances. When $\lambda = 0$, the constraint is inactive. The constraint is equivalent to imposing a Bayesian prior distribution on the segment proportions of the form $\mathrm{Beta}(\lambda+1, \lambda+1)$.[1]

We propose three distinct approaches to implementing this partitioning. The first two use decision tree-inspired techniques on the full set of features to attempt to find a partition that maximises the difference in survival curves. The methods are essentially an inversion of the formation of decision trees, insofar as the split determines the labels instead of the labels determining the split. A tree thus formed is a 'decision stump' (that is, a decision tree with a single binary split) [25].

In a standard decision tree, the split of the data at each node is defined by a single feature, with different features defining the splits at different nodes. Finding the best such split at each node is a simple optimisation problem. Each split divides the feature space into two parts, with the hyperplane partitioning the two parts lying perpendicular to the axis of the feature defining the split and parallel to the hyperplane defined by the axes of the remaining features. Such a split is sometimes called an *axis-parallel* split.

The first two methods we use search for a single *oblique* split, one that uses a combination of multiple features to partition the data [26, 27]. The result is an oblique decision stump with two leaf nodes.

A. Method 1: ICARE Model

Our first partitioning method is a naïve version of an oblique split, the hyperplane defined by the unweighted disjunction of the highest-ranked axis-parallel splits. This is essentially the ICARE model described in [28]. To implement this approach, we first rank the single-feature axis-parallel splits with respect to the constrained ABSC. This ranking is then used to select features in order. Each step adds instances to the set defined by the disjunction of the ranked features up to that point and defines a ABSC value that declines with each step. The result is an ordered set of pairs of increasing instance counts and decreasing area differences, forming a trade-off curve between the two measures.

In order to find an optimal point on this trade-off curve, we find the point that, when left out, minimises the stair-step hyperrectangle area formed by the union of the rectangles determined by all other points. This technique is discussed in [29, 30] in the context of multi-objective optimisation in evolutionary algorithms.

The resulting optimised oblique hyperplane is a baseline against which to compare the weighted oblique hyperplanes found by our full oblique algorithm, described next.

B. Method 2: Inverse Oblique Split

Because oblique decision tree algorithms employ non-exhaustive searches that are not guaranteed to find the global optimal oblique hyperplane and are subject to convergence to a local optimum, we looked for an oblique tree algorithm that attempts to address this limitation. One algorithm that employs methods designed to escape local optima is OC1 [31]. The second method for generating the decision stump for our segmentation is based on this algorithm. Our implementation of OC1 seeks a single oblique split of the dataset that maximises the objective function defined by the ABSC for the split groups, with the Beta constraint pushing the size of the two groups away from extreme proportions of the dataset.

In order to make the coding of the features amenable to the weighted sums specified by oblique splits, the first step is to recode the best axis-parallel splits from $\{0, 1\}$ to $\{-1, 1\}$ and centre the numeric features using their medians.

The algorithm continues with the best axis-parallel split. The best axis-parallel split specifies a hyperplane where all the feature weights but one (that of the single feature defining the split) are 0. The value of the objective function is calculated based on the partitioned groups determined by this hyperplane. The feature split becomes the sole member of the set of attempted features. The algorithm continues by randomly selecting from the remaining features (without replacement), updating the hyperplane when the feature with an optimised weight increases the value of the objective function when incorporated into the hyperplane. We use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm to find the optimum weight [32, Chapter 9].

This continues until all the features have been selected as candidates for improvement or until a specified maximum of m features have been incorporated into the hyperplane. If the maximum has improved by adding at least one feature and its weight to the hyperplane, and the limit of m features has not been reached, the set of attempted features is emptied, and the cycle of selecting features one at a time to update the hyperplane is repeated.

For the first specified count of k cycles with no improvement, a feature in the hyperplane is chosen at random to knock out of the hyperplane—setting its weight to 0—after which the cycling resumes. Cycling through the features continues until

[1] The Beta distribution has the probability density function (PDF) $f(p|a,b) \propto p^{a-1}(1-p)^{b-1}$. If $a = b = \lambda + 1$, the PDF simplifies to $p^{\lambda}(1-p)^{\lambda} = (p(1-p))^{\lambda}$.

the objective function shows no improvement at the end of the cycle for the k^{th} time.

The addition of feature knockouts is an innovation not present in the original version of OC1 [31]. Knockouts perturb interim solutions that cannot be improved by adding more features. In our experiments, the temporary decline in the objective function value resulting from the knockout of a feature is almost always overcome in subsequent cycles.

This process ends with an optimum based on the final hyperplane $\mathcal{H}_*$. But this $\mathcal{H}_*$ optimum may still be local, not global. To address this possibility, after finding the local optimum, we perform a second step that perturbs $\mathcal{H}_*$ by finding the factor α that when multiplied by the random hyperplane $\mathcal{H}_r$ and added to $\mathcal{H}_*$ maximizes the objective function. If the maximised objective function value is greater than the maximum from the first step, $\mathcal{H}_*$ is updated and the second step ends. After this step, we repeat the first step but with the hyperplane initialised with the results of the second step. After this, the last two steps can be alternated as many times as desired. In our experiments, we cycled through the steps until no improvement was found in 5 successive cycles.

C. Method 3: Supervised Principal Components

The third method is based on the supervised principal component analysis (SPCA) described by [33] and [34]. The rationale for SPCA is outlined in [33], in the context of using gene expressions to estimate a latent cell type related to a higher survival rate. That rationale is recast here for the problem of partitioning the imminent death and eventual death groups. Our implementation of this method assumes that a latent partition can be approximated by the first principal component of a set of the features that best separate the $\mathcal{I}$ and $\mathcal{E}$ groups individually. The rationale is as follows.

Suppose the separation of two groups contained in Y is related to an underlying latent (that is, not directly observable) feature U by a linear model:

$$Y = \beta_0 + \beta_1 U + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_Y^2)$.

There is a set of p features P available, a subset $\mathcal{P}$ of which are related to the latent variable by:

$$X_j = \alpha_{0j} + \alpha_{1j} U + \epsilon_j, \quad j \in \mathcal{P}$$

where $\epsilon_j \sim \mathcal{N}(0, \sigma_X^2)$ and ε and ϵ_j are independent.

Presumably, for some $k \in P$, $k \notin \mathcal{P}$. The latent feature U can be regarded as some quality of instances that distinguishes symptomatic early death from the rest of the dataset. Knowing U would make the separation of the $\mathcal{I}$-group from the $\mathcal{E}$-group possible. The set $\mathcal{P}$ represents manifest indicators of the presence of U. To estimate U, we first need to estimate which features belong to $\mathcal{P}$ and which do not.

As with the prior methods, each individual feature is used to split the instances into two groups for which the ABSC is a maximum, by finding the optimal split point (or group, for categorical features). The maximum areas found can be used to rank the features, where the features ranking highest are hypothesised to be the best candidates for membership in $\mathcal{P}$. A cut-off is chosen so that features ranking higher comprise the principal component features $\hat{\mathcal{P}}$, the estimated set of features related to U. The latent feature U can then be estimated with a PCA of the member features of $\hat{\mathcal{P}}$. The first principal component is then used to classify instances into the $\mathcal{I}$ and $\mathcal{E}$ groups by finding the optimal point on the trade-off curve between cut-off values of the first principal component and the corresponding areas between survival curves of the partitions for each cut-off.

The optimal threshold is found by iterating through threshold values of the first principal component between the k^{th} and the $100 - k^{th}$ percentile (where k is chosen to ensure that neither group is too small) and choosing the threshold that maximises the ABSC.

Because many of the features in the ELSA dataset are Boolean or categorical, which are not appropriate for PCA [35, 36], some additional pre-processing is done before conducting the SPCA. We use the method of [36], *factor analysis of mixed data* (FAMD), as implemented in the Prince Python package [37].

IV. Results

When the feature ranking with a specified λ value was done on the entire dataset, separate survival curves were constructed for the fitted $\mathcal{I}$ and $\mathcal{E}$ groups using the methods described in section III. In order to gauge the degree of separation for each method, two-sample t tests based on the mean and variance of survival time for each survival curve were conducted, and two-tailed p-values were calculated.

Figure 2 (*left*) shows the result of the ICARE method and maximum hyperrectangle loss on the trade-off curve to select the optimal point. The partition was defined by the disjunctive combination of the 12 highest-ranking features, using an objective function control parameter setting of $\lambda = 0.00$ (no constraint), with the $\mathcal{I}$ group comprising 6.7% of the dataset (408 instances). There is a good separation ($p = 0.0013$), with 16% of the $\mathcal{I}$-group dying within the first two years (median 75 months), but only 2% of the $\mathcal{E}$-group dying in the same period.

Figure 2 (*centre*) shows the result of the OC1 algorithm, with the maximum number of features set to 7 and the objective function control parameter set to $\lambda = 0.32$. The partition is defined by an optimised oblique split, with the $\mathcal{I}$ group comprising 6.0% of the dataset (366 instances), somewhat smaller in size than the group found using the ICARE method. The two curves show better separation than with ICARE ($p = 0.0002$), with 18% of the $\mathcal{I}$-group dying within the first two years (median 62 months), and still only 2% of the $\mathcal{E}$-group dying in the same period.

Figure 2 (*right*) shows the result of the SPCA method, with the objective function control parameter set to $\lambda = 0.28$, using the top 17% of ranked features. The partition used the first principal component, which is the dimension of maximum variation. The partition is defined by the SPCA derived from 112 features, with the $\mathcal{I}$ group comprising 6.0% of the dataset

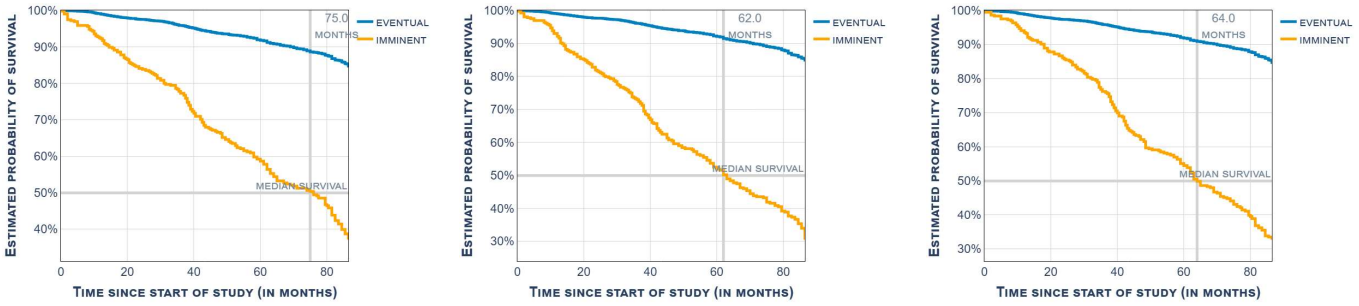


Fig. 2: *Kaplan-Meier survival curves for the partitioning of the training set into imminent and eventual death groups, using the ICARE (left), OC1 (centre), and SPCA (right) algorithms.*

(368 instances), approximately the same size found in the OC1 method. The curves show a separation stronger than ICARE but weaker than OC1 ($p = 0.0004$), which was also built with fewer features. The partition shows 14% of the $\mathcal{I}$-group (median 64 months) and 3% of the $\mathcal{E}$-group dying within the first two years.

These results show that all three methods are capable of producing a partition of the dataset that can be used to move to the next phase of our analysis, the construction of separate models for the two partitions. The results of these early experiments suggest that the OC1 algorithm is the most promising, because it provides the greatest separation.

To demonstrate the benefit of partitioning, we trained Cox PH elastic net, AFT, and RSF models on 5-fold cross-validation training datasets for the unpartitioned data and also on the partitions from each of the partitioning methods. Table I shows the integrated Brier scores for the unpartitioned and combined partitioned models for the three partitioning methods and three model types on test data from the 5-fold cross-validation. In each case, the proposed combined two-partition model outperformed the baseline unpartitioned model with positive improvement for each partitioning/model combination. Significance of improvement was estimated from normal approximations of bootstrap distributions of differences between IBS for partitioned versus non-partitioned models for each combination of partition and model, based on 200 bootstrap samples of model partitioning for each partition/model combination, where none showed significant deviation from normality on Shapiro-Wilk tests.

TABLE I: Cross-validated (5-fold) integrated Brier scores for partitioned models

Partition	Model	IBS (No Partition)	IBS (Two Partition)	Improvement (%)	One-tailed p-value
ICARE	Cox PH	0.066	**0.062**	7.0	< 0.0001
ICARE	AFT	0.074	**0.070**	4.5	0.0006
ICARE	RSF	0.057	**0.056**	2.5	< 0.0001
OC1	Cox PH	0.066	**0.059**	10.5	< 0.0001
OC1	AFT	0.074	**0.070**	4.6	0.0001
OC1	RSF	0.057	**0.055**	3.2	< 0.0001
S-PCA	Cox PH	0.066	**0.060**	9.8	< 0.0001
S-PCA	AFT	0.074	**0.072**	3.0	0.0019
S-PCA	RSF	0.057	**0.056**	2.8	< 0.0001

V. Conclusions

Our results suggest that the partitioning of survival data into imminent and eventual risk groups, followed by separate model training in the partitions—where the drivers of risk in each of the groups can be identified separately—holds promise for improving prediction. The OC1 method holds the most promise among the three we propose, with both the greatest improvement in the partitioned over the unpartitioned model and the fewest features used to define the partition.

Future work will focus on refining the partitioning methods, as well as developing more sophisticated survival modelling methods, namely, heterogeneous ensembles of the three models used in this study.

All code for partitioning, model training, and significance tests is available at https://github.com/KitDarkfeather/survival-bipartition. For steps needed to access the ELSA data, contact the first author directly.

Acknowledgment

Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number R01AG017644. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ELSA is funded by the NIHR Policy Research Programme (HEI) 198_1074_03. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

References

[1] A. F. Abbasi, M. N. Asim, S. Ahmed, S. Vollmer, and A. Dengel. Survival prediction landscape: An in-depth systematic literature review on activities, methods, tools, diseases, and databases. *Frontiers in Artificial Intelligence*, 7, July 2024.

[2] D. R. Cox and D. Oakes. *Analysis of Survival Data*, volume 21. CRC Press, 1984.

[3] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2002.

[4] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[5] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, January 1950.

[6] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17-18):2529–2545, September 1999.

[7] M. Sill, T. Hielscher, N. Becker, and M. Zucknick. c060: Extended inference with lasso and elastic-net regularized Cox and generalized linear models. *Journal of Statistical Software*, 62(5), 2014.

[8] K. Bannangkoon, K. Hongsakul, and T. Tubtawee. Validation of the ALBI-TAE model and comparison of seven scoring systems for predicting survival outcome in patients with intermediate-stage hepatocellular carcinoma undergoing chemoembolization. *Cancer Imaging*, 23(1), May 2023.

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

[10] H. Kvamme and Ø. Borgan. The Brier Score under administrative censoring: Problems and solutions. 2019.

[11] G. Heller. The added value of new covariates to the Brier score in Cox survival models. *Lifetime Data Analysis*, 27(1):1–14, October 2020.

[12] U. B. Mogensen, H. Ishwaran, and T. A. Gerds. Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, 50(11), 2012.

[13] F. Xiong, X. Cao, X. Shi, Z. Long, Y. Liu, and M. Lei. A machine learning-based model to predict early death among bone metastatic breast cancer patients: A large cohort of 16,189 patients. *Frontiers in Cell and Developmental Biology*, 10, December 2022.

[14] R. B. Walter, M. Othus, G. Borthakur, F. Ravandi, J. E. Cortes, S. A. Pierce, F. R. Appelbaum, H. A. Kantarjian, and E. H. Estey. Prediction of early death after induction therapy for newly diagnosed acute myeloid leukemia with pretreatment risk scores: A novel paradigm for treatment assignment. *Journal of Clinical Oncology*, 29(33):4417–4424, November 2011.

[15] C. Errani, M. Cosentino, G. Ciani, L. Ferra, P. A. Alfaro, B. Bordini, and D. M. Donati. C-reactive protein and tumour diagnosis predict survival in patients treated surgically for long bone metastases. *International Orthopaedics*, 45(5):1337–1346, January 2021.

[16] L. Thomas, J. Crook, and D. Edelman. *Credit Scoring and its Applications*. SIAM, 2017.

[17] K. Idbenjra, K. Coussement, and A. De Caigny. Investigating the beneficial impact of segmentation-based modelling for credit scoring. *Decision Support Systems*, 179:114–170, April 2024.

[18] NatCen Social Research. English Longitudinal Study of Ageing, 2023.

[19] UK Data Service. UK Data Service, 2024. Accessed: 2024-10-16.

[20] R. Crawford and P. Mei. An overview of the ELSA 'End of Life' data. *London: The Institute for Fiscal Studies*, 2018.

[21] D. G. Kleinbaum and M. Klein. *Survival Analysis: A Self-learning Text*. Springer New York, 2012.

[22] J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, 2006.

[23] S. Selvin. *Survival Analysis for Epidemiologic and Medical Research*. Cambridge University Press, 2008.

[24] V. Damuzzo, L. Agnoletto, L. Leonardi, M. Chiumente, D. Mengato, and A. Messori. Analysis of survival curves: Statistical methods accounting for the presence of long-term survivors. *Frontiers in Oncology*, 9, June 2019.

[25] W. Iba and P. Langley. Induction of one-level decision trees. *Machine Learning Proceedings 1992*, pages 233–240, 1992.

[26] L. Breiman, J. H. Friedman, R. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, 1984.

[27] S. K. Murthy, S. Kasif, and S. Salzberg. A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2:1–32, August 1994.

[28] L. Rebaud, N. Capobianco, N. Captier, T. Escobar, B. Spottiswoode, and I. Buvat. Similar performance of 8 machine learning models on 71 censored medical datasets: A case for simplicity. *medRxiv*, September 2024.

[29] M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. *Evolutionary Multi-Criterion Optimization*, pages 62–76, 2005.

[30] J. Bader and E. Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, March 2011.

[31] S. K. Murthy, S. Kasif, S. Salzberg, and R. Beigel. OC1: A randomized algorithm for building oblique decision trees. In *Proceedings of AAAI*, volume 93, pages 322–327. Citeseer, 1993.

[32] J. E. Dennis Jr and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, 1996.

[33] E. Bair, T. Hastie, D. Paul, and R. Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473):119–137, March 2006.

[34] E. Bair and R. Tibshirani. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biology*, 2(4):e108, April 2004.

[35] Gilbert Saporta. Simultaneous analysis of qualitative and quantitative data. In *Societa Italiana di Statistica. XXXV riunione scientifica*, volume 1, pages 62–72. CEDAM, 1990.

[36] Jérôme Pagès. Analyse factorielle de données mixtes. *Revue de statistique appliquée*, 52(4):93–111, 2004.

[37] Max Halford. Prince. https://github.com/MaxHalford/prince, 2023. MIT license.